



Université Mohamed Khider de Biskra
Faculté des sciences exactes et des sciences de la nature
et de la vie
Département des sciences de la nature et de la vie
Filière : Sciences biologiques

Référence / 2022

MÉMOIRE DE MASTER

Spécialité : Parasitologie

Présenté et soutenu par :
GUIDOUM BOUZIANI Lyes et HASSENI Rania

Le: mercredi 29 juin 2022

Phylogénie moléculaire et délimitation des espèces du genre *Scorpio* (Scorpionidae, Scorpiones) basée sur le marqueur d'ADN mitochondrial COI.

Jury :

Mme. MEDJADBA Aicha	MAB	Université de Biskra	Président
Pr. MOUSSI Abdelhamid	Pr	Université de Biskra	Rapporteur
Mme. AOURAGH Hayat	MAA	Université de Biskra	Examineur

Année universitaire: 2021/2022

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Remerciements

Nous tenons à saisir cette occasion et adresser nos profonds reconnaissances à **Allah** le tout puissant et le tout miséricordieux, qui nous a donné la force, le courage, la volonté et surtout la patience d'accomplir ce modeste travail.

Un grand merci à notre encadreur **Mr le professeur MOUSSI Abdelhamid** qui nous a donné la chance de faire partie de ce projet. Son positivisme, ses nombreux conseils ainsi que ses nombreuses corrections ont été pour nous une grande aide, nous le remercions infiniment pour son encadrement et d'avoir dirigé ce travail et surtout pour sa disponibilité malgré ses nombreuses occupations. Nous espérons que ce travail sera à la hauteur de ses espérances.

Nos remerciements chaleureux aux membres de jury, c'est pour nous le grand honneur que vous faites en acceptant de présider et d'examiner notre travail et de l'attribuer des remarques et des corrections très intéressantes.

Sans oublier de remercier **Mr le professeur LAIADI Ziane** pour nous avoir accueillis au sein de son laboratoire de recherche et de nous permet la réalisation d'une partie de ce modeste travail. Sincères remerciements à tous les membres « du laboratoire de génétique, biotechnologie et valorisation des ressources vitales », en particulier la doctorante **ZEKRI Wissame** de nous avoir aidés.

Nous tenons à remercier Mr **BEN DAHMANE Abedelhafedh** pour ces nombreux conseils en Bioinformatique durant nos études et son aide.

Nos remerciements s'adressent aussi à **Mr SADINE Salah Eddine** pour ses nombreux conseils sur l'étude des scorpions.

C'est l'occasion à adresser nos reconnaissances envers nos familles et nos proches qui nos ont accompagné, aidé, soutenu et encouragé surtout au long de la réalisation de ce mémoire.

Nos profonds remerciements vont également à nos enseignants, nos collègues d'étude et toutes les personnes qui nos ont aidé ne serait-ce qu'avec un mot d'encouragement.

En fin, nos plus sincères remerciements vont également aux l'ensemble du personnel département de Science de la Nature et de la Vie.

Dédicace

Je dédie ce modeste travail

À mon héros qui me donne de la force quand j'en ai besoin de plus. Ton sourire, ta patience, ta sagesse et tendresse de cœur me donneront à jamais une raison de sourire. À la lumière de ma vie, à mon exemple éternel, mon soutien moral qui s'est toujours sacrifié pour me voir réussir que dieu le garde, **mon père « HASSENI Mohamed »**.

À la reine de ma vie qui a su m'encourager et me pousser à aller toujours plus loin dans mes études, qui m'a toujours soutenu de manière inconditionnelle depuis le début de mes études, qui m'a accordé sa totale confiance, son amour et toute son affection. Sans jamais faillir, **ma chère maman « Noura »**.

« Aucun hommage ne pourrait être à la hauteur de l'amour Dont ils ne cessent de me combler. J'espère que j'ai pu atteindre leurs espérances. Que dieu les gardes et les protège, Leur procure bonne santé et longue vie. »

À ma sœur **Ibtissem** et son mari **Ismail**, et ces filles, mes chères **Biba, Tasnim** et **Ranim**. Que dieu les gardes et les protège.

À mes frères : **Farouk, Okba** et **Djalel** qui mon toujours encouragé.

À mes cousines : **Nahla, Hana, Fairous** et **Salsabil**

A tous les membres de ma famille grands et petits qui étaient toujours a mes cotés. Mes oncles paternels et leurs femmes, mes tantes et leurs enfants, mes oncles maternels, mes tantes, leurs maris et leurs enfants.

À mes chères et proches amies : **Kaouther, Nouria, Chahinez, Raida, Roukaia, Amani** je les remercie pour leur soutien moral, leur encouragement incessant ... qu'*Allah* vous garde pour moi.

A toute la promotion de 2^{ème} **master Parasitologie 2021/2020**.

A mon binôme et toute sa famille.

A tous ceux qui m'ont toujours encouragé dans mes moments les plus difficiles et soutenu dans mes choix et tout au long de mes études.

Et enfin je dédie ce travail à **moi-même** d'avoir eu le courage de finaliser ce travail.

RANIA

Table des matières

Remerciements

Dédicace

Table des matières I

Liste des Tableaux II

Liste des Figures III

Liste des Abréviations IV

Introduction **Error! Bookmark not defined.**

Première partie : Synthèse bibliographique

Chapitre 1: Généralité sur les scorpions et le genre *Scorpio*

1.1. Généralité sur les scorpions 3

1.1.1. Historique 3

1.1.2. Description des scorpions 3

1.1.2.1. Définition 3

1.1.2.2. Reproduction 3

1.1.2.3. Régime alimentaire 3

1.1.2.4. Venins des scorpions 3

1.1.2.5. Habitat 4

1.1.3. Classification 4

1.1.4. Morphologie 4

1.1.5. Distribution mondiale 5

1.2. Généralité sur le genre *Scorpio* 5

1.2.1. Position systématique du genre *Scorpio* dans le monde de scorpion 6

1.2.2. Liste des espèces de *Scorpio* 6

Chapitre 2: Barcoding moléculaire

2.1. Introduction 7

2.2. Le problème de la classification des êtres vivants 7

2.3. Apparition du terme de DNA Barcoding 7

2.4. L'intérêt du DNA Barcoding dans la taxonomie animale 8

2.5. Barcode moléculaire 8

2.5.1. Définition 8

2.5.2. Principe	8
2.5.3. Avantage du gène mitochondrial.....	9
2.5.4. DNA Barcoding et la résolution de la diversité d'espèces.....	10
2.6. Applications de l'ADN Barcoding	10
2.6.1. Surveillance des ravageurs agricoles	10
2.6.2. Identification des vecteurs de maladies	10
2.6.3. Préservation des ressources naturelles	10
2.6.4. Protection des espèces menacées.....	11
2.6.5. Surveillance de la qualité de l'eau.....	11
2.6.6. Authentification de routine des produits de santé naturels	11
2.6.7. Identification des plantes	11

Deuxième partie : Partie expérimentale

Chapitre 3: Matériel et Méthodes

3.1. Matériel	12
3.1.1. Appareillage	12
3.1.2. Origine des séquences.....	12
3.2. Méthodes.....	13
3.2.1. Outil de la bioinformatique	13
3.2.2. Banques de données.....	14
3.2.2.1. BOLDsystems	14
a. Recherche et la sélection des séquences (données nucléiques).....	14
b. Outil d'identification sur BOLDsystems	16
3.2.2.2. GenBank	17
a. Identification et vérification des séquences sélectionnées sur GenBank.....	17
b. Outil d'identification sur BLAST.....	18
c. Options du BLAST	18
3.2.3. Exploitation des données	19
3.2.3.1. Alignement.....	19
a. Principe	20
b. Alignement multiple	20
c. Raffinement	21
3.2.3.2. Construction de l'arbre phylogénique	23
3.2.3.3. Délimitation des espèces.....	24

a. modèle multi-rate Poisson Tree Processes (mPTP).....	25
b. Barcode Index Numbers (BINs).....	25
c. Assemble Species by Automatic Partitioning (ASAP).....	26

Chapitre 4: Résultats et discussion

4.1. Résultats	27
4.1.1. Collection et vérification des séquences	27
4.1.2. Arbre phylogénétique.....	28
4.1.3. Résultat de la délimitation des espèces avec les trois méthodes	35
4.1.3.1. Résultat du la méthode de délimitation des espèces (ASAP)	35
4.1.3.2. Résultat du la méthode de délimitation des espèces (mPTP)	39
4.2. Discussion	44
Conclusion	47
Bibliographie	49
Annexes	
Résumés	

Liste des Tableaux

Tableau 1: les caractéristiques des ordinateurs utilisées	12
Tableau 2: Le nombre des séquences avant et après l'identification sur BOLD et BLAST. ...	28
Tableau 3: Liste des espèces de genre <i>Scorpio</i> dans le monde (Rein, 2020)	(Annexe 2)
Tableau 4: Le code de nucléotide selon UIB	(Annexe 3)
Tableau 5: L'origine géographique des espèces de <i>Scorpio</i> utilisées dans notre étude et leurs numéros d'accès GenBank de leurs séquences IDS	(Annexe 5)

Liste des Figures

Figure 1 : Anatomie d'un scorpion (Abbassi <i>et al.</i> , 2018)	5
Figure 2 : Schéma présentatif sur le principe du code-barres ADN.....	9
Figure 3 : exemple sur le résultat obtenu dans la plateforme BOLDsystems	15
Figure 4 : exemple d'une séquence utilisée sous format FASTA.	16
Figure 5 : exemple sur le résultat d'outil d'identification sur BOLDsystems.	17
Figure 8 : exemple sur l'identification et la vérification des séquences sur GenBank.	18
Figure 7 : Les 4 différents colonnes de comparaison dans le plateforme BLAST	19
Figure 8 : Service web NGphylogeny.fr.....	22
Figure 9 : Outil Gbloks sur Service web NGphylogeny.fr	23
Figure 10 : Exemple sur une séquence mal identifiée prouvée par l'application d'identification de BOLDsystems.....	27
Figure 11 : Phylogramme de vraisemblance maximale ML construit par logiciel MEGA X, déduit pour les spécimens du genre <i>Scorpio</i> . Les nombres sur les branches représentent le support bootstrap. L'arbre est enraciné sur <i>Hottentotta saulcyi</i> comme out groupe.	34
Figure 12 : Comparaison des résultats de la délimitation des 6 espèces de <i>Scorpio</i> à l'aide de deux méthodes différentes et comparaison avec les BINs, l'arbre construit par ML sur le serveur web NGphylogeny.fr	43

Liste des Abréviations

ABGD: Automatic Barcode Gap Discovery

ADN : Acide Désoxyribo Nucléique

ADNm: Acide Désoxyribo Nucléique Mitochondrial

ASAP: Assemble Species by Automatic Partitioning

BIN: Barcode Index Number

BLAST: Basic Local Alignment Search Tool

BOLD: Barcode Of Life Data base

bPTP: bayesian Poisson Tree Processes

COI: Cytochrome c Oxydase sous-unité I

GMYC: Generalized Mixed Yule-Coalescent

IDS: The BOLD Identification System

ML: Maximum Likelihood

MP: Maximum Parsimony

mPTP: multirate Model the Poisson Tree Processes

MSA: Multiple Sequence Alignements

NCBI: National Centre for Biotechnology Information

NIG/CIB: National Institute of Genetics, Center for Information Biology

NIH: National Institutes of Health

Nj: Neighbor-joining

NLM: National Library of Medicine

OTUs: Operational Taxonomic Units

Pb: paire de base

PTP: Poisson Tree Processes

TCS: Statistical Parsimony Analysis

UPGMA: Unweighted Pair Groupe Method with Arithmetic Mean

Introduction

Les scorpions comprennent plus de 2700 espèces en 2022 dont une cinquantaine sont considérées comme d'importance médicale (Rein, 2022). Chaque année, environ 1.5 million de personnes sont piquées par des scorpions (Feola *et al.*, 2020), et gravement touchées par leur venin, et entraînent 3 250 décès (0.27% des envenimations) (Marie-France *et al.*, 1999; Maria *et al.*, 2016; Vaucel, 2020). Les espèces de cette animale sont terrestres et elles peuvent être retrouvées dans tout le monde sauf les zones polaires (Polis, 1990; Fet, 2010). La plus part des espèces de scorpions se vivre dans l'hémisphère sud, et on les trouve surtout dans les régions chaudes notamment dans les déserts (Ismail *et al.*, 2004). Les scorpions possèdent une grande potentialité d'adaptation qui leur a permis de se placer sur tous les étages bioclimatiques et dans tous les biotopes (Vachon, 1952; Saddine, 2018).

Les difficultés d'identification des espèces à partir de techniques morphologiques (Maria *et al.*, 2016) , ainsi que le faible nombre d'études, sont deux des raisons de l'incertitude sur le nombre d'espèces de scorpions existant dans le monde. L'utilisation de marqueurs moléculaires en plus des méthodes morphologiques pourrait être utile pour une évaluation plus précise de la diversité des espèces dans le monde. Un exemple où une identification erronée potentielle des espèces de scorpions pourrait être importante, concerne les espèces du genre *Scorpio*. Les scorpions du genre *Scorpio*, dans la famille Scorpionidae, comprennent 17 espèces et 5 sous espèces à la fois des animaux de couleur brun foncé avec des pattes plus claires ainsi qu'une couleur complètement jaunâtre à jaune rougeâtre (Rein, 2022). Le genre est limité à l'Afrique et à l'Asie occidentale, à l'est de l'Iran. Il vit en Afrique du Nord au nord du Sahara, de la côte atlantique à l'Égypte, sur la péninsule arabique et au levant, au nord au sud de la Turquie (Fet , 2000).

Depuis que Hebert *et al.* (2003) ont introduit l'approche moléculaire du DNA barcoding, basée sur un fragment du gène mitochondrial cytochrome c oxydase subunit I (COI), elle constitue un outil moléculaire puissant pour l'identification des espèces et l'évaluation de la biodiversité (Hebert *et al.*, 2003a). Cependant, pour clarifier la diversité cryptique ou le polymorphisme des espèces, il peut être nécessaire d'appliquer des méthodes moléculaires modernes pour des approches analytiques plus sophistiquées (Jin *et al.*, 2018). Ces dernières années, les analyses phylogénétiques moléculaires et les approches de délimitation des espèces sont considérées comme des outils fondamentaux pour comprendre les modèles de diversité des espèces (Zhang *et al.*, 2013 ; González-Serna *et al.*, 2018). Par conséquent, plusieurs algorithmes et méthodes ont été proposés pour les méthodes de délimitation des espèces moléculaires.

Cette étude a eu pour objectif d'étudier les techniques fournies par la biologie moléculaire pour déterminer le rôle de la code barre moléculaire et mettre en valeur la bioinformatique dans ce domaine sur la richesse spécifique de genre *Scorpio*. D'autre part de rechercher les possibilités de décrire des hypothèses pour l'identification moléculaire de séquences disponibles dans les bases de données du genre *Scorpio* par les méthodes de délimitations.

Le manuscrit est divisé en deux parties, il s'agit d'une étude théorique résulte d'une recherche bibliographique en relation avec notre thématique recherchée et une partie expérimentale résulte d'une analyse des travaux en relation avec ce travail.

- L'étude bibliographique de ce manuscrit (la première partie) comprend deux chapitres complémentaires :

Chapitre 1 : est consacrée à une étude bibliographique dans laquelle nous avons résumé des généralités et quelques informations sur les scorpions et le genre *Scorpio* étudiées.

Chapitre 2 : est centré sur le Barcode moléculaire (code-barres d'ADN) et leurs avantages dans le domaine taxonomique et ses applications.

- Ensuite la deuxième partie qui concerne étude expérimentale comprend deux chapitres :

Chapitre 3 : Matériel et méthodes décrire les analyses phylogénétiques des séquences nucléotidiques par des méthodes explicatives de la bioinformatique pour obtenir l'arbre phylogénétique.

Chapitre 4 : Les résultats obtenus à partir des analyses phylogénétiques sous forme d'arbre phylogénétique et une discussions sur les groupes obtenus après la délimitation des espèces étudiées.

Première partie :
Synthèse
Bibliographique

Chapitre 1:
Généralité sur les
scorpions et le genre
Scorpio

1.1. Généralité sur les scorpions

1.1.1. Historique

Les scorpions sont les plus anciens arthropodes chélicérates connus. Ils sont apparus dans les milieux aquatiques siluriens, dont les premiers sont des fossiles aquatiques ou du moins amphibies datant de **425** à **450** millions d'années. Ils sont largement distribués, parce qu'il est animaux lents avec des mouvements limités et très dépendants du biotope (Vachon, 1952; Polis, 1990; Fet, 2010; Sadine *et al.*, 2020).

1.1.2. Description des scorpions

1.1.2.1. Définition

Les scorpions sont des arachnides présents dans toutes les régions chaudes ou tempérées du monde (Delbarre et Haro, 2014). Ce sont des invertébrés nocturnes venimeux renvoyés d'un exosquelette chitineux articulé qui les protège et les rend résistants aux conditions écosystémiques les plus extrêmes (Vachon, 1952).

1.1.2.2. Reproduction

Après une parade nuptiale comparable à une « danse », le mâle entraîne la femelle vers un lieu propice au dépôt de son spermatophore (« capsule » qui contient les gamètes mâles). Une fois la femelle fécondée, il peut lui arriver de dévorer le mâle, c'est pourquoi la fuite de ce dernier est prompte. Les scorpions sont ovovivipares et peuvent donner naissance à dizaines de jeunes (appelés « pullus »), qui montent rapidement sur le dos de la mère une fois qu'ils sont nés (Lourenço, 2000 ; Rouschmeyer, 2015; Lourenço, 2018).

1.1.2.3. Régime alimentaire

Les scorpions consomment essentiellement divers arthropodes (araignées, insectes, crustacés ...) qu'ils chassent en visitant leur territoire ou « à l'affût » (Polis, 1990; Rouschmeyer, 2015).

1.1.2.4. Venins des scorpions

Toutes les espèces de Scorpions sont venimeuses (Goyffon, 2002). Le venin de scorpion est un mélange complexe de toxines qui varie selon les taxons et parfois selon les individus d'une même espèce (Marie-France *et al.*, 1999). Certaines espèces sont nocifs pour l'homme à cause de leur piqûre mortelle. En raison de leur venin mortel, ils ont toujours été sous l'attention et la haine dans le monde entier (Rouhullah *et al.*, 2016). Toutes les espèces de Scorpions sont toxiques bien que tous ne soient pas mortels pour l'homme (Tikader et Bastawade, 1983).

La recherche sur le venin de scorpion est abondante et prometteuse, tant le venin contient diverses substances pouvant être utilisées sur un plan pharmacologique (Fet, 2010).

1.1.2.5. Habitat

En général, Ils vivent en groupe (Vachon, 1952), Ils cherchent les coins obscurs où ils creusent des terriers (Ismail, 2003 ; Sadine *et al.*, 2014). Ils existent dans une variété d'habitats : pierres, rochers, écorces et sous les vieux bâtiments. Ils cherchent des coins sombres où ils creusent des terriers (Guo *et al.*, 2013).

1.1.3. Classification

Les scorpions sont des animaux appartenant à la branche des arthropodes, sous-embanchement des chélicérates, classe des Arachnides. Ils sont une ancienne lignée d'arthropodes terrestres et l'un des ordres les plus cosmopolites de la classe des Arachnides après les araignées (Goyffon, 2002 ; Coelho, 2013).

L'ordre des scorpions se nommés aussi Scorpionida. Il y a quelques années, il était appelé Scorpionida. La classification d'après le Catalog of scorpions of the world (Fet *et al.*, 2000) et reprise par Prendini et Wheeler (2005) l'ordre des scorpionides contient 19 familles, et selon Soleglad et Fet (2005) l'ordre des scorpionides compose par 15 familles (Fet, 2010).

Aujourd'hui, Il existe 2704 espèces de scorpions dans le monde ont été décrites par les zoologistes en 2022 (Rein, 2022). Et plusieurs espèces sont encore décrites chaque année, car bien de nombreuses espèces sont rarement en contact avec l'homme, soit par leur mode de vie discret, soit par leur faible densité humaine (Goyffon, 2002; Fet, 2010).

1.1.4. Morphologie

La littérature sur l'anatomie et la morphologie du scorpion a commencé sérieusement au début des années 1800 (John, 1990). Ces arachnides, de taille variable entre 3 à 20 cm et de poids variant entre 3 et 60 grammes, avec un corps divisé en trois parties : Prosoma ou céphalothorax, Mesosoma ou pré-abdomen, Metasoma ou post abdomen ou queue (Abassi *et al.*, 2018). Les scientifiques décrivent généralement les couleurs des scorpions en termes de brun, jaune, orange, rouge et des variantes telles que le brun orangé, le jaune vif et le rouge foncé (Bannister *et al.*, 1977).

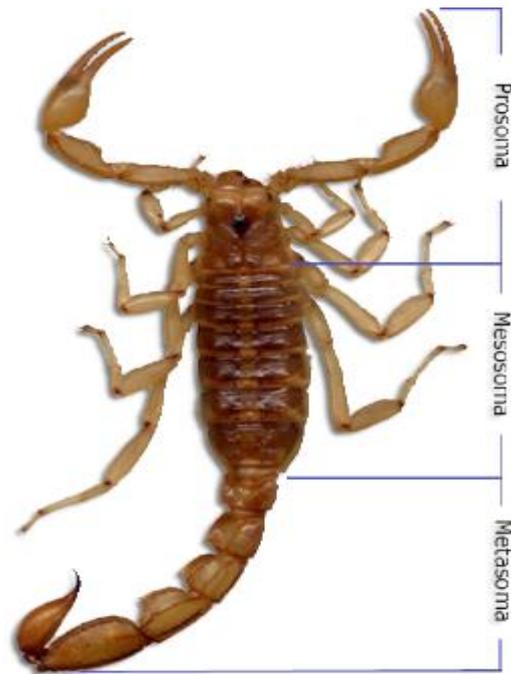


Figure 1 : Anatomie d'un scorpion (Abbassi *et al.*, 2018)

1.1.5. Distribution mondiale

Les scorpions figurent parmi les animaux qui ont une large répartition à travers le monde, notamment dans les déserts et les zones arides (Polis, 1990 ; Sadine, 2018).

