République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A. MIRA - Béjaia

Faculté des Sciences de la Nature et de la Vie
Département des sciences biologiques de l'environnement
Spécialité : Biologie de la conservation

Réf :.....

Mémoire de Fin de Cycle
En vue de l'obtention du diplôme

MASTER

Thème

Exploitation des ressources et logiciels de la bioinformatique pour la reconstruction phylogénétique de certaines espèces du genre : *Genista L.*

Présenté par :
ZOUGAGH HICHAM & GRAIT WALID

Soutenu le : 02/07/2024

Devant le jury composé de :

Mme. HAMAIDI Ghania	MAA	Présidente
Melle. BENMOUHOUB Hassina	MAA	Examinatrice
Mr.ADJRAD smail	MAA	Encadrant

Année universitaire : 2023 / 2024

Remerciements

Nous remercions le bon dieu de nos avoir accordé santé, énergie et volonté pour réaliser ce travail

Nous tenons à exprimer toute notre reconnaissance envers nos parents.

Nous remercions via ces quelques lignes toutes les personnes qui ont contribué de près à ou de loin l'élaboration de ce travail.

Nous remercions vivement notre promoteur monsieur ***ADJRAD Smail*** *pour son aide et son orientation set son conseil nécessaire à la réalisation de ce projet.*

Nous remercions également la présidente du jury ***Mme HAMAÏDI G****. et l'examinatrice* ***Melle BENMOUHOUBH.*** *pour avoir accepté d'évaluer ce travail et pour toutes leurs remarques et critiques.*

Enfin, que toutes les personnes ayant contribué à la réalisation de ce travail et qui n'ont Pas été citées ci-dessus sachent que je ne les oublie pas et que je les remercie chaleureusement.

Dédicace

Je dédie ce mémoire à

Mes très chers parents

qui sont toujours dans mon cœur, qui ont consacré leur vie à mon éducation et ma réussite, qui m'ont encouragé dans les moments les plus difficiles de ma vie.

Que Dieu les garde et les protège pour nous.

Mes sœurs, mes frères

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

Toutes mes amies.

Toutes les étudiantes de la faculté SNV.

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

HICHAM

Dédicace

Je dédie ce mémoire à

A mais chers Parents

Je vous remercie pour tout le soutien et l'amour que vous me

Portez depuis mon enfance et j'espère que votre bénédiction

m'accompagne toujours.

Un vraiment spécial dédicace pour

Mes frères et Mes sœurs

Mes cousins et cousines

Ainsi qu'à Mon binôme HICHAM et sa famille.

Toutes mes amies

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

WALID

Liste des abréviations

- **ADN: A**cide **D**ésoxyribo**N**ucléique.
- **ARN**: **A**cide **R**ibo**N**ucléique.
- **ATDB :A**rabidopsis **T**haliana DataBase
- **BIC :B**ayesian **I**nformation **C**riterion.
- **CAIC** : **C**orrected **A**kaike **I**nformation **C**riterion
- **DDBJ: D**NA **D**ata **B**ank of **J**apan.
- **DAMBE**: **D**ata Analysis in **M**olecular **B**iology and **E**volution.
- **EMBL: E**uropean **M**olecular **B**iology **L**aboratory.
- **EBI : I**nstitut **E**uropéen de **B**ioinformatique.
- **ECDC : E**uropean **C**entre for **D**isease **P**revention and **C**ontrol
- **ITS :I**nternal **T**ranscribed **S**pacer.
- **K2P :K**imura à **2 P**aramètres.
- **LANL : Los** Alamos **N**ational **L**aboratory.
- **NCBI:N**ational **C**enter for **B**iotechnology **I**nformation.
- **UPGMA**: **U**nweighted **P**air-**G**roup **M**ethod using **A**rithmetic Averages.
- **PAUP: P**hylogenetic **A**nalysis **U**sing **P**arsimony.
- **PDB: P**rotein **D**ata **B**ank.
- **PGD :P**lant **G**enome **D**atabase.
- **RDP :R**ibosomal **D**atabase **P**roject.
- **AIC :C**ritère d'**i**nformation d'**A**kaike
- **GTR: G**eneral **t**ime **r**eversible.
- **TN** :**T**ajima-**N**ei .
- **HKY** : **H**asegawa, **K**ishino et **Y**ano
- **NJ :N**eighbor-**J**oining.
- **PDB** : **P**rotein **D**ata **B**ank.
- **PAUP:P**hylogenetic **A**nalysis **U**sing **p**arsimony.
- **MV** : **M**aximum**V**raisemblance.
- **OTUs** :**O**perationelle **T**axonomic **U**nits

.

Listes des figures

Liste des tableaux

Remercîments
Dédicace
Liste des Abréviations
List des Figures
Liste des Tableau

Sommaire

Introduction

Introduction

La bioinformatique a joué un rôle crucial dans le développement de bases de données et d'outils essentiels pour les chercheurs en systématique phylogénétique. Ces ressources permettent aux scientifiques d'exploiter de vastes ensembles de données, de les analyser, de comparer des séquences génomiques et d'effectuer des analyses informatiques complexes. En fournissant des plateformes centralisées telles que Boldsystems pour le partage et l'accès aux informations biologiques, la bioinformatique a accéléré les découvertes scientifiques et favorisé la collaboration entre chercheurs à l'échelle mondiale **(Ratnasingham et Hebert, 2007).**

Les outils de la bioinformatique comprennent des programmes informatiques qui aident à révéler les mécanismes fondamentaux à la base des problèmes biologiques liés à la structure et fonction des macromolécules, des voies biochimiques, des processus pathologiques et évolutifs **(Korba ., 2020).**

L'utilisation de tels programmes devient de plus en plus une nécessité dans le domaine de la recherche biologique, plus partéculièrement dans la systématique phylogénétique, leur emploi permet une interprétation plus adéquate et plus rationnelle des données moléculaires (Saïb et Habibatni, 2014). Ainsi, plusieurs espèces végétales, animales et bactériennes ont du être reclassés dans de nouveaux genres auxquels ils n'appartenaient pas, et de nouvelles classifications basées sur des données moléculaires ont récemment été proposées **(Lecointre et Le Guyader, 2016).**

Donc, à l'heure où la disponibilité des sources de données utilisées dans les phylogénétiques augmente à un rythme exponentiel, il est devenu de plus en plus critique de comprendre les avantages et les désavantages de l'utilisation de diverses méthodes de reconstruction phylogénétique, et quelle méthode utiliser pour quel type de données, les mesures de la validation statistique des phylogénies et, enfin, les mesures de comparaison d'arbres phylogénétiques **(Nylander, 2001).**

Ainsi, l'objectif de ce présent travail est dans un premier temps, de traiter de manière aussi détaillée que possible la démarche de la systématique phylogénétique, depuis la

définition du problème et le choix et l'acquisition des données jusqu'à l'inférence des arbres, leur évaluation, leur comparaison et leur interprétation, et dans un deuxième temps, de maîtriser l'aspect technique de cette démarche en appliquant les méthodes les plus utilisées dans la reconstruction phylogénétique sur des données séquences nucléotidiques de 34 espèces du genre *Genista*.

Ce document est organisé en trois chapitres dont le premier est théoriques et comportent, respectivement, une introduction aux notions de bases de la phylogénie , les méthodes d'inférence d'arbres phylogénétiques, les méthodes d'évaluation et de comparaison des arbres phylogénétiques et les principales banques de données biologiques.

Nous présenterons également quelques logiciels de reconstruction phylogénétique, et la présentation du genre *genista* .

Les deux derniers chapitres concernent l'étude expérimentale effectuée, qui consiste à récupérer des séquences nucléotidiques de l'ADN ribosomique ITS1 et ITS2 à partir d'une banque de données, puis reconstruire, évaluer et comparer les phylogénies de 34 espèces du genre *Genista* en utilisant quatre grandes méthodes de reconstruction phylogénétique : UPGMA, Neighbor-Joining, la méthode de maximum de parcimonie et la méthode de maximum de vraisemblance. Enfin, nous terminons par une conclusion générale et perspective.

Synthèse bibliographiques

I.1. Définition de La phylogénie

La phylogénie est une science qui a pour objectif d'étudier les relations de parentés entre les espèces. Elle mêle des disciplines aussi variées que l'anatomie interne et externe **(Darlu *et*Tassy, 2004)**.

La phylogénie est une discipline importante en biologie car elle permet de comprendre l'origine et l'évolution de la biodiversité sur Terre. En étudiant les relations évolutives entre les espèces, les scientifiques peuvent également identifier les facteurs qui ont conduit à leur diversification, tels que les changements climatiques ou les pressions de sélection environnementale **(Futuyma, 2013).**

D'après **Gattoliat(2002),**le but d'une reconstruction phylogenitique ne se limite pas à vouloir reconstruire un arbre aussi proche que possible de l'histoire des taxa,mais égalment à estimer la manière dont les transformations des caractéres se répartissent dans l'arbre.

En résumé, la phylogénie est une méthode puissante pour étudier les relations évolutives entre les êtres vivants et comprendre comment la biodiversité sur Terre a évolué au fil du temps. Les scientifiques qui travaillent dans ce domaine utilisent différentes méthodes et données pour reconstruire l'histoire évolutive des espèces et établir les liens de parenté entre elles **(Wiley *et* Lieberman, 2011).**

I.2. Les arbres phylogénétiques

Un arbre phylogénétique appelé dendrogramme**(Figure1)** est une représentation graphique de la phylogénie. Il exprime les liens entre les taxa sous la forme d'une succession debranchements **(Rasmont, 1997 ; Schmidt, 2003).**

- Les sommets externes sont appelés **feuilles** (OTUs) ils représentent les unités Taxonomiques ; c'est la seule partie basée sur l'observation.
- Les sommets internes appelés **nœuds**, et ils représentent l'ancêtre communhypothétique dans le sens ou leur existence n'est pas fondée sur l'observation,mais sur le processus de reconstruction.
- La relation entre deux nœuds appelés **branche**, les branches peuvent être évaluées,C'est-à-dire que l'on peut leur associer une mesure (une distance, une quantitéd'évolution, un nombre de mutation).

- La **racine** représente un ancêtre commun des espèces traitées. Les liens entre les nœuds et les feuilles sont orientées.

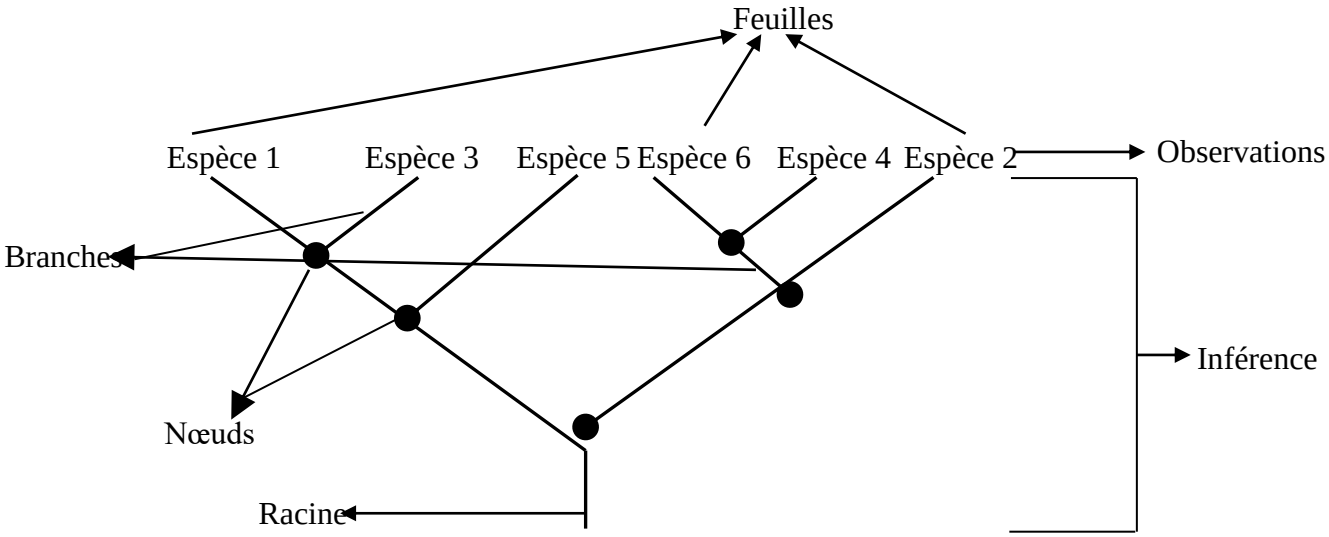


Figure 1 : Schéma des différentes notations concernant la représentation d'un arbre phylogénétique **(Daniel Richard *et al*.2014).**

I.3. Les arbres enracinés et les arbres non enracinés

On distingue deux grands types d'arbres (**Tahiri, 2019) :**

a) Un arbre enraciné est un arbre dans lequel un des nœuds est désigné pour être la racine (c.-à-d., un ancêtre commun), Cet arbre est orienté et cette orientation correspond au temps d'évolution des taxons **(Figure 2-a).**

b) Un arbre non enraciné est une représentation intemporelle des relations phylogénétiques, les liens entre nœuds ne sont pas orientés, et un seul et unique chemin permet de passer d'un sommet à l'autre. Cet arbre n'induit donc aucune hiérarchie **(Figure 2- b).**

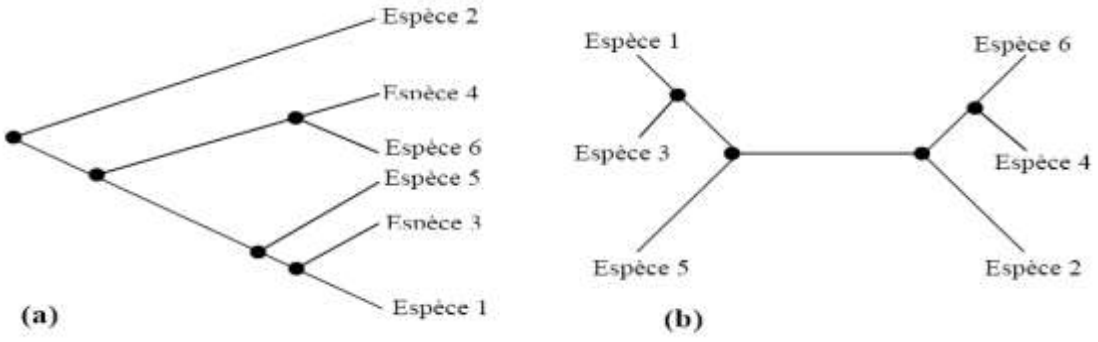


Figure 2 : Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné (**Tahiri, 2019).**

I.3.1 Enracinement par un groupe externe

La racine désigne le nœud le plus ancien d'une reconstruction phylogénétique, c'est une représentation de l'ancêtre de l'ensemble de taxons étudiés. C'est également la reconnaissance de l'origine et l'orientation. Une analyse phylogénétique ne peut pas résoudre la position du nœud racine sans l'aide d'un groupe externe qui signifie un groupe de taxons extérieurs. Toutefois, pour enraciner un arbre, il faut que le groupe externe ne soit pas trop éloigné sur le plan évolutif pour permettre de détecter les caractères homologues **(Thomas et al, 2016).**

I.4 Les groupes taxonomiques

D'après **Rasmont (1997)** et **Tourasse (1992),** dans le domaine de la reconstruction phylogénétique, on distingue trois groupes taxonomiques : le groupe monophylétique, le groupe paraphylétique et le groupe polyphylétique :

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

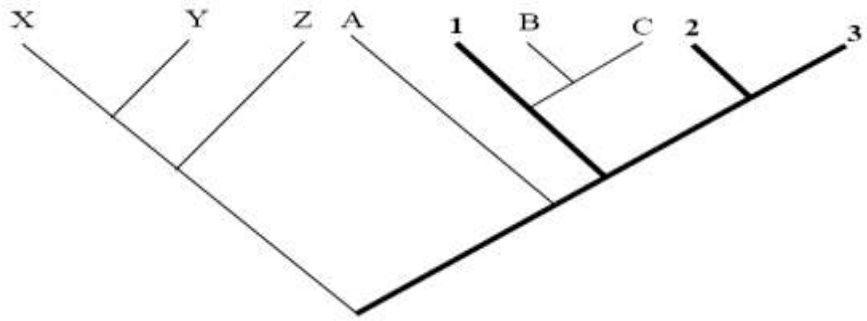


Figure 3 : Diverses catégories de groupes taxonomiques. Les taxa X, Y, Z forment un groupe monophylétique. Les taxa 1, 2 et 3 forment un groupe paraphylétique. Les taxa A, B et C forment un groupe polyphylétique **(Tourasse, 1992).**

I.5 Différentes représentations graphiques pour les arbres

Selon les méthodes par lesquelles ils ont été construits, il existe trois types d'arbres phylogénétiques (dendrogrammes - **Figure 4**)**(Erwan Corrce, 2013).**

- **Phénogramme** : Dendrogramme obtenu par les méthodes de distance (méthodes phénétiques), où les relations entre taxa expriment des degrés de similitude globale.
- **Cladogramme** : Dendrogramme exprimant les relations phylogénétiques entre taxa et construit à partir de l'analyse cladistique, elle-même basée sur la parcimonie.
- **Phylogramme** : Dendrogramme dont la longueur des branches est proportionnelle au nombre de changements évolutifs, il est obtenu par la méthode de maximum de vraisemblance.

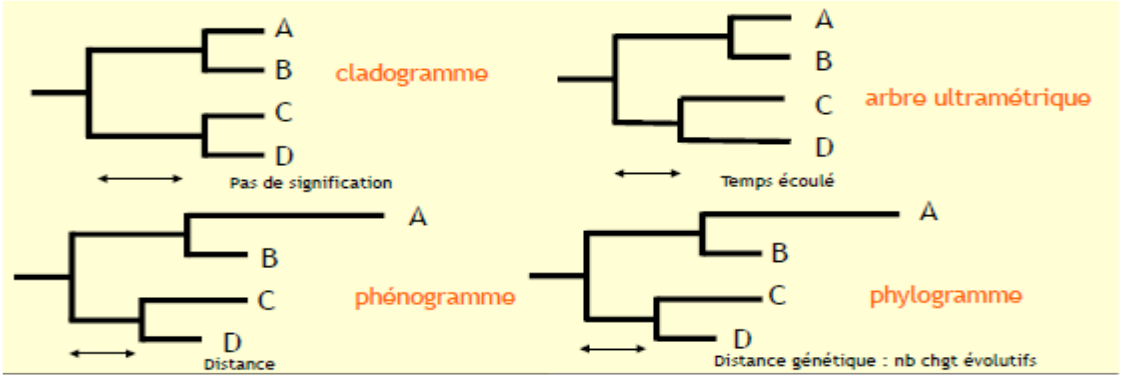


Figure 4 : Schéma représentant les différents types de dendrogrammes **(Erwan Corre ,2013)**

I.6 Les différents types de caractères utilisés en phylogénie

Initialement, la classification n'était basée que sur les caractéristiques morphologiques, qui comprennent les caractères observables (liés aux différents états : morphologiques, biochimiques et physiologiques) et les patterns binaires (présence d'un caractère donné / absence de ce même caractère). Cependant, suite au développement de la théorie de l'évolution, la systématique phylogénétique prend en compte tous les caractères héritables, depuis ce qui est visible (anatomie et morphologie, fondement de la classification traditionnelle, etc.) jusqu'aux séquences d'ADN et d'ARN, en passant par les protéines et les données de la paléontologie **(Peeters *et* al, 2009).**

Plusieurs régions du génome sont proposées pour l'utilisation en phylogénie chez les plantes. Il s'agit notamment de Ribulose bisphosphate carboxylase large chain (rbcL), Megakaryocyte-Associated Tyrosine Kinase (matK) et des régions d'espaceurs intergéniques ribosomiques (ITS), séquences des cytochromes (**Chase *et al.* 2005; Kress *et* Erickson, 2007; Lahaye *et al.* 2008).**

Les régions ITS sont les plus fréquemment utilisées pour les analyses phylogénétiques au niveau des genres et des espèces **(Coleman, 2003),** Les ITS sont des régions variables, relativement courtes et faciles à séquencer. Elles sont utiles pour l'identification des plantes.**(Baldwin *et al*. 1995).**

Les deux espaceurs transcrits (interne et externe) sont comme leurs noms l'indiquent, des séquences séparant les zones codantes pour les ARN ribosomiques (5,8S, 18S, et 26S) **(Figure 5)** qui se trouvent transcrits lors de l'expression de ces gènes **(Baldwin *et al.* 1995 ; Bodo Slotta, 2000).**

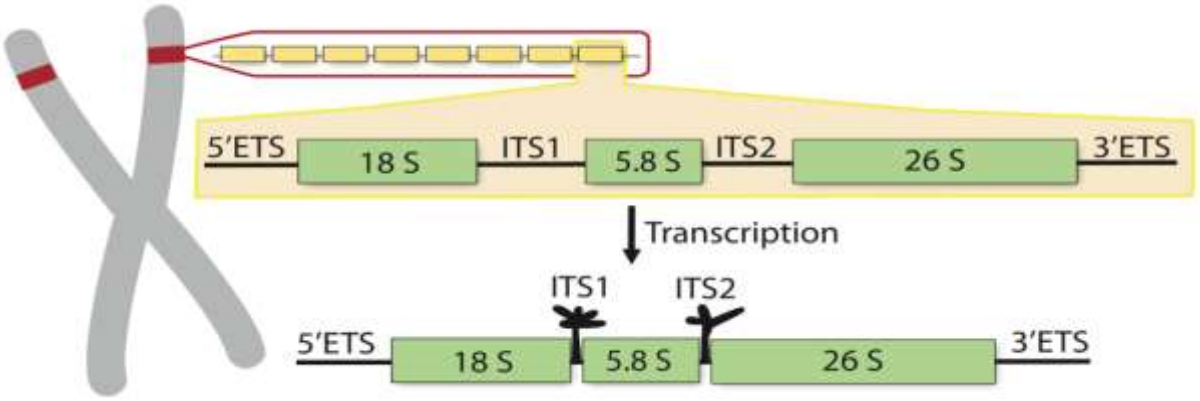


Figure 5 : Structure de la région de l'ADNr chez les plantes **(Edger *et al.*, 2014)**

I.7 La reconstruction phylogénétique

I.7.1. Alignement des séquences

L'alignement des séquences est une étape fondamentale pour toute reconstruction phylogénétique à partir des caractères moléculaires (ADN ou protéines) **(Tahiri, 2019 ; Sater, 2021).**

En bioinformatique, aligner deux séquences consiste à réarranger des chaines d'ADN, ARN ou de protéines dans le but de trouver les zones de fortes similarités entre ces séquences. L'alignement se fait par l'ajout de «gaps » dans chacune des deux séquences afin d'avoir un maximum de concordance entre les éléments des séquences se trouvant aux mêmes positions **(Bouchair, 2014).**

En pratique, la reconstruction phylogénétique nécessite l'alignement de plusieurs séquences **(Guindon, 2003).** Dans ce cas, on parle d'alignement multiple, c'est-à-dire alignement de plus de deux séquences, il permet de repérer des parties bien conservées entre toutes les séquences mises en jeu **(Varré, 2000).**

I.7.2. Nettoyage de l'alignement multiple des séquences

Le nettoyage des séquences est une étape qui vise à éliminer les positions mal alignées et les régions divergentes ce qui permet de supprimer le bruit d'alignement.après avoir terminé l'alignement multiple à l'aide de différentes méthodes **(Edgar, 2004)**

I.8 Les méthodes de reconstruction phylogénétiques

On distingue trois méthodes de reconstruction phylogénétiques: les méthodes de distances, de parcimonie et probabilistes utilisant la vraisemblance **(Hassel, 2015).**

I.8.1. Les méthodes de distances

Aussi appelées méthodes phénétiques, elles se proposent de reconstruire des arbres sans racine en se basant sur les ressemblances observées entre chaque paire d'unités évolutives. Ces ressemblances sont exprimées à travers des mesures de distances basées sur le nombre de substitutions qui se sont produites sur les lignées évolutives depuis l'ancêtre **(Zein Eddine, 2014).** Pour ces méthodes, plus la ressemblance entre deux unités est importante, plus leurs liens de parenté sont étroits **(Cheikh rouhou, 2006).**

I.8.1. 1 Calcul des distances

Un indice de similarité à partir d'un alignement de deux séquences peut être obtenu par la mesure de nombre de sites où l'on retrouve une correspondance entre deux résidus divisé par le nombre total de sites. Une mesure de dissimilarité peut être obtenue en mesurant cette fois le rapport entre le nombre de sites mutés et le nombre total de sites **(Darlu *et* Tassy, 1993 ; Salemi et Vandamme, 2003).**

Toutefois, ce simple comptage des différences observées sous-estime le nombre réel d'événements de substitutions (transitions et transversions (**Figure 6**) qui ont eu lieu parce qu'il ignore la possibilité de changements multiples successifs (cachés) en un même site **(Tableau I).** C'est pourquoi plusieurs modèles représentant le processus de substitution ont été développés afin d'estimer la « vraie » distance. Ces modèles sont appelés modèles d'évolution et la distance calculée à partir de ces modèles est appelée distance évolutive **(Tourasse, 1992, Cassan, 2017)**

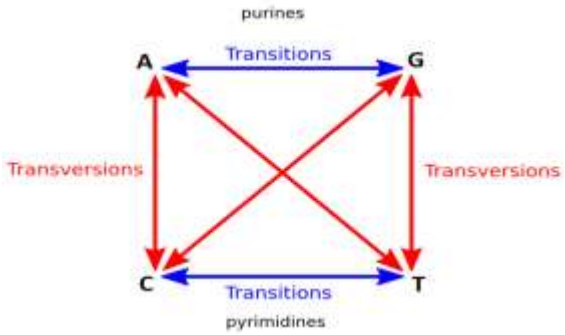


Figure 6 : Définition de la transversion et de la transition **(Tourasse, 1992)**

La distance évolutive entre les séquences correspond au nombre moyen de mutations par site ayant eu lieu depuis que les séquences ont divergé de leur ancêtre commun (**Cassan, 2017).**

Tableau 1 : Comparaison entre le nombre de substitutions réelles et observées, d'après Graur et **Li (2000)(Anne Denoual, 2017)**

	Séquence 1	**Séquence 2**	**Nombre de substitutions observées**	**Nombre de substitutions réelles**
Substitution unique	**T**	**T➔A**	**1**	**1**
Substitutions multiples	**A**	**T➔G➔T**	**1**	**2**
Substitutions coïncidentes sur le même site	**A ➔C**	**A➔G**	**1**	**2**
Substitutions parallèles	**T➔A**	**T➔A**	**0**	**2**
Substitutions convergentes	**A➔C➔T**	**A➔T**	**0**	**3**
Substitutions inversées	**T➔A➔T**	**T**	**0**	**2**

I.8.1.2 Modèles d'évolution des séquences

Les modèles d'évolution ont pour but de modéliser les divers types de mutation affectant l'ADN, et d'expliquer ainsi certaines différences fonctionnelles ou structurales entre les organismes. Ils permettent les calculs de distances évolutives et surtout la construction d'arbres phylogénétiques exhibant les relations de parenté entre espèces. Les modèles d'évolution les plus utilisés sont : **(Zein Eddine, 2014; Cassan, 2017)** :

a) Modèle de Jukes-Cantor (JC)

Ce modèle, souvent noté JC69, suppose que les sites évoluent indépendamment les uns des autres et que chaque substitution a le même taux d'occurrence, c'est-à-dire les quatre bases ont les mêmes fréquences ($\pi A = \pi T = \pi G = \pi C$). Les transitions ($\alpha$) et les transversions ($\beta$) sont équiprobables ($\alpha = \beta$).

b) Modèle de Kimura (K2P)

D'après ce modèle, souvent noté K80, les quatre bases ont les mêmes fréquences ($\pi A = \pi T = \pi G = \pi C$). Par contre les transitions (α) et les transversions (β) ne sont pas équiprobables ($\alpha \neq \beta$), car dans la réalité les transitions (T C et A G) sont plus fréquentes que les transversions (T ou C A ou G).

c) Modèle de Tajima-Nei (TN)

Il est basé sur l'hypothèse selon laquelle les quatre bases n'ont pas les mêmes fréquences ($\pi A \neq \pi T \neq \pi G \neq \pi C$). Par contre, les transitions (α) et les transversions (β) sont équiprobables ($\alpha = \beta$).

d) Modèle de Hasegawa, Kishino et Yano (HKY 85)

Suivant ce modèle, les quatre bases n'ont pas les mêmes fréquences ($\pi A \neq \pi T \neq \pi G \neq \pi C$), et les transitions ($\alpha$) et les transversions ($\beta$) ne sont pas équiprobables ($\alpha \neq \beta$).

e) Modèle GTR (General time reversible)

Dans ce modèle également, les quatre bases ont des fréquences différentes ($\pi A \neq \pi T \neq \pi G \neq \pi C$), et il existe 6 types de substitutions (AC, AT, AG, CT, CG, TG).

I.8.1.3. Le choix du modèle

Avec autant de modèles disponibles, la question se pose de savoir lequel s'ajuste le mieux à nos données. Dans la littérature il y a trois critères qui sont les plus fréquemment utilisés pour la comparaison et choix du modèle **(Botero-Castro, 2014).**

a) Le critère d'information d'Akaike (AIC) :

C'est le premier critère qui a été proposé par Akaike (1973), compare les vraisemblances de deux modèles alternatifs tout en pénalisant le nombre de paramètres du modèle d'évolution **(Botero-Castro, 2014).**

Le modèle d'évolution qui présente la valeur AIC la plus petite est considéré comme le plus probable. Les modèles les plus riches en paramètres auront nécessairement une meilleure adéquation aux données **(Hassel, 2015).**

b) critère d'information d'Akaike corrigé (AICc)

Le deuxième critère est apparu quelques années plus tard, comme une correction à l'AIC (AICc) qui cherche à tenir compte en plus des paramètres du modèle d'évolution, de la taille de l'échantillon **[Hurvich et Tsai, 1989], (Botero-Castro, 2014, El Alaoui, 2017).**

c) Le critère d'information bayésienne (BIC)

Le critère d'information bayésien (en anglais bayesian information criterion ou BIC) est un critère d'information dérivé du critère d'Akaike. À la différence de ce dernier, le BIC dans son calcul va pénaliser de manière plus importante le nombre de paramètres du modèle et la taille de l'échantillon (nombre de sites dans un alignement) **(Botero-Castro, 2014).**

I.8.1.4. Les différents types de méthodes de distances

Plusieurs méthodes ont été développées pour construire un arbre phylogénétique à partir d'une matrice de distance. Parmi ces méthodes, on peut citer :

a) La méthode d'UPGMA

C'est la plus simple pour construire des phénogrammes, elle suppose que les taux de substitution entre séquences sont à peu près homogènes et permet d'estimer des arbres ultra métriques. C'est une technique qui utilise un algorithme regroupant les séquences séparées par les distances les plus courtes, par ordre de similarité **(Brochier, 2007)**.

La première étape consiste à identifier les deux taxons ayant la plus petite distance. Ces deux taxons sont ensuite traités comme une seule unité et l'on recherche parmi les taxons restants celui ayant la distance la plus petite avec cette unité, et ainsi de suite **(Bouzaza, 2018).**

b) La méthode de Neighbor-Joining

Elle est basée sur le calcul de distances évolutives séparant des séquences homologues. Elle n'implique pas l'hypothèse d'horloge moléculaire, et tient compte des différences de vitesse d'évolution entre les différentes branches de l'arbre phylogénétique. Par conséquent, elle constitue la méthode de distances la plus souvent utilisée **(Zein Eddine, 2014; Cassan, 2017).**

I.8.1.5.Avantages et désavantages des méthodes de distance

- **Avantages :** les méthodes de distance ont l'avantage d'être rapides et permettent d'analyser de grandes bases de données, de tester un grand nombre d'hypothèses alternatives, et aussi d'intégrer des modèles de changements évolutifs **(Botero-Castro, 2014 ; Zein Eddine, 2014).**

- **Désavantages :** les méthodes de distance réduisent la matrice de caractères à une matrice de distances ce qui induit une perte d'information. Toutefois, elles ne permettent pas de combiner dans une même matrice des caractères de nature différente (morphologique ou moléculaire). Elle produit des arbres sans racines **(Botero-Castro, 2014; Zein Eddine, 2014).**

I.8.2 Les méthodes de parcimonie

La méthode de maximum de parcimonie est l'une des premières méthodes utilisées pour la construction d'arbre phylogénétique. Celle-ci considère que l'arbre reflétant le mieux la réalité est celui qui minimise le nombre de mutations le long des branches.

Il faut donc dans un premier temps, estimer les séquences des noeuds internes induisant le moins de mutations possibles, puis calculer le score de parcimonie prenant en compte lesmutations le long de l'arbre et enfin chercher la topologie disposant du score le plus faible dans l'ensemble des topologies possibles **(Cassan, 2017).**

I.8.2.1 Avantages et désavantages des méthodes de parcimonie

- **Avantages :** la parcimonie permet l'analyse des caractères morphologiques ou des données fossiles, de même que l'évaluation des différents arbres **(Zein Eddine, 2014).**
- **Désavantages :** cette méthode ne s'applique qu'à des caractères discrets. Les résultats obtenus dépendent du taux de substitution, qui ne doit pas être trop élevé. De plus, le calcul des substitutions ne tient pas compte de la longueur des branches («attraction des longues branches»)**(Felsenstein 1978) (Botero-Castro, 2014; Zein Eddine, 2014).**

I.8.3 La méthode de maximum de vraisemblance

C'est une méthode qui repose sur une approche probabiliste basée sur un modèle d'évolution prédéfini. Grâce à ce modèle, le maximum de vraisemblance permet de calculer pour chaque site d'un alignement de séquences une probabilité d'observer un état de caractère donné. Parmi l'ensemble des topologies ainsi considérées, celle qui présente la plus grande vraisemblance est alors retenue. (**Cavé-radet, 2018**).

I.8.3.1 Avantages et désavantages de la méthode de maximum de vraisemblance

- **Avantages :** la méthode de MV est considérée comme l'une des plus fiables, car elle permet d'obtenir le résultat le plus proche de l'arbre évolutif réel. En plus, elle permet d'appliquer différents modèles d'évolution et d'estimer la longueur des branches en fonction des changements évolutifs **(Zein Eddine, 2014, Cassan, 2017).**

- **Désavantages :** cette méthode est la plus longue (temps de calcul lourds) vu le nombre important d'estimations nécessaires à faire **(Zein Eddine, 2014, Cassan, 2017).**

I.9 Fiabilité des arbres phylogénétiques

Plusieurs tests statistiques ont été élaborés pour évaluer la fiabilité des arbres phylogénétiques. Les deux méthodes les plus fréquemment utilisées sont : bootstrap ou de jackknife **(Tahiri, 2012).** Cette mesure indiquera le niveau de robustesse de l'arbre à étudier. Si les modifications de ré échantillonnage ont une faible influence, alors cela indiquera que l'arbre estimé est robuste. Dans le cas contraire, l'arbre sera peu robuste, et de ce fait les analyses qui se baseront sur cet arbre seront peu fiables **(Tahiri, 2012).**

Les scores de robustesse, mesurés en utilisant les techniques de bootstrap ou de jackknife, sont d'une grande importance puisqu'ils expriment formellement un taux de confiance statistique associé à l'analyse phylogénétique effectuée **(Tahiri, 2019).**

I.9.1 Le bootstrapping

C'est une méthode statistique utilisée pour évaluer la robustesse d'une reconstruction phylogénétique **(Zein Eddine, 2014).**C'est un tirage avec remise des caractères de la matrice de départ permettant de constituer de nouvelles matrices de taille identique. En d'autres termes, ce test permet d'évaluer la résistance de la reconstruction suite à une perturbation du jeu de données initiales par pondération aléatoire des différents caractères. Chacune de ces nouvelles matrices est soumise à une recherche de l'arbre le plus parcimonieux. Ces arbres sont ensuite combinés en un seul arbre de consensus majoritaire. Les clades apparaissent dans plus de 50 % des cas, recevant alors une valeur de bootstrap comprise entre 50 et 100 % (un soutien supérieur à 70 % est considéré comme bon selon **(Hillis et Bull, 1993)**

I.9.2 Le Jackknife

Cette méthode a été appliquée aux problèmes de phylogénie par **(Mueller et Ayala (1982).** Supposons une matrice de données constituée de k caractères. Le jackknife consiste à effectuer k reconstructions phylogénétiques différentes. Chacune d'elles ayant été obtenue en supprimant un caractère différent **(Gattoliat, 2002).**

Par exemple, si la matrice X est constituée de k caractères, le jackknife crée k matrices de taille inférieure à la taille de la matrice originale X. Les k matrices sont obtenus par soustraction à chaque fois d'un seul caractère à partir de la matrice initiale. Ensuite un arbre est construit à partir de chacune des k matrices différentes. Enfin, comme la technique bootstrap, pour chaque branche, on obtient une proportion jackknife en relevant le

pourcentage de fois où elle apparaît dans les différentes phylogénies inférées à partir des matrices issues de ce tirage **(Darlu et Tassy, 1993).**

I.10. Comparaison des arbres phylogénétiques

Il existe plusieurs mesures de comparaison d'arbres phylogénétiques parmi les plus populaires, nous retrouvons : la distance topologique de robinson et foulds "RF" (1981) **(Tahiri, 2019).**

I.10.1. La distance de Robinson et Foulds (1981) (RF)

Cette dernière méthode de comparaison est une mesure de la distance topologique entre deux arbres phylogénétiques. Elle est égale au nombre minimal d'opérations élémentaires nécessaires pour transformer un arbre en un autre. On entend par opérations élémentaires la fusion ou la séparation de deux noeuds **(Tahiri, 2019, Randriamihamison, 2021).** La **Figure 7**présente deux arbres T et T'ayant tous deux 5 feuilles identiques.