Dans le monde, les scorpions sont absents des zones les plus froides, c'est-à-dire grossièrement le tiers nord du globe (Tikader et Bastawade, 1983). Ils sont partout en Afrique, en Australie, en Amérique du sud et centrale, et atteignent leur limite septentrionale dans le sud du Canada, de la France ou du Kazakhstan par exemple (Rouschmeyer, 2015).

1.2. Généralité sur le genre *Scorpio*

Le genre *Scorpio* a été créé par Linnaeus en 1758 (en partie), et a *Scorpio maurus* Linnaeus, 1758 comme espèce type, définie par désignation postérieure (Fet, 2000).

Birula (1910) a d'abord passé en revue toutes les populations de *S. maurus* en Afrique du Nord et au Moyen-Orient. « Ses recherches ont indiqué que toutes les formes de *S.maurus* sont des sous-espèces, mais toutes les formes ont été classées en deux groupes comme « *sectio maurus* » [(*S. m. maurus* L. (= *S. maurus* L.), *S. m. tunetanus* Birula, 1910 (= *S. punicus* Fet, 2000), *S. m. palmatus* (Ehrenberg, 1828), *S. m. fuscus* (Ehrenberg, 1829), *S. m. subtypicus* Birula, 1910 (= *S. birulai* Fet, 1997), *S. m. magadorensis* Birula, 1910 (= *S. magadorensis* Birula, 1910), *S. m. hesperus* Birula, 1910 (= *S. hesperus* Birula, 1910)] et « *sectio*

propinquus » [*S. M.arabicus* (Pocock, 1900), *S. m. propinquus* (Simon, 1872), *S. m. Kruglovi* Birula, 1910, *S. m. townsendi* (Pocock, 1900)] » (Halil et al., 2017).

1.2.1. Position systématique du genre *Scorpio* dans le monde de scorpion

Le genre *Scorpio* a partaient de la famille des Scorpionidae qui décrit par Latreille en 1802 (Lourenço, 2009). Cette famille se compose de 18 genres et 185 espèces (seuls les taxons existants sont comptés) (Rein, 2022).

- **Règne** : Animalia
- **Embranchement** : Arthropoda
- **Sous-embranchement** : Chelicerata
- **Classe** : Arachnida
- **Ordre** : Scorpionida
- **Famille** : Scorpionidae
- **Genre** : *Scorpio*

1.2.2. Liste des espèces de *Scorpio*

Ce genre contient 17 espèces jusqu'à présentes qui sont: *Scorpio birulai*, *Scorpio ennedi*, *Scorpio fuliginosus*, *Scorpio fuscus*, *Scorpio hesperus*, *Scorpio kruglovi*, *Scorpio mogadorensis*, *Scorpio niger*, *Scorpio occidentalis*, *Scorpio palmatus*, *Scorpio propinquus*, *Scorpio punicus*, *Scorpio savanicola*, *Scorpio sudanensis*, *Scorpio tassili*, *Scorpio weidholzi* et en fin *Scorpio maurus*. La liste en détails avec les sources et leurs d'origine géographique est présentée dans (Annexe 1), L'espèce de *Scorpio maurus* porte 5 sous espèces qui sont : *S. maurus maurus*, *S. maurus arabicus*, *S. maurus behringi*, *S. maurus stemmleri*, *S. maurus townsendi* (Rein, 2022).

Chapitre 2:

Barcoding moléculaire

2.1. Introduction

C'était en 25 avril 1953 que la biologie a connu un grand élan grâce à la découverte de deux jeunes scientifiques James Watson et Francis Crick la double structure hélicoïdale de la molécule d'ADN qui à un intérêt biologique considérable et ce fut la naissance de la biologie moléculaire (Watson et Crick, 1953; Halloran, 1984). C'est le résultat d'une recherche de plus d'un siècle à la définition de la structure du matériel génétique dans les êtres vivants après les lois de Gregor Mendel et la découverte de la nature chimique de l'ADN par Jhoann Friedrich Miesher (Lunardi, 2003). Cette découverte est la première pierre de cette nouvelle science. D'autres on dit qu'une nouvelle ère est née, l'ère de la biologie moléculaire (Halloran, 1984).

La découverte de la structure de l'ADN a donné un coup de pouce aux scientifiques pour comprendre d'autres domaines notamment les mécanismes de la réplication et l'expression des informations génétiques (Lunardi, 2003).

La biologie moléculaire comprend d'autres outils enzymatiques qui permettent de manipuler, lier, couper, synthétiser et encore lire l'ADN. L'expérimentation *in vivo* a donné naissance à des voies de recherches et des techniques très avancées dans le cadre de la thérapie génique et de la cartographie génomique (Laudenbach *et al.*, 1999).

2.2. Le problème de la classification des êtres vivants

L'une des méthodes la plus importante et la plus complexe pour les biologistes et les taxonomistes est la classification des êtres-vivants (Ankita *et al.*, 2016).

Récemment, Diethard *et al.* (2003) ont plaidé en faveur d'un système taxonomique basé sur l'ADN. L'analyse des séquences d'ADN est utilisée depuis 30 ans pour faciliter l'identification des espèces. Différentes séquences ont été utilisées pour différents groupes taxonomiques et dans différents laboratoires (Robert *et al.*, 2005).

2.3. Apparition du terme de DNA Barcoding

De ce point de vue et selon une approche scientifique, Paul Hébert en 2003 a orienté cette idée en proposant une technique fiable, rentable et accessible au problème actuel de la diversité des espèces appelée **DNA-Barcoding** (Hebert *et al.*, 2003a).

La technique de DNA-Barcoding tombe à pic car la discrimination du règne animale qui constitue environ 10 à 15 millions d'espèces n'est pas une tâche facile (Hammond, 1992; Hawksworth et Kalin-Arroyo, 1995 ; Hebert *et al.*, 2003a), en plus cette technique a trouvé la solution aux erreurs courantes de diagnostic morphologique qui ne sont souvent efficaces que

pour un stade de vie ou pour un sexe particulier. Les systèmes d'identification microgénomique permettent la discrimination à travers l'analyse d'un petit segment d'ADN qui représente une approche extrêmement prometteuse au diagnostic de la diversité biologique (Hebert *et al.*, 2003a; Raupach *et al.*, 2010).

2.4. L'intérêt du DNA Barcoding dans la taxonomie animale

DNA-Barcoding est maintenant un outil taxonomique accepté pour déterminer à quelle espèce appartient un organisme. Les codes-barres ADN sont maintenant utilisés et promus pour une variété d'applications biologiques, notamment ; l'identification d'espèces cryptiques, la découverte d'espèces et la révision taxonomique (Smith *et al.*, 2006 ; Bickford *et al.*, 2007 ; Burns *et al.*, 2008 ; Craig *et al.*, 2011).

2.5. Barcode moléculaire

2.5.1. Définition

DNA Barcoding est une méthode conçue pour rendre l'identification d'espèces plus facile, précise et automatisable en utilisant des marqueurs génétiques courtes et standardisées d'ADN mitochondriale comme étiquettes d'espèces. En conséquence, elle rendra le rythme de la découverte d'espèces par les taxonomistes plus rapide. Elle offre aussi aux taxonomistes l'occasion d'élargir considérablement, et éventuellement de compléter, une banque mondiale de la diversité de la vie (Hebert et Gregory, 2005).

Les premiers tests de codes-barres génétiques utilisant des marqueurs mitochondriaux sur des animaux ont rapporté une précision de près de 100 % (Meyer et Paulay, 2005).

2.5.2. Principe

Hébert *et al.* (2003) ont proposé qu'une seule séquence de gène sera suffisante pour différencier toutes les espèces animales, ou du moins la grande majorité d'entre eux, et ils ont aussi conseillé l'utilisation du gène de la sous-unité I du cytochrome c oxydase (COI) de l'ADN mitochondrial comme système global de bio-identification pour les animaux (Ward *et al.*, 2005; Taylor et Harris, 2012).

Un fragment de 650 paires de bases de l'extrémité 5' du gène mitochondrial code pour la première sous unité de la cytochrome c oxydase (COI) a été proposé comme norme mondiale, le soi-disant « code-barres région » pour les animaux (Hebert *et al.*, 2003a ; Hebert *et al.*, 2003b; Raupach *et al.*, 2010).

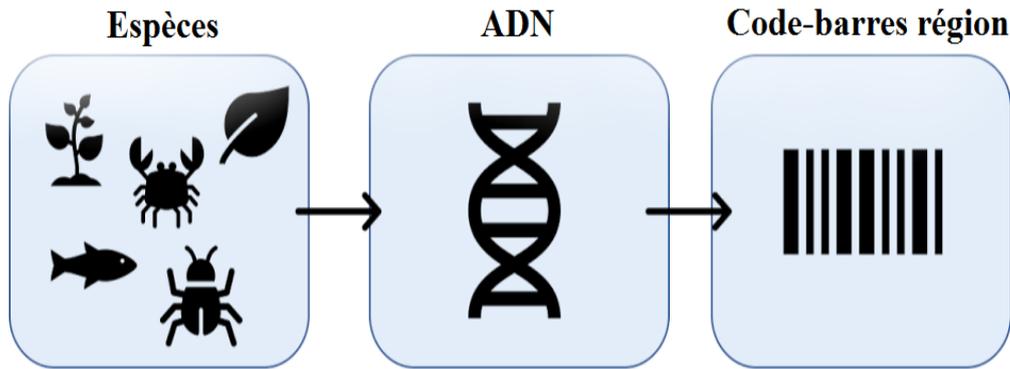


Figure 2 : Schéma présentatif sur le principe du code-barres ADN

2.5.3. Avantage du gène mitochondrial

Il est clair que le génome mitochondrial des animaux est une meilleure cible d'analyse que le génome nucléaire en raison de ses plusieurs caractéristiques qui le rend intéressant pour la taxonomie moléculaire, à savoir les taux de substitution généralement élevés les indels (insertions et délétions) sont rares, l'hérédité exclusivement maternelle et son manque d'introns, de son exposition limitée aux recombinaison (Saccone *et al.*, 1999 ; Hebert *et al.*, 2003a ; Ballard et William 2004 ; Raupach *et al.*, 2010) . Ce gène est considéré comme approprié car la diversité des séquences nucléotidiques d'une même région génique permettait régulièrement la discrimination d'espèces étroitement proche (Hebert *et al.*, 2003b). En outre, la raison de l'hérédité uni-parentale et de l'haploïdie, l'ADN mitochondrial a une taille de population efficace quatre fois plus petite par rapport à l'ADN nucléaire, conduisant à une lignée plus rapide (Birky *et al.*, 1983; Raupach *et al.*, 2010).

En plus parmi ses avantages importants, les amorces universelles pour ce gène sont très robustes, permettant la récupération de son extrémité 5' à partir de représentants de la plupart, ou la totalité des phylums animaux (Folmer *et al.*, 1994 *in* Hebert *et al.*, 2003a). Encor, le COI semble posséder une plus grande gamme de signaux phylogénétiques que tout autre gène mitochondrial. En fait, l'évolution de ce gène est suffisamment rapide pour permettre la discrimination non seulement d'espèces étroitement apparentées, mais aussi de groupes phylogéographiques au sein d'une même espèce (Cox et Hebert 2001 ; Wares et Cunningham 2001 ; Hebert *et al.*, 2003a). De plus, ce gène est disponible en très grande quantité dans les tissus et les cellules, ce qui rend son extraction, son analyse, et son séquençage disponible en laboratoire. Il est l'un des marqueurs le plus populaire pour la génétique des populations et la phylo-géographie et l'identification d'organismes (Avisé, 2000 ; Derycke *et al.*, 2010 ; Hebert *et al.*, 2011)

Cette approche du code-barres a été appliquée avec succès chez divers vertébrés et taxons d'invertébrés pour la délimitation et l'identification des espèces (Smith *et al.*, 2006 ; Raupach *et al.*, 2010).

2.5.4. DNA Barcoding et la résolution de la diversité d'espèces

Des travaux récents ont conduit qu'un système d'identification basé sur l'ADN et sur le gène mitochondrial, le cytochrome c oxydase la sous-unité I (COI) peut aider à la résolution de la diversité d'espèces. Ils ont établi que la diversité dans les séquences des amino-acides codées par ce gène mitochondrial est suffisante pour placer de manière fiable les espèces en catégories taxonomiques supérieures (d'embranchements aux ordres) (Hebert *et al.* 2003b).

2.6. Applications de l'ADN Barcoding

Cette technologie a connu de nombreuses utilisations dans différents domaines, dont par exemple :

2.6.1. Surveillance des ravageurs agricoles

Les codes-barres ADN peuvent aider à identifier les ravageurs à n'importe quel stade de vie, ce qui facilite leur contrôle, et permet aux agriculteurs d'économiser des milliards de dollars en dommages causés par ces ravageurs. L'initiative mondiale de codage à barres des téphritidés contribue à la gestion des mouches des fruits en fournissant des outils pour identifier et arrêter les mouches des fruits à la frontière (Kaur, 2015).

2.6.2. Identification des vecteurs de maladies

Le code-barres ADN permet aux non-écologistes d'identifier les espèces de vecteurs pouvant provoquer des maladies infectieuses graves chez les animaux et les humains, de comprendre ces maladies et de les guérir. Une initiative mondiale de codage à barres des moustiques dans la création d'une bibliothèque de codes à barres de référence qui peut aider les responsables de la santé publique à contrôler ces maladies causant par ces espèces vectrices plus efficacement et avec très moins d'insecticides (Bass *et al.*, 2015 ; Allard, 2016 ; Elsiddig *et al.*, 2022).

2.6.3. Préservation des ressources naturelles

En utilisant le code-barres ADN, les gestionnaires des ressources naturelles peuvent surveiller le commerce illégal de produits fabriqués à partir de ressources naturelles comme les feuillus. Fishbol est une bibliothèque de code-barres de référence pour les feuillus pour

améliorer la gestion et la conservation des ressources naturelles (Fish, 2012 ; Stat *et al.*, 2017).

2.6.4. Protection des espèces menacées

La population de primates est réduite de 90 % en Afrique à cause de la chasse à la viande de brousse. Le code-barres ADN peut être utilisé par les forces de l'ordre pour la viande de brousse sur les marchés locaux qui est obtenue à partir de viande de brousse (Kaur, 2015 ; site web 1).

2.6.5. Surveillance de la qualité de l'eau

L'eau potable est une ressource de processus pour les êtres vivants. En étudiant les organismes vivant dans les lacs, les rivières et les ruisseaux, leur santé peut être mesurée ou déterminée. Le code-barres ADN est utilisé pour créer une bibliothèque de ces espèces qui peuvent être difficiles à identifier. Les codes à barres peuvent être utilisés par les agences environnementales pour améliorer la détermination de la qualité et pour créer de meilleures politiques qui peuvent assurer un approvisionnement sûr en eau potable (Kaur, 2015 ; site web 2).

2.6.6. Authentification de routine des produits de santé naturels

L'authenticité des produits de santé naturels est une importante question juridique, économique, sanitaire et de conservation. Les produits de santé naturels sont souvent considérés comme sûrs en raison de leur origine naturelle (Kaur, 2015).

2.6.7. Identification des plantes

L'identification des feuilles des plantes même si les fleurs ou les fruits ne sont pas disponibles, et l'identification des plantes médicinales (Kaur, 2015).

Deuxième partie :
Partie expérimentale

Chapitre 3:

Matériel et méthodes

Cette section fait l'objet d'une enquête scientifique qui consiste à faire collecter, sélectionner, vérifier, traiter et analyser une série des séquences nucléotidiques de genre *Scorpio* puis synthétisés de façon d'un arbre phylogénétique et objective à partir des données moléculaire probantes et pertinentes afin d'atteindre nos objectifs soulignés dans l'introduction

3.1. Matériel

Cette étude a été effectuée sur des outils en bioinformatiques et en biostatistiques soit sur des plateformes en ligne soit des logiciels téléchargeables.

3.1.1. Appareillage

Les appareils utilisés dans ce travail sont de simples ordinateurs, ses caractéristiques sont détaillées dans le tableau ci-dessous :

Tableau 1: les caractiristiques des ordinateurs utilisées

Les caractéristiques	Ordinateur 1	Ordinateur 2
Nom de l'ordinateur	DELTA- PC	DELTA- PC
Fabricant de système	LENOVO	Acer
Processeur	Intel ® Celeron ® CPU N3060 @ 1.60GHz (2CPUs)	Intel ® Celeron ® 2957U @ 1.4GHz (2CPUs)
Mémoire installée RAM	4096MB RAM	4096MB RAM
Stockage	500 GB HDD	500 GB HDD
Système d'exploitation	Windows 7 édition intégrale 64 bits	Windows 7 édition intégrale 64 bits
Type de système	Système d'exploitation 64 bits	Système d'exploitation 64 bits
Version du système	(6.1, version 7601)	(6.1, version 7601)

3.1.2. Origine des séquences

Toutes les espèces de genre *Scorpio* retenues dans cette étude ont été sélectionnées sur la base de disponibilités de leurs séquences nucléotidiques soumises dans BOLDsystems, et vérifiées dans la banque des données nucléotidiques GenBank et BLAST hébergé sur le portail du National Center Biotechnology Information NCBI.

L'étude a été menée sur des séquences ADN de 6 espèces de genre *Scorpio* à partir d'un marqueur unique, nommé Barcode, constitué d'une séquence d'environ 650 pb correspondant à l'extrémité 5' du gène du cytochrome c oxydase 1 du génome mitochondrial.

3.2. Méthodes

Cette partie décrit les méthodes utilisées afin d'aboutir à l'objectif de l'approche.

3.2.1. Outil de la bioinformatique

Le code-barres ADN est un système d'identification moléculaire récent et largement utilisé qui vise à identifier des spécimens biologiques et à les attribuer à une espèce donnée. Cependant, le code-barres ADN est encore plus que cela, ainsi que de nombreuses utilisations pratiques, il peut être considéré comme le noyau d'un système taxonomique intégré, où la bioinformatique joue un rôle primordial (Casiraghi *et al.*, 2010).

Dans ce contexte, nous pouvons dire que DNA barcoding est un outil moléculaire et bioinformatique qui vise à identifier les espèces biologiques (Hebert *et al.*, 2003a). L'idée de base est assez simple : grâce à l'analyse de la variabilité d'un ou plusieurs marqueurs moléculaires standards, il est possible de discriminer des entités biologiques au rang taxonomique de l'espèce. Cette méthode repose sur l'hypothèse que la variation génétique entre les espèces dépasse celle au sein des espèces. Par conséquent, l'analyse idéale des codes-barres ADN reflète les distributions des variabilités intra et inter spécifiques séparées par une distance appelée « écart de code-barres ADN » (Meyer et Paulay, 2005; Martin et Wiemers, 2007; Casiraghi *et al.*, 2010).

De quoi dépend la précision? Le « gap » du codage à barres. Le problème critique du codage à barres est la précision. Dans quelle mesure une seule séquence de gène réussit-elle à délimiter et à identifier les espèces? La précision dépend en particulier de l'étendue et de la séparation entre la variation intra-spécifique et la divergence inter-spécifique dans le marqueur sélectionné. Plus il y a de chevauchement entre la variation génétique au sein des espèces et la divergence séparant les espèces sœurs, moins le code-barres est efficace (Meyer et Paulay, 2005).

Nous utilisons ces méthodes pour tester les performances et employons une nouvelle approche phylogénétique pour calculer la variation intraspécifique et la divergence interspécifique. Les taux d'erreur sont estimés pour :-1 l'identification des échantillons par rapport à une phylogénie bien caractérisée et -2 l'aide à la découverte d'espèces pour des

groupes partiellement connus. Nous constatons que l'erreur globale la plus faible pour l'identification des espèces est de 4 % (Meyer et Paulay, 2005).

Après avoir effectué toutes ces analyses, la bibliothèque nous met à disposition un moyen de créer l'arbre phylogénétique sur la base des analyses précédentes suivant la méthode du Neighbor-joining. Le principe de cette méthode est de trouver des paires d'unités taxonomiques opérationnelles (OTU = voisins =Neighbor) qui minimisent la longueur totale des branches à chaque étape du cluster des OTU en commençant par une étoile arbre. Les longueurs des branches ainsi que la topologie d'un arbre parcimonieux peuvent être obtenues rapidement en utilisant cette méthode (Saitou et Nei, 1987).

Enfin, nous obtenons un phylogramme sous forme de clusters, c'est derniers nous permettent de connaître le degré de parenté des organismes c'est-à-dire quand les individus sont regroupée dans un même cluster ça signifie qu'ils appartiennent à une même espèce.

3.2.2. Banques de données

3.2.2.1. BOLDsystems

Barcode Of Life Data Systems (BOLDsystems) est une plate-forme Web créée en 2005, développée au Centre for Biodiversity Genomics au Canada, qui fournit un environnement intégré pour le stockage et l'analyse de données de codes-barres ADN. Elle se compose de quatre modules principaux : un portail de données, un portail éducatif, un registre de NIB (espèces putatives) et un atelier de collecte et d'analyse de données (Megan *et al.*, 2011).

En assemblant les données moléculaires morphologiques et de distribution, il comble un fossé traditionnel en bioinformatique. BOLD est disponible gratuitement pour tout chercheur intéressé par les codes-barres ADN. En fournissant des services spécialisés, il facilite l'assemblage des enregistrements qui répondent aux normes nécessaires pour obtenir la désignation codes-barres dans les bases de données mondiales de séquences (Sujeewan et Hebert, 2007).

Au cours des dernières années, ce portail est devenu un puissant atelier en ligne et un centre informatique central pour la communauté des codes-barres ADN (Megan *et al.*, 2011).

a. Recherche et la sélection des séquences (données nucléiques)

Le portail de données publiques prend en charge les requêtes basées sur la taxonomie (noms scientifiques uniquement), la géographie, les champs d'attribution (c'est-à-dire les collectionneurs et les taxonomistes), les dépositaires de spécimens, les codes de projet ou

d'ensemble de données, les identifiants de spécimen et de séquence (sampleids, museumids, traités, etc.). Les recherches en texte libre sont autorisées. Le système essaiera d'extraire des termes significatifs de la recherche. Les termes de recherche doivent être séparés par un espace, les termes de recherche contenant plus d'un mot ("États-Unis" ou "Mus musculus") doivent être entourés de guillemets.