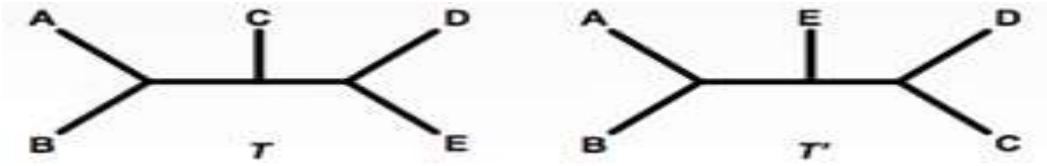


Figure 7 : Deux arbres T et T' de taille 5, avec des taxons **(Tahiri, 2019).**

On applique la distance RF en se basant sur les couples obtenus par la séparation au niveau des branches internes. Pour passer de T à T' on voit qu'il est nécessaire de fusionner C à D et E puis de séparer E de D et C, soit deux opérations donc DRF = 2 **(Tahiri, 2019).**

I.10.2 La Branch Score Distance (Bs)

La Branche Score Distance **(Kuhner et Felsenstein, 1994)**"Bs" est basée sur une représentation des arbres par leurs longueurs de branches (Randriamihamison, 2021).Les deux arbres ayant les mêmes ensembles de feuilles, chaque arbre est représenté par l'ensemble des longueurs de branches pour toutes les branches possibles, même celles n'existant pas forcément dans les arbres considérés. Si la branche n'est pas présente dans l'arbre, alors la valeur associée est 0 sinon, il s'agit de sa longueur dans l'arbre. La Branche Score Distance est alors la somme des carrés des écarts des longueurs des branches des deux arbres (un exemple de calcul de la distance de Bs) est représenté dans la**(Figure 8) (Aboukhalil, 2016):**

$$Bs = (x_a - y_a)^2 + (x_b - y_b)^2 + (x_c - y_c)^2 + (x_d - y_d)^2 + (x_{ab} - 0)^2 + (y_{ac} - 0)^2$$

Figure 8 : Méthode de calcul de la distance Bs entre deux arbres T et T' de taille 4, avec des taxons similaires.

I.11. La recherche du meilleur arbre

Le choix d'une méthode de reconstruction phylogénétique est laissé au libre arbitre du biologiste, mais il est recommandé de tester plusieurs méthodes différentes pour un jeu de données. Les progrès réalisés en biologie moléculaire permettent à présent d'obtenir rapidement de grandes quantités de séquences. De nombreux jeux de données peuvent ainsi

II.1. Définition des bases des données

Depuis que les biologistes travaillent avec des séquences en grande quantité, c'est-à-dire depuis le développement et la généralisation de l'utilisation des méthodes rapides de séquençage, la nécessité d'organiser et d'accéder aisément à ces données s'est fait ressentir. Les premières banques de données qui ont été créées concernent les informations structurales sur les protéines puis, très rapidement après, les séquences protéiques et nucléotidiques **(Perière, 2000).**

II.2. Les différents types de banques de données

Aujourd'hui on peut distinguer les banques dites généralistes des banques spécialistes **(Vergnaud, 2002 ; Pillet, 2000).** Les premières ont pour vocation d'être les plus exhaustives possibles, c'est-à-dire rassembler la totalité des séquences ou rappels bibliographique informations connues pour tout l'ensemble des espèces avec ou sans expertise particulière. Les secondes se sont constituées autour d'une thématique biologique, afin de réunir les séquences d'une famille protéique pour toutes les espèces. Elles dérivent le plus souvent des banques généralistes et nécessitent l'intervention d'experts qui sont la plupart du temps les auteurs de la banque **(Dorkeld, 1994 ; Blanchet, 1999).**

II.2.1. Les banques généralistes

Lesprincipales collections généralistes de séquences nucléotidiques et protéiques sont : Boldsystems, GenBank (NCBI), EMBL et DDBJ, Celles-ci sont régulièrement mises à jour **(Pillet, 2000).**

a)BoldSystems

C'est une plateforme web qui offre un environnement intégré pour l'assemblage et L'utilisation de donnes de codes à barres ADN. Il fournit une base de données en ligne pour la collecte et la gestion de données sur les spécimens. Le lien d'identification en BOLD permet d'accéder aux moteurs d'identification des animaux, des plantes et des champignons basés sur le COI, matK, rbcL et ses gènes. Cette ressource est disponible sans avoir besoin d'un compte d'utilisateur, bien que des fonctionnalités avancées soient disponibles pour ceux qui sont enregistrés auprès du système **(Ratnasingham *et al.* 2013).**

b) GenBank

La GenBank est une collection annotée de toutes les séquences d'ADN publiquement disponibles. Cette banque est mise à jour régulièrement grâce à des échanges quotidiens de séquences avec la banque européenne EMBL et la banque japonaise DDBJ **(Perrière, 2000).**

c) EMBL

La European Molecular Biology Laboratory (EMBL) a été créée en 1980 à Hiedelberg **(Perrière, 2000 ; Guindon, 2003).** Elle est gérée et maintenue depuis 1994 par l'Institut Européen de Bioinformatique (EBI) à Cambridge. C'est la première source européenne de données sur les séquences nucléiques **(Pillet, 2000).**

d) DDBJ

La DDBJ (Dna Data Bank of Japan) est une banque Japonaise de données créée en 1984 et maintenue au NIG à Mishima **(Perrière, 2000).** Les données présentes dans ces banques proviennent désormais en quasi-totalité de soumissions directes effectuées par les auteurs, par l'intermédiaire du réseau Internet **(Golding et *al.*2003).**

II.2.2 Les Banques spécialistes

Du fait de l'augmentation exponentielle du nombre de séquences nucléotidiques et protéiques publiées, la nécessité de développer des banques spécialisées est rapidement apparue. Ainsi, aujourd'hui, il en existe de nombreuses. Les informations disponibles dans ces banques et leur mise à jour sont très variables **(Blanchet, 1999).**

D'après **(Perrière, 2000**), on distingue deux types de banques spécialisées : les banques thématiques et les banques génomiques.

a)Les banques thématiques

Ces banques se consacrent à une thématique biologique bien précise (par exemple : banques sur les structures moléculaires des protéines, des complexes protéines-acides nucléiques, de virus et de polyosides, banques sur les structures de l'ARN, banques sur des familles de gènes…etc.). Ces banques intègrent donc des séquences et des données qui sont spécifiques à ce domaine. On peut par exemple citer :

- **La base PDB (Protein Data Bank):** est la plus connue des banques des données sur les structures moléculaires. Elle rassemble les structures tridimensionnelles des macromolécules obtenues par cristallographie aux rayons x. Elle contient aussi des références bibliographiques et des données sur la structure primaire et secondaire des protéines **(Golding et al. 2003).**

- **La base RDP (Ribosomal Database Project):** est une banque américaine qui contient non seulement les séquences des ARN ribosomiques de la grande et de la petite sous-unité du ribosome, mais aussi des alignements et des arbres phylogénétiques **(Perrière, 2000).**

b) Les bases de données génomiques

Ce sont des banques de données centrées sur les génomes des organismes étudiés en biologie, c'est-à-dire l'ensemble des séquences d'ADN contenues dans les chromosomes d'organismes spécifiques. Ces organismes modèles peuvent être des virus, bactéries, champignons, plantes et animaux. En plus des données de séquences, ces bases rassemblent aussi d'autres types de données (description de phénotypes, mutations, expression de gènes, bibliographies…etc.) **(Pillet, 2000).** Dans cette catégorie on peut citer :

- **ECDC (European Centre for Disease Prevention and Control):** est une banque qui présente une compilation des séquences *d'E. Coli* contenues dans EMBL et dans GenBank.
- **PGD (Plant Genome Database):**est une base de données qui compile des données sur le génome de plusieurs plantes céréalières. Par ailleurs, ATDB« *Arabidopsis thaliana* DataBase » rassemble toutes les informations sur les données génétiques, le clonage et le séquençage des molécules de la plante crucifère *Arabidopsis thaliana.*

II.2.3 Format de stockage des données

Le format d'une séquence peut être défini par l'ensemble des règles (contraintes) de présentation auxquelles sont soumises la ou les séquences dans un fichier donné**(Vergnaud, 2002).** Ainsi, le format permet donc :

- Une mise en forme automatisée.
- Le stockage homogène de l'information.
- Le traitement informatique ultérieur de l'information.

a) Format EMBL, Genbank et DDBJ

Le format de stockage utilisé par EMBL **(Annexe 1, Tableau 9**) est différent de celui utilisé conjointement par GenBank **(Annexe 1, Tableau 10)** et la DDJB **(Annexe 1, Tableau11).** Cette différence ne porte que sur la façon de représenter les données mais la philosophie générale de ces systèmes est la même **(Perrière, 2000).** Le format utilisé par EMBL étant cependant plus « lisible » si l'on se place dans la perspective du développement de programmes accédant aux données des banques.

Dans le cas de GenBank et de la DDBJ, ces champs sont indiqués par des identificateurs organisés sur deux niveaux dont la séparation est repérable par des espaces différents. Le

niveau 1 correspond aux colonnes des champs dans le cas de format EMBL ; dans ce niveau on trouve aussi des informations d'ordre général concernant la séquence en question. Par exemple les champs ID, AC, KW et OS dans le format EMBL correspondent respectivement aux : Locus, Accession, Keywords, Organism dans le format GenBank et DDBJ **(Golding et Morton, 2000).**

b)Formats spécifiques aux logiciels d'analyse phylogénétiques

D'après **Blanchet (1999),** il existe plusieurs autres formats pour les fichiers de séquences, allant de plus simple (Staden et Fasta) aux beaucoup plus compliqués. Certains de ces formats sont spécifiques aux logiciels qui les utilisant. Par exemple le format Phylip **(Annexe 2, Tableau 12)** est spécifique au package PHYLIP de **Felsenstein (2004)**. Le format NEXUS est spécifique aux logiciels PAUP de **Swoford (1997).**

Le format Staden est le plus ancien et le plus simple : c'est une suite de lettres (bases nucléotidiques ou acides aminés) de la séquence organisée en chacune terminée par un retour à la ligne (80 caractères maximum). Ce format n'autorise qu'une séquence par fichier.

Le format FASTA **(Annexe 2, Tableau 13)** est également simple. En effet, ce format de séquence ne contient qu'un minimum d'informations. Il contient le signe «>» tout au début de l'entrée (pour indiquer le début d'une nouvelle séquence) suivi du numéro d'accession auxbanques de données, et ensuite le nom de l'espèce, et enfin la nature de la molécule séquencée et, en revenant à la ligne, la séquence proprement dite disposée en lignes **(Vergnaud, 2002).**

II.3 Logiciels d'analyse phylogénétique

Aujourd'hui, il existe une myriade de logiciels, implantés sur différentes machines depuis le microordinateur jusqu'à des ordinateurs les plus puissants, écrits dans tel ou tel langage, traitant tel ou tel problème.

Dans le domaine de la reconstruction phylogénétique, **Felsenstein (2004)** a énuméré sur son site web, plus de 392 paquets de programmes et 54 serveurs web gratuits traitant les différents aspects de la phylogénie, allant des problèmes de comparaison des séquences, calcul des distances évolutives entre des séquences nucléotidiques ou entre des séquences protéiques, reconstruction et visualisation d'arbres phylogénétiques et l'évaluation de la fiabilité des arbres phylogénétiques inférés par les différentes méthodes. Dans cette section, nous décrivons les logiciels les plus utilisés dans les analyses phylogénétiques.

II.3.1 Logiciels d'alignement multiple des séquences

a) Clustal

ClustalW (Des Higgins) est le programme d'alignement multiple le plus employé dans la reconstruction phylogénétique, Il y a eu de nombreuses versions de Clustal au cours du développement de l'algorithme : Clustal V, ClustalW, ClustalX(c'est ClustalW avec interface graphique**)** et CLUSTAL Ω (Omega)**(Bouarour, 2022).**CLUSTALW est fondé sur l'utilisation d'un algorithme d'alignement progressif.Les séquences les plus similaires sont alignées en premier puis l'alignement vers les séquences les plus distantes. C'est également un programme de construction d'arbre phylogénétique.**(Bouarour, 2022).**

b) T-COFFEE

T-COFFEE est fondé également sur un alignement progressif .En plus de realiser un alignement global entre chacune des paires de séquence, il procède à un alignement local afin d'optimiser l'alignement entre les séquences très divergentes **(Noterdame,Higgins ,2000)**

II.3.2 Logiciels de nettoyage de l'alignement multiple des séquences

a) Gblocks

Gblocks est un outil de nettoyage d'alignement, c'est à dire capable de sélectionner un sous-ensemble de sites dans l'alignement qui augmentera théoriquement la qualité du signal, en comparaison de l'alignement bruité complet. GBlocks considère la sélection des sites par blocs conservés, et non pas en considérant chaque site indépendamment. C'est un avantage théorique intéressant de considérer qu'un ensemble de sites conservés consécutifs a une valeur plus importante que des sites potentiellement isolés **(Larivière, 2016).**

II.4.Logiciels de reconstruction phylogénétique

II.4.1 Le paquet Phylip

Phylip (PHYLogeny Inference Package), est un ensemble d'environ 34 programmes informatiques permettant de faire de la reconstruction phylogénétique. Il a été développé par Joseph Felsenstein entre 1980 à 1995 **(Golding et Morton, 2003 ; Felsenstein, 2004).**

II.4.2. Le logiciel PAUP

PAUP (Phylogenetic Analysis Using Parsimony) est un logiciel de reconstruction phylogénétique mis au point par D.L Swofford entre 1989 et 1998. C'est le logiciel le plus cité dans la littérature scientifique moderne. Il a été conçu spécialement pour faire desanalyses phylogénétiques selon la méthode de parcimonie, puis il a été élargi pour inclure d'autres types d'analyses phylogénétiques telles que les méthodes de distances et les méthodesde

maximum de vraisemblance, et pour réaliser des tests statistiques telles que : le bootstrap et le jackknife **(Swofford, 1998).**

II.4.3.MEGA

MEGA est un logiciel gratuit et convivial qui permet d'effectuer de multiples tâches liées à l'évolution moléculaire et à la phylogénétique. IL peut gérer de grands ensembles de données de séquences d'ADN, d'ARN ou de protéines, et effectuer l'alignement, l'édition, l'annotation et la traduction. IL peut également calculer des distances évolutives, tester des hypothèses et déduire des arbres phylogénétiques à l'aide de différentes méthodes, telles que la vraisemblance maximale, la parcimonie maximale ou la jonction de voisins. MEGA dispose d'une interface utilisateur graphique qui vous permet de visualiser et de manipuler vos données et vos résultats, ainsi que d'une interface en ligne de commande qui permet le traitement par lots et la création de scripts. **(Tamura *et al.*, 2021).**

II.5. Logiciels de visualisation d'arbres phylogénétiques

Actuellement, Il existe plusieurs programmes qui permettent la visualisation des arbres phylogénétiques, par exemple, les programmes : PhyloDraw, NJplot, GeneTree, Phylip (DRAWTREE et DRAWGRAM), Genedoc, Dambe, Treecon, TreeView, et Spectrum **(Choi *et al.* 2000).**

II.5.1.Le logiciel PhyloDraw

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

II.5.2. Le logiciel TreeView

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

II.6. Présentation du genre *Genista*

Le genre *Genista*dont le nom dérive du mot latin genesta qui signifie arbuste, appartient à la famille des Fabaceae et à la sous-famille cosmopolite des Faboideae. Cette sous-famille est la plus importante des Legumineuses, et compte 478 genres, 28 tribus et 13800 espèces **(Kacem, 2015).** Les genres les plus importants sont: Astragalus (2000), Indigofera (700), Tephrosia (400), Trifolium (300), Lupinus (200) **(Judd, 2002; Spichiger, 2004).** Le genre *Genista*, compte environ 200 espèces. Le genre *Genista* constitue un matériel intéressant qui mérite d'être mieux connu afin de mettre en lumière ses avantages et ses potentialités **(Lograda, 2010; Kacem, 2015).** Ce genre se distingue également par ses effectifs élevés d'espèces sous espèces et ces variétés endémiques et rares **(Lograda, 2010).**

II.7. Distribution et aire géographique

L'aire de distribution du genre *Genista*est circumméditerranéen et s'étend jusqu'au nord-est de l'Europe. Le genre est également très répandu à l'ouest de la Russie, en Turquie, en Syrie et au Caucase **(Gibbs, 1966).** Les espèces du genre ont une grande plasticité écologique, elles sont présentes dans des territoires soumis à des conditions bioclimatiques très différentes, depuis les zones semi-arides jusqu'aux zones très humides **(Azzioui, 2000, Kacem, 2015).** En Afrique du Nord, la concentration des plantes du genre est observée en Algérie et au Maroc. En effet, l'Algérie est représentée par 23 espèces, parmi lesquelles 11 sont endémiques **(Quezel, 1962; Maire, 1987),** il est localisé dans la région est, sud et au grand Sahara **(Quezel et Santa 1963)**.