On a trouvé 184 enregistrements publiés de genre *Scorpio*, avec 184 enregistrements avec séquences, formant 31 BIN (clusters), avec des spécimens de 4 pays, déposés dans 1 institut (Figure3).

Parmi ces enregistrements, 184 ont des noms d'espèces représentés en 6 espèces, qui sont :

- *Scorpio fuscus* : 68 séquences
- *Scorpio palmatus* : 58 séquences
- *Scorpio maurus* : 26 séquences
- *Scorpio propinquus* : 23 séquences
- *Scorpio Kruglovi* : 7 séquences
- *Scorpio fuliginosus* : 2 séquences

The screenshot shows the BOLD Systems search interface. At the top, there is a navigation bar with links for 'BASES DE DONNÉES', 'IDENTIFICATION', 'TAXONOMIE', 'TABLE DE TRAVAIL', 'RESSOURCES', and 'CONNEXION'. A search bar contains the text 'scorpio' and a 'RECHERCHE' button. To the right of the search bar, there are options for 'Spécimens' (DWC, XML, VST), 'Séquences' (FASTA, TRACE), 'Combiné' (XML, VST), and a 'Carte' dropdown menu set to 'Générer à partir de'. Below the search bar, the results are displayed in a list format. The first three results are for *Scorpio maure* with different identifiers (GACAC1124-12, GACO1801-19, GACO1802-19). Each result includes taxonomic information, identifiers, and deposit details. On the right side, a 'Résumé des résultats' section provides a summary: 'Trouvé 184 enregistrements publiés, avec 184 enregistrements avec séquences, formant 31 BIN (clusters), avec des spécimens de 4 pays, déposés dans 1 institut. Parmi ces enregistrements, 184 ont des noms d'espèces et représentent 6 espèces.'

Figure 3 : exemple sur le résultat obtenu dans la plateforme BOLDsystems

Tous les séquences du gène à fragmenter ont été téléchargées sur la base de données «BOLDsystems » sous format de fichier FASTA.

- Le format FASTA est utilisé pour stocker des séquences nucléiques ou protéiques, il est le format le plus utilisé dans le domaine de bioinformatique.

Dans notre travail, nous avons utilisés ce format pour télécharger et stocker les séquences nucléotidiques de genre *Scorpio*. Voici un exemple de l'une des séquences FASTA utilisées (figure 4) :

Entrez les séquences au format fasta dans l'orientation vers l'avant :

```
>GBA27939-15|Scorpio fuscus|COI-5P|KT188253
GACTATATATTTAATTTAGGTGGATGGGCTCAATGGTTGGGACTGCTTTGAGAATGATAATTCGGATTGAGACTGGGAGACTGGTTCTTTTATTGGTGATGATCAGCTTTATAATGTT
GTAGTAACAGCTCATGCTTTTGAATGATTTTTTTATGGTAATGCCTATTATGATTGGAGGTTTTGAAAATTGGTTGGCCCTTAATATTAGGAGCTCCGGATATGGCTTTTCCTCGTTTG
AATAATATGAGATTTTGGTTACTCCTCTGCTTTTTTTGCTTTAGGGTCTGCTGCTTTGAAAGAGGAGCTGGGACTGGGTGGACTGTTTATCCTCCTATCTTCTGGTATGTTTCAT
TCAGGTGGTCTGTAGATATGACTATTTTTCTTACATTTGGCTGGGGTATCATCAATTTAGGAGCTATTAATTTTACTACTATTGTAATATGCGTAGAGATGGAATGATTATGGAT
CGAGTTCCTTATTTGTATGGTCTGTTCTTATTACGGCTGTGTTGTTACTTCTTTTCCTGCTTTAGCTGGGGCTATTACTATGTTGTTGACTGACCGTAATTTAATACTCTTTTTTTG
ATCCTGCTGGTGGAGGGGATCCTATTTTGTATCAGCATTGTTTTGGTTTTTGGACACCCTGAGGTTTATATTTAATTTCTCCTGGGTTTGGGATGATTTCTCATATTGTAAGTCATCATA
CAGGAAAAAGGAGCCTTTTGGGACTATTGGGATGATTTATGCTATAGTATCTATTGGTTTTTGGGATTTGAGTATGGGCTCATATATTTACTGTGGGAATGGATGGGATACTCG
GGCTATTTTACGGCAGCTACTATGGTATTGCGGTGCCAAGTGGTATTAGGGTTTTAGATGAATTGCTACTATATATGGATCCCGAGTTTATTTTTCTCCACCTATGATGGGGCTGTAG
GATTTATTTTTTATTACTGTTGGTGGTTGACTGGTGAATGTTGGCTAATCCTCTTTGGATGAATTCCTCATGACTTATTATGTTGTT|
```

SOUMETTRE

Figure 4 : exemple d'une séquence utilisée sous format FASTA.

b. Outil d'identification sur BOLDsystems

Le système d'identification BOLD (**IDS**) pour COI accepte les séquences de la région 5' du gène mitochondrial de la sous-unité I du cytochrome c oxydase et renvoie une identification au niveau de l'espèce lorsque cela est possible. Chaque enregistrement de code à barres COI avec une identification au niveau de l'espèce et une longueur de séquence minimale de 500 bp. Cela comprend de nombreuses espèces représentées par seulement un ou deux spécimens ainsi que toutes les espèces avec une taxonomie provisoire (Sujevan et Hebert, 2007).

Pour utiliser l'outil d'identification afin d'identifier n'importe quel individu, il suffit de copier et coller sa séquence se format fasta dans un champ prévu à cet effet.

- Cet outil nous retourne ensuite un résultat comportant plusieurs éléments (ci-dessous) :
 - A. Il comprend également un graphique (**A**) qui montre les 20 premiers individus dont le pourcentage de similarité avec la séquence soumise est le plus élevé. La valeur de ce

pourcentage de similarité est présentée dans l'ordre. Ce graphique nous permet de voir si la séquence que nous avons soumise est déjà représentée dans la bibliothèque de référence ou non.

- B.** Un deuxième élément (B) indique les noms d'espèces éventuellement associés à chacun de ces individus de référence et leurs individus de référence ainsi que leur pourcentage de similarité avec la séquence soumise. Voilà, ce qui nous permettra d'attribuer un nom d'espèce à notre individu. Par exemple, nous pouvons noter ici que l'espèce *Scorpio fuscus* est un très bon candidat car elle présente une similarité totale (100%) entre le gène COI et la séquence que nous avons soumise expliqué dans (Figure 5).

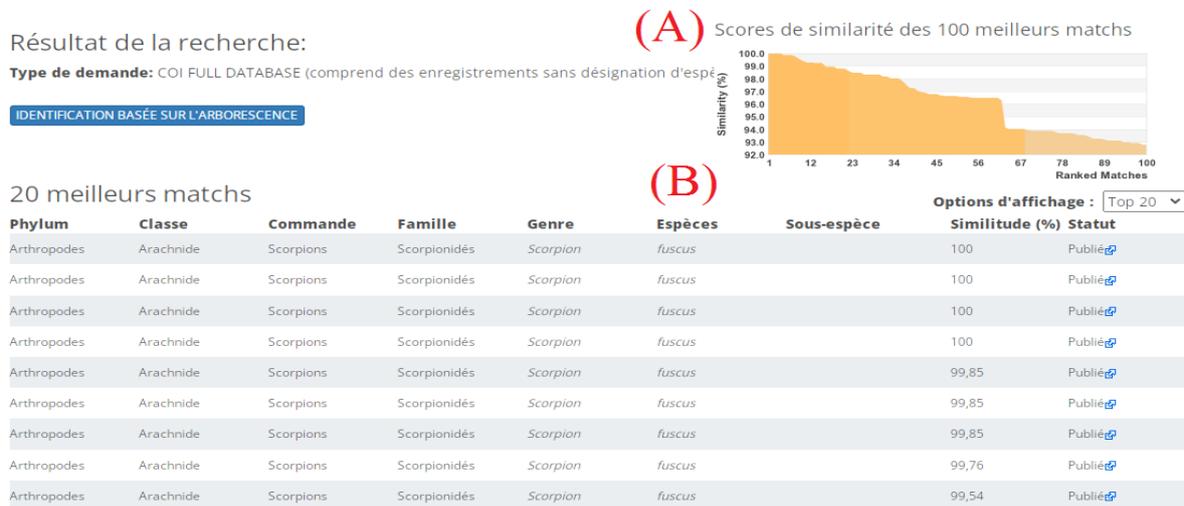


Figure 5 : exemple sur le résultat d'outil d'identification sur BOLDsystems.

3.2.2.2. GenBank

GenBank (<http://www.ncbi.nlm.nih.gov/>) est une base de données publique complète de séquences de nucléotides et d'annotations bibliographiques et biologiques. GenBank est construit et distribué par le National Center for Biotechnology Information (NCBI), une division de la National Library of Medicine (NLM), située sur le campus des National Institutes of Health (NIH) des États-Unis à Bethesda, MD, États-Unis. Il contient des séquences de nucléotides accessibles au public pour près de 260 000 espèces formellement décrites (Dennis *et al.*, 2017).

a. Identification et vérification des séquences sélectionnées sur GenBank

Afin de confirmer l'identification des toutes les séquences de genre étudiées, il faut vérifier l'origine géographique des espèces de *Scorpio* étudiées et leurs numéros d'accès GenBank de leurs séquences IDS.

Toutes les séquences nucléotidiques utilisées dans notre étude et leur localité ont été vérifiées dans le site web GenBank, Les vérification sont réalisées par copie et on colle le numéro d'accès GenBank dans un champ prévu à cet effet. Il est expliqué dans la figure suivante (figure 6) :



Figure 6 : exemple sur l'identification et la vérification des séquences sur GenBank.

Après l'identification et la vérification des séquences utilisées sur GenBank, nous avons tablé l'origine de pays (le nom de pays) de chaque séquence sur le format Fasta téléchargées. Pour faciliter la lecture des régions géographiques dans l'arbre phylogénétique à la fin.

b. Outil d'identification sur BLAST

L'outil de recherche d'alignement local de base (BLAST) est un programme de recherche de similarité de séquence. Le Center for Biotechnology Information (NCBI) gère un serveur BLAST avec une page d'accueil à l'adresse : <http://www.ncbi.nlm.nih.gov/BLAST/>. Gènes

Le programme compare les séquences de nucléotides ou de protéines aux séquences de la base de données et calcule la signification statistique des correspondances. BLAST trouve des zones de similitude locale entre les séquences biologiques et peut être utilisé pour déduire des relations fonctionnelles et évolutives entre des séquences ainsi que pour aider à identifier les membres de familles de gènes (Jian *et al.*, 2006).

c. Options du BLAST

BLAST est divisé à 4 colonnes principales qui sont présentées dans la figure suivante (figure 7):

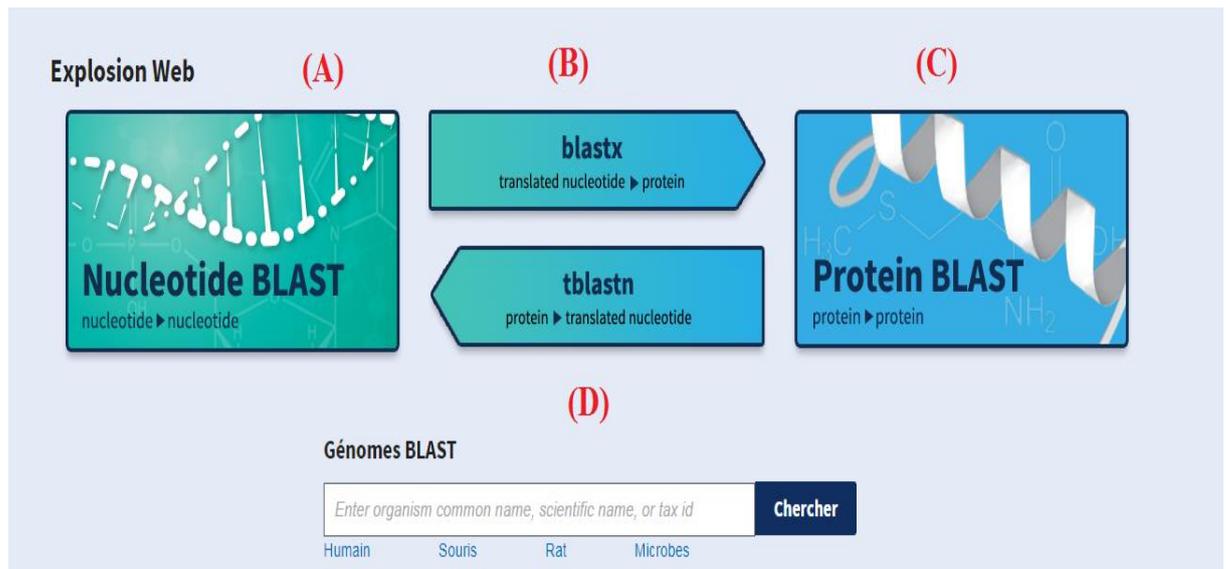


Figure 7 : Les 4 différents colonnes de comparaison dans la plateforme BLAST

(A). Compare une séquence nucléique à une banque nucléique (Blastn): utile pour étudier une séquence qui ne code pas une protéine, ou localiser un ARNm sur un génome.

- C'est l'outil qui nous avons choisie pour la comparaison de toutes les séquences nucléiques collectées dans cette étude.

(B). Compare une séquence nucléique traduite dans les 6 phases de lecture à une banque protéique: utile pour savoir si une séquence nucléique code une protéine et éventuellement localiser la partie codante.

(C). Compare une séquence protéique à une banque protéique (Blastp): recherche les homologues d'une protéine.

(D). Compare une séquence protéique à une banque nucléique traduite dans les 6 phases de lecture: utile pour identifier le gène et/ou l'ARNm qui code une protéine (Marzouk, 2018).

3.2.3. Exploitation des données

3.2.3.1. Alignement

L'alignement de séquences est une manière de dessiner deux ou plusieurs séquences (d'ADN, d'ARN ou de protéines) l'une en dessous de l'autre, afin de mettre en évidence les régions similaires. L'objectif principal de l'alignement est d'arranger les composés (nucléotides ou acides aminés) pour identifier les zones de concordance. Ces alignements sont réalisés par des programmes informatiques avec différentes méthodes dont l'objectif est de

maximiser le nombre de coïncidences entre les nucléotides ou les acides aminés des différentes séquences (Vincent *et al.*, 2005).

a. Principe

L'alignement des séquences (nucléotide ou protéine) d'ADN consiste à comparer une séquence d'ADN prélevée sur un individu à une ou plusieurs séquences identifiées, stockées dans une base de données dont les propriétés sont déjà connues (Vincent *et al.*, 2005 ; Jammali *et al.*, 2022)

b. Alignement multiple

Un MSA (Multiple Sequence Alignments) se construit grâce à un algorithme qui a pour but de maximiser le nombre de coïncidences entre nucléotides ou acides aminés des différentes séquences. Une matrice de distances entre les paires de séquences est ainsi obtenue. Celle-ci est alors utilisée pour calculer le score minimum ou idéal de chaque séquence présente dans le fichier d'entrée, qui est en général un fichier au format fasta, avec chacune des autres séquences présentes. L'idée est d'obtenir, en fichier de résultat, toutes nos séquences alignées entre elles de la meilleure façon possible (Pevzner, 2006).

Une représentation d'un ensemble de séquences, dans lesquelles les résidus équivalents (d'un point de vue fonctionnel ou structural) sont alignés en colonnes (Jammali *et al.*, 2022).

L'alignement multiple peut être réalisé grâce à plusieurs programmes de la bioinformatique comme : BioEdit, Multalin, Geneious, Emboss, ClustalW, ClustalΩ, Phylip, Méga,...etc.

Dans notre étude, on a utilisé la version 11 du logiciel Méga X téléchargeable à l'adresse www.mega.software.net/mega3/ (Kumar *et al.*, 2018), est qui revient au programme de création d'alignements multiples de séquences MUSCLE.

Le programme MEGA X version 11 (Molecular Evolutionary Genetics Analysis) a été utilisé pour l'alignement multiple des séquences nucléotidiques entre les différents taxons. C'est un logiciel libre de droit et téléchargeable à partir de plusieurs source d'internet disponible sur déférent moteur de recherche (Google, Yahoo...). Il propose deux algorithmes différents : ClustalW et MUSCLE (Hall , 2013).

c. Raffinement

Cette étape est facultative. Elle élimine les positions mal alignées et les régions divergentes (supprime le bruit d'alignement).

Après avoir terminé l'alignement multiple, nous passons à l'étape suivante qui est le raffinement "ou nettoyage des séquences" soit manuellement en utilisant le logiciel BIOEdit, soit automatiquement en ligne en utilisant l'une des méthodes disponibles sur NGphylogeny.fr.

NGphylogeny.fr Service web gratuit et simple d'utilisation dédié à la reconstruction et à l'analyse des relations phylogénétiques entre séquences moléculaires (figure 8).

Phylogeny.fr, créé en 2008, a été conçu pour faciliter l'exécution de flux de travail phylogénétiques, aujourd'hui, il est largement utilisé. Il intègre de nombreux outils dans leur dernière version (e.g. TNT, FastME, MrBayes, etc.) ainsi que de nouveaux outils conçus au cours des dix dernières années (e.g. PhyML, SMS, FastTree, trimAl, BOOSTER, etc.). Ces outils couvrent un large éventail d'utilisations (recherche de séquences, alignement de séquences multiples, sélection de modèles, inférence d'arbres et dessin d'arbres) et un large panel de méthodes standard (distance, parcimonie, maximum de vraisemblance et bayésien). Ils sont intégrés dans des flux de travail, qui ont déjà été configurés ('One click'), peuvent être personnalisés ('Advanced'), ou sont construits à partir de zéro ('A la carte'). NGPhylogeny.fr est accessible gratuitement sur l'adresse <https://ngphylogeny.fr> (Frédéric *et al.*, 2019).



Figure 8 : Service web NGphylogeny.fr

Dans notre travail nous avons utilisé la méthode Gblocks via le site web NGphylogeny.fr, c'est une méthode très facile et plus utilisée dans l'alignement des séquences nucléotidiques.

Gblocks est un programme informatique qui élimine les positions mal alignées et les régions divergentes d'un alignement de séquences d'ADN ou de protéines. Ces positions peuvent ne pas être homologues ou avoir été saturées par de multiples substitutions et il convient de les éliminer avant l'analyse phylogénétique. Il sélectionne les blocs de la même manière que cela se fait habituellement à la main, mais en suivant un ensemble reproductible de conditions. Il est très rapide dans le traitement des alignements et convient donc parfaitement aux analyses phylogénétiques à grande échelle (Van der *et al.*, 2021 ; Castresana, 2022).

The screenshot shows the NG Phylogeny.fr website. At the top, there is a navigation bar with links for 'Maison', 'Analyse de la phylogénie', 'Outils', 'Espace de travail', 'Documentation', 'À propos de', and 'Connexion'. Below the navigation bar, there is a section titled 'A La Carte' with a subtitle: 'À l'aide du créateur de flux de travail, personnalisez votre flux de travail en sélectionnant les bons outils'. Underneath, there is a 'Nom' section with a text input field labeled 'Nom du flux de travail..'. Below that is the 'Outils' section, which is divided into four categories: 1. Alignement multiple (MAFT, LE MUSCLE, Oméga clustal), 2. Curation d'alignement (BMGE, Gblocks, Bruyant, trimAl), 3. Inférence d'arbre (FastME, TNT, PhyML+SMS, PhyML, FastTree, Monsieur Bayes), and 4. Rendu d'arbre (Affichage Newick). The 'Gblocks' option in the 'Curation d'alignement' category is circled in red.

Figure 9 : Outil Gbloks sur Service web NGphylogeny.fr

3.2.3.2. Construction de l'arbre phylogénique

Après avoir discuté de l'alignement multiple (MSA) et de la correction de toutes les séquences étudiées, nous sommes passés à l'étape suivante : la construction d'un arbre phylogénétique. Les objectifs des arbres sont de visualiser les relations de parenté, et son un moyen de mettre en évidence l'évolution d'une ou plusieurs espèces. En effet, plusieurs méthodes de génération d'arbres utilisant différents algorithmes :

- la méthode UPGMA (*Unweighted Pair Groupe Method with Arithmetic Mean*)
- la méthode du Neighbour Joining (*Neighbor Joining- NJ*)
- la méthode du maximum de vraisemblance (*Maximum Likelihood- ML*)
- la méthode du maximum de parcimonie (*Maximum Parsimony*)

D'autres méthodes apparaissent régulièrement comme celles basées sur de l'apprentissage artificiel ou encore de l'inférence bayésienne comme dans le logiciel *MrBayes* qui est très utilisé. Néanmoins, les méthodes principalement utilisées restent la NJ et la ML.

Et c'est la méthode du maximum de vraisemblance (*Maximum Likelihood*- ML) qu'on à utilisée dans notre travail. Il s'agit d'une méthode probabiliste qui nécessite un modèle d'évolution. Le choix de ce modèle est crucial pour la qualité de l'arbre obtenu. On dit qu'il convient de l'utiliser à partir du moment où le nombre de caractères analysés est supérieur à la moitié du nombre de séquences analysées, sinon la reconstruction est considérée comme incorrecte. Elle est souvent décrite comme étant la meilleure méthode, c'est-à-dire la plus efficace pour trouver l'arbre le plus proche de la réalité. Son désavantage se situe au niveau des temps de calculs qui sont extrêmement longs (il m'est arrivé d'avoir des travaux tournant sur le cluster pendant plusieurs semaines pour des fichiers contenant plusieurs centaines de séquences) (Meyer, 1991 ; Gould et Pollock, 1997) <https://bioinfo-fr.net/alignements-multiples-quels-logiciels-choisir?hilité=alignement>.