II.8. Classification du genre Genista

Règne : Plantae
Division : Magnoliophyta Cronquist
Subdivision Magnoliophytina Frohne & U. Jensen
Classe : Rosopsida Batsch
Subclasse : Rosidae Takht.
Superordre : Fabanae R. Dahlgren
Ordre : Fabales Bromhead
Famille : Fabaceae Lindl

Tribu : Genisteae (Adans.) Benth.

II.9. Les intérêts du genre *Genista*

Le genre *Genista* présente plusieurs intérêts, notamment :

a)Intérêts médicinaux

Certaines espèces de *Genista*, telles que *G. tinctoria* et *G. canariensis*, contiennent des composés photochimiques aux propriétés anti-inflammatoires, antioxydants et anti-tumorales potentielles **(Barrajón-Catalán *et al.*2011 ; Fernández-Arroyo *et al.*, 2012)**. Des extraits de Genista sont également étudiés pour leurs effets hypoglycémiants et leur potentiel bénéfique dans le traitement du diabète **(Ferreira *et al.*, 2010).** De plus, certaines espèces de *Genista* sont utilisées en médecine traditionnelle pour traiter divers problèmes de santé, comme les troubles digestifs, les infections urinaires et les douleurs articulaires **(Sciriha *et al.*2019).**

b) Intérêts écologiques

Le genre *Genista* comprend des espèces jouant un rôle important dans la fixation de l'azote atmosphérique grâce à leurs nodules racinaires symbiotiques avec des bactéries fixatrices d'azote **(Custodio *et al,* 2016).** Certaines espèces de Genista sont également utilisées dans la restauration d'écosystèmes dégradés, en raison de leur capacité à coloniser rapidement les sols pauvres et à améliorer leur fertilité **(Rodríguez-Echeverría et Pérez-Fernández, 2005)**. Enfin, certaines espèces de Genista sont considérées comme des plantes de couverture efficaces pour prévenir l'érosion des sols et maintenir la biodiversité **(Galliano *et al.* 2018).**

Partie Pratique

Matériel et méthodes

Matériels et Méthodes

II.1. Choix des espèces à analyser

34 espèces de *Genista* dont 12 sont présentes en Algérie ont fait l'objet de cette étude. Le choix des espèces a été fait de telle sorte que toutes les sections de *Genista* telles que décrites par Gibbs (1966) soient représentées. La liste complète des espèces utilisées dans nos analyses, leurs numéros d'accession à la banque des données NCBI et leurs origines géographiques, sont données dans le **(Tableau 2).**

II.2. Choix de l'extra-groupe

Le meilleur critère pour choisir des extra-groupes et de prendre ceux-ci parmi les taxons proches de celui à analyser **(Darlu et Tassy, 1993).** Parmi les genres proches de *Genista* sont *Cytisus*, *Laburnum* et *Cytisophyllum***(Gaëlle Auvray, Pardo, 2004).** Nous avons donc sélectionné comme extra-groupe pour nôtre analyse les espèces : *Laburnum anagyroides, Cytisus arboreus* et *Cytisophyllum sessilifolium***(Tableau 2).**

II.2.1. Les données moléculaires

Les données moléculaires que nous avons utilisées sont des séquences des espaces internes transcrits 1 et 2 (ITS1 et ITS2) d'ADN ribosomique, L'Espaceur interne transcrit est une portion de l'ADN ribosomiaux située entre les gènes de la petite et de la grande sous-unité de l'acide ribonucléique ribosomique. Il est utilisé en phylogénie des eucaryotes et en barcoding moléculaire des champignons ou des cyanobactéries.

Les espaceurs internes transcrits ont un taux de mutations sensiblement plus élevé sous l'effet d'insertions, de délétions et de mutations ponctuelles, ce qui rend inopérante la comparaison de séquences d'ITS entre espèces très éloignées, comme les humains et les amphibiens par exemple. Il est toutefois utilisé avec succès pour des analyses phylogénétiques entre espèces voisines **(White *et al.* 1990)**

II.3. Récupération des séquences nucléotidiques dans les banques de données

Toutes les séquences nucléotidiques utilisées dans notre étude ont été récupérées dans la banque de données NCBI le 23 Mai 2024sous deux formats : le format NCBI pour extraire certaines informations concernant les séquences comme l'origine géographique des espèces le format Fasta (format très simple) pour manipuler facilement les séquences.

Pour l'interrogation de la banque, nous avons utilisé la plate-forme web BoldSystems **(Ratnasingham and Hebert, 2007),** les mots clés utilisés pour la consultation de la banque sont : *Genista,* Internal Transcribed Spacer 1, Internal Transcribed Spacer 2.

Les résultats de la recherche des séquences de *Genista* sur la plate-forme Boldsystems sont représentés dans le **(Tableau 3),** ces résultatsnous ont amené à choisir GenBank comme source pour télécharger nos données.

Tableau 2 : Liste complète des espèces de *Genista* utilisées dans notre étude, classées conformément au système de Gibbs leurs numéros d'accession à la banque de données Genbank et leurs origines géographiques.

Sous- genre	Section	Espèces	Numéros d'accession à la banque de données NCBI.		Origine géographique
			ITS1	ITS2	
Genista	Section 1 : *Genista*	1) *G. tinctoria*	AF351095	AF351095	Espagne
		2) *G. januensis*	AY263650	AY263650	Espagne
		3) *G. elata*	Z72262	Z72263	Allemagne
	Section 2 : *Spartioïdes*	4) *G cinerea*	AJ699036	AJ699037	Espagne
		5) *G. ramosisima*	AY263662	AY263662	Espagne
		6) *G. pseudopilosa*	AJ971613	AJ971614	Espagne
		7) *G.florida*	AF330660	AF330660	Maroc
	Section 3 : *Erinacoïdes*	8) *G. pumila*	AJ971617	AJ971618	Espagne
		9) *G. aspalathoïdes*	KT381198	KT381198	Algérie
		10) *G. polyanthus*	AJ698970	AJ698971	Espagne
		11) *G. lobelii*	AJ971593	AJ971594	Spain
	Section 4 : *Scorpioïdes*	12) *G. ferox*	KT381205	KT381205	Algérie
		13) *G.scoprpius*	AJ699026	AJ699027	Espagne
		14) *G.Corsica*	AJ971565	AJ971566	Sardaigne
		15) *G. carpetana*	AJ699050	AJ699051	Spain
Phyllobotrys	Section 1 : *Phyllobotrys*	16) *G.falcate*	AJ971571	AJ971572	Espagne
		17) *G. anglica*	AJ698984	AJ698985	Maroc
		18) *G.berberida*	AJ698972	AJ698973	Espagne
	Section 2 : *Volgera*	19) *G. triacanthos*	AJ971627	AJ971628	Espagne
		20) *G. tricuspidata*	KT381217	KT381217	Algérie
		21) *G .erioclada*	AJ971569	AJ971570	Algérie
		22) *G. ulicina*	KT381215	KT381215	Algérie
Spartocarpus	Section 1: *Acanthospartum*	23) *G .acanthoclada*	AJ698974	AJ698975	Grèce
	Section 2 : *spartocarpus*	24) *G. numidica*	KT381211	KT381211	Algérie
		25) *G. sparthoïdes*	AJ402882	AJ402883	Algérie
		26) *G.radiate*	AJ294511	AJ294512	Italie
		27) *G. aetnensis*	AJ402884	AJ402885	Italie
	Section 3 : *Fasselospartum*	28) *G.enista fasselata*	AJ699048	AJ699049	Chypre
	Section 4 : *Cephalospartum*	29) *G. cephalantha*	AJ698992	AJ698993	Algérie
		30) *G. quadriflora*	AJ699054	AJ699055	Algérie
		31) *G. capitellara*	AJ698982	AJ698983	Algérie
		32) *G. microcephala*	AJ698980	AJ698981	Algérie
		33) *G. umbellata*	AJ698986	AJ698987	Algérie
	Section 5 : *Spartidium*	34) *G. saharae*	EF457729	EF457729	Royaume-Uni
Espèces de l'extra-groupe.		*1) Laburnum anagyroïdes*	AY263679	AY263679	Espagne
		2) Cytisus arboreus	AF351123	AF351123	Espagne
		3) Cytisophyllum sessilifolium	AF351104	AF351104	Espagne

Tableau 3 : Résultats de la recherche sur le genre Genista sur la plateforme BOLDsystems (les diagrammes circulaires ont été retravaillés à l'aide de STATISTICA 7.0).

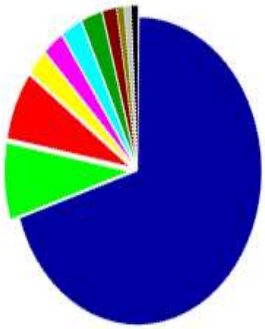
Results Summary :

Found **272** published records,
with **272** records with sequences,
with specimens from **26** countries,
deposited in **11** institutions.
Of these records, **272** have species names, and represent **109** species.

Specimen Distribution :

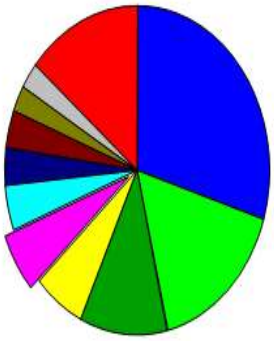
Institutions :

- Mined from GenBank, NCBI: 238
- National Museum Wales: 9
- Musee National d'histoire Naturelle Luxembourg: 8
- San Diego Natural History Museum, Herbarium: 3
- University of Guelph, OAC Herbarium: 3
- Tromso University Museum: 3
- Royal Botanic Gardens, Kew: 3
- Palermo Botanical Garden, Herbarium Mediterraneum: 2
- Royal Ontario Museum, Green Plant Herbarium (TRT): 1
- National Institute for Agricultural and Veterinarian Research, Portugal: 1
- Others: 1



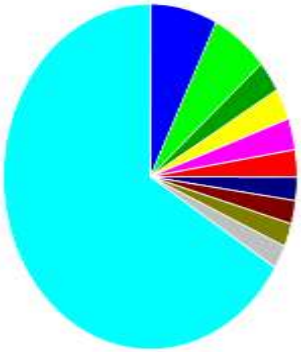
Countries :

- Unspecified*: 81
- Spain: 45
- Italy: 29
- United States: 17
- Algeria: 16
- United Kingdom: 12
- Morocco: 10
- France: 10
- Greece: 7
- Luxembourg: 7
- Others: 38



Species :

- Genista anglica: 20
- Genista tinctoria: 17
- Genista germanica: 8
- Genista cf. spachiana AK-2011: 8
- Genista pilosa: 8
- Genista sericea var. rigida: 7
- Genista monspessulana: 6
- Genista tridentata: 6
- Genista cinerea: 6
- Genista sericea var. sericea: 6
- Others: 180



II.4. Assemblage et alignement des séquences

Les analyses phylogénétiques que nous avons effectuées ont été réalisées en plusieurs étapes, le résumé de ces étapes est représenté dans la**(Figure 9).**

Les séquences d'ADN obtenues pour chaque espèce ont été combinées pour former une seule séquence. L'organisation des séquences a été réalisée selon l'ordre suivant : ITS1 puis ITS2. Toutes les séquences d'ADN ainsi produites ont été combinées dans un seul jeu de données, c'est-à-dire combinées dans une seule matrice de caractères moléculaires. Notre jeu de données contient donc 34 espèces de *Genista* plus trois espèces prises comme extra groupes (*Laburnum anagyroïdes, Cytisus arboreus, Cytisophyllum sessilifolium).*

Les séquences présentes dans le jeu de données ont été alignées en utilisant le logiciel d'alignement multiple ClustalX2 **(Larkin *et al.* 2007**), le programme étant utilisé avec tous ses paramètres par défaut (pénalité d'un gap est de 15 et pénalité d'extension d'un gap est de 6,66).

II.5. Nettoyage de l'alignement des séquences

Afin de ne pas altérer la qualité de l'alignement des séquences lors de nettoyage des séquences, nous avons calculer à l'aide de logiciel MEGA11 **(Tamura et al., 2021).**Les fréquences des différentes bases nucléotidiques, le nombre des sites analysés, le nombre des sites variables entre les différentes séquences et le nombre des sites phylogénétiquement informatifs.

Dans un premier temps le nettoyage de l'alignement des séquences a été réalisé manuellement, en supprimant les grands blocs de gaps et les colonnes qui en possèdent une grande quantité, puis nous l'avons vérifié à l'aide de programme Gblocks 0.91b **(Castresana 2000).**

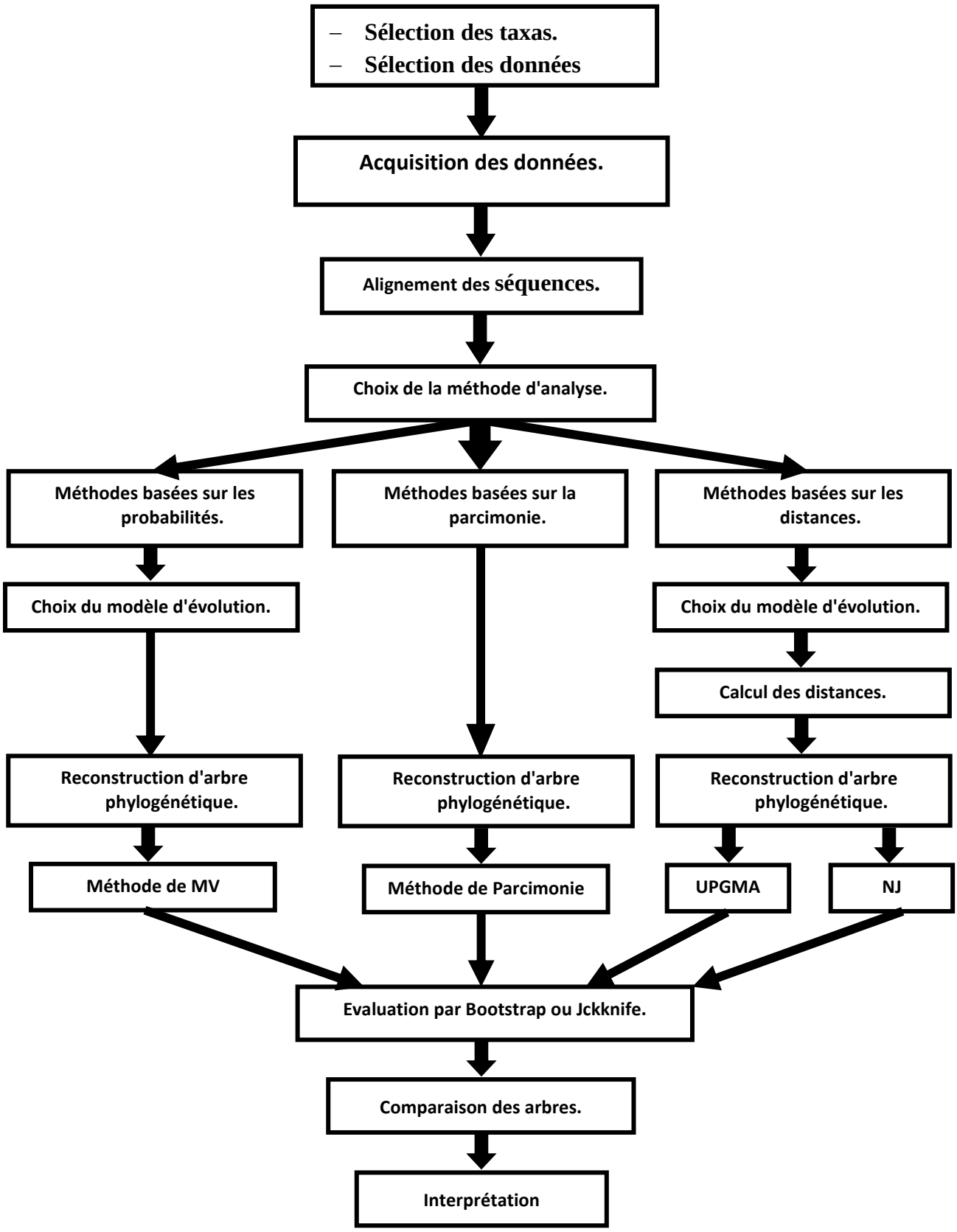


Figure 9 : Schéma résumant les différentes étapes d'une reconstruction phylogénétique **(Anne Denoual, 2017)**

II.6. Choix du modèle d'évolution

Il existe plusieurs tests statistiques permettant de mesurer l'adéquation des données à un modèle d'évolution particulier, les plus utilisés sont : le test basé sur le critère d'Akaike (Corrected Akaike Information Criterion «cAIC ») et le test basé sur le critère d'information bayésienne (Bayesian Information Criterion « BIC ») **(Botero-Castro, 2016).**

Le meilleur modèle d'évolution adaptée à notre jeu de données a été sélectionné à l'aide de logiciel MEGA (**Annexe 3**) et le modèle de Kimura à 2 paramètres a ainsi été choisi pour la reconstruction phylogénétique, car c'est ce modèle qui a présenté les valeurs de cAIC et BIC les plus faibles (cAIC = 7153.417, BIC =7664.894).

II.7. Les Reconstructions phylogénétiques

Parmi les différentes méthodes disponibles pour inférer des phylogénies, nous avons choisi de tester quatre principales méthodes : Deux méthodes de distance ont été utilisées, (UPGMA et Neighbour-Joining), la méthode cladistique (la Parcimonie) et la méthode probabiliste (le Maximum de vraisemblance). Les différentes analyses phylogénétiques ont été effectuées en utilisant plusieurs logiciels disponibles et téléchargeables sur le net, les principaux programmes de reconstructions d'arbres sont présentés dans le **(Tableau 4).**

Tableau 4 : Liste des logiciels bioinformatiques utilisés dans les analyses phylogénétiques ce travail.