L'analyse ML a été réalisée à l'aide du modèle Kimura-2 paramètres (K2P) avec l'option de suppression par paires. La prise en charge du bootstrap a été déduite à l'aide de l'algorithme de bootstrap rapide, basé sur 5 000 répliques. Les arbres phylogénétiques ont été visualisés à l'aide de et iTOL ver. 5 (Letunic et Bork, 2021).

3.2.3.3. Délimitation des espèces

La délimitation des espèces est le processus qui consiste à déterminer quels groupes d'organismes individuels constituent des populations différentes d'une même espèce et lesquels constituent des espèces différentes. Les données génomiques peuvent jouer un rôle important dans la délimitation des espèces dans le cadre de nombreux concepts d'espèces existants. La délimitation des espèces moléculaires a été réalisée par plusieurs méthodes différentes (BINs : Barcode Index Number ; ABGD : Automatic Barcode Gap Discovery ; ASAP : Assemble Species by Automatic Partitioning ; mPTP : Multi-rate Poisson Tree Processes ;). Toutes ces méthodes ont regroupé les séquences COI en unités taxonomiques opérationnelles (Operational Taxonomic Units : OTUs) en fonction de la similarité des séquences (Yang et Ziheng, 2020).

Les différentes méthodes de délimitation des espèces :

- ✓ BINs : Un outil d'attribution BIN (*Barcode Index Number*) sur le serveur BOLD en utilisant l'algorithme raffiné de liaison unique (RESL) (Ratnasingham et Hebert, 2013) http://www.boldsystems.org/index.php/Public_BarcodeIndexNumber_Home.
- ✓ Deux méthodes basées sur la distance :
 - ABGD (Automatic Barcode Gap Discovery ; Puillandre *et al.*, 2012).

- ASAP (Assemble Species by Automatic Partitioning ; Puillandre *et al.*, 2021).
- ✓ Trois méthodes basées sur un arbre phylogénétique :
 - GMYC (modèle Generalized Mixed Yule Coalescent; Pons *et al.*, 2006 ; Fujisawa et Barraclough, 2013).
 - bPTP (le modèle Bayesian Poisson Tree Processes ; Zhang *et al.*, 2013).
 - mPTP (le modèle multi-rate Poisson Tree Processes ; Kapli *et al.*, 2017).

Parmi les multiples méthodes de délimitation des espèces, la plus populaire est le modèle GMYC (Pons *et al.*, 2006 ; Fujisawa et Barraclough 2013).

Le modèle bPTP peut être effectués en ligne via le serveur : (<http://species.h-its.org/>)

De nombreux tests montrent que bPTP surpasse GMYC sur les données de simulation, et les résultats PTP sont comparables à GMYC sur des ensembles de données réels. <https://cme.h-its.org/exelixis/web/software/PTP/index.html>

Dans notre étude, nous avons utilisée trois méthodes : le mPTP, ASAP et BINs proposé par BOLDsystems comme des méthodes de délimitations des espèces.

a. modèle multi-rate Poisson Tree Processes (mPTP)

La méthode « Poisson Tree Processes » (PTP) est une approche sensible à la phylogénie, le PTP multi-débit (mPTP) est une méthode améliorée qui atténue les lacunes théoriques et techniques du PTP. Les résultats sur les données empiriques suggèrent que le mPTP est supérieur au PTP et aux méthodes populaires basées sur la distance, car il produit systématiquement des délimitations plus précises en ce qui concerne la taxonomie (Kapli *et al.*, 2017).

b. Barcode Index Numbers (BINs)

Le système "Barcode Index Numbers" est un cadre en ligne via BOLDsystems qui regroupe les séquences de codes à barres de manière algorithmique, ce système peut être utilisé pour vérifier les identifications des espèces ainsi que pour documenter la diversité lorsque les informations taxonomiques font défaut. Ce système se compose de trois parties :

1-Un algorithme de regroupement utilisant des méthodes théoriques des graphes pour générer des unités taxonomiques opérationnelles (OTUs) et des espèces putatives à partir de données de séquence sans informations taxonomiques préalables.

2-Un registre organisé de grappes de codes-barres intégré à une base de données en ligne de spécimens et de données taxonomiques avec prise en charge des annotations communautaires.

3-Un cadre d'annotation qui permet aux chercheurs d'examiner et de critiquer les identifications taxonomiques associées à chaque BIN et d'informer les propriétaires de données des erreurs.

L'algorithme BIN a été testé efficacement sur un large éventail de groupes taxonomiques et montre un potentiel d'applications dans les études sur l'abondance des espèces et les codes à barres.

https://v3.boldsystems.org/index.php/resources/handbook?chapter=2_databases.html§ion=bins

c. Assemble Species by Automatic Partitioning (ASAP)

Assemble Species by Automatic Partitioning c'est une méthode de construire des partitions d'espèces à partir d'alignements de séquences. ASAP est suffisamment efficace pour diviser des ensembles de données en tant que grandes séquences de 10^4 en espèces putatives en quelques minutes (Nicolas *et al.*, 2020).

Chapitre 4:

Résultats et discussion

Dans cette section, nous présentons l'ensemble de nos résultats obtenus, qui sont également présentés par un arbre phylogénétique délimité par des méthodes de délimitations des espèces.

4.1. Résultats

Nous avons étudié 106 séquences d'ADN du genre *Scorpio* téléchargées des bases de données par les méthodes des outils bioinformatiques.

4.1.1. Collection et vérification des séquences

Après l'identification des toutes les 184 séquences de genre *Scorpio* sur BOLDsystems et BLAST, nous avons trouvé qu'il y a des séquences mal identifiées. Par exemple, Il convient de noter ici que nous avons trouvé une séquence enregistrée comme espèce *S. maurus* dans la famille des Scorpionidae, alors qu'après la vérification dans l'outil d'identification sur BOLDsystems, nous avons confirmé qu'il s'agit d'une espèce très loin qui est (*Buthus occitanus* à 90.27%) appartient à la famille Buthidae» l'exemple sous-dessous (figure 10).

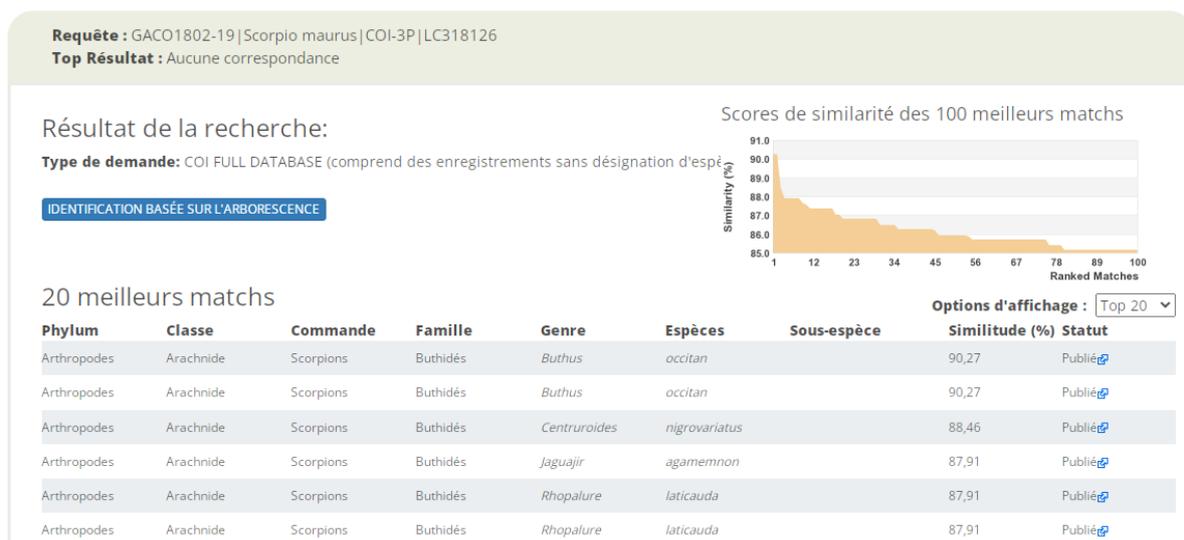


Figure 10 : Exemple sur une séquence mal identifiée prouvée par l'application d'identification de BOLDsystems

Afin de confirmer l'identification des 184 séquences du genre étudié, nous avons éliminé les 78 séquences suspectes et il ne reste que 106 séquences qui nous semblent bien identifiées. Ces séquences sont présentées dans le tableau ci-dessous avec leur taxonomie (Tableau 2) :

Tableau 2: Le nombre des séquences avant et après l'identification sur BOLD et BLAST.

Les espèces	Le nombre des séquences avant l'identification sur BOLD ET BLAST	Le nombre des séquences après l'identification sur BOLD ET BLAST
<i>Scorpio fuscus</i>	68	30
<i>Scorpio palmatus</i>	58	27
<i>Scorpio maurus</i>	26	22
<i>Scorpio propinquus</i>	23	18
<i>Scorpio kruglovi</i>	7	7
<i>Scorpio fuliginosus</i>	2	2
Total	184	106

Nous avons choisie une seule séquence comme out-groupe de genre déférent de la famille des Buthidae qui est *H. saulcyi* du genre *Hottentotta*. Ce genre le plus proche du genre étudiée *Scorpio*, pour enraciner nos arbres.

Les 106 séquences du genre *Scorpio* avec la séquence d'out-groupe et ses origines géographiques et leurs numéros d'accession GenBank de leurs séquences IDS, sont présentées sur le tableau suivant (Annexe : Tableau 3).

L'alignement multiple et le raffinement ont montré que les séquences différentes dans leur structure primaire et dans leurs longueurs. Ces résultats indiquent la présence de gaps dans les différentes séquences alignées.

4.1.2. Arbre phylogénétique

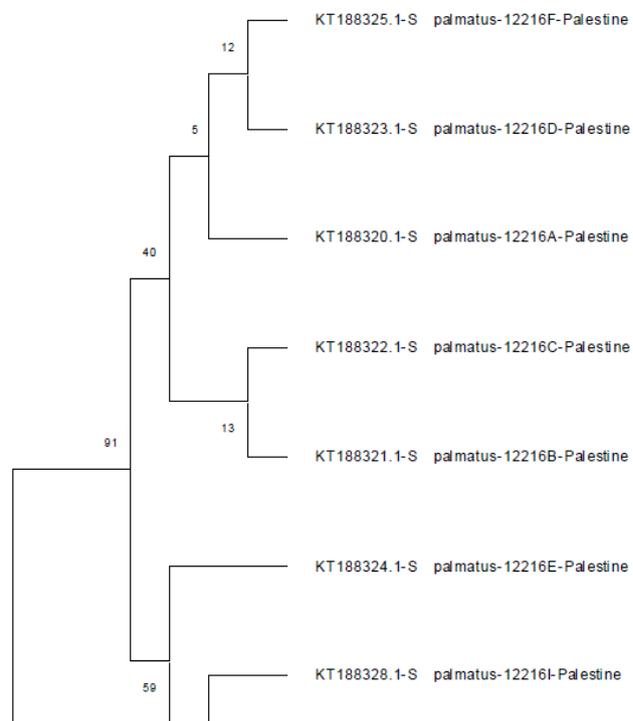
Nous avons construit un arbre phylogénétique dérivé de toutes les données nucléiques que nous avons collectées après la confirmation sur les outils des identifications BOLD et BLAST.

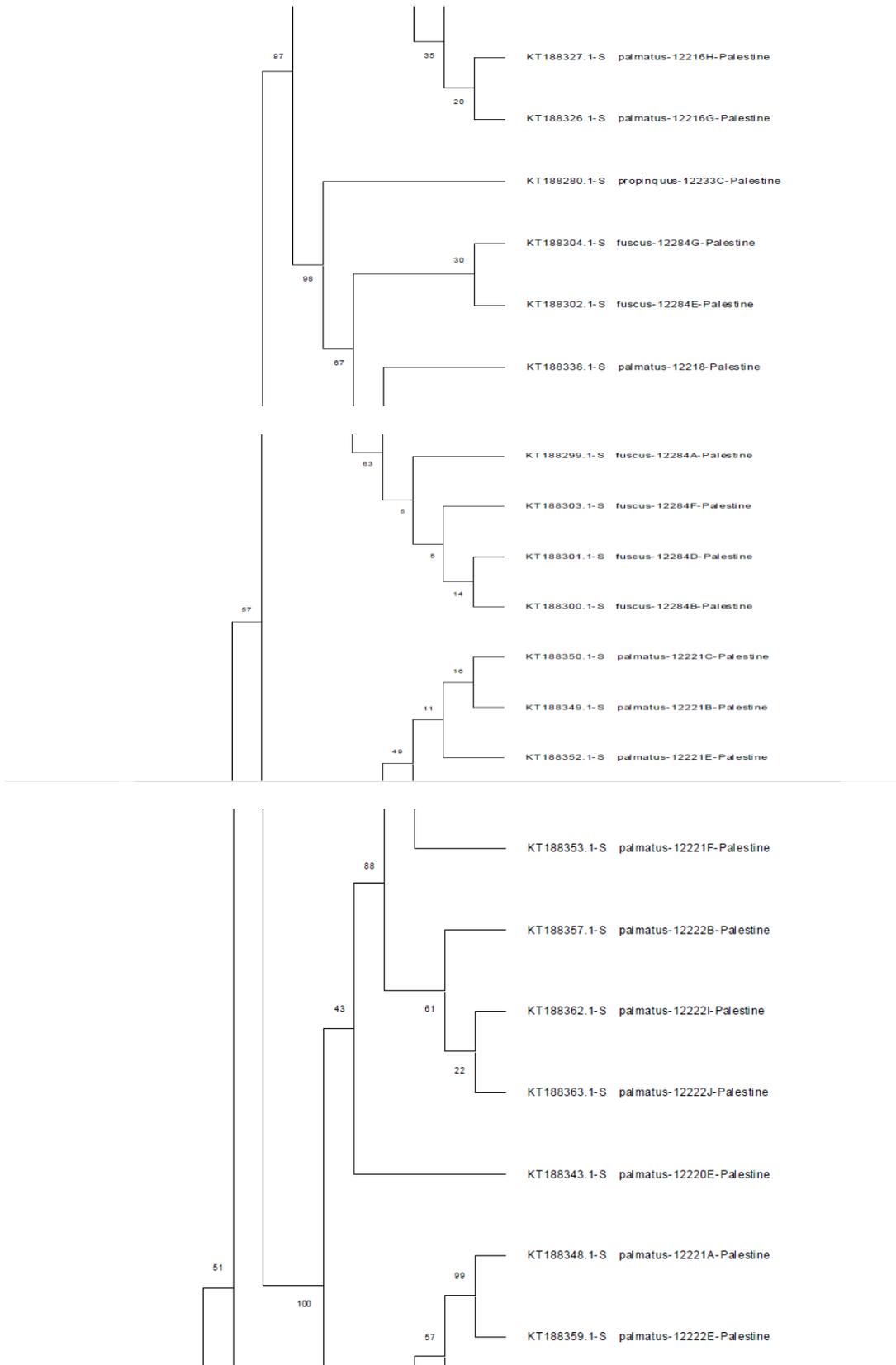
Reconstruction phylogénétique, l'arbre phylogénétique Maximum Likelihood (ML) a été reconstruit par MEGA X version 11 (Tamura et Nei, 1993 ; Kumar *et al.*, 2018), et le site web NGphylogeny.fr (Junier *et al.*, 2010 ; Guindon *et al.*, 2010 ; Lemoine *et al.*, 2018 ; Lemoine *et al.*, 2019).

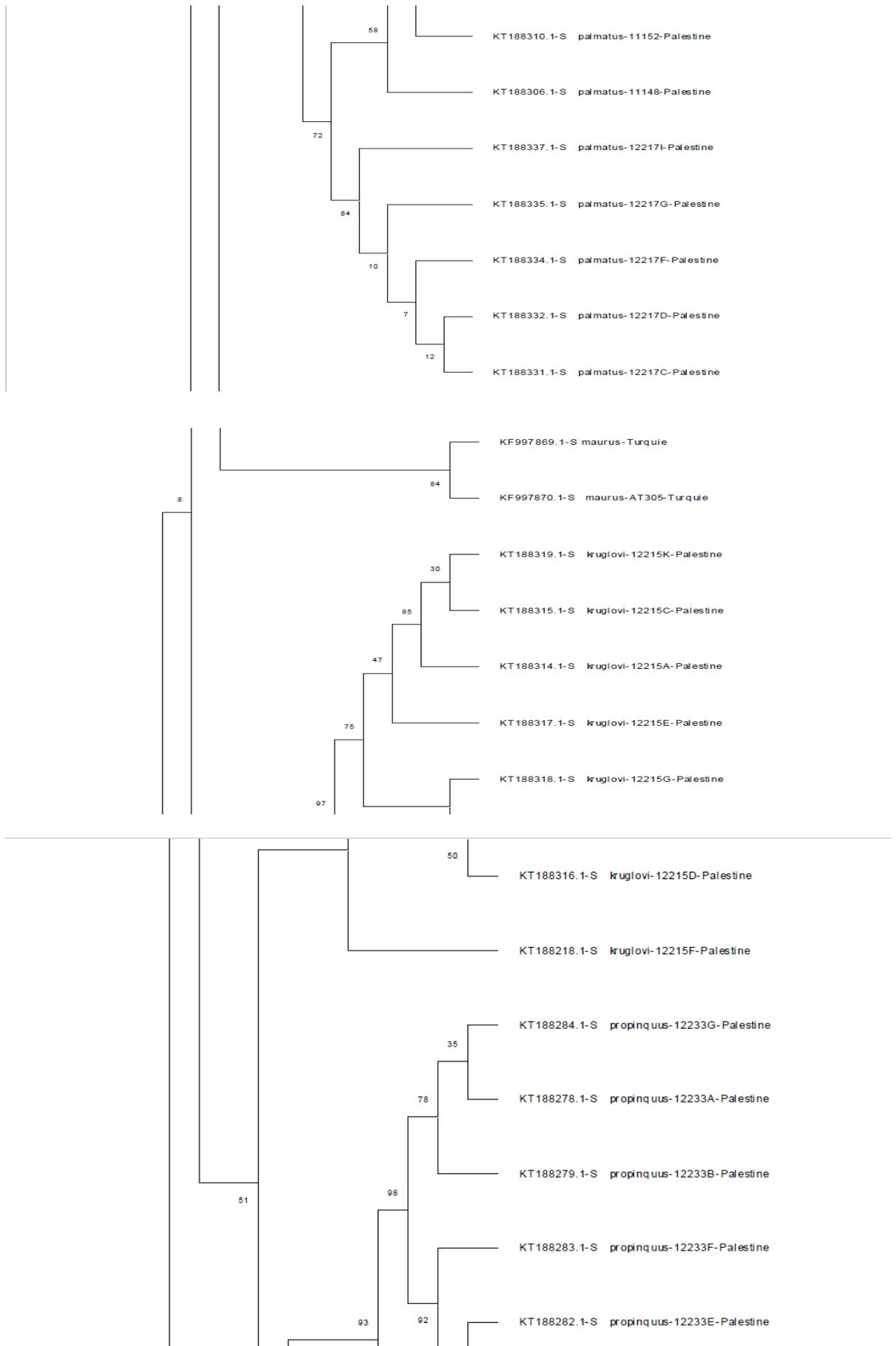
Malgré les deux arbres phylogénétique sont réalisés avec l'algorithme ML l'une avec le logiciel MEGA X version 11 (figure 11) et l'autre avec le site NGPhylogeny.fr (figure 12) on à constater que les clades sont presque identiques.

Analyse évolutive par la méthode du maximum de vraisemblance

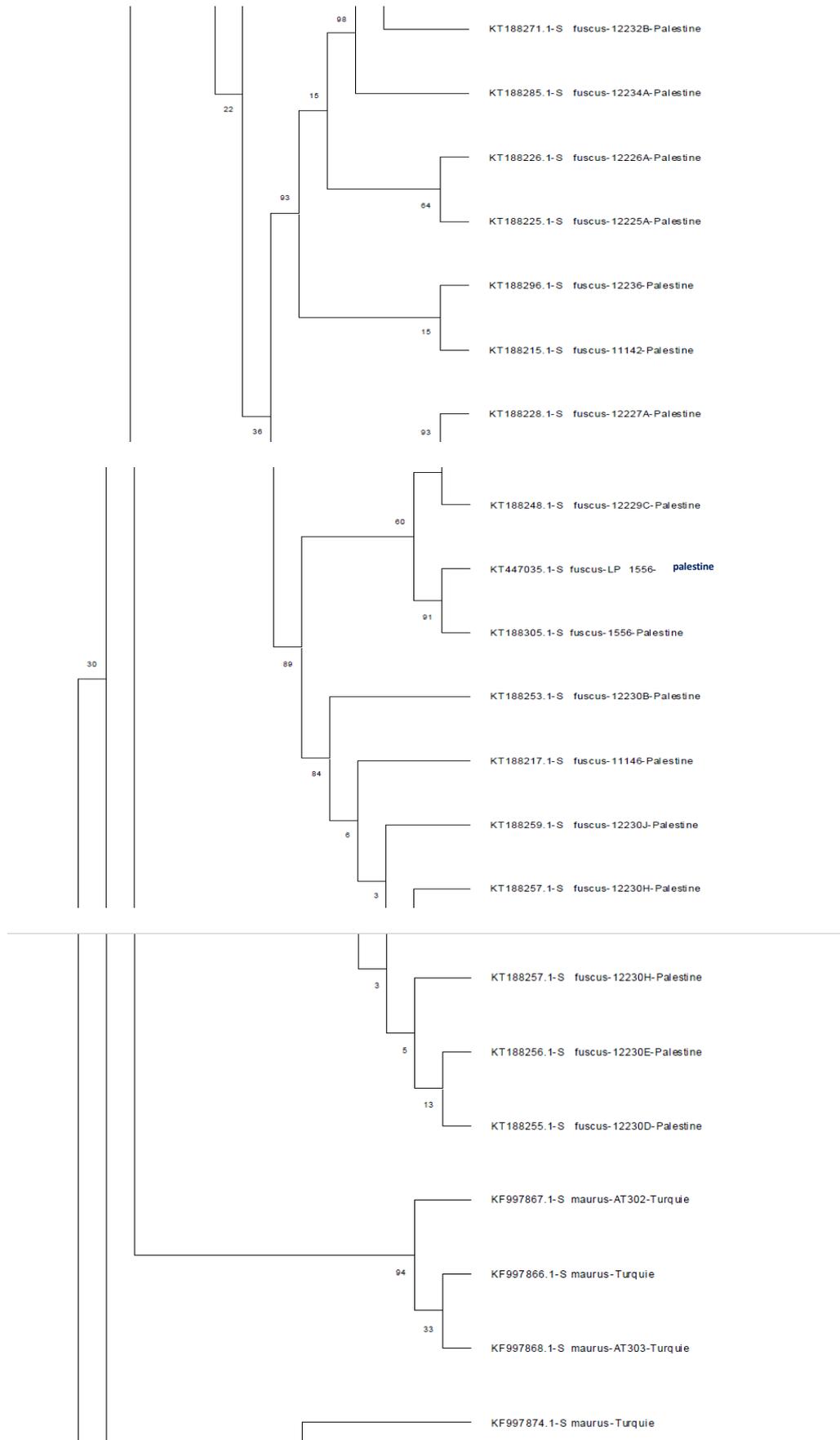
L'histoire de l'évolution a été déduite en utilisant la méthode du maximum de vraisemblance et le modèle Tamura-Nei (Tamura et Nei, 1993). L'arbre présentant la plus forte vraisemblance (-3809,46) est présenté. Le pourcentage d'arbres dans lesquels les taxons associés se regroupent est indiqué à côté des branches. Le ou les arbres initiaux pour la recherche heuristique ont été obtenus automatiquement en appliquant les algorithmes Neighbor-Joining et BioNJ à une matrice de distances par paires estimées à l'aide de l'approche MCL (Maximum Composite Likelihood), puis en sélectionnant la topologie présentant la valeur de vraisemblance logarithmique supérieure. Cette analyse a porté sur 107 séquences nucléotidiques. Les positions de codons incluses étaient les suivantes : 1er+2e+3e+non codant. Il y avait un total de 518 positions dans le jeu de données final. Les analyses évolutives ont été réalisées dans MEGA X (Kumar *et al.*, 2018).











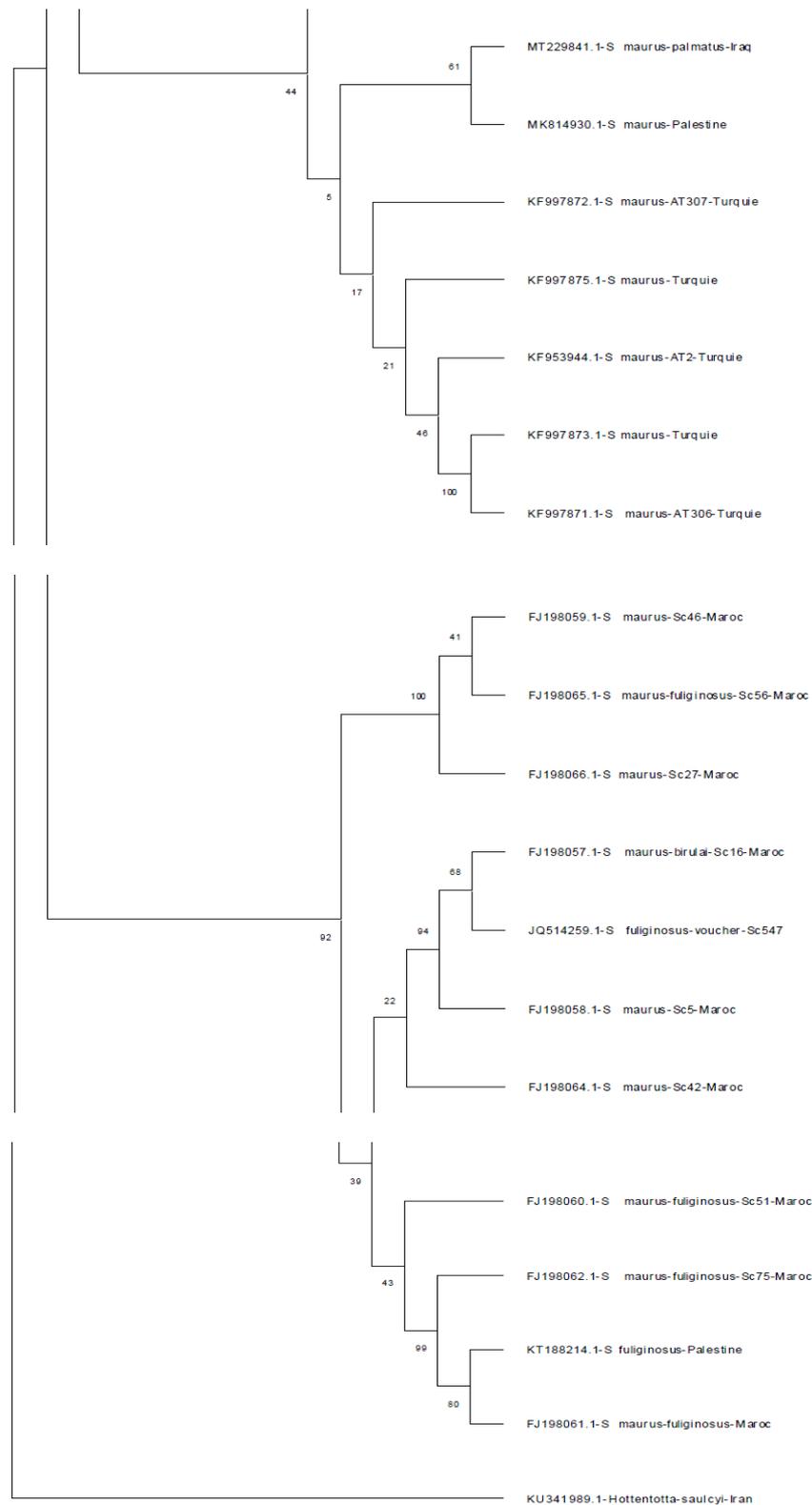


Figure 11 : Phylogramme de vraisemblance maximale ML construit par logiciel MEGA X (Tamura et Nei, 1993 ; Kumar *et al.*, 2018), déduit pour les spécimens du genre *Scorpio*. Les nombres sur les branches représentent le support bootstrap. L'arbre est enraciné sur *Hottentotta saulcyi* comme out groupe.

4.1.3. Résultat de la délimitation des espèces avec les trois méthodes

De plus, nous avons utilisé trois outils différents de délimitation des espèces (BINs disponible sur BOLDsystems, mPTP, ASAP). Afin d'estimer le nombre de mOTUs et de faire correspondre les identifications morphologiques des espèces avec les délimitations génétiques (figure 12).

Les 107 séquences de 6 espèces ciblées dans notre travail, la méthode de délimitations des espèces mPTP montrent que l'arbre se divise en 9 groupes principaux avec le 10^{ème} groupe qui est l'out-Groupe *Hottentotta saulcyi*. Par contre la méthode de délimitations des espèces proposée par la base de données BOLD montre une répartition des séquences en 30 groupes (BINs) principaux avec l'out-Groupe, ainsi que la méthode de délimitations des espèces ASAP a regroupées l'arbre en 31 groupes principaux avec l'out-Groupe.

4.1.3.1. Résultat du la méthode de délimitation des espèces (ASAP)

Partition 1

Score: 1

Proba: 6.196841e-02

Nombre des Groupes: **31** (29)

- **Groupe [1]:** nombre des séquences: 3

-id:FJ198066.1 ; *S. maurus*-Sc27-Maroc ; FJ198059.1-*S. maurus* -Sc46-Maroc ; FJ198065.1-*S. maurus fuliginosus*-Sc56-Maroc.

- **Groupe [2] :** nombre des séquences: 2

-id:FJ198058.1-*S. maurus* -Sc5-Maroc ; FJ198057.1-*S. maurus -birulai*-Sc16-Maroc.

- **Groupe [3] :** nombre des séquences: 1

-id:FJ198064.1-*S. maurus* -Sc42-Maroc.

- **Groupe [4] :** nombre des séquences: 1

-id:FJ198060.1-*S. maurus -fuliginosus*-Sc51-Maroc.

- **Groupe [5] :** nombre des séquences: 3

-id:FJ198062.1-*S. maurus fuliginosus*-Sc75-Maroc ; KT188214.1-*S. fuliginosus*-Palestine ; FJ198061.1-*S. maurus -fuliginosus*-Maroc.

- **Groupe [6] :** nombre des séquences: 1

-id:KF997874.1-*S. maurus* -Turquie

- **Groupe [7]** : nombre des séquences: 1

-id:KF997866.1-*S. maurus* -Turquie.

- **Groupe [8]** : nombre des séquences: 1

-id:KF997875.1-*S.maurus*-Turquie.

- **Groupe [9]** : nombre des séquences: 1

-id:KF997869.1-*S. maurus* -Turquie

- **Groupe [10]** : nombre des séquences: 2

-id:KF997873.1-*S. maurus* –Turquie ; KF997871.1-*S. maurus* -AT306-Turquie.

- **Groupe [11]** : nombre des séquences: 1

-id:KF997868.1-*S. maurus* -AT303-Turquie.

- **Groupe [12]** : nombre des séquences: 3

-id:KT188293.1-*S. fuscus* -12235A-Palestine ; KT188295.1-*S. fuscus* -12235C-Palestine ; KT188294.1-*S. fuscus* -12235B-Palestine

- **Groupe [13]** : nombre des séquences: 10

-id:KT188253.1-*S. fuscus* -12230B-Palestine; KT188217.1-*S. fuscus* -11146-Palestine; KT188259.1-*S. fuscus* -12230J-Palestine; KT188257.1-*S. fuscus* -12230H-Palestine; KT188256.1-*S. fuscus* -12230E-Palestine; KT188255.1-*S. fuscus* -12230D-Palestine; KT447035.1-*S. fuscus* -LP_1556-Palestine; KT188305.1-*S. fuscus* -1556-Palestine; KT188228.1-*S. fuscus* -12227A-Palestine; KT188248.1-*S. fuscus* -12229C-Palestine

- **Groupe [14]** : nombre des séquences: 9

-id:KT188320.1-*S. plamatus* -12216A-Palestine ; KT188325.1-*S. plamatus* -12216F-Palestine ; KT188323.1-*S. plamatus* -12216D-Palestine ; KT188322.1-*S. plamatus* -12216C-Palestine ; KT188321.1-*S. plamatus* -12216B-Palestine ; KT188328.1-*S. plamatus* -12216I-Palestine ; KT188327.1-*S. plamatus* -12216H-Palestine ; KT188326.1-*S. plamatus* -12216G-Palestine ; KT188324.1-*S. plamatus* -12216E-Palestine.

- **Groupe [15]** : nombre des séquences : 8

-id:KT188338.1-*S. plamatus* -12218-Palestine ; KT188299.1-*S. fuscus* -12284A-Palestine ; KT188303.1-*S. fuscus* -12284F-Palestine ; KT188301.1-*S. fuscus* -12284D-Palestine ; KT188300.1-*S. fuscus* -12284B-Palestine ; KT188304.1-*S. fuscus* -12284G-Palestine ; KT188302.1-*S. fuscus* -12284E-Palestine ; KT188280.1-*S. propinquus*-12233C-Palestine

- **Groupe [16]** : nombre des séquences: 1

-id:JQ514259.1-*S. fuliginosus*-voucher-Sc547

- **Groupe [17]** : nombre des séquences: 1

-id:KT188297.1-*S. fuscus* -12237-Palestine.

- **Groupe [18]** : nombre des séquences: 1

-id:MT229841.1-*S. maurus -palmatus*-Iraq.

- **Groupe [19]** : nombre des séquences: 7

-id:KT188314.1-*S. kruglovi*-12215A-Palestine ; KT188319.1-*S. kruglovi*-12215K-Palestine ; KT188315.1-*S. kruglovi*-12215C-Palestine ; KT188317.1-*S. kruglovi*-12215E-Palestine ; KT188318.1-*S. kruglovi*-12215G-Palestine ; KT188316.1-*S. kruglovi*-12215D-Palestine ; KT188218.1-*S. kruglovi*-12215F-Palestine.

- **Groupe [20]** : nombre des séquences: 1

-id:KT188219.1-*S. fuscus* -12223A-Palestine

- **Groupe [21]** : nombre des séquences: 4

-id:KT188296.1-*S. fuscus*-12236-Palestine ; KT188215.1-*S. fuscus*-11142-Palestine ; KT188226.1-*S. fuscus*-12226A-Palestine ; KT188225.1-*S. fuscus* -12225A-Palestine.

- **Groupe [22]** : nombre des séquences: 5

-id:KT188287.1-*S. fuscus*-12234C-Palestine ; KT188289.1-*S. fuscus*-12234E-Palestine ; KT188274.1-*S. fuscus*-12232E-Palestine ; KT188285.1-*S. fuscus*-12234A-Palestine ; KT188271.1-*S. fuscus*-12232B-Palestine.

- **Groupe [23]** : nombre des séquences: 6

-id:KT188279.1-*S. propinquus*-12233B-Palestine ; KT188284.1-*S. propinquus*-12233G-Palestine ; KT188278.1-*S. propinquus*-12233A-Palestine ; KT188283.1-*S.*

propinquus-12233F-Palestine ; KT188282.1-*S. propinquus*-12233E-Palestine ; KT188281.1-*S. propinquus*-12233D-Palestine.

- **Groupe [24]** : nombre des séquences: 1

-id:KF953944.1-*S. maurus* -AT2-Turquie.

- **Groupe [25]** : nombre des séquences: 1

-id:KF997867.1-*S. maurus* -AT302-Turquie.

- **Groupe [26]** : nombre des séquences: 10

-id:KT188261.1-*S. propinquus*-12231B-Palestine ; KT188264.1-*S. propinquus*-12231E-Palestine ; KT188263.1-*S. propinquus*-12231D-Palestine ; KT188268.1-*S. propinquus*-12231I-Palestine ; KT188267.1-*S. propinquus*-12231H-Palestine ; KT188266.1-*S. propinquus*-12231G-Palestine ; KT188262.1-*S. propinquus*-12231C-Palestine ; KT188265.1-*S. propinquus*-12231F-Palestine ; KT188260.1-*S. propinquus*-12231A-Palestine ; KT188298.1-*S. propinquus*-12241-Palestine.

- **Groupe [27]** : nombre des séquences: 1

-id:KF997872.1-*S. maurus* -AT307-Turquie.

- **Groupe [28]** : nombre des séquences: 1

-id:KT188222.1-*S. propinquus*-12224C-Palestine

- **Groupe [29]** : nombre des séquences: 1

-id:MK814930.1-*S. maurus* -Iran.

- **Groupe [30]** : nombre des séquences: 17

-id:KT188310.1-*S. plamatus* -11152-Palestine ; KT188306.1-*S. plamatus* -11148-Palestine ; KT188337.1-*S. plamatus* -12217I-Palestine ; KT188335.1-*S. plamatus* -12217G-Palestine ; KT188334.1-*S. plamatus*-12217F-Palestine ; KT188332.1-*S. plamatus* -12217D-Palestine ; KT188331.1-*S. plamatus* -12217C-Palestine ; KT188348.1-*S. plamatus* -12221A-Palestine ; KT188359.1-*S. plamatus* -12222E-Palestine ; KT188357.1-*S. plamatus* -12222B-Palestine ; KT188362.1-*S. plamatus* -12222I-Palestine ; KT188363.1-*S. plamatus* -12222J-Palestine ; KT188353.1-*S. plamatus* -12221F-Palestine ; KT188352.1-*S. plamatus* -12221E-Palestine ; KT188350.1-*S. plamatus* -12221C-Palestine ; KT188349.1-*S. plamatus* -12221B-Palestine ; KT188343.1-*S. plamatus* -12220E-Palestine

- **Groupe [31]** : nombre des séquences: 1

-id: KF997870.1-*S. maurus* -AT305-Turquie.

4.1.3.2. Résultat de la méthode de délimitation des espèces (mPTP)

Commande : /bin/mptp mptp --ml --multi --outgroup KU341989-1-*Hottentotta-saulcyi*-Iran --outgroup_crop --tree_file ../uploads/qo09j0a82g1slll482tse4v673.newick --output_file ../uploads/qo09j0a82g1slll482tse4v673.

1 Nombre d'arêtes supérieures à la longueur minimale de la branche : 101 / 210

Score du modèle nul : 312.008192

Meilleur score pour le taux de coalescence multiple : 312.008192

Nombre d'espèces délimitées : 9

- **Groupe [1]** : nombre des séquences: 24

KT188295-1-*S. fuscus*-12235C-Palestine ; KT188294-1-*S. fuscus*-12235B-Palestine ;
 KT188293-1-*S. fuscus*-12235A-Palestine ; KT188297-1-*S. fuscus*-12237-Palestine ;
 KT188219-1-*S. fuscus*-12223A-Palestine ; KT188215-1-*S. fuscus*-11142-Palestine ;
 KT188296-1-*S. fuscus*-12236-Palestine ; KT188226-1-*S. fuscus*-12226A-Palestine ;
 KT188225-1-*S. fuscus*-12225A-Palestine ; KT188271-1-*S. fuscus*-12232B-Palestine ;
 KT188285-1-*S. fuscus*-12234A-Palestine ; KT188289-1-*S. fuscus*-12234E-Palestine ;
 KT188274-1-*S. fuscus*-12232E-Palestine ; KT188287-1-*S. fuscus*-12234C-Palestine ;
 KT188248-1-*S. fuscus*-12229C-Palestine ; KT188228-1-*S. fuscus*-12227A-Palestine ;
 KT188305-1-*S. fuscus*-1556-Palestine ; KT447035-1-*S. fuscus*-LP_1556-Palestine ;
 KT188256-1-*S. fuscus*-12230E-Palestine ; KT188257-1-*S. fuscus*-12230H-Palestine ;
 KT188259-1-*S. fuscus*-12230J-Palestine ; KT188217-1-*S. fuscus*-11146-Palestine ;
 KT188255-1-*S. fuscus*-12230D-Palestine ; KT188253-1-*S. fuscus*-12230B-Palestine

- **Groupe [2]** : nombre des séquences: 7

KT188218-1-*S. kruglovi*-12215F-Palestine ; KT188316-1-*S. kruglovi*-12215D-Palestine
 ; KT188318-1-*S. kruglovi*-12215G-Palestine ; KT188317-1-*S. kruglovi*-12215E-Palestine ;
 KT188319-1-*S. kruglovi*-12215K-Palestine ; KT188315-1-*S. kruglovi*-12215C-Palestine ;
 KT188314-1-*S. kruglovi*-12215A-Palestine.

- **Groupe [3]** : nombre des séquences: 7

KT188222-1-*S. propinquus*-12224C-Palestine ; KT188284-1-*S. propinquus*-12233G-Palestine ; KT188278-1-*S. propinquus*-12233A-Palestine ; KT188279-1-*S. propinquus*-12233B-Palestine ; KT188282-1-*S. propinquus*-12233E-Palestine ; KT188281-1-*S. propinquus*-12233D-Palestine ; KT188283-1-*S. propinquus*-12233F-Palestine

Groupe [4] : nombre des séquences: 10

KT188298-1-*S. propinquus*-12241-Palestine ; KT188264-1-*S. propinquus*-12231E-Palestine ; KT188263-1-*S. propinquus*-12231D-Palestine ; KT188261-1-*S. propinquus*-12231B-Palestine ; KT188260-1-*S. propinquus*-12231A-Palestine ; KT188265-1-*S. propinquus*-12231F-Palestine ; KT188266-1-*S. propinquus*-12231G-Palestine ; KT188267-1-*S. propinquus*-12231H-Palestine ; KT188262-1-*S. propinquus*-12231C-Palestine ; KT188268-1-*S. propinquus*-12231I-Palestine.