Programme	**Fonction du logiciel**
Clustalx	– Alignements Multiple des séquences.
Gblocks	– Nettoyage de l'alignement multiple des séquences.
MEGA	– Choix de Modèle d'évolution – Reconstruction et visualisation des arbres phylogénétiques. – Evaluation des branches selon la méthode de bootstrap
Programme Treedist du Paquet Phylip :	– Comparaison des arbres phylogénétiques à l'aide de deux distances : – La distance de Robinson et Foulds (1981). – La branche score distance de Kohn et Felseinstein (1994).

II.7.1 Par les méthodes de distances

Le calcul des distances évolutives entre les séquences nucléotidiques alignées effectué suivant le modèle de Kimura à 2 paramètres à l'aide de MEGA.

La matrice des distances a été utilisée pour la construction arbres phylogénétiques par les méthodes : UPGMA et Neighbor-Joining à l'aide de MEGA.

II.7.2 Par la méthode de parcimonie

La construction des arbres parcimonieux à partir de notre jeu de données a été réalisée à l'aide de MEGA.

II.7.3. Par la méthode de maximum de vraisemblance

L'arbre de maximum de vraisemblance pour nos données, a été obtenu à l'aide de Logiciel MEGA, le modèle d'évolution utilisé est celui de Kimura à 2 paramètres (K2P).

II.8. Fiabilité, arbres consensus et visualisation des arbres construits

La robustesse des nœuds a été évaluée par la méthode bootstrap pour 1000 réitérations, elle été réalisée à l'aide de logiciel MEGA.

II.8.1. Visualisation des arbres phylogénétiques

Tous les arbres ont été visualisés à l'aide des logiciels MEGA et Phylodraw. Seuls les bootstraps d'une valeur supérieure à 50% ont été indiqués dans les arbres.

II.9.Comparaison des arbres phylogénétiques

Quelle que soit la méthode de construction de l'arbre, les données ont été obtenues sous format Newick **(Annexe 4)** pour les utilisées dans la comparaison des différents arbres.

La comparaison des arbres consensus obtenus par les quatre méthodes utilisées a été faite avec le programme Treedist du package phylip versions 3.65 (**Felsenstein, 2004).** Deux types de distances ont été utilisés : La distance topologique de **Robinson etFoulds(1981)** (symmetric difference notée RF) et la branche score distance «Branch Score Distance, notée Bs» de Kuhner et **Felsenstein (1994).**

Résultats et discussion

Résultats et discussions

III.1. Résultats

La combinaison des séquences ITS 1 et ITS2 de chaque espèce considérée dans notre analyse nous donne des séquences dont la longueur varie entre 460 à 611 bases nucléotidiques. Après la réalisation d'un alignement complet de toutes les séquences avec le programme ClustalX, l'introduction des gaps dans le résultat a fait augmenter le nombre des sites comparés à 739 sites. Après nettoyage de l'alignement multiple des séquences, nous avons constaté que 261/461 sont des sites variables, dont 136 sont des sites phylogénétiquement informatifs et 200/461 sites sont des sites conservés.

III.1. 1. Arbre phylogénétique obtenu par la méthode UPGMA

Le traitement phylogénétique des distances corrigées par le modèle de Kimura à 2 paramètres (K2) par la méthode UPGMA, a produit le phénogramme représenté dans la (**Figure 10)**. De même, la méthode Neighbor-Joining a donné le phénogramme représenté dans la **(Figure 11).**

III.1. 2. Arbre phylogénétique obtenu par la méthode de parcimonie

La méthode de Parcimonie appliquée à des séquences nucléotidiques résulte en deux arbres les plus parcimonieux avec un nombre de pas de 1171, l'arbre consensus majoritaire est donné dans la **Figure 12**.

III.1. 3. Arbre phylogénétique obtenu par la méthode de maximum de vraisemblance

L'arbre phylogénétique des 34 espèces de *Genista* construit par la méthode du maximum de vraisemblance sous le modèle d'évolution de l'ADN du Kimura à 2 paramètres (K2) est donné dans la **Figure13**.

III.1. 4. Comparaison des pourcentages

Le résultat de comparaison des pourcentages bootstrap entre les différents arbres est représenté dans le **tableau 5.**

Tableau 5 : Comparaison des pourcentages bootstrap entre les différents arbres phylogénétiques.

Méthode	Nombre de nœuds dont le pourcentage de bootstrap est supérieur à 50%	Pourcentage moyen de bootstrap
UPGMA	20	71.7%
N-J	24	79.97%
Parcimonie	23	77.47%
M-V	26	80%

III.1.5. Comparaison des arbres phylogénétiques par la distance topologique de Robinson et Foulds (1981) (RF)

Le résultat de la comparaison des arbres phylogénétiques par la distance topologique de Robinson et Foulds (1981) (RF) est représenté dans le **tableau 6.**

Tableau 6 : Distances topologiques de Robinson et Foulds "RF" entre les arbres phylogénétiques construits par les quatre méthodes (UPGMA, NJ, Maximum de parcimonie et MV).

	UPGMA	NJ	Parcimonie	MV
UPGMA	0	29	35	31
NJ	29	0	22	14
Parcimonie	35	22	0	22
MV	31	14	22	0

III.1.6. Comparaison des arbres phylogénétiques par la branche score distance de Kuhner et Felsenstein (1994)

Le résultat de comparaison des arbres phylogénétiques par la branche score distance de Kuhner et Felsenstein (1994) est représenté dans le **Tableau 7**.

Tableau 7 : La branche score distance de Kuhner et Felsenstein (1994) "Bs" entre les arbres phylogénétiques construits par les trois méthodes (UPGMA, NJ et MV)..

	UPGMA	**NJ**	**MV**
UPGMA	0	0.0488342	0.0646541
NJ	0.0488342	0	0.0304689
MV	0.0646541	0.0304689	0

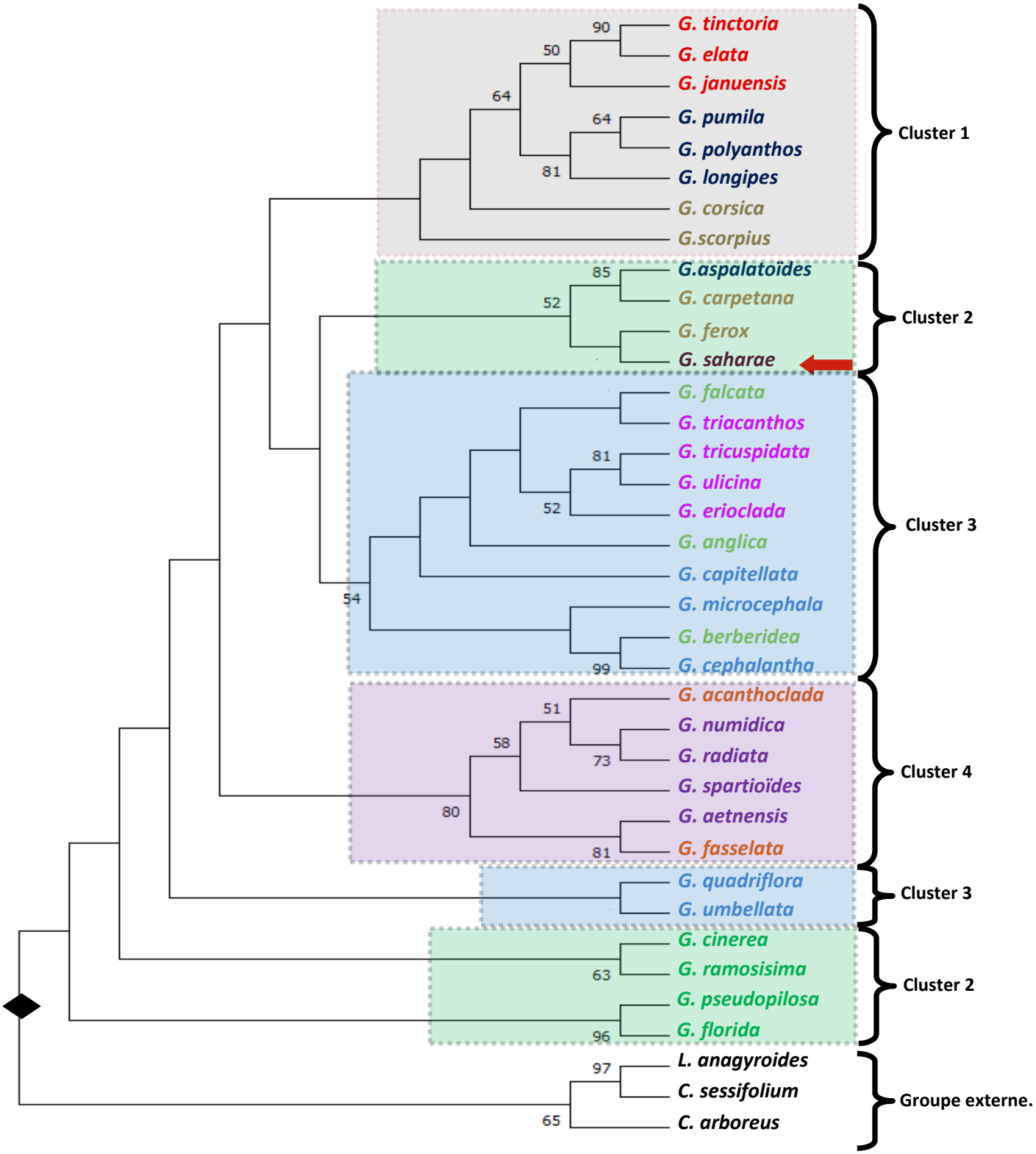


Figure 30 : Phénogramme UPGMA basé sur des séquences ITS1 et ITS2, reconstruisant les relations phylogénétiques de 34 espèces de genre *Genista*. Les nombres sur les branches indiquent les pourcentages bootstrap pour 1000 répliques, Seuls les nœuds soutenus par des valeurs de bootstrap d'au moins 50% ont été retenus. Les noms des sections tels que définit par Gibbs (1966) sont représentés par différentes couleurs (voir le tableau 8).

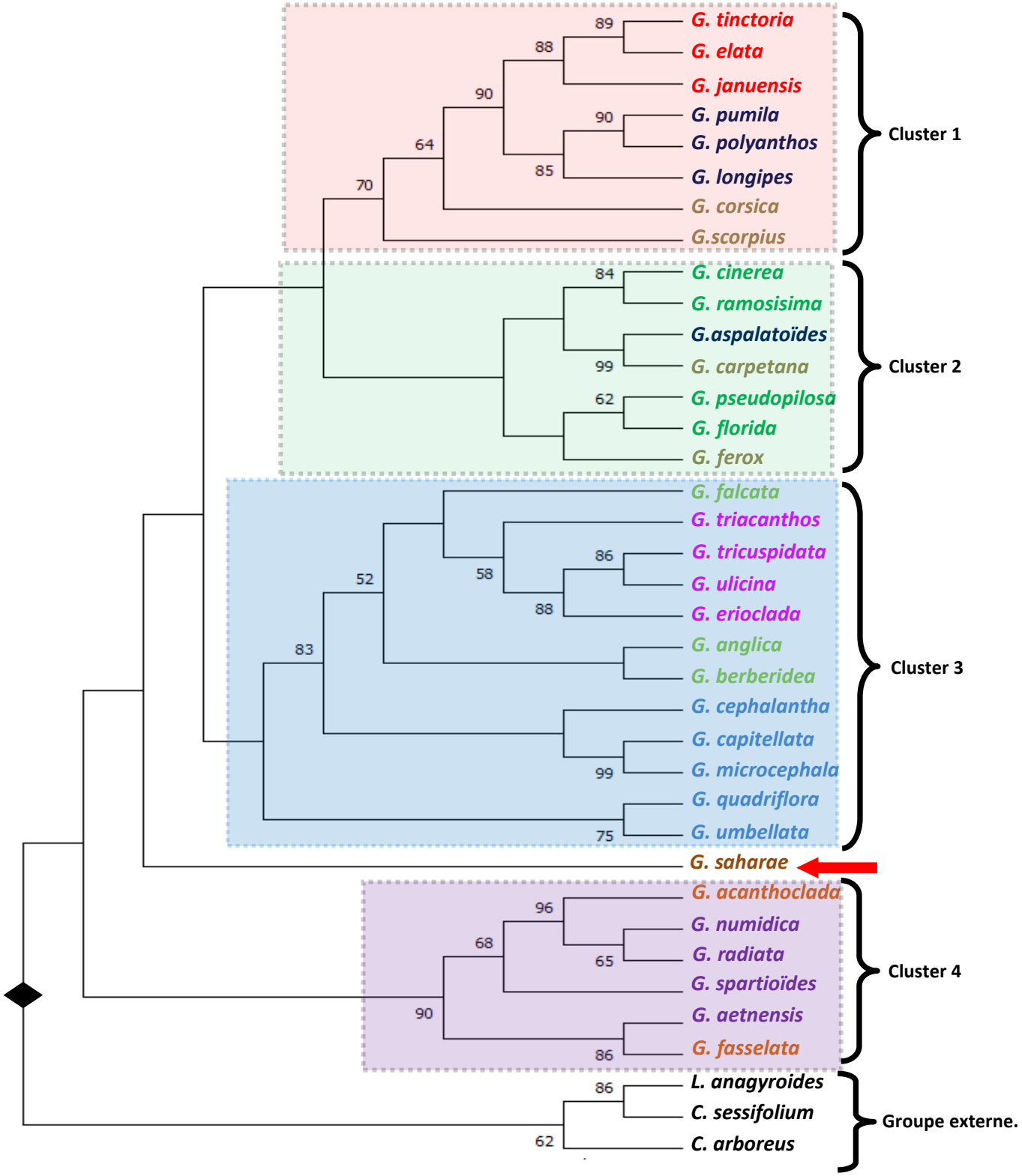


Figure 34 : Phénogramme NJ basé sur des séquences ITS1 et ITS2, reconstruisant les relations phylogénétiques de 34 espèces de genre *Genista*. Les nombres sur les branches indiquent les pourcentages bootstrap pour 1000 répliques, Seuls les nœuds soutenus par des valeurs de bootstrap d'au moins 50% ont été retenus. Les noms des sections tels que définit par Gibbs (1966) sont représentés par différentes couleurs **(voir le tableau 8)**.

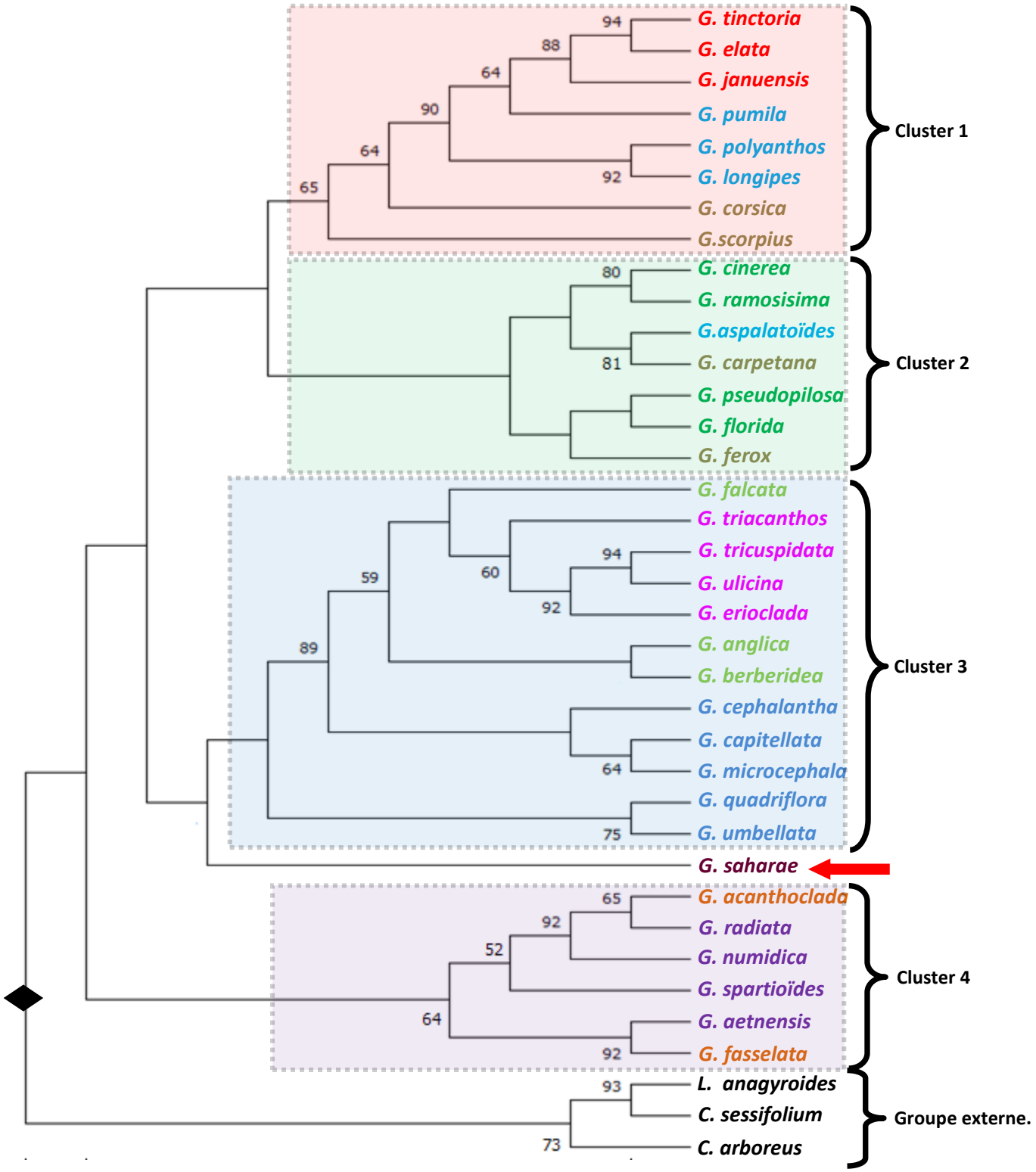


Figure 58 : Cladogramme de consensus majoritaire des 2 arbres les plus courts obtenus par la méthode de Parcimonie basé sur les séquences ITS1 et ITS2 d'ADN ribosomique, reconstruisant les relations phylogénétiques de 34 espèces de genre *Genista*. Seuls les nœuds soutenus par des valeurs de bootstrap d'au moins 50% ont été retenus. Les noms des sections tels que définit par Gibbs (1966) sont représentés par différentes couleurs **(voir le tableau 8).**

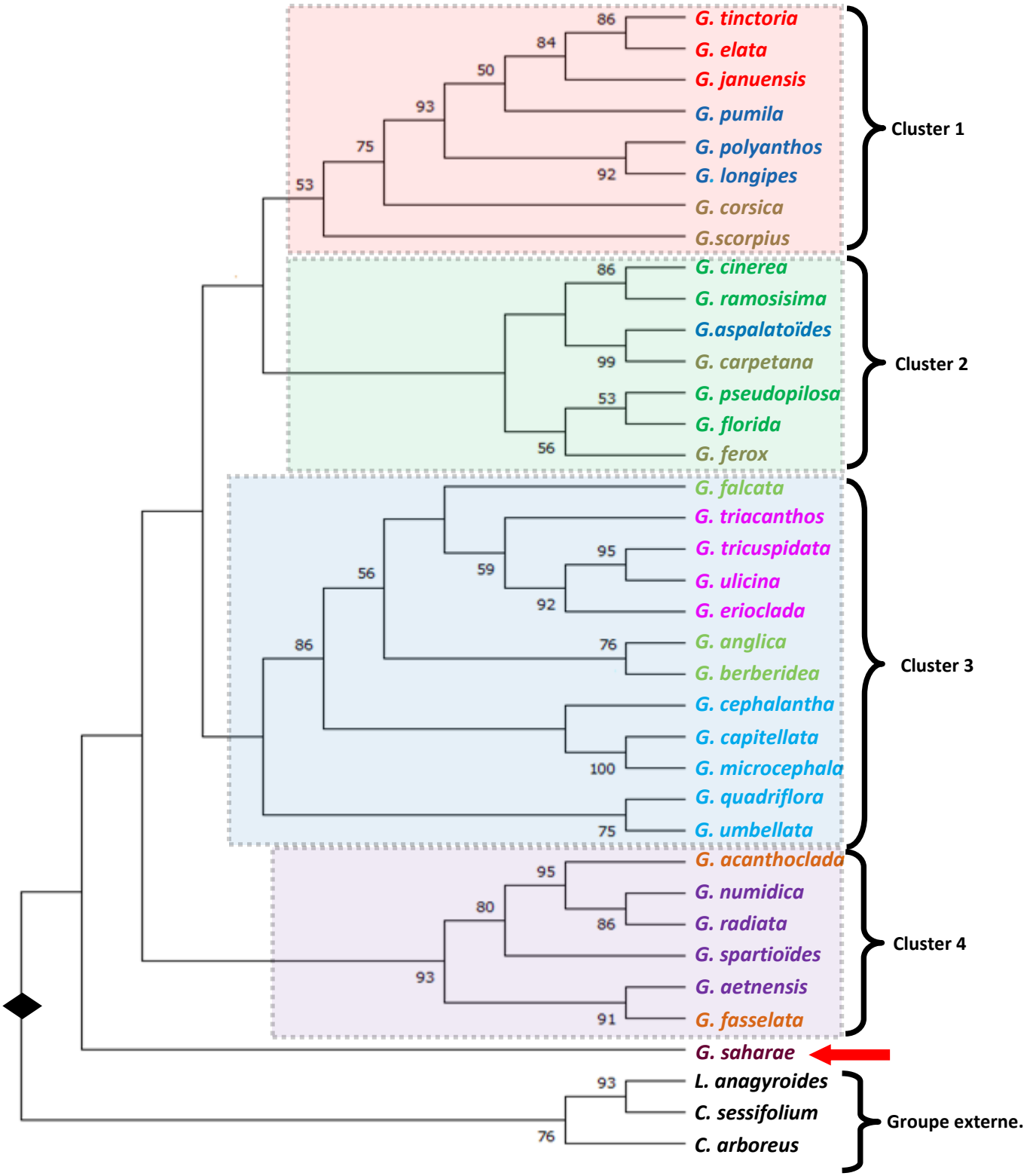


Figure 82 : Phylogramme obtenu selon la méthode de Maximum de vraisemblance à partir des séquences ITS1 et ITS2, reconstruisant les relations phylogénétiques de 34 espèces de genre *Genista*. Seuls les nœuds soutenus par des valeurs de bootstrap d'au moins 50% ont été retenus. Les noms des sections tels que définit par Gibbs (1966) sont représentés par différentes couleurs (voir le tableau 8).

Tableau 8 : Tableau comparatif de la classification morphologique selon le système de Gibbs (1966) avec les résultats de ce présent travail

Sous genre	Section	Espèces	Groupe	Cluster
Genista	*Genista*	1. *Genista tinctoria* 2. *Genista januensis* 3. *Genista elata*	Monophylétique	**1**
	Spartioïdes	1. *Genista cinerea* 2. *Genista ramosisima* 3. *Genista pseudopilosa* 4. *Genista florida*	Paraphylétique avec les *2espèces : G. ferox et G. aspalathoïdes*	**2** **2** **2** **2**
	Erinacoïdes	1. *Genista pumila* 2. *Genista aspalathoïdes* 3. *Genista polyanthus* 4. *Genista longipes*	Polyphélitique	**1** **2** **1** **1**
	Scorpioïdes	1. *Genista ferox* 2. *Genista scoprpius* 3. *Genista Corsica* 4. *Genista carpetana*	Polyphélitique	**2** **1** **1** **2**
Phyllobotrys	*Phyllobotrys*	1. *Genista falcate* 2. *Genista anglica* 3. *Genista berberida*	Polyphélitique	**3** **3** **3**
	Volgera	1. *Genista triacanthos* 2. *Genista tricuspidata* 3. *Genista erioclada* 4. *Genista ulicina*	Monophylétique	**3** **3** **3** **3**
Spartocarpus	*Acanthospartum*	1. *Genista acanthoclada*		**4**
	spartocarpus	1. *Genista numidica* 2. *Genista sparthoïdes* 3. *Genista eadiata* 4. *Genista aetnensis*	Paraphylétique avec les 2 espèces *: G. acanthoclada et G. fasselata*	**4** **4** **4**
	Fasselospartum	1. *Genista fasselata*		**4**
	Cephalospartum	1. *Genista cephalantha* 2. *Genista quadriflora* 3. *Genista capitellara* 4. *Genista microcephala* 5. *Genista umbellata*	Polyphélitique	**3** **3** **3** **3** **3**
	Spartidium	1. *Genista saharae*		
	Espèces de l'extra-groupe.	1) *L. anagyroïdes* *2) C. arboreus* *3) C. sessilifolium*	Extra-groupe.	

III.2. Discussions

La plateforme Boldsystems (Barcode Of Life Data Systems) **(Ratnasingham andHebert, 2007)** est l'une des ressources bioinformatiques les plus développées disponibles pour les chercheurs et les scientifiques sur le Net **(Lopez-Vaamonde et al. 2021).**

Il sert de passerelle vers de nombreuses ressources, notamment PubMed, GenBank, EMBL, DDBJ, musées et établissements de recherche scientifiques et bien d'autres **(Ratnasingham and Hebert, 2007).** Pour utiliser cette ressource bioinformatique dans la recherche des séquences nucléotidique, il suffit de faire une recherche en texte libre : comme par exemple le nom de genre, ou de l'espèce et la nature de la séquence nucléotidique **(Bouteleux, 2012).**

Cet outil nous retourne ensuite un résultat comportant plusieurs éléments **(Tableau 2),** comme par exemple le nombre de séquences déposés sur Genista, les institutions qui ont déposé ces données, l'origine géographique des séquences déposées.

Dans le domaine de la reconstruction phylogénétique, il semble possible d'affirmer aujourd'hui que la méthode la plus fiable pour reconstruire une phylogénie à partir de séquences nucléotidiques, est la méthode du maximum de vraisemblance **(Lopoez et all, 2002; Strimmer et Robertson, 2001**; **Felsenstein, 2004 etVincent Ranwez, 2013).**

C'est pour cette raison que, ici, nous avons prix l'arbre obtenu par cette méthode **(Figure 13 et Tableau 8)** comme référence d'autant plus qu'il présente des branches relativement les plus soutenues par les pourcentages de bootstrap (80%, voir **Tableau 5**).

Sur cet arbre, en prenant comme référence la classification du genre *Genista* réalisée par **Gibbs (1966)** sur la base des caractères morphologiques, nous avons fait ressortir 4 clusters de niveau de section au sous-genre (**Figure 13**). Tous les 4 clusters, à quelques différences d'agencement des espèces près, se retrouvent dans tous les arbres **(Figures 10, 11, 12 et 13).**

Cependant, comme nous avons tenté du mieux visualiser sur **Tableau 8**, des différences surgissent dans les branches les plus internes. Il en ressort que, les arbres les plus congruents sont ceux obtenus respectivement par le Maximum de vraisemblance le Neighbor-Joining et la Parcimonie (**Tableau 6 et 7**), avec, tous les trois un pourcentage bootstrap moyen supérieur à 75%. Quant à l'arbre UPGMA (**Figure 10** et **Tableau 5**), comme on peut le constater, il se montre fortement incongruent (RF entre UPGMA et MV = 31 et Bs entre UPGMA et MV = 0.0646541) avec les autres arbres avec un pourcentage bootstrap moyen de seulement 71,7%. Parmi ces arbres, le couple le plus congruent est celui de maximum de vraisemblanceNeighbor-Joining, ce couple présente les valeurs RF et Bs les plus faibles ((RF = 14 et Bs = 0.0304689) (**Tableaux 6 et 7**).

D'après **(Vincent Ranwez, 2013),** La capacité des méthodes de distances à reconstruire un arbre ayant une topologie correcte est nettement inférieure à celle des méthodes de maximum de vraisemblance. Les méthodes de distances permettent d'obtenir de manière rapide une histoire évolutive raisonnable d'un ensemble de séquences, alors que les méthodes de maximum de vraisemblance permettent d'obtenir une histoire évolutive beaucoup plus fiable, mais nécessitent un temps de calcul nettement plus important.

La méthode du maximum de parcimonie est très appréciée car plus rapide en temps de calcul, mais pas aussi précise que la méthode de Maximum de vraisemblance (MV). En plus, elle ne donne aucune information sur la longueur des branches et ne fait pas de corrections pour les substitutions multiples **(Cheikh rouhou, 2006).**

Si nous comparons la composition en espèces des 4 clusters obtenus dans cette présente analyse avec la révision du genre *Genista* réalisée par **Gibbs (1966)** ou il reconnait l'existence de trois sous-genres distincts dans le genre *Genista* sur la base d'un certain nombre de caractères morphologiques. Nous constatons dans le **Tableau 8**, que hormis pour les espèces appartenant à la section *Genista* du sous-genre *Genista* (Cluster 1 et cluster 2 sur l'arbre MV) et à la section *Volgera*du sous-genre *Phyllobotrys*, qui forment des groupes monophylytiques, les espèces appartenant aux autres sections (Cluster 3 et cluster 4 sur l'arbre MV) se sont montrées polyphylétiques.

Nous résultats sont très conformes à ceux obtenus par **Pardo *et al.*, (2004)** sur les relations phylogénétiques de *Genista* et des genres apparentés (*Teline*,*Chamaespartium*, *Pterospartum*, *Echinospartum*,*Ulex*, *Stauracanthus* et *Retama*) qui ont été évalués par l'analyse des séquences de l'espaceur interne transcrit nrADN (région ITS), et le ADNcp espaceur trnL-trnF intergéniques.

Nos résultats ne sont pas conformes avec ceux obtenus par **Lograda (2010)** dans son étude caryologique et phytochimique de six espèces endémiques du genre *Genista* en Algérie, en effet, dans l'arbre UPGMA basée sur les distances de linkage obtenu, l'espèce *G. saharae* forme un groupe monophylétique avec *G. microcephala*, et *G. numidica* se trouve éloignée du cluster des cinq autres espèces endémiques étudiées par **Lograda (2010),** ce qui n'est pas le cas dans notre étude.

Pour expliquer les incongruences des reconstructions basées sur les données molécualires avec celles basées sur d'autres types de données, **(Darlu et Tassy, 1993)** pensent que cela signifie que l'histoire de l'évolution des gènes peut être différente de l'histoire de l'évolution des espèces.

Enfin, comme on peut le constater dans tous les arbres construits ici (**Figures 10, 11, 12** et **13**), nos résultats démontrent la monophylie du genre *Genista* vis-à-vis de l'extra-groupe, et que *G. saharae*, dont l'appartenance au genre est controversée **(Meriane, 2018)**, en effet, cette espèce a été identifiée comme *Spartidium saharae* par **Pomel (1874)** et **Bourquin et al (1987)**, et comme *Calobota saharae* par **Boatwright et al., (2009**), et par *Genista saharae* par Cosson. et Durieu, 1855, dans notre travail, cette espèce se rapproche plus de l'extra-groupe que du cluster de l'ensemble des espèces analysées ici (l'arbre MV, **Figure 13**).

Conclusion

Conclusion

La bioinformatique a joué un rôle crucial dans le domaine de la systématique phylogénétique en améliorant la compréhension des relations inter-espèces. L'analyse des caractères moléculaires, tels que les séquences nucléotidiques et protéiques, permet aux chercheurs de reconstruire l'histoire évolutive des organismes et d'explorer les mécanismes sous-jacents à la biodiversité **(Ratnasingham et hebert, 2007).**

La connaissance approfondie des méthodes de reconstruction d'arbres phylogénétiques pour réaliser des analyses phylogénétiques est primordiale, La plupart des méthodes de reconstruction phylogénétique consistent à identifier les meilleurs arbres selon un critère d'optimalité (distance minimale, la vraisemblance ou le score de parcimonie)(**Rodríguez-Ezpeleta, 2007),** Chacune de ces méthodes que nous avons présentées possède des avantages et des désavantages différents. Le choix de l'une d'elles est souvent conditionné par le type de données que l'on souhaite traiter (données morphologiques, données moléculaires, taille des données et taux d'évolution...etc.). D'une manière générale, il faudrait en utiliser plusieurs afin de pouvoir confronter les résultats **(Zein Eddine, 2014).**

Dans notre travail, les résultats donnés par les méthodes de Neighbor-Joining, de Parcimonie et celle du Maximum de vraisemblance sont hautement congruents et leurs arbres phylogénétiques tous soutenus par des valeurs bootstrap supérieur à 75% contre 71,7% pour l'arbre UPGMA fortement incongruent avec les trois premiers, les valeurs des distances de robinson et foulds (1981)"RF" et la branche score distance de Kuhner et **Felsenstein (1994)** "Bs" de cet arbre avec les autres arbres sont les plus élevées. A la lumière de nos résultats, toutes les sections de la classification traditionnelle selon **Gibbs (1966)** se sont montrés plus ou moins polyphylétiques excepté pour la section *Genista* du sous-genre *Genista* et la section Volgera de la section Phyllobotrys. De plus la monophylie du genre *Genista* et la position de *G.saharae* à l'égard de ce genre ont été confirmées.

A l'avenir, il serait intéressant d'élargir cette étude à toutes les espèces appartenant au genre *Genista*. Il serait également très intéressant, de revoir la position phylogénétique de l'espèce très controversée *G. saharae* vis-à-vis du genre *Genista* sur la base d'autres caractères moléculaires.

Références bibliographiques

- **Auvray G., 2011** .Les relations phylogénétiques au sein d'un système réticulé : cas particulier de *Cytisus scoparius* L. (*Genista*, Fabaceae) et des espèces, hybrides et cultivars apparentés. Biologie végétale. Université d'Angers, 2011. Français. NNT.
- **Azzioui,** O., ES-Sgaouri, A., et Fennane, M. (**2000**). Valeur ecologique et biogeographique du genre *Genista L*. du Maroc. Lagascalia, 21(2), 263-278
- **Baldwin B.G., Sanderson M.J., Porter J.M., Wojciechowski M.F., Campbell C.S. et Donoghue M.J. (1995).** The ITS region of nuclear ribosomal DNA: a valuable source of evidence on Angiosperm ahylogeny. Ann Mo Bot Gard 82: 247-277.
- **Barrajón-Catalán, E., Fernández-Arroyo, S., Saura, D., Guillén, E., Fernández-Gutiérrez, A., Segura-Carretero, A., & Micol, V. (2011).**Cistaceae aqueous extracts containing ellagitannins show antioxidant and antimicrobial capacity, and cytotoxic activity against human cancer cells. Food and Chemical Toxicology, 49(9), 2312-2323.
- **Blanchet C., (1999)-** Logiciel MPSA et ressources bioinformatiques client-serveur Web dédiés à l'analyse de séquences de protéine. Thèse de doctorat en Bioinformatique. Dir. Thèse : Deleage G. Univ. Claude Bernard-Lyon I : 23-36. 158p.
- **Blanchet C., 1999.** Logiciel MPSA et ressources bioinformatiques client-serveur Web dédiés à l'analyse de séquences de protéine. Thèse de doctorat en Bioinformatique. Dir. Thèse : Delage G. Univ. Claude Bernard-Lyon I : 23-36. 158p.
- **Boatwright J.S., Tilney P.M. and Van Wyk B.-E.2009.** The generic concept of Lebeckia (Crotalarieae, Fabaceae): Reinstatement of the genus Calobota and the new genus Wiborgiella. South African Journal of Botany,75: 546–556.
- **Bodo Slotta T.A, (2000) -** Phylogenetic Analysis of Iliamna (Malvaceae) Using the Internal Transcribed Spacer Region. Thesis of Sciences in Biology. Faculty of the Virginia Polytechnic Institute and State University. 1-14. 76p.
- **Bouarour C., (2022)** – Analyse des performances de l'algorithme d'alignement de séquences CLUSTAL. Mémoire de Master en Bioinformatique. Dirigé par le Dr. Daas Mohamed Skander. Université Frères Mentouri Constantine 1.
- **Bouchair N (2014).** Diagnostic de systèmes complexes par comparaison de listes d'alarmes : application aux systèmes de contrôle du LHC. Traitement du signal et de l'image [eess.SP]. Université de Grenoble.
- **Bourquin D., Brenneissen, R. and Wicky .K.1987.** Zur Alkaloid fuhrung von Spartidium (Coss.et Dur) Pomel (Fabaceae). pharm acta hel,62 :10 -11.