- **Groupe [5] :** nombre des séquences: 20

KF997866-1-*S. maurus*-Turquie ; KF997867-1-*S. maurus*-AT302-Turquie ; KF997868-1-*S. maurus*-AT303-Turquie ; KF997875-1-*S. maurus*-Turquie ; KF953944-1-*S. maurus*-AT2-Turquie ; KF997871-1-*S. maurus*-AT306-Turquie ; KF997873-1-*S. maurus*-Turquie ; KF997874-1-*S. maurus*-Turquie ; MT229841-1-*S. maurus-palmatus*-Iraq ; MK814930-1-*S. maurus*-Palestine ; KF997872-1-*S. maurus*-AT307-Turquie ; FJ198066-1-*S. maurus*-Sc27-Maroc ; FJ198059-1-*S. maurus*-Sc46-Maroc ; FJ198065-1-*S. maurus-fuliginosus*-Sc56-Maroc ; FJ198060-1-*S. maurus-fuliginosus*-Sc51-Maroc ; FJ198062-1-*S. maurus-fuliginosus*-Sc75-Maroc ; FJ198061-1-*S. maurus-fuliginosus*-Maroc ; KT188214-1-*S. fuliginosus*-Palestine ; FJ198064-1-*S. maurus*-Sc42-Maroc ; FJ198058-1-*S. maurus*-Sc5-Maroc ; FJ198057-1-*S. maurus-birulai*-Sc16-Maroc ; JQ514259-1-*S. fuliginosus*-voucher-Sc547

- **Groupe [6] :** nombre des séquences: 2

KF997870-1-*S. maurus*-AT305-Palestine ; KF997869-1-*S. maurus*-Turquie

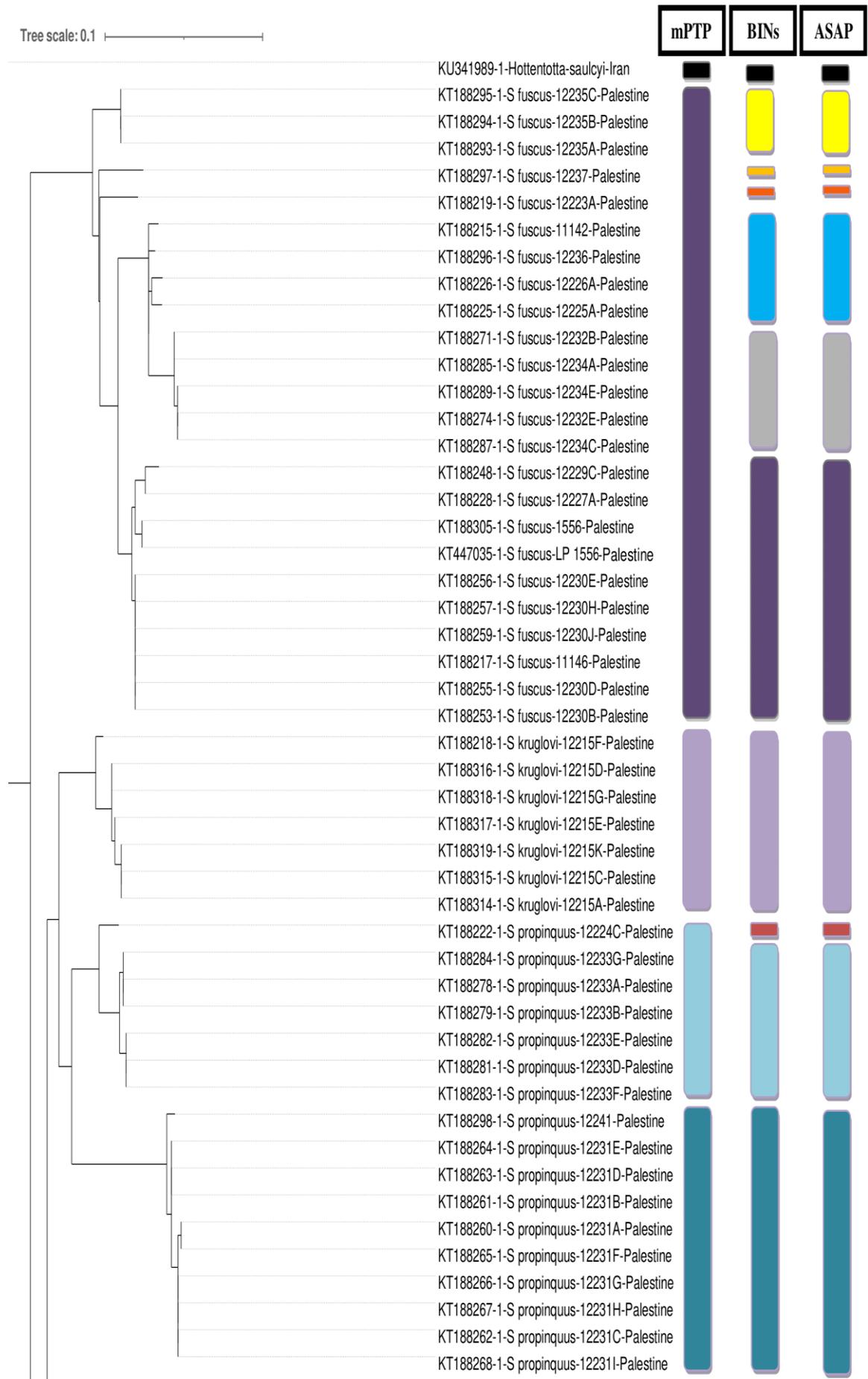
- **Groupe [7] :** nombre des séquences: 8

KT188280-1-*S. propinquus*-12233C-Palestine ; KT188302-1-*S. fuscus*-12284E-Palestine ; KT188304-1-*S. fuscus*-12284G-Palestine ; KT188301-1-*S. fuscus*-12284D-Palestine ; KT188303-1-*S. fuscus*-12284F-Palestine ; KT188299-1-*S. fuscus*-12284A-Palestine ; KT188300-1-*S. fuscus*-12284B-Palestine ; KT188338-1-*S. palmatus*-12218-Palestine

- **Groupe [8] :** nombre des séquences: 9

KT188324-1-S. *palmatus*-12216E-Palestine ; KT188327-1-S. *palmatus*-12216H-Palestine ; KT188326-1-S. *palmatus*-12216G-Palestine ; KT188328-1-S. *palmatus*-12216I-Palestine ; KT188322-1-S. *palmatus*-12216C-Palestine ; KT188323-1-S. *palmatus*-12216D-Palestine ; KT188325-1-S. *palmatus*-12216F-Palestine ; KT188321-1-S. *palmatus*-12216B-Palestine ; KT188320-1-S. *palmatus*-12216A-Palestine

Groupe [9] : nombre des séquences: 17 ; KT188343-1-S. *palmatus*-12220E-Palestine ; KT188362-1-S. *palmatus*-12222I-Palestine ; KT188363-1-S. *palmatus*-12222J-Palestine ; KT188357-1-S. *palmatus*-12222B-Palestine ; KT188350-1-S. *palmatus*-12221C-Palestine ; KT188352-1-S. *palmatus*-12221E-Palestine ; KT188349-1-S. *palmatus*-12221B-Palestine ; KT188353-1-S. *palmatus*-12221F-Palestine ; KT188306-1-S. *palmatus*-11148-Palestine ; KT188310-1-S. *palmatus*-11152-Palestine ; KT188359-1-S. *palmatus*-12222E-Palestine ; KT188348-1-S. *palmatus*-12221A-Palestine ; KT188332-1-S. *palmatus*-12217D-Palestine ; KT188334-1-S. *palmatus*-12217F-Palestine ; KT188335-1-S. *palmatus*-12217G-Palestine ; KT188331-1-S. *palmatus*-12217C-Palestine ; KT188337-1-S. *palmatus*-12217I-Palestine



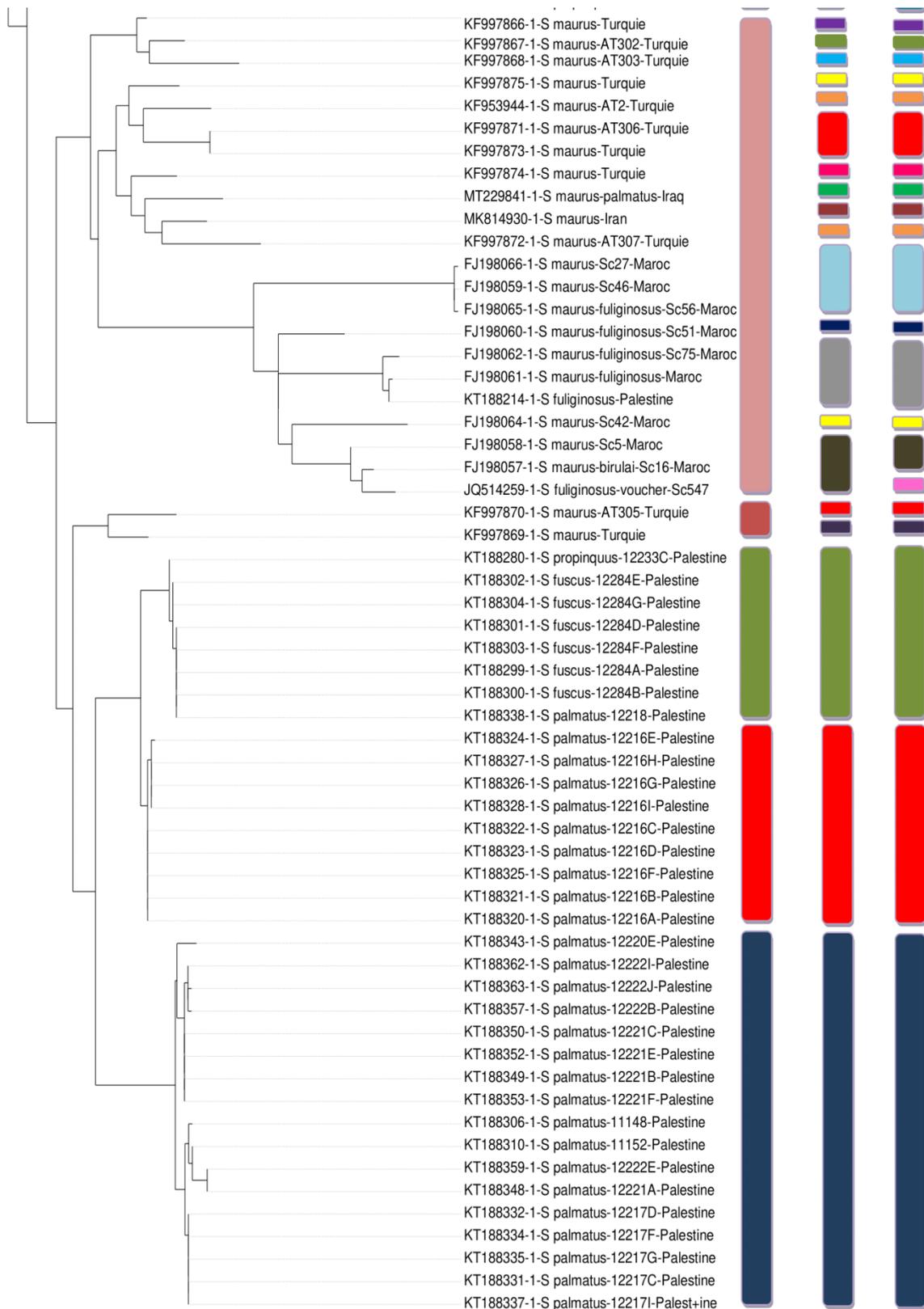


Figure 12 : Comparaison des résultats de la délimitation des 6 espèces de *Scorpio* à l'aide de deux méthodes différentes et comparaison avec les BINs, l'arbre construit par ML sur le serveur web NGphylogeny.fr (Junier *et al.*, 2010 ; Guindon *et al.*, 2010 ; Lemoine *et al.*, 2018 ; Lemoine *et al.*, 2019).

- Pour BINs:
http://www.boldsystems.org/index.php/Public_BarcodeIndexNumber_Home

4.2. Discussion

Cette partie discute les analyses bioinformatiques réalisées pour estimer la richesse spécifique du genre *Scorpio* présentée au sein des séquences qui nous avons collectées. Nous ont donné de nombreux résultats que nous avons tenté de les interpréter ici.

En tant qu'outil rapide et précis pour distinguer et identifier les espèces, l'ADN code barres été largement utilisé dans de nombreuses disciplines depuis qu'il a été proposé, en particulier dans la classification des animaux, ce qui améliore considérablement l'efficacité et la précision de la classification des espèces (Li *et al.*, 2018).

Notre étude a fourni des preuves d'un nombre substantiel d'espèces négligées d'importance dans l'identification moléculaire chez le genre *Scorpio*.

L'apparition de l'espèce *H. saulcyi* qui s'appartienne au genre *Hottentotta* dans un clade séparé dans toutes les méthodes de délimitation des espèces qu'on a réalisées (mPTP, BINs et ASAP), confirme notre bon choix de cette espèce comme out groupe pour enraciner nos arbres.

Pour les 6 espèces analysées, la délimitation moléculaire ne correspondait pas à l'identification morphologique, L'ensemble de ces données a permis de construire un arbre phylogénétique (Figure 12) qui confirme la paraphylie du genre *Scorpio*. Ainsi que les serveurs mPTP, ASAP et BINs proposé par BOLD sont utilisés comme outil de délimitation des espèces pour estimer le nombre de mOTU et faire correspondre l'identification morphologique des espèces avec 106 séquences génétiques. En conséquence, l'algorithme mPTP a identifié 9 mOTU, tandis que BINs a révélé 30 mOTU, outre que ASAP montrent que l'arbre se divisés en 31 mOTU. Ce qui est utilisé comme estimation la plus conservatrice pour la discussion où certains mOTU ont été fusionnés ou divisés par rapport aux 9 mOTU de mPTP. La différence entre les trois méthodes est due à l'algorithme dans lequel ils travaillent ces méthodes.

Pour mPTP, comme dans l'analyse phylogénétique, le 1^{er} clade de *S. fuscus* (Groupe 1) regroupé en grand groupe de 24 séquences qui représentent une série de 24 espèces sœurs largement distribuées en s'étendant sur Palestine. Par contre, BINs et ASAP se divisent le clade en 6 groupes.

Bien que le 2^{ème} clade de *S. kruglovi* (Groupe 2) qui représente 7 séquences distribuées en Palestine. Dans les trois méthodes de délimitations (mPTP, BINs et ASAP), qui confirme que le clade de cette espèce est monophylique.

Ensuite, les 3^{ème} et 4^{ème} clades mPTP de *S. propinquus* (Groupe 3 et 4) ils sont divisés en deux clades respectivement. Le 3^{ème} représente 7 séquences et le 4^{ème} composé de 10 séquences, ces deux clades (Groupe 3 et 4) distribuées en Palestine. Alors que le BINs et ASAP ont montré que le 3^{ème} clade est subdivisé en deux sous-clade : un clade regroupée une séquence (KT188222) à part, et l'autre clade comprend les séquences restants. Les données phylogénétiques et géographiques dans les clades ont montré la présence d'une même lignée génétique dans la même région géographique Palestine.

Le 5^{ème} clade *S. maurus* (Groupe 5) qui constitue un branche, se regrouper ensemble du 22 séquences représenter une série de 20 séquence de *S. maurus* et deux séquences exotique (intrus) qui est *S. fuliginosus* (JQ514259.1) du Maroc et (KT188214.1) du Palestine, qui sont peut être des séquences mal identifiées (putatives). Au sein de ce clade est divisé en 17 sous-clades représentés 22 séquences (17 sous-groupe) par la méthode de délimitation des espèces BINs. Alors que l'ASAP faire embrancher le clade en 18 sous-clades représentés 22 séquences (18 sous-groupe). Ces espèces ont été réparties dans ce groupe à des régions éloignées les uns des autres comme : Palestine, Maroc et Turquie (la région du bassin méditerranéen). C'est la raison probablement pour laquelle le groupe (dans mPTP) a été divisé en plusieurs sous groupes (dans BINs et ASAP). Nos résultats confirment les suggestions antérieures selon lesquelles *Scorpio maurus* est un complexe d'espèces (Froufe *et al.*, 2008 ; Talal *et al.*, 2015 ; Halil *et al.*, 2017)

Par contre on trouve un 6^{ème} clade mPTP séparé de la branche de *S. maurus* (Groupe 4) qui représente 2 séquences distribuées en Turquie. Tandis que ces deux séquences sont divisées en deux groupes par le BINs et ASAP, cela est dû à la différence entre les 3 méthodes.

Alors que, nous avons montré un 7^{ème} clade dans les trois méthodes mPTP, BINs et ASAP, qui constitue un autre groupe de *S. fuscus* (Groupe 7), qui étaient paraphylétiques dans l'arbre. Toutes les séquences de ce clade sont été réparties de Palestine, ce regroupement copine ensemble de 8 séquences représenter une série de 6 séquences de *S. fuscus* et deux séquences de différentes. En effet, une séquence identifiée comme appartenant à l'espèce *S. propinquus* (KT188280), et l'autre séquence identifiée comme appartenant à l'espèce *S.*

palmatus (KT188338), actuellement placées dans la section *S. fuscus* sur des arguments moléculaires, peut être les séquences sont mal identifiées ou mal traitées.

En fin, les trois méthodes mPTP, BINs et ASAP montre que l'espèce *S. palmatus* est divisé en deux sous-clades respectivement (Groupe 8 et 9), L'une de (Groupe 8) *S. palmatus*, semble représenter une série de 9 séquences sœurs, L'une d'autre (Groupe 9) *S. palmatus*, semble représenter une série de 17 séquences sœurs. L'espèce a été répartie dans les deux groupes dans la région de Palestine et la localités adjacentes à la mer Morte (Talal *et al.*, 2015) . Ces deux groupes sont paraphylétiques dans l'arbre.

Dans l'ensemble des résultats obtenus, nous avons montré dans cette enquête scientifique que le seul marqueur CO1 d'ADN mitochondrial est très efficace pour l'identification moléculaire, mais on propose d'utiliser d'autres marqueurs génétiques pour obtenir des résultats fiables et aussi pour améliorer l'effet de code barre moléculaire sur l'identification et la taxonomie moderne.

Conclusion

Conclusion

L'objectif de cette présente étude consiste à la mise en évidence de l'effet de la Barcoding moléculaire sur l'amélioration de l'identification moléculaire des espèces de genre *Scorpio* et le domaine taxonomique.

Malheureusement, nous n'avons pas eu la chance de réaliser ce travail avec des échantillons collectés sur sites. À cause du manque des produits et l'indisponibilité de séquençage.

Notre travail a été mené dans le but de construire de l'arbre phylogénétique du genre *Scorpio* pour l'identification moléculaire contribuer à la construction d'une base de données (code-barres d'ADN) des 6 espèces qui sont : *S. fuscus*, *S. palmatus*, *S. maurus*, *S. propinquus*, *S. kruglovi*, *S. fuliginosus*.

Nous avons fait les inférences phylogénétiques, à partir des séquences collectées à BOLDsystems et vérifier et identifier sur BLAST et GenBank.

Les séquences génétiques ont été alignées à l'aide du programme MUSCLE dans le logiciel MEGA 11, et nous avons utilisées Gbloks disponible sur site NGphylogeny.fr pour éliminer les positions mal alignées et nettoyées les séquences. L'arbre phylogénétique a été reconstruit à l'aide de MEGA 11 et sur site NGphylogeny.fr, grâce à la méthode du maximum de vraisemblance (ML).

En effet, bien que les collectes de ce genre fussent réalisées à trois méthodes de délimitations des espèces. Ils été effectués en ligne via mPTP et ASAP, et la méthode de délimitations des espèces BINs proposée par la base de données BOLD, les résultats de la délimitation des espèces BINs et ASAP sont presque identique. Cela confirme l'efficacité de ces méthodes et que cet arbre est bien correct.

Si la capacité du DNA Barcoding à identifier les organismes s'est révélée souvent très efficace, son utilisation pour délimiter et découvrir potentiellement les espèces, en revanche, elle a été rarement controversée, cela nécessite l'utilisation de plusieurs marqueurs moléculaires pour avoir des résultats plus exacts.

Ainsi, cette étude nous a permis de caractériser la richesse spécifique de genre *Scorpio* en montrant notamment des résultats un peu différente entre la taxonomie traditionnelle et la taxonomie moléculaire.

Un autre point intéressant de cette étude, c'est de mettre en évidence la variation spatiale des 6 espèces examinées, notamment en réalisant des estimations de richesse spécifique pour les individus collectés sur BOLDsystems. Cependant l'effort d'échantillonnage actuel pour ce portail n'est pas encore suffisant. Et il a besoin d'une vérification minutieuse de l'authenticité et la viabilité des séquences en sa possession sur cette plateforme.

Nous considérons les nouvelles données comme une ressource importante pour les futures études sur ce genre de scorpion d'une importance médicale considérable, indiquent que des travaux similaires sont nécessaires pour aborder pleinement la diversité au sein de ce genre d'espèce.

Bibliographie

Bibliographie

1. Abbassi Nadia, H. R. (2018). Méta-analyse des données des envenimations scorpioniques au Maroc. *SMEtax journal* , 115-122.
2. Allard, M. W. (2016). The future of Whole-Genome Sequencing for public health and the clinic. . *Journal of Clinical Microbiology*, 54 , 1946–1948.
3. Ankita Rajpoot, P. (2016). Dna Barcoding And Traditional Taxonomy: An Integrative Approach. *International Journal of Current Research*, Vol. 8, Issue, 11 , 42025-42031.
4. Avise, J. C. (2000). *Phylogeography: The History and Formation of Species*. Cambridge london ENGLAND: Harvard University Press UK.
5. bannister J.L., B. W. (1977). *records of the western australian museum*. Australie: Publications officer A.F LOVELL.
6. Bass, D. S. (2015). Diverse applications of environmental DNA methods in parasitology. . *Trends in Parasitology*, 31 , 499–513.
7. Bastawade, B. I. (1983). SCORPIONS SCORPIONIDA : ARACHNIDA. *THE FAUNA OF INDIA* , 1-669.
8. Birky CW Jr, M. T. (1983). : An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. . *Genetics (103)* , 513-527.
9. Castresana, J. (2022). About Gblocks Version0.91b, January2002 Copyright ©. *Institut de Biologia Evolutiva (CSIC-UPF)* .
10. Christopher P. Meyer, G. P. (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS BIOLOGY* 3 (12), e422 , 2229-2238.
11. Coelho, P. J. (2013). Phylogeography of *Androctonus* scorpions from the Maghreb Region. *Mestrado em Biodiversidade, Genética e Evolução, UNIVERSIDADE DO PORTO FACULDADE DE CIÊNCIAS* , 3-80.
12. Craig Costion1, 2. A. (2011). plant DNA Barcodes Can Accurately Estimate Species Richness in Poorly Known Floras. *PLoS ONE* 6(11) , 6, e26841.
13. Das, D. B. (2007). Les espèces cryptiques comme fenêtre sur la diversité et la conservation. *Tendances en écologie et évolution, Volume 22, numéro 3* , 148-155.
14. Dennis A. Benson, M. C.-M. (2017). GenBank. *Nucleic Acids Research* , 41-47.
15. Diethard Tautz1, P. A. (2003). A plea for DNA taxonomy. *TRENDS in Ecology and Evolution Vol.18 No.2 February 2003* , 70-74.

16. Elsiddig Nouredin 1, *. D. (2022). DNA Barcoding of Potential Mosquito Disease Vectors (Diptera, Culicidae) in Jazan Region, Saudi Arabia. *pathogens MDPI*, 11, 486 , 2-15.
17. Feola, A., Perrone, M., Piscopo, A., Casella, F., Della Pietra, B., & Di Mizio, G. (2020). Autopsy Findings in Case of Fatal Scorpion Sting: A Systematic Review of the Literature. *Healthcare*. 8(3):325 <https://doi.org/10.3390/healthcare8030325> , 325.
18. Fet, V. (2000). Family SCORPIONIDAE Latreille, 1802. *Catalog of the Scorpions of the world (1758-1998)* , 430-486.
19. Fet, V. (2010). *Scorpions du Monde*. FRANCE: N.A.B Edition.
20. Fish-bol. (2012). Fish Barcode of Life World Conference . *International Pavilion C at the Expo exhibition area in the city of Yeosu* , 1-14.
21. Folmer O, B. M. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. . *Mol Mar Biol Biotech* (3) , 294-299.
22. Frédéric Lemoine, D. C.-A.-B. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research, Volume 47, Issue W1* , , w260-w265.
23. Fujisawa, T. &. (2013). Délimitation des espèces à l'aide de données à un seul locus et de l'approche Generalized Mixed Yule Coalescent : une méthode révisée et une évaluation sur des ensembles de données simulées. . *Biologie systématique* , 62 (5) , 707-724.
24. González-serna, M. J. (2018). A review of cross-backed grasshoppers of the genus *Dociostaurus* Fieber (Orthoptera: Acrididae) from the western Mediterranean: insights from phylogenetic analyses and DNA-based species delimitation. *Systematic Entomology*, 43(1) , 139-146.
25. Gould, W. R. (1997). Catch-effort maximum likelihood estimation of important population parameters. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(4) , 890-897.
26. Goyffon, M. (2002). Le scorpionisme en Afrique sub-saharienne. *Bull Soc Pathol Exot, Laboratoire d'études et de recherches sur les arthropodes irradiés (LERAI), Muséum national d'histoire naturelle, 57 rue Cuvier, 75005 Paris, France (mgoyffon@cimrs1.mnhn.fr)* , 191-193.
27. Guindon, S. a. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. . *Systematic Biology* , 59:307–321.

28. Halil KOÇ1, *. H. (2017). Genetic diversity within *Scorpio maurus* (Scorpiones: Scorpionidae) from Turkey . *NORTH-WESTERN JOURNAL OF ZOOLOGY* 13 (1) , 27-33.
29. Hall G., B. (2013). Construire des arbres phylogénétiques à partir de données moléculaires avec MEGA. *Biologie moléculaire et évolution* , volume 30, numéro 5, mai 2013 , 1229-1235.
30. Halloran, S. M. (1984). The Birth of Molecular Biology: An Essay in the Rhetorical Criticism of Scientific Discourse. *Rhetoric Review* 3.1, 70-83. Print. . *Rhetoric Review* 3.1, Print. , 70-83.
31. Hammond, P. (1992). Species inventory . *Global biodiversity* , 17-39.
32. Haro, N. D. (2014). Piqûre de scorpion *Isometrus maculatus* à la Réunion *Isometrus maculatus* scorpion sting in Reunion Island. *SFMU et Springer-Verlag France* , 191-193.
33. Hawksworth, D. L.-A. (1995). Magnitude and distribution of biodiversity. In *Global biodiversity assessment* (ed. V. H. Heywood). *Cambridge University Press.* , 107–191.
34. Hebert Paul D. N., C. A. (2003a). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond.* , B 270, 313–321 313 DOI.
35. Hebert, A. J. (2001). Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. *Molecular Ecology* (10) , 371–386.
36. Hebert, P. D., Ratnasingham, S., et De Waard, J. R. (2003b). Barcoding Animal Life: Cytochrome c Oxidase Subunit 1 Divergences among Closely Related Species. *Biological Sciences, Vol.270* , S96-S99.
37. Hebert, P. D., S. R. (2001). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society Lond. B* 2003 270, S96-S99 doi: 10.1098/rsbl.2003.0025 , s96-99.
38. Hebert, P. D., S. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *the royal society, biology letters* , 96-99.
39. Hebert, P. D., Sujeevan, R., & R., D. W. (2003b). Barcoding Animal Life: Cytochrome c Oxidase Subunit 1 Divergences among Closely Related Species. *Biological Sciences, Vol. 270* , pp.S96-S99.
40. Hebert, P. D., T. R. (2005). La promesse du codage à barres ADN pour la taxonomie . *Systematic Biology, Volume 54, Issue 5* , 852–859.
41. hebert, S. R. (2007). BARCODING, BOLD: The Barcode of Life Data System. *Molecular Ecology Notes* , 1-10.

42. Ismail M. Abdel-Nabi 1*, A. M.-R. (2004). Intraspecific diversity of morphological characters of the burrowing scorpion *Scorpio maurus palmatus* (Ehrenberg, 1828) in Egypt (Arachnida: Scorpionida: Scorpionidae). *Serket vol. 9(2)* , 41-67.
43. Ismaïl, M. (2003). Traitement du syndrome d'envenimation scorpionique : 12 ans d'expérience avec la sérothérapie. . *Journal international des agents antimicrobiens* , 21 (2) , 170-174.
44. J. William O. Ballard, M. C. (2004). The incomplete natural history of mitochondria. *MOLECULAR ECOLOGY, Volume13, Issue4* , 729-744.
45. Jian Ye, S. M. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Research, Vol. 34, Web Server issue* , 6-9p.
46. Jin, Q. H. (2018). A two-step DNA barcoding approach for delimiting moth species: moths of Dongling Mountain (Beijing, China) as a case study. *Scientific reports, 8(1)* , 1-12.
47. John M. Bruns, D. H. (2008). Codes-barres ADN et espèces cryptiques de papillons skipper du genre *Perichares* dans la zone de conservation Guanacaste, Costa Rica. *PNAS* , 6350–6355.
48. John P. WARES1, 2. A. (2001). PHYLOGEOGRAPHY AND HISTORICAL ECOLOGY OF THE NORTH ATLANTIC INTERTIDAL. *Evolution, 55(12)* , 2455–2469.
49. John T., H. (1990). Anatomy and Morphology. Dans G. A. POLIS, *the biology of scorpion* (pp. 2-200). Stanford: Stanford university Press.
50. Junier, T. a. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* , 26:1669–1670.
51. Kapli, P. L. (2017). Processus d'arbre de Poisson multi-taux pour la délimitation d'espèces à un seul locus sous maximum de vraisemblance et chaîne de Markov Monte Carlo. *Bioinformatique* , 33 (11) , 1630-1638.
52. Kaur, S. (2015). DNA Barcoding and Its Applications. *International Journal of Engineering Research and General Science Volume 3, Issue 2, Part 2* , 602-604.
53. Kumar S., S. G. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. . *Molecular Biology and Evolution (35)* , 1547-1549.
54. Laudénbach V., M. J. (1999). To Understand Molecular Biology. *Annales Françaises d'Anesthésie et de Réanimation* , 18, 725-747.
55. Lemoine, F. a.-A.-B. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic acids research* , 47:W260-W265.

56. Lemoine, F. a.-B. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data . *Nature* , 556:452–456.
57. Letunic, I. &. (2021). Interactive Tree Of Life (iTOL) v5 : un outil en ligne pour l'affichage et l'annotation d'arbres phylogénétiques. *Recherche sur les acides nucléiques*, 49 (W1) , W293-W296.
58. Li-ping chen, Y.-G. C.-M.-L.-C.-J. (2018). Application of DNA Barcoding in the Classification of Grasshoppers (Orthoptera: Acridoidea)—A Case Study of grasshoppers from Hebei Province, China . *ResearchGate"Article in Zootaxa"* , 99-110.
59. Lourenço, W. R. (2000). Reproduction in scorpions, with special reference to. *European Arachnology 2000 (S. Toft & N. Scharff eds.)* , 71-85.
60. Lourenço, W. R. (2009). Reanalysis of the genus *Scorpio* Linnaeus 1758 in sub-Saharan Africa and description of one new species from Cameroon (Scorpiones, Scorpionidae). *Entomol. Mitt. zool. Mus. Hamburg* , 99-113.
61. Lourenço, W. R. (2018). Scorpions and life-history strategies: from evolutionary dynamics toward the scorpionism problem. *Journal of Venomous Animals and Toxins including Tropical Diseases* , 1-12.
62. Lunardi, J. (2003). Fifty years ago, the double helix gave birth to the molecular biology. *Ann Biol Clin 2003* , 23-33.
63. Maria S.V. Santos, P., Cláudio G.L. Silva, P., Basílio Silva Neto, P., Cícero R.P. Grangeiro Júnior, M., Victor H.G. Lopes, M., Antônio G. Teixeira Júnior, M., et al. (2016). Clinical and Epidemiological Aspects of Scorpionism in the World: A Systematic Review. *WILDERNESS & ENVIRONMENTAL MEDICINE*, 27 , 504–518 .
64. Marie-France, M.-E. C. (1999). Les toxines des venins de scorpion. *Annales de l'Institut Pasteur / Actualités*10(2) , 207-222.
65. Martin Wiemers, K. F. (2007). Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* (4:8) , 1-16.
66. Marzouk, Z. (2018). ATELIER BIOINFORMATIQUE: comparaison des séquences et des génomes. *ISB-MONASTIR* , 2-35.
67. Maurice A Smith, A. G. (2006). L'interaction de processus adaptatifs avec différentes échelles de temps sous-tend l'apprentissage moteur à court terme. *PLOSS BIOLOGY*4(6): e179. <https://doi.org/10.1371/journal.pbio.0040179> , 1-6.
68. Maurizio Casiraghi, M. L. (2010). DNA barcoding: a six-question tour to improve users' awareness about the method . *Briefings in Bioinformatics, Volume 11, Issue 4* , 440–453.

69. Megan Milton, M. V. (2011). Barcode of Life Data Systems. *www.boldsystems.org* , 1-34.
70. Meier R, K. S. (2006). DNA barcoding and taxonomy in Diptera: a tale of high Diptera: a tale of highintraspecific variability and low identification success. *Syst Biol* (55) , 715–728. .
71. Meyer, K. (1991). Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetics Selection Evolution*, 23(1) , 67-83.
72. Naruya Saitou, M. N. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Saitou, N., & Nei, M. (1987). The neighbor-joining method: a n Molecular biology and evolution*, 4(4) , 406-425.
73. Nicolas Puillandre, S. B. (2021). ASAP : assembler les espèces par partitionnement automatique. *molecular ecology resources* , 609-620.
74. Pevzner, P. A. (2006). Aligement multiple. *Bio-informatique moléculaire* , 125-135.
75. Polis, G. A. (1990). *The biology of scorpions*. Stanford: Stanford University Press.
76. Pons, J. B.-Z. (2006). Délimitation des espèces basée sur la séquence pour la taxonomie de l'ADN des insectes non décrits. . *Biologie systématique* , 55 (4) , 595-609.
77. Puillandre, N. B. (2021). ASAP: assemble species by automatic partitioning. . *Molecular Ecology Resources*, 21(2) , 609-620.
78. Puillandre, N. L. (2012). ABGD, Automatic Barcode Gap Discovery pour la délimitation primaire des espèces. *Écologie moléculaire* , 21 (8) , 1864-1877.
79. Ratnasingham, S. &. (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. . *PloS one*, 8(7) , e66213.
80. Raupach Michael J, A. J. (2010). Molecular species identification of Central European ground beetles (Coleoptera:Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* 2010 , 7-26.
81. Raven, P. H., Williams, T.. (1997). The dimensions of life of earth in Raven PH, editor.Nature and human society: The quest for a sustainable world. *Washington,D.C.: National Academy Press.* , 30–45.
82. Rein, J. O. (2020). Taxonomical updates in The Scorpion Files for Scorpionidae (2008). *The Scorpion Files* , 56-61.
83. Rein, J. O. (2022). The scorpion files, SCORPIONIDÉS Latreille 1802. <https://www.ntnu.no/ub/scorpion-files/scorpionidae.php> , (consulté le 9 avril 2022).

84. Rein, J. O. (2022). The Scorpion Files. <https://www.ntnu.no/ub/scorpion-files> / , (consulté le 18 juin 2022).
85. Robert D. Ward, . T. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360 , 1847–1857.
86. Rouhullah Dehghani, F. M. (2016). Review study of scorpion classification in Iran. *Journal of Entomology and Zoology Studies* , 440-444.
87. Rouschmeyer, L. (2015). Clé de détermination simplifiée des scorpion . *L'outil Faune-paca "LPO PACA"* , 3-15.
88. Saccone, C. D. (1999). Evolutionary genomics in the Metazoa: the mitochondrial DNA as a model system. *Gene* 238 , 195–210.
89. Saddine, S. E. (2018). La faune scorpionique de Sahara septentrional algérien: diversité et ecologie. *Thèse de doctorat en biologie, Université Kasdi Merbah Ourgela* , 6-19.
90. Sadine S. E., A. (2014). Aperçu sur les scorpions de Ghardaïa (Algérie). *Journal of Advanced Research in Science and Technology* , 12-13.
91. Sadine S. E.. D. (October 2020). Overview on Scorpions of Algeria. *ALGERIAN JOURNAL OF HEALTH SCIENCES.VOL. 2 SUPPLÉMENT 1 (2020) S8-S14* , 9.
92. Sadine, S. e. (2018). *La faune scorpionique du Sahara Septentrional algérien: Diversité et écologie. Thèse de doctorat.* Ouergla. Algérie: Université de Ouergla.
93. Safa Jammali, A. D.-Y. (2022). From pairwise to multiple spliced alignment. *Bioinformatics Advances, OXFORD* , 1-11.
94. Sofie Derycke1, 2. J. (2010). Exploring the Use of Cytochrome Oxidase c Subunit 1 (COI) for DNA Barcoding of Free-Living Marine Nematodes. *PLoS ONE* www.plosone.org 1 October 2010, Volume 5 , Issue 10, e13716 , 1-9.
95. Stat, M. H. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports* 25(12240) , 1-11.
96. Stav Talal a, I. T.-S. (2015). Scorpion speciation in the Holy Land: Multilocus phylogeography corroborates diagnostic differences in morphology and burrowing behavior among *Scorpio* subspecies and justifies recognition as phylogenetic, ecological and biological species. *Molecular Phylogenetics and Evolution journal homepage: www.elsevier.com/locate/ympev* , 226–237.
97. Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees . *Molecular Biology and Evolution* (10) , 512-526.

98. Taylor, H. R., et Harris, W. E. (2012). An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources* , doi: 10.1111/j.1755-0998.2012.03119.x.
99. Vachon, M. (1952). Etude sur les scorpions. *Institut Pasteur d'Algérie*, (pp. 477-479). alger .
100. Van der Straeten, J. D. (2021). Validation of a PCR-Based Next-Generation Sequencing Approach for the Detection and Quantification of Minimal Residual Disease in Acute Lymphoblastic Leukemia and Multiple Myeloma Using gBlocks as Calibrators. *The Journal of Molecular Diagnostics*, 23(5) , 599-611.
101. Vaucel, J. (2020). Les piqûres de scorpions chez l'enfant en Guyane française: une étude rétrospective de 2003 à 2018. *Sciences du Vivant [q-bio]*. 2020. ffdumas-02950932 , 6-7.
102. Vincent Derrien, J.-M. R.-k. (2005). Plasma, un nouvel algorithme progressif pour l'alignement multiple de séquence. *HAL open science* , 39-48.
103. Watson, J. D., et Crick, F. H. (1953). A Structure for Deoxyribose Nucleic Acid Nature 171, 737-738. *resonance november 2014* , 96-98.
104. Wilson, E. O. (2003). The encyclopedia of life. *Trends Ecol Evol* 18 , 77–80.
105. Xiaoxiao Guo¹, C. M. (2013). Two peptides, TsAP-1 and TsAP-2, from the venom of the Brazilian yellow scorpion, *Tityus serrulatus*: Evaluation of their antimicrobial and anticancer activities. *journal homepage"biochimie"* , 1784-1794.
106. Yang, B. R. (2020). *Species Delimitation*. In Scornavacca, C., Delsuc, F., and Galtier, N. . france : editors, Phylogenetics in the Genomic Era,.
107. Zhang, H. L. (2013). The phylogeny of the Orthoptera (Insecta) as deduced from mitogenomic gene sequences. *Zoological Studies*, 52(1) , 1-13.
108. Zhang, L. e. (2013). Neuroprotection by tetrahydroxystilbene glucoside in the MPTP mouse model of Parkinson's disease. *Toxicology Letters* 222.2 , 155-163.

➤ **Les sites web :**

Site web 1: <https://www.barcodeofwildlife.org/>.

Site web 2: <https://www.sccwrp.org/about/research-areas/bioassessment/dna-barcoding/>.

Site web 3: <https://www.ntnu.no/ub/scorpion-files/scorpionidae.php>

Site web 4: <https://www.boldsystems.org/>

Site web 5: <https://www.ncbi.nlm.nih.gov/nucleotide>

Site web 6: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Site web 7: <https://ngphylogeny.fr/>

Site web 8: <https://mptp.h-its.org/#/tree>

Site web 9: <https://bioinfo.mnhn.fr/abi/public/asap/asapweb.htm>

Annexes

Annexes

Annexe 1 : Photo originale de *Scorpio maurus*



Annexe 2 :

Tableau 3: Liste des espèces de genre *Scorpio* dans le monde (Rein, 2020)

Taxons (espèce)	Statut	Distribution	commentaires
<i>Scorpio birulai</i> Fet, 1997	Nouveau statut	Maroc	Statut précédent de la sous-espèce : <i>Scorpio maurus subtypicus</i> Birula, 1910
<i>Scorpio ennedi</i> Lourenco, Duhem & Cloudsley Thompson, 2012	Nouveau sp.	Tchad	
<i>Scorpio fuliginosus</i> Pallary, 1928	Nouveau statut	Maroc	Statut précédent: <i>Heterometrus</i> <i>fuliginosus</i> (Pallary, 1928)
<i>Scorpio fuscus</i>	Nouveau statut		Statut précédent de la sous-espèce : <i>Scorpio maurus fuscus</i> Ehrenberg,

(Ehrenberg, 1829)		Palestine	1829
<i>Scorpio hesperus</i> Birula, 1910	Nouveau statut	Maroc	Statut précédent de la sous-espèce : <i>Scorpio maurus hesperus</i> Birula, 1910
<i>Scorpio kruglovi</i> (Ehrenberg, 1829)	Nouveau statut	Palestine, Jordanie, Syrie	Statut précédent de la sous-espèce : <i>Scorpio maurus kruglovi</i> Ehrenberg, 1829
<i>Scorpio mogadorensis</i> Birula, 1910	Nouveau statut	Maroc	Statut précédent de la sous-espèce : <i>Scorpio maurus mogadorensis</i> Birula, 1910
<i>Scorpio niger</i> Lourenco, Duhem & Cloudsley Thompson, 2012	Nouveau sp.	Niger	
<i>Scorpio occidentalis</i> Werner, 1936	Nouveau statut	Mali, Mauritanie et le Sénégal	Statut précédent de la sous-espèce : <i>Scorpio maurus occidentalis</i> Werner, 1936
<i>Scorpio palmatus</i> (Ehrenberg, 1829)	Nouveau statut	Palestine, Egypte	Statut précédent de la sous-espèce : <i>Scorpio maurus palmatus</i> Ehrenberg, 1829
<i>Scorpio propinquus</i> (Simon, 1872)	Nouveau statut	Palestine, Syrie	Statut précédent de la sous-espèce : <i>Scorpio maurus propinquus</i> (Simon, 1872)
<i>Scorpio punicus</i> Fet, 2000	Nouveau statut	Algérie, Libye, Maroc & Tunisie	Statut précédent de la sous-espèce : <i>Scorpio maurus tunetanus</i> Birula, 1910
<i>Scorpio savanicola</i>	Nouveau		

Lourenco, 2009	sp.	Cameroun	
<i>Scorpio sudanensis</i> Lourenco & Cludsley- Thompson, 2009	Nouveau sp.	Soudan	
<i>Scorpio tassili</i> Lourenco & Rossi, 2016	Nouveau sp.	Algérie	
<i>Scorpio weidholzi</i> Werner, 1929	Nouveau statut	Maroc	Statut précédent de la sous-espèce : <i>Scorpio maurus weidholzi</i> Werner, 1929

Annexe 3 :**Tableau 4:** Le code de nucléotide selon UIB

Symbole	Signifiant	Acide nucléique
A	A	Adénine
C	C	Cytosine
G	G	Guanine
T	T	Thymine

Annexe 4: Exemple d'une séquence nucléotidique au Format **GenBank**

LOCUS KT188338 1078 bp ADN linéaire INV 02-JUL-2015
 DÉFINITION Isolat de *Scorpio palmatus* 12218 sous-unité I de la cytochrome
 oxydase (COI)
 gène, CD partiels ; mitochondriale.
 ADHÉSION KT188338
 MODÈLE KT188338.1
 MOTS CLÉS .
 SOURCE mitochondrie *Scorpio palmatus*
 ORGANISME [Scorpio palmatus](#)
 Eucaryote; Métazoaires ; Ecdysozoa; Arthropodes ; Chélicerata ;
 arachnide ;
 Scorpiones ; Iurida ; Scorpionoïdes ; Scorpionidés ;
 Scorpioninae;
 Scorpion.
 REFERENCE 1 (bases 1 à 1078)
 AUTEURS Talal., S., Tesler, I., Sivan, J., Ben-Shlomo, R., Muhammad
 Tahir, H.,
 Prendini, L., Snir, S. et Gefen, E.
 TITRE Spéciation des scorpions en Terre Sainte : Phylogéographie
 multilocus

corrobore les différences diagnostiques de morphologie et d'enfouissement
comportement parmi les sous-espèces de Scorpion et justifie la reconnaissance comme
espèces phylogénétiques, écologiques et biologiques
JOURNAL Mol. Phylogénète. Évol. (2015) Sous presse [Publié](#)
25987530
REMARQUE Statut de la publication : disponible en ligne avant l'impression

REFERENCE 2 (bases 1 à 1078)
AUTEURS Talal., S., Tesler, I., Sivan, J., Ben-Shlomo, R., Tahir, M., Prendini, L., Snir, S. et Gefen, E.
TITRE Soumission directe
JOURNAL Soumis (19-JUN-2015) Biologie et environnement, Université de Haïfa, Oranim, Tivon 36006, Israël

COMMENTAIRE ##Assembly-Data-START##
Technologie de séquençage :: Séquençage didésoxy Sanger
##Assembly-Data-END##