- **Bouteleux O., (2012) -** Introduction au DNA-Barcoding.Master II "Sciences de l'Insecte". Université François Rabelais Tours.
- **BOUZAZA Z. ET MEZALI K (2018)**. Evolution and pattern of connectivity in some species of the genus Patella from the Algerian coast Advances in Bioresearch, , vol **9**(5) (In press).
- **Bouzaza Z., (2018)-**Etude systématique, phylogénétique, phylogéographique et démographique de Patella ferruginea (Gmelin, 1791), Patella caerulea (Linnaeus, 1758) et Cymbula safiana (Lamarck, 1819) de la frange côtière algérienne. Universite Abdelhamid Ibn Badis de Mostaganem.
- **Brochier C. (2007).** Cours introductifs à la reconstruction phylogénétique, EGEE : Evolution, Génome et Environnement, Université d'Aix en Provence, Marseille, France. 98p.
- **Cassan E., (2016)-**Bioinformatique des gènes chevauchants; application à la protéine antisens ASP du VIH-1. Médecine humaine et pathologie. Université Montpellier.
- **CAVÉ-RADET A., (2018) -** Évolution de la tolérance aux Hydrocarbures Aromatiques Polycycliques (HAPs) chez les spartines polyploïdes : Analyses Physiologiques et régulations transcriptomiques par les micro-ARNs Thèse présentée et soutenue à Rennes Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex.
- **Cheikh Rouhou M., (2006)**- évaluation des classifications phylogénétiques des bacillaceae basees sur les genes de l'operon rrn et de genes de menage. Mémoire de la maitrise en biologie. universite du quebec a montreal.
- **Chervin Hassel,(2015)-** Epidémiologie moléculaire et évolution de l'entérovirus A71 et interactions génétiques avec les autres entérovirus de l'espèce A responsables de la maladie pied-main-bouche.. Médecine humaine et pathologie. Université d'Auvergne - Clermont-Ferrand I.
- **Choi J.H., Jung H-Y., Kim H-S., and Cho H.G, 2000**- PHYLODRAW: A Phylogenetic Tree Drawing System. Department of computer science, Pusan National University, Pusan, Korea. Classification du vivant. P153.
- **Coleman AW. (2003)**. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet 19: 370-375.

- **Custodio, L., Ferreira, A. C., Pereira, H., Silvestre, L., Vizetto-Duarte, C., Barreira, L. ... et Varela, J. (2016).**The marine halophytes Carpobrotus edulis L. and Arthrocnemum macrostachyum L. are potential sources of nutritionally important PUFAs and metabolites. Food Chemistry, 193, 1-8.
- **Darlu P. et Tassy P. (2004).** La reconstruction phylogénétique, concepts et méthodes. Masson Ed., 258 pp.
- **Darlu P., Tassy P., 1993-** La reconstruction phylogénétique : concepts et méthodes. Ed. Masson, Paris. 245p.
- **Denoual M., (2017)** – Etude de la diversité génétique des souches d'Anaplasma ovis chez les chèvres en corse. Thèse de doctorat vétérinaire. Faculté de Médecine de Nantes. ONIRIS : Ecole nationale vétérinaire, agroalimentaire et de l'alimentation Natntes atlantique. 94p.
- **Dorkeld F., (1994)**- MultiMap : un modèle objet dédié à la cartographie comparée des génomes de mammifères. Thèse de doctorat en Biométrie. Dir. Thèse : Gautier C. N° d'ordre : 267-94. Univ. Claude Bernard – Lyon I. 40-50. 266p.
- **Duvignon M., (2012)**-utilisation de marqueurs moléculaires pour une avancée dans la résolution de la phylogénie du genre vanilla. Mémoire master. Univ-languedoc.
- **Edgar, R.C., (2004),** MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 11.
- **El Alaoui W., (2008)** - Estimation des longueurs de branche et artefact sur la datation moléculaire. Mémoire en vue de l'obtention du grade de Maître ès sciences (M. Sc). Université de Montréal.
- **Elodie Cassan., (2016**) Bioinformatique des gènes chevauchants ; application à la protéine antisens ASP du VIH-1. Médecine humaine et pathologie. Université Montpellier.
- **Erwan C. (2013).**Introduction aux méthodes de phylogénie.erwan.corre@sb-roscoff.fr.Rennes 5-6/12/2013.p373.
- **Felsenstein J. (1981).** Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution, 17**:** 368-376.
- **Felsenstein J., (2004)** - Inferring Phylogenies, University ofWashington. Sinauer Assoc., 2004, pp. xx + 664.
- **Fernández-Arroyo, S., Barrajón-Catalán, E., Micol, V., Segura-Carretero, A., et Fernández-Gutiérrez, A. (2012).**High-performance liquid chromatography with diode array detection coupled to electrospray time-of-flight and ion-trap tandem mass spectrometry to identify phenolic compounds from a Cistaceae plant extract. Phytochemical Analysis, 23(6), 660-670.

- **Ferreira, A., Proença, C., Serralheiro, M. L., & Araújo, M. E. (2010).**The in vitro screening for acetylcholinesterase inhibition and antioxidant activity of medicinal plants from Portugal. Journal of Ethnopharmacology, 108(1), 31-37.
- **Fidel Botero-C.,(2014)**- Systématique, phylogénie et évolution moléculaires des Phyllostomidae (Mammalia, Chiroptera) : une approche mitogénomique comparative. Biologie moléculaire. Université Montpellier II - Sciences et Techniques du Languedoc, 2014. Français. NNT : 2014MON20053.
- **Futuyma, D. J. (2013).** Evolution (3rd ed.). Sinauer Associates.
- **Gattolliat J-L.,(2002)-** Etude systématique, cladistique et biogéographique des Baetidae (Ephemeroptera) de Madagascar, Thèse de doctorat en Zoologie et Ecologie Animale. Dir. Thèse : J-M Elouard. Univ. Lausanne. 111-145. 279p.
- **GIBBS, P.E.,(1966)-** A revision of the genus Genista L. - in Notes Royal Bot. Gard., Edinburgh 27
- **Golding B., Morton D., (2003)**. Elementary Sequence Analysis. Department of Biology. McMaster. Univ. Hamilton, Ontario. 190 p.
- **Guindon S., (2003)-** Méthodes et algorithmes pour l'approche statistique en phylogénie, Thèse de Doctorat en Biologie. Dir. Thèse : O Gascuel. Univ. Montpellier II, 9-50. 155p.
- **Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., & Donoghue, M. J. (2016).** Plant Systematics: A Phylogenetic Approach (4th ed.). Sinauer Associates.
- **Koichiro Tamura, Glen Stecher, Sudhir Kumar**, 2021 - MEGA11: Molecular Evolutionary Genetics Analysis Version 11, Molecular Biology and Evolution, Volume 38, Issue 7, July 2021, Pages 3022–3027.
- **Korba R .,(2020)** polycopie DE cours Master II faculté snv .université mohamed el bachir el ibrahimi-borj bou arréridj.
- **Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S., Barraclough TG. & Savolainen V. (2008)**. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA 105: 2923.
- **Larivière D.,(2016).** Méthodes bioinformatiques D'analyse de l'histoire évolutive des Familles de gènes Préparée au sein de l'école doctorale SIBAGHE Et de l'unité de recherche AGAP.
- **Larkin, M.A., Blackshields G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D.,**

Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X Version 2.0. Bioinformatics, 23, 2947-2948.

- **Lecointre, G., et Le Guyader, H. (2016).** Classification phylogénétique du vivant Tome 1.
- **Lograda T.,(2010).** Variabilités cariologiques et biochimiques de quatre espèces endémiques du genre *Genista* L. Thèse de magister en biologie végétale. Université Ferhat Abbas-Sétif. Algerie.
- **Lopez P., Casane D., Herve P., (2002)-** Bio-informatique (5) Phylogénie et évolution moléculaires. Collection Médecine/Sciences; 18: 1146-54.
- **MERIANE D., (2018) -** Etude biologique et phytochimique de Calobota saharae (Coss. et Dur.) Boatwr. et B.E. van Wyk. Thèse pour l'obtention du diplôme de Doctorat en Sciences. Université Ferhat Abbas Sétif 1.
- **NASSIRA K., (2015)** Etude phytochimique et valorisation biologique Des deux plantes, Genista quadriflora Munby (Fabaceae) et Convolvulus tricolor L (Convolvulaceae). Soutenue le 25/10/2015. Présentée pour l'obtention du titre de docteur sciences université constantine 1.
- **Nathanaël Randriamihamison(2021).** Classification Ascendante Hiérarchique sous Contrainte de Contiguïté pour l'Analyse de données Hi-C. Génétique. Université Paul Sabatier - Toulouse III, 2021. Français. NNT : 2021TOU30108. tel-03424118v2 .
- **Nylander J.A.A., 2001**- Taxon Sampling in Phylogenetic Analysis: Problems and Strategies Reviewed. Department of Systematic Zoology Evolutionary Biology Centre. Uppsala University. 28p.
- **Otsen M., Verkaar E. L., Ruijter C., Hanekamp E., (2003).** Hybridization of banteng (Bos javanicus) and zebu (Bos indicus) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. Heredity, **90:** 10–16.
- **PARDO C., CUBAS P. and TAHIRI H., 2004**. Molecular phylogeny and systematics of Genista (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL- trnF intergenic spacer) Plant Systematics and evolution; 244(1-2): 93-119.
- **Perière G., (2000).** Bases de données et outils d'analyse pour la génomique bactérienne. Mémoire de l'habilitation à diriger des recherches. Univ. Claude Bernard, LYON 1. 100 p.
- **Pillet V., (2000).** Méthodologie d'extraction automatique d'information à partir de la littérature scientifique en vue d'alimenter un nouveau système d'information. Thèse de

doctorat en Sciences de l'Information et de la communication. Dir. Thèse : L. Quoniam et B. Jac. Univ. Aix-Marseille III. 20-43. 134p.

- **Quézel, P., et Santa, S. (1962)**-Nouvelle flore de l'Algérie. TI. ED., Centre National de la Recherche Scientifique, Paris. *16*(4), 459-459.
- **Quezel, P.et Santa, S. (1963)**. Nouvelle flore de l'Algérie et des régions désertiques méridionales. Vol 2, Éditions du Centre National de la Recherche Scientifique, 565 p
- **Ranwez V., 2013.** Méthodes efficaces pour reconstruire de grandes phylogénies suivant le principe du maximum de vraisemblance. Thèse de doctorat en Bioinformatique. Université Montpellier II - Sciences et Techniques du Languedoc, 2002. Français : 49-50.122p.
- **Rasmont R., 1997-** Evolution Biologique. Traduction de la 2ême édition anglaise. Ed. Départ. DeboeckUniv.Paris. Bruxelles. 371-507.
- **Ratnasingham S., Hebert D.N., (2007).** Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada N1G 2W1.
- **Ratnasingham, S., et Hebert, P. D. (2013).**BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Barcode Of Life Data Systems Handbook.
- **Richard D., Nattier R., Richard G., Soubaya T. (2014).**Atlas de phylogénie: la
- **Robert A., (2016)**- Elucidating Cancer Evolution using Single-Cell Sequencing and Comparative Genomics. The Watson School of Biological Sciences at Cold Spring Harbor Laboratory. Cold Spring Harbor Laboratory.
- **Rodríguez-Echeverría, S., & Pérez-Fernández, M. A. (2005).**Potential use of Iberian shrubby legumes and Rhizobium for revegetation and desertification control in Mediterranean.
- **Saitou N, Nei M. (1987).**The Neighbor-joining method: a new method for Reconstructing phylogenetic trees. Mol. Biol. Evol. 4 : 406-425.
- **Salemi M., Vandamme A.M., (2003)-** The phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. Cambridge University Press. 72-133.
- **Salemi M., Vandamme A.M., (2003)-** The phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. Cambridge University Press. 72-133
- **Schmidt H-A., (2003)-** Phylogenetic Trees from Large Datasets Inaugural. Thèse de Doctorat en Mathématique. Dir. Thèse : Von Haeseler A. Univ. Heinrich–Heine–Düsseldorf vorgelegt von Heiko. 123p.
- **Strimmer K., Robertson D.L., (2001)**- Inference and Applications of Molecular Phylogenies: An Introductory Guide. Chapter 4, in: C. Sansom and R. M. Horton (Eds.). The

Internet for Molecular Biologists (Practical Approach Series). Oxford University Press, Oxford, UK. To appear. 1-28.

- **Swofford D.L., (1998)-** PAUP*. Phylogenetic Analysis Using Parsimony and other methods. Version 4.0 (beta version)**.** Laboratory of Molecular Systematics Smithsonian.
- **Tahiri N., (2012)** - Un nouvel algorithme pour retrouver les relations phylogénétiques entre la distribution géographique des espèces et leurs compositions génétiques. Mémoire de la maîtrise Informatique. Université du Québec à Montréal.
- **Tahiri N., (2019)** – Algorithmes bioinformatiques pour la reconstruction d'arbres consensus et de super-arbres multiples. Thèse du doctorat en informatique. Université du Québec à Montréal.
- **Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., et Kumar, S. (2011).** MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28(10), 2731-2739.

- **Thomas, B. S., et Gupta, R. C. (2016).**Properties of high strength concrete containing scrap tire rubber. Journal of Cleaner Production, 113, 86-92.
- **Tourasse N-J., (1992)-** Développement d'une distance évolutive entre séquences prenant en compte la variabilité du taux de substitution entre sites et application à la reconstruction de phylogénies moléculaires anciennes. Thèse Doctorat en Génétique et Biologie des Populations.Univ. Claude Bernard - Lyon 1. 1-61. 186p.
- **Varré J-S., (2000)-** Concepts et algorithmes pour la comparaison de séquences génétiques : une approche informationnelle. Thèse de Doctorat en Informatique. Dir. Thèse : Delahaye J.-P. et Rivals E. Univ. Sciences et Technologies de Lille. 19-37. 193p.
- **Vergnaud N., (2002)-** Bioinformatique : un état des lieux. Ed. Ecole doctorale ERIC – Université de Lyon 2. Département SILR. Univ. Nantes. 26p.
- **Vergnaud, G. (2002).** Piaget visité par la didactique. Intellectica-La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo), 33, 107-123.
- **Vicente M. C., Fulton T., (2003)**. Utilisation des marqueurs moléculaires pour l'étude de la diversité génétique végétale. International Plant Genetic Institute for Genomic Diversity Resources Institute.
- **Vincent R., (2002)** -Méthodes efficaces pour reconstruire de grandes phylogénies suivant le principe du maximum de vraisemblance. Bio-informatique Méthodes efficaces pour

reconstruire de grandes phylogénies suivant le principe du maximum de vraisemblance. Bio-informatique [q-bio.QM Université Montpellier II - Sciences et Techniques du Languedoc, NNT : tel-00843175. Thèse de Doctorat de l'Université de Montpellier.

- **Vincent Sater.,(2021)-** Développement de nouvelles méthodes algorithmiques pour le traitement des UMI à partir des données de séquençage haut débit.. Base de données [cs.DB]. Normandie Université,
- **White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990).** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), PCR protocols: a guide to methods and applications (pp. 315-322). Academic Press.
- **Wiley, E. O., & Lieberman, B. S. (2011). Phylogenetics:** Theory and Practice of Phylogenetic Systematics (2nd ed.). Wiley-Blackwell.
- **Zein Eddine R., 2014** - Bulinus sp : Epidémiologie moléculaire, structure génétique et phylogénie dans trois pays africains. Interactions avec le genre Schistosoma. Thèse de doctorat en Biologie des organismes : Phylogénie, Structure génétique des populations, Evolution. Thèse co-dirigée par Gilles Dreyfuss, Félicité Flore Djuikwo-Teukeng.Ecole Doctorale n° 523 Gay Lussac.