CARACTÉRISTIQUES Localisation/
Source des qualificatifs 1..1078
/organism="Scorpio palmatus"
/organelle="mitochondrie"
/mol_type="ADN génomique"
/isole="12218"
/db_xref="taxon : [1662106](#) "
/collection_date="2012"
[gène](#) <1..>1078
/gene="COI"
[CDS](#) <1..>1078
/gene="COI"
/codon_start=2
/transl_table= [5](#)
/product="sous-unité I de la cytochrome oxydase"
/protein_id=" [AKN62967.1](#) "

/translation="TMYLIMGGWASMVGTALSMIRIETGSPGFIGDDQLYNVVVTA
HAFVMIFFVMVPIIMIGGFNWLVPMLLGLAPDMAFPRLNNMSFWLLPPAFFMLLGSAAL
ESGAGTGWTVYPPLSSGMFHSGGSVDMTIFSLHLAGVSSILGAINFITIVNMRSDGM
IMDRVPLFVWSVLITAVLLLLSLPVLGAIITMLLTDRNFNTSFFDPAGGGDPILYQHL
FWFFGHPEVYILILPGFGMISHIVSHHTGKKEPFGTIGMIYAMVSI GFLGFVVAHHM
FTVGMVDVTRAYFTAATMVIAPVTGIKVFSWIATMYGSRVYFSPMMWAVGFILFTV
GGLTGVMLANSSLDVILHDTYYVVA"

ORIGINE
1 gactatgtat ttaattatag gtgggtgggc ttcaatggtt ggaactgctt taagaatgat
61 aattcggatt gagactggaa gaccaggttc ttttattggt gatgatcagc tttacaatgt
121 tgtagtact gcgcatgctt ttgtaataat tttttttatg gtaataccta ttatgattgg
181 aggttttggg aattgattgg ttctttaat attaggggct cggatatag cttttcctcg
241 tttaaataat atgagatatt gattgcttcc tctgctttt tttatgcttt tagggctctg
301 tgctttggaa agaggggagg ggactgggtg aactgtttat cctcctttgt cttctggtat
361 atttcattca ggtggttctg tagatatgac tattttttct ttacatttag ctggggttct
421 atcaatttta ggagctatta atttattac aactattgtg aatatacgtg gagatggaat
481gattatggat cgagttcctt tatttgatg gtctgttctt attacggctg tattgttact
541 tctttctttg cctgttttgg ctggggcaat tactatattg ttaactgatc ggaattttaa
601 cacttctttt ttgatcctg ctggtggtgg ggatcctatt ttgtatcagc atttgttttg
661 attttttggg catcctgagg tttatatttt gattcttctt gggtttggaa tgattttctca
721 tattgtaagt catcatacag gaaaaaagga gccttttggg actattggga tgatttatgc

781 tatagtgtct attgggtttt tagggtttgt agtatgggct catcatatgt ttactgtggg
 841 aatggatggt gatactcgag cttattttac agcagctact atggtaattg cggttcctac
 901 tggattaag gtgttttagat gaattgctac tatataggg tctcggggtt atttctctcc
 961gcctatgatg tgggctgtgg ggtttatttt tttatttact gtaggaggtt tgactgggtg
 1021 tatattagct aatttttctt tagatgtaat tcttcatgac acttattatg ttgttgc

Annexe 5 :

Tableau 5: L'origine géographique des espèces de *Scorpio* utilisées dans notre étude et leurs numéros d'accès GenBank de leurs séquences IDS

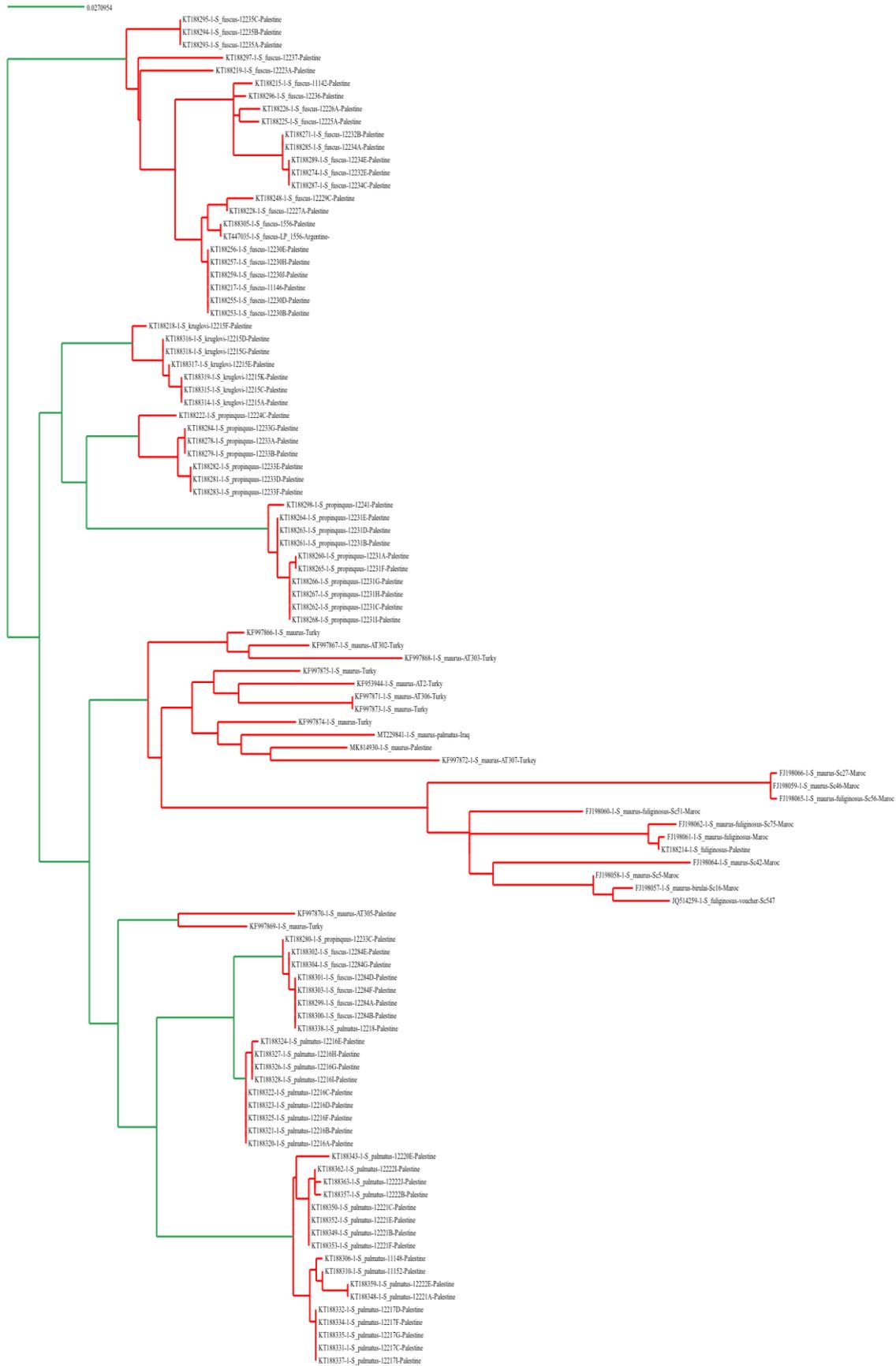
N°	Toxon	Numéro d'accès GenBank (IDS Séquence)	Localité	Source
01	<i>S. maurus</i>	FJ198066	Maroc	Froufe <i>et al.</i> , 2008
02	<i>S. maurus</i>	FJ198059	Maroc	Froufe <i>et al.</i> , 2008
03	<i>S. maurus</i>	FJ198065	Maroc	Froufe <i>et al.</i> , 2008
04	<i>S. maurus</i>	FJ198058	Maroc	Froufe <i>et al.</i> , 2008
05	<i>S. maurus</i>	FJ198057	Maroc	Froufe <i>et al.</i> , 2008
06	<i>S. maurus</i>	FJ198064	Maroc	Froufe <i>et al.</i> , 2008
07	<i>S. maurus</i>	FJ198060	Maroc	Froufe <i>et al.</i> , 2008
08	<i>S. maurus</i>	FJ198062	Maroc	Froufe <i>et al.</i> , 2008
09	<i>s. fuliginosus</i>	KT188214	Palestine	Talal <i>et al.</i> , 2015
10	<i>S. maurus fuliginosus</i>	FJ198061	Maroc	Froufe <i>et al.</i> , 2008
11	<i>S. maurus</i>	KF997874	Turquie	Sipahi <i>et al.</i> , 2013
12	<i>S. maurus</i>	KF997866	Turquie	Sipahi <i>et al.</i> , 2013
13	<i>S. maurus</i>	KF997875	Turquie	Sipahi <i>et al.</i> , 2013
14	<i>S. maurus</i>	KF997869	Turquie	Sipahi <i>et al.</i> , 2013
15	<i>S. maurus</i>	KF997873	Turquie	Sipahi <i>et al.</i> , 2013
16	<i>S. maurus</i>	KF997871	Turquie	Sipahi <i>et al.</i> , 2013
17	<i>S. fuscus</i>	KT188293	Palestine	Talal <i>et al.</i> , 2015
18	<i>S. fuscus</i>	KT188253	Palestine	Talal <i>et al.</i> , 2015
19	<i>S. fuscus</i>	KT188217	Palestine	Talal <i>et al.</i> , 2015
20	<i>S. palmatus-iso</i>	KT188320	Palestine	Talal <i>et al.</i> , 2015
21	<i>S. fuscus</i>	KT188228	Palestine	Talal <i>et al.</i> , 2015
22	<i>S. fuliginosus-voucher</i>	JQ514259		van der Meijden <i>et al.</i> , 2012
23	<i>S. palmatus-iso</i>	KT188338	Palestine	Talal <i>et al.</i> , 2015
24	<i>S. fuscus -iso</i>	KT188299	Palestine	Talal <i>et al.</i> , 2015
25	<i>S. fuscus -iso</i>	KT188297	Palestine	Talal <i>et al.</i> , 2015

26	<i>S. maurus-iso-palmatus</i>	MT229841	Iraq	Hussen <i>et al.</i> , 2020
27	<i>S. palmatus-iso</i>	KT188324	Palestine	Talal <i>et al.</i> , 2015
28	<i>S. fuscus</i>	KT188259	Palestine	Talal <i>et al.</i> , 2015
29	<i>S. fuscus</i>	KT188257	Palestine	Talal <i>et al.</i> , 2015
30	<i>S. fuscus</i>	KT188256	Palestine	Talal <i>et al.</i> , 2015
31	<i>S. fuscus</i>	KT188255	Palestine	Talal <i>et al.</i> , 2015
32	<i>S. fuscus</i>	KT447035	Palestine	Ojanguren-Affilastro <i>et al.</i> , 2015
33	<i>S. Kruglovi</i>	KT188314	Palestine	Talal <i>et al.</i> , 2015
34	<i>S. fuscus</i>	KT188305	Palestine	Talal <i>et al.</i> , 2015
35	<i>S. fuscus</i>	KT188219	Palestine	Talal <i>et al.</i> , 2015
36	<i>S. fuscus</i>	KT188295	Palestine	Talal <i>et al.</i> , 2015
37	<i>S. fuscus</i>	KT188294	Palestine	Talal <i>et al.</i> , 2015
38	<i>S. fuscus</i>	KT188296	Palestine	Talal <i>et al.</i> , 2015
39	<i>S. fuscus</i>	KT188248	Palestine	Talal <i>et al.</i> , 2015
40	<i>S. fuscus</i>	KT188226	Palestine	Talal <i>et al.</i> , 2015
41	<i>S. fuscus</i>	KT188225	Palestine	Talal <i>et al.</i> , 2015
42	<i>S. Kruglovi</i>	KT188218	Palestine	Talal <i>et al.</i> , 2015
43	<i>S. palmatus</i>	KT188328	Palestine	Talal <i>et al.</i> , 2015
44	<i>S. palmatus</i>	KT188327	Palestine	Talal <i>et al.</i> , 2015
45	<i>S. palmatus</i>	KT188326	Palestine	Talal <i>et al.</i> , 2015
46	<i>S. palmatus</i>	KT188325	Palestine	Talal <i>et al.</i> , 2015
47	<i>S. palmatus</i>	KT188323	Palestine	Talal <i>et al.</i> , 2015
48	<i>S. palmatus</i>	KT188322	Palestine	Talal <i>et al.</i> , 2015
49	<i>S. palmatus</i>	KT188321	Palestine	Talal <i>et al.</i> , 2015
50	<i>S. fuscus</i>	KT188304	Palestine	Talal <i>et al.</i> , 2015
51	<i>S. fuscus</i>	KT188302	Palestine	Talal <i>et al.</i> , 2015
52	<i>S. fuscus</i>	KT188287	Palestine	Talal <i>et al.</i> , 2015
53	<i>S. propinquus</i>	KT188279	Palestine	Talal <i>et al.</i> , 2015
54	<i>S. fuscus</i>	KT188215	Palestine	Talal <i>et al.</i> , 2015
55	<i>S. maurus</i>	KF953944	Turquie	Sipahi <i>et al.</i> , 2013
56	<i>S. maurus</i>	KF997867	Turquie	Sipahi <i>et al.</i> , 2013
57	<i>S. Kruglovi</i>	KT188319	Palestine	Talal <i>et al.</i> , 2015
58	<i>S. Kruglovi</i>	KT188317	Palestine	Talal <i>et al.</i> , 2015
59	<i>S. Kruglovi</i>	KT188315	Palestine	Talal <i>et al.</i> , 2015

60	<i>S. fuscus</i>	KT188303	Palestine	Talal et al., 2015
61	<i>S. fuscus</i>	KT188301	Palestine	Talal et al., 2015
62	<i>S. fuscus</i>	KT188300	Palestine	Talal et al., 2015
63	<i>S. propinquus</i>	KT188280	Palestine	Talal et al., 2015
64	<i>S. fuscus</i>	KT188285	Palestine	Talal et al., 2015
65	<i>S. fuscus</i>	KT188271	Palestine	Talal et al., 2015
66	<i>S. propinquus</i>	KT188261	Palestine	Talal et al., 2015
67	<i>S. maurus</i>	KF997872	Turquie	Talal et al., 2015
68	<i>S. Kruglovi</i>	KT188318	Palestine	Talal et al., 2015
69	<i>S. Kruglovi</i>	KT188316	Palestine	Talal et al., 2015
70	<i>S. propinquus</i>	KT188283	Palestine	Talal et al., 2015
71	<i>S. propinquus</i>	KT188282	Palestine	Talal et al., 2015
72	<i>S. propinquus</i>	KT188281	Palestine	Talal et al., 2015
73	<i>S. fuscus</i>	KT188289	Palestine	Talal et al., 2015
74	<i>S. propinquus</i>	KT188284	Palestine	Talal et al., 2015
75	<i>S. propinquus</i>	KT188278	Palestine	Talal et al., 2015
76	<i>S. fuscus</i>	KT188274	Palestine	Talal et al., 2015
77	<i>S. propinquus</i>	KT188268	Palestine	Talal et al., 2015
78	<i>S. propinquus</i>	KT188267	Palestine	Talal et al., 2015
79	<i>S. propinquus</i>	KT188266	Palestine	Talal et al., 2015
80	<i>S. propinquus</i>	KT188262	Palestine	Talal et al., 2015
81	<i>S. propinquus</i>	KT188298	Palestine	Talal et al., 2015
82	<i>S. propinquus</i>	KT188265	Palestine	Talal et al., 2015
83	<i>S. propinquus</i>	KT188264	Palestine	Talal et al., 2015
84	<i>S. propinquus</i>	KT188260	Palestine	Talal et al., 2015
85	<i>S. propinquus</i>	KT188263	Palestine	Talal et al., 2015
86	<i>S. propinquus</i>	KT188222	Palestine	Talal et al., 2015
87	<i>S. maurus-palmatus</i>	AY156585	Palestine	Talal et al., 2015
88	<i>S. maurus</i>	MK814930	Palestine	Talal et al., 2015
89	<i>S. palmatus</i>	KT188310	Palestine	Talal et al., 2015
90	<i>S. maurus</i>	KF997870	Palestine	Talal et al., 2015
91	<i>S. palmatus</i>	KT188357	Palestine	Talal et al., 2015
92	<i>S. palmatus</i>	KT188306	Palestine	Talal et al., 2015
93	<i>S. palmatus</i>	KT188362	Palestine	Talal et al., 2015
94	<i>S. palmatus</i>	KT188348	Palestine	Talal et al., 2015

95	<i>S. palmatus</i>	KT188363	Palestine	Talal <i>et al.</i> , 2015
96	<i>S. palmatus</i>	KT188353	Palestine	Talal <i>et al.</i> , 2015
97	<i>S. palmatus</i>	KT188352	Palestine	Talal <i>et al.</i> , 2015
98	<i>S. palmatus</i>	KT188350	Palestine	Talal <i>et al.</i> , 2015
99	<i>S. palmatus</i>	KT188349	Palestine	Talal <i>et al.</i> , 2015
100	<i>S. palmatus</i>	KT188343	Palestine	Talal <i>et al.</i> , 2015
101	<i>S. palmatus</i>	KT188337	Palestine	Talal <i>et al.</i> , 2015
102	<i>S. palmatus</i>	KT188335	Palestine	Talal <i>et al.</i> , 2015
103	<i>S. palmatus</i>	KT188334	Palestine	Talal <i>et al.</i> , 2015
104	<i>S. palmatus</i>	KT188332	Palestine	Talal <i>et al.</i> , 2015
105	<i>S. palmatus</i>	KT188331	Palestine	Talal <i>et al.</i> , 2015
106	<i>S. palmatus</i>	KT188359	Palestine	Talal <i>et al.</i> , 2015
107	<i>Hottentotta saulcyi</i>	KU341989	Iran	Oshaghi <i>et al.</i> , 2015

Annexe 7: Résultat de délimitation des espèces via mPTP



Résumés

جنس *Scorpio* هو عقرب ينتمي الى عائلة (Scorpionidae) ومهم من الناحية الطبية ومع ذلك لا يُعرف سوى القليل جدًا من الأبحاث حول التنوع الجيني داخل هذا الجنس من العقارب. تم تطوير تطبيق التصنيف الجزيئي والمعلومات التي يتيحها لنا الحمض النووي وتكنولوجيا الرمز الشريطي و تطور وسائل المعلوماتية الحيوية كل هذا طبق لتثمين التنوع البيولوجي لمختلف الكائنات و تدقيق مجال تصنيفها. تركز هذه الدراسة العلمية على اختبار مدى جدوى الرمز الشريطي للحمض النووي من خلال العلامة الوحيدة COI ميتوكوندريا في التعرف الجزيئي على جنس *Scorpio*. تم جمع ما مجموعه 106 تسلسل جيني COI ل 6 أنواع من جنس *Scorpio* الذي حدده الباحثون حسب نمطه الظاهري و الوراثة و شكله الخارجي. يتم تقديم البيانات في شجرة الوراثة العرقية، والتي تشير إلى عدم واحدية النمط التصنيفي لمعظم الأنواع بالمقابل واحدية النمط التصنيفي لـ *S. kruglovi*. استخدمنا مجموعة البيانات الجديدة الخاصة بنا لإنشاء قائمة بوحدة MOTU المحدثة للتسلسلات. عملت التحديدات الجزيئية للأنواع mPTP، BINs، ASAP بشكل جيد لمعظم الأنواع، باستثناء تلك التي هي جزء من *S. maurus* ما يؤكد النظريات السابقة كونها نوع أولي مركب من عدة أنواع.

الكلمات المفتاحية: *Scorpio*؛ الرمز الشريطي للحمض النووي؛ COI الميتوكوندريا؛ شجرة التطور؛ تحديد الأنواع.

Résumé

Le genre *Scorpio* appartenant à la famille des Scorpionidae est parmi les genres qui ont une importance médicale, malgré ça on sait très peu de recherches concernant diversité génétique au sein de ce genre. Le développement de la systématique moléculaire et l'exploitation des données fournis par l'ADN et les codes-barres ainsi l'amélioration des outils de la bioinformatique a été appliquée dans la valorisation de la biodiversité des différents organismes et la précision du statut taxonomique de différentes espèces. Cette étude scientifique s'attache à tester les faisabilités du code-barres d'ADN par un marqueur unique COI mitochondrial dans l'identification moléculaire des espèces du genre *Scorpio*. Nous avons collectées un total de 106 séquences du gène COI pour 6 espèces du genre *Scorpio* identifiées morphologiquement et phylogénétiquement par les chercheurs. Nous présentons les données dans un arbre phylogénétique, qui suggère la paraphélie pour la plupart des espèces à l'exception de l'espèce *S. kruglovi* qui était monophylique. Nous avons utilisé notre nouvel ensemble de données pour générer une liste du mOTUs actualisée pour les séquences. Les délimitations moléculaires des espèces (mPTP, BINs, ASAP) ont bien fonctionnés pour la plupart des espèces, à l'exception de ceux qui font partie des *S. maurus* ce qui confirme les théories antérieures selon lesquelles *Scorpio maurus* est une espèce primitive qui contient un complexe d'espèces.

Mot clé : *Scorpio* ; Code-barres ADN ; COI mitochondrial ; Arbre phylogénétique ; Délimitation des espèces.

Abstract

The genus *Scorpio* belonging to the family Scorpionidae is among the genera of medical importance, yet very little research is known about genetic diversity within this genus. The development of molecular systematics and the exploitation of DNA and barcode data as well as the improvement of bioinformatics tools has been applied in the enhancement of the biodiversity of different organisms and the precision of the taxonomic status of different species. This scientific study aims to test the feasibility of DNA barcoding by a single mitochondrial COI marker in the molecular identification of species of the genus *Scorpio*. We collected a total of 106 COI gene sequences for 6 species of the genus *Scorpio* identified morphologically and phylogenetically by the researchers. We present the data in a phylogenetic tree, which suggests paraphyly for most species except for *S. kruglovi* which was monophyletic. We used our new dataset to generate an updated mOTUs list for the sequences. Molecular species delineations (mPTP, BINs, ASAP) worked well for most of the species, with the exception of those that are part of *S. maurus*, confirming previous theories that *Scorpio maurus* is a primitive species that contains a species complex.

Keyword: *Scorpio*; DNA barcoding; mitochondrial COI; phylogenetic tree; species delineation