ANNEXES

Annexe 1

Exemples des Formats des Banques de données

Tableau 9 : Exemple d'une séquence nucléotidique au Format EMBL

```
ID   KT381234; SV 1; linear; genomic DNA; STD; PLN; 448 BP.
XX
AC   KT381234;
XX
DT   31-JAN-2017 (Rel. 131, Created)
DT   31-JAN-2017 (Rel. 131, Last updated, Version 1)
XX
DE   Genista numidica subsp. numidica voucher MARS03823 trnL-trnF intergenic
DE   spacer, partial sequence; chloroplast.
XX
KW   .
XX
OS   Genista numidica subsp. numidica
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae;
OC   rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50 kb inversion clade;
OC   genistoids sensu lato; core genistoids; Genisteae; Genista.
OG   Plastid:Chloroplast
XX
RN   [1]
RP   1-448
RA   Baumel A., Tatoni T., Vela E.;
RT   "Molecular phylogenetics and origin of Genista L. Sect. Erinacoides Spach
RT   based on nucleotide sequences of nuclear and chloroplastic DNA";
RL   Unpublished.
XX
RN   [2]
RP   1-448
RA   Gasmi A.;
RT   ;
RL   Submitted (12-AUG-2015) to the INSDC.
RL   Mediterranean Institute of Marine and Terrestrial Biodiversity and Ecology,
RL   Aix-Marseille Universite UMR-CNRS 7263/IRD237 Facultes des Sciences St
RL   Jerome Boite 421 Av. Escadrille Normandie-Niemen 13397, Marseille cedex 20,
RL   France
XX
DR   MD5; 78bffdf2974ca861f6669cab686e1f4c.
XX
CC   ##Assembly-Data-START##
CC   Sequencing Technology :: Sanger dideoxy sequencing
CC   ##Assembly-Data-END##
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..448
FT                   /organism="Genista numidica subsp. numidica"
FT                   /organelle="plastid:chloroplast"
FT                   /sub_species="numidica"
FT                   /mol_type="genomic DNA"
FT                   /country="Algeria"
FT                   /isolation_source="dried leaf"
FT                   /specimen_voucher="MARS03823"
FT                   /db_xref="taxon:1784776"
FT   misc_feature    <1..>448
FT                   /note="trnL-trnF intergenic spacer"
XX
SQ   Sequence 448 BP; 153 A; 73 C; 59 G; 163 T; 0 other;
     tttgttcaag tccctctatc cccaaaagtc cgggttcact ctctaattat ttctcctaaa       60
     tcctcttttt ttttattcgt tatgtgtctt attcagtcca ttctttcaca aatggatctg      120
     attggaattt ttctttttt tatcacaatc acaagtcttg ggatatgtaa ttgaaatatt       180
     aaatatatat attaaacata caatttttt atggtaaacc tacaaacgaa catcttatcc       240
     ttgagcaagc aagcttcata ttaatgatta acaatacata atgattacta ctactgaaaa      300
     taaaaaacaa aattaattta aaaagaaata aaaaaagtct ttttttattt agttgacata      360
     gattgattga catatattca agtaatctct taaaaggaga tgctgcgcca agaatggtcg      420
```

Tableau 10 : Exemple d'une séquence au Format GenBank

```
LOCUS       AJ294515                 239 bp    DNA     linear   PLN 09-APR-
   2002
DEFINITION  Genista numidica var. supravillosa internal transcribed spacer
   1
(ITS1).
ACCESSION   AJ294515
VERSION     AJ294515.1
KEYWORDS    internal transcribed spacer 1; ITS1.
SOURCE      Genista numidica var. supravillosa
ORGANISM  Genista numidica var. supravillosa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50
kb inversion clade; genistoids sensu lato; core genistoids;
Genisteae; Genista.
REFERENCE   1
AUTHORS   De Castro,O., Cozzolino,S., Jury,S.L. and Caputo,P.
TITLE     Molecular relationships in Genista L. Sect. Spartocarpus Spach
(Fabaceae)
JOURNAL   Plant Syst. Evol. 231 (1-4), 91-108 (2002)
REFERENCE   2  (bases 1 to 239)
AUTHORS   De Castro,O.
TITLE     Direct Submission
JOURNAL   Submitted (04-SEP-2000) De Castro O., Dipartimento di Biologia
vegetale, Universita' degli Studi di Napoli Federico II, Via Foria,
223 Napoli, I-80139, ITALY
FEATURES             Location/Qualifiers
source          1..239
/organism="Genista numidica var. supravillosa"
/mol_type="genomic DNA"
/variety="supravillosa"
/specimen_voucher="herbarium sheet RNG, A. Dubuis s.n."
/db_xref="taxon:136803"
/country="Algeria: Akfadou, Tizi Ouzou"
misc_feature    1..239
/note="internal transcribed spacer 1, ITS1"
ORIGIN
1 tcgaagcctc acaagcagtg cgaccctgtg aatatgtttt actactcagg ggtggctaga
61 ggtgttcgtc acctcggtcc ccctcgtgtc gggaggcgcc ccaccctgtg tggtctcctc
121 ctggcccaat aacaaaaccc cggcgccgaa cgcgccaagg aaattttaat tgtctagtgc
181 gccccccgtcg gcccggagac ggtgcccgtg cgggtggcgt tgcgacacat gtatcctaa
//


LOCUS       AJ294515                 239 bp    DNA     linear   PLN 09-APR-
   2002
```

Tableau 11 : Exemple d'une séquence nucléotidique au Format DDBJ:

```
LOCUS       JF338215                494 bp    DNA     linear   PLN 04-NOV-2011
DEFINITION  Cf. Genista monspessulana x Genista stenopetala AK-2011 tRNA-Leu
            (trnL) gene, intron; chloroplast.
ACCESSION   JF338215
VERSION     JF338215.1
KEYWORDS    .
SOURCE      chloroplast cf. Genista monspessulana x Genista stenopetala AK-2011
  ORGANISM  cf. Genista monspessulana x Genista stenopetala AK-2011
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50
kb inversion clade; genistoids sensu lato; core genistoids;
            Genisteae; Genista.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Kleist,A. and Jasieniuk,M.
  TITLE     A molecular phylogenetic analysis of invasive and ornamental brooms
            and their relationships within the Genistoid legumes
  JOURNAL   Mol. Phylogenet. Evol. 61 (3), 970-977 (2011)
   PUBMED   21925614
REFERENCE   2  (bases 1 to 494)
  AUTHORS   Kleist,A. and Jasieniuk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2011) Department of Plant Sciences, University of
            California, Davis, One Shields Avenue, Davis, CA 95616, USA
FEATURES             Location/Qualifiers
source          1..494
                     /organism="cf. Genista monspessulana x Genista stenopetala
                     AK-2011"
                     /organelle="plastid:chloroplast"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:988322"
/pop_variant="population ASRA B"
/country="USA"
                     /collected_by="A.Kleist"
                     /PCR_primers="fwd_name: c, fwd_seq: cgaaatcggtagacgctacg,
                     rev_name: d, rev_seq: ggggatagagggacttgaac"
                     /note="invasive plant, putatively Genista monspessulana x
                     Genista stenopetala"
gene<1..>494
                     /gene="trnL"
/note="tRNA-Leu; UAA"
intron<1..>494
/gene="trnL"
BASE COUNT          192 a           74 c           90 g          138 t
ORIGIN
        1 aattggattg agccttggta tggaaactta ccaagcgata attttcaaat tcagagaaac
       61 cctggaattg acaatgggca atcctgagcc aaatcccgtt tttcccaaaa acaaagaaaa
      121 gttcagaaat cgaaaataaa aaggataggt gcagagactc aatggaagct gttctaacaa
      181 atggaattga cgacatttcc tttcgcgttg ggttaggaaa gtaatccttt cgtcgaaatt
      241 gcggaaagga tcaagaataa acgtatatac atatatatac gtatatgtac tgaaatatta
      301 tttcaattga ttaataaaaa taaagactga aaatctctat ttgttgaagg aggaattgaa
      361 tattcattga tcaaatcatt cattccatga aaatctgata gatcttttta agagctgact
      421 aatcagacga gaataaagat agagtcccat tctacatgtc aataccgaca acaatgaaat
      481 ttatagtaag agga
//
```

Annexe 2

Tabeau 12 : Exemple de séquences nucléotidiques alignées au format Phylip :

```
     8    138
Genista cinerea    TCGAAGCCTC ACAAGCAGTG CGACCC-GTG AATTTGTTTG ACTACTCAGG
Genista ramosissim TCGAAGCCTC ACAAGCAGTG CGACCC-GTG AATTTGTTCT ACTAATCAGG
Genista pseudopilo TCGAAGCCTC ACAAGCAGTG CGACCC-GTG AATTTGTTTG ACTACTCAGG
Genista florida    TCGAAGCCTC ACAAGCAGTG CGACCC-GTG AATTTGTTTG ACTACTCAGG
Genista pumila     TCGAAGCCTC ACAAGCAATG CGACCC-GTG AATTTGTTTG ACTACTCAGG
Genista aspalathoides TCGAAGCCTC ACAAGCAATG CGACCC-GTG AATTTGTTTG ACTACTCAGG
Genista tinctoria  TCGAAGCCTC ACAAGCAATG CGACCC-GTG AATTTGTTTG ACTACTCAGG
Enista elata      TCGAAGCCTC ACAAGCAATG CGACCCCGTG AATTTGTTTG ACTACTCAGG

 GGTGGCTAGA GGTGTTCG-G CACCTCGGTC CCCCTCGTGT CGGGAGGT-C
GGTGGCTAGA GGTGTTTG-G CACCTCGGTC CCCCTCGTGT CGGGAGGTGC
   GGTGGCTAGA GGTGTTTG-G CACCTCGGTC CCCCTCGTGT CGGGAGGCGC
   GGTGGCTAGA GGTGTTCT-G CACCTCGGTC CCCCTCGTGT CGGGAGGCGT
  GGTGGGTAGA GGTGTTCT-G CACCTCGATC CCCCTCGTGT CGGGAGGCGT
     GGTGGCTAGA GGTGTCTG-A CACCTTGGTC CCCCTTGTGT CGGGAGGCGT
         GGCGGCTTGA GGTGTTCCCA CACCTTGGTC CCCCTCGTGC CGGGAGGCGT
         GGCGGCTTGA G-TGTTCCCA CACCTTGGTC CCCCTCGTGC CAGTTGGCGT

         TCCTCCTGGC CAAAT----- ---------- --------
         CCCACCTCGC GTGGTCTCCT CCTGGCCAAA TAATAAAA
         CCCACCTTGC GTGGTCTCCT CCTGGCCTAA TAACAAAA
         CCC------- ---------- ---------- --------
      CCCACCTCGT GTGGTC---- ---------- --------
         CCCACCTCG- ---------- ---------- --------
     C--------- ---------- ---------- --------
         CCCACCTCGT GTGGTCTCCT TCTGGCCCAA TA------
```

Tableau 13 : Exemple d'une séquence nucléotidique au format FASTA :

```
>AJ294515.1 Genista numidica var. supravillosa internal transcribed spacer 1 (ITS1)
TCGAAGCCTCACAAGCAGTGCGACCCTGTGAATATGTTTTACTACTCAGGGGTGGCTAGAGGTGTTCGTCACCTCGGTCCC
    CCTCGTGTCGGGAGGCGCCCCACCCTGTGTGGTCTCCTCCTGGCCCAATAACAAAACCCCGGCGCCGAACGCGCCAA
    GGAAATTTTAATTGTCTAGTGCGCCCCCGTCGGCCCGGAGACGGTGCCCGTGCGGGTGGCGTTGCGACACATGTATC
    CTAA
```

Annexe 3

Tests statistiques de l'adéquation des données à un modèle d'évolution particulier réalisés à l'aide de logiciel MEGA11. Le test basé sur le critère d'Akaike (Corrected Akaike Information Criterion «cAIC ») et le test basé sur le critère d'information bayésienne (Bayesian Information Criterion « BIC »).

Table. Maximum Likelihood fits of 24 different nucleotide substitution models.

Model	Parameters	BIC	AICc
K2+G	**67**	**7664.894**	**7153.417**
K2+G+I	68	7674.480	7155.378
GTR+G	74	7694.500	7129.653
GTR+G+I	75	7699.354	7126.884
TN93+G	71	7702.876	7160.901
T92+G	68	7723.071	7203.970
HKY+G	70	7726.158	7191.807
T92+G+I	69	7732.455	7205.729
K2+I	67	7735.400	7223.924
TN93+I	71	7779.647	7237.672
GTR+I	74	7786.360	7221.514
T92+I	68	7803.450	7284.349
HKY+I	70	7803.728	7269.377
K2	66	7856.148	7352.297
JC+G	66	7871.710	7367.859
JC+G+I	67	7881.352	7369.876
TN93	70	7888.355	7354.004
GTR	73	7897.136	7339.914
T92	67	7922.067	7410.591
HKY	69	7924.882	7398.156
JC+I	66	7938.417	7434.566
JC	65	8055.489	7559.264

NOTE: Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected). This analysis involved 34 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 460 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.

Abbreviations: TR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

Annexe 4

Les arbres phylogénétiques : UPGMA, Neighbor Joining(NJ), Maximum de parcimonie et Maximum de vraisemblance au format Newick.

1. L'arbre UPGMA :

((((((((((G._anagyroides:0.00753794,C._sessilifolium:0.00753794)0.9710:0.01359458,C._arboreus:0.02113252)0.6400:0.00561234,(G._pseudopilosa:0.01003085,G._florida:0.01003085)0.9830:0.01671401)0.1740:0.00780268,G._scorpius:0.03454754)0.0340:0.00172970,(G._cinerea:0.02724033,G._ramosissima:0.02724033)0.5830:0.00903692)0.0200:0.00188084,(G._quadriflora:0.02973598,G._umbellata:0.02973598)0.3880:0.00842210)0.0340:0.00320990,(G._fasselata:0.03052273,(G._spartioides:0.02449052,(G._acanthoclada:0.02143483,(G._numidica:0.01560541,G._radiata:0.01560541)0.7260:0.00582943)0.5050:0.00305569)0.7020:0.00603221)0.8150:0.01084525)0.0430:0.00282154,(G._corsica:0.03642659,((G._tinctoria:0.01240132,G._elata:0.01240132)0.9600:0.01488002,(G._januensis:0.02604543,(G._pumila:0.01111719,G._polyanthos:0.01111719)0.9860:0.01492824)0.3010:0.00123591)0.5010:0.00914526)0.3880:0.00776293)0.4230:0.00874737,(((G._tricuspidata:0.02613014,G._ulicina:0.02613014)0.8220:0.00784063,G._erioclada:0.03397077)0.5680:0.01166086,(G._berberidea:0.04376143,(G._cephalantha:0.04169679,((G._capitellata:0.00109830,G._microcephala:0.00109830)1.0000:0.03739878,(G._falcata:0.03518473,(G._anglica:0.03322236,G._triacanthos:0.03322236)0.1840:0.00196237)0.1500:0.00331235)0.0820:0.00319971)0.0750:0.00206465)0.1410:0.00187020)0.5190:0.00730525)0.4690:0.00866098,G._ferox:0.06159787)0.4960:0.00404397,G._saharae:0.06564184)0.0000:0.00000000,G._aspalathoides:0.07047510);

2. L'arbre Neighbor Joining (NJ) :

(((((((G._tricuspidata:0.02596671,G._ulicina:0.02629358)0.8630:0.00746233,G._erioclada:0.03434907)0.9000:0.00968459,G._triacanthos:0.02648042)0.6970:0.00318249,G._falcata:0.03619905)0.4180:0.00233992,(G._anglica:0.03284203,G._berberidea:0.04301892)0.3560:0.00187365)0.4940:0.00559434,(G._cephalantha:0.04462774,(G._capitellata:0.00000000,G._microcephala:0.00233876)1.0000:0.03488218)0.3280:0.00240642)0.7670:0.01235782,(G._quadriflora:0.02132095,G._umbellata:0.03815100)0.7600:0.01004912)0.2690:0.00175309,((((G._anagyroides:0.00081218,C._sessilifolium:0.01426371)0.9360:0.00628753,C._arboreus:0.02843957)0.7110:0.00796425,G._saharae:0.08011048)0.2340:0.00137748,(G._spartioides:0.02374891,(G._fasselata:0.03325656,(G._acanthoclada:0.02610137,(G._numidica:0.01614837,G._radiata:0.01506245)0.6280:0.00116289)0.9350:0.00702232)0.4850:0.00280087)0.8440:0.01750706)0.2570:0.00318823,(((G._cinerea:0.02189838,G._ramosissima:0.03258227)0.7770:0.01365898,G._aspalathoides:0.08202178)0.1890:0.00215900,(((G._florida:0.01011503,G._ferox:0.06394074)0.3800:0.00094223,G._pseudopilosa:0.01024065)0.5610:0.00747664,(G._scorpius:0.03676762,(G._corsica:0.03778927,((G._pumila:0.01418653,G._polyanthos:0.00804785)0.9090:0.00590401,(G._januensis:0.02635307,(G._tinctoria:0.01126852,G._elata:0.01353411)0.9640:0.01724166)0.9130:0.00794918)0.9140:0.00606445)0.6600:0.00543138)0.5770:0.00672813)0.1940:0.00090819)0.1680:0.00327000);

3. L'arbre de Maximum de parcimonie :

(((((((G._tinctoria,G._elata)0.9590,G._januensis)0.8310,G._corsica)0.1290,(G._pumila,G._polyanthos)0.6400)0.6950,G._scorpius)0.6560,((G._pseudopilosa,(G._florida,G._ferox)0.3100)0.5660,((((G._acanthoclada,G._numidica)0.0520,G._radiata)0.6690,G._spartioides)0.3780,G._fasselata)0.9030)0.0400)0.0060,(G._quadriflora,G._umbellata)0.7990)0.0080,(((G._cinerea,G._ramosissima)0.7470,G._aspalathoides)0.3570,(((G._falcata,(G._triacantho

s,((G._tricuspidata,G._ulicina)0.9540,G._erioclada)0.9500)0.6290)0.2950,(G._anglica,G._berberidea)0.2660)0.5750,(G._cephalantha,(G._capitellata,G._microcephala)1.0000)0.4680)0.8560)0.0700,G._saharae)0.0080,((G._anagyroides,C._sessilifolium)0.9410,C._arboreus)0.7440);

4. L'arbre de Maximum de vraisemblance :

(((((((((G._tinctoria:0.00801345,G._elata:0.01659668)0.9200:0.02108657,G._januensis:0.02851646)0.8000:0.01208236,G._polyanthos:0.00675735)0.3500:0.00206786,G._pumila:0.01350795)0.8700:0.01264237,G._corsica:0.03731942)0.6400:0.00636331,G._scorpius:0.03797776)0.5600:0.00959414,(G._aspalathoides:0.08454775,((G._cinerea:0.02297877,G._ramosissima:0.03652661)0.7200:0.02150363,(G._pseudopilosa:0.01087725,(G._florida:0.00896052,G._ferox:0.06447747)0.3500:0.00001672)0.4900:0.00564654)0.2500:0.00624106)0.1400:0.00434034)0.1300:0.00503803,(((G._anagyroides:0.00001672,C._sessilifolium:0.01513420)0.9500:0.00619032,C._arboreus:0.02738897)0.7700:0.00752846,(G._fasselata:0.03822861,(G._spartioides:0.02897547,(G._acanthoclada:0.02687877,(G._numidica:0.01545659,G._radiata:0.01536174)0.5900:0.00001672)0.7300:0.00001672)0.4600:0.00480172)0.9100:0.02181609)0.2900:0.00435949)0.0700:0.00276598,G._saharae:0.07935634,((G._quadriflora:0.02500225,G._umbellata:0.03575434)0.7700:0.01479947,(((G._capitellata:0.00001672,G._microcephala:0.00232029)1.0000:0.03602151,G._cephalantha:0.04434680)0.3500:0.00472159,(G._anglica:0.03208912,(G._berberidea:0.05008572,(G._falcata:0.03671120,(G._triacanthos:0.02741204,(G._erioclada:0.02898006,(G._tricuspidata:0.02572101,G._ulicina:0.02751602)0.9900:0.01410138)0.8800:0.01417543)0.6600:0.00182525)0.3900:0.00233850)0.1400:0.00165759)0.4900:0.01168430)0.8500:0.01918419)0.3800:0.00561145);

Résumé

Ce travail a pour ambition d'exposer d'une manière aussi complète que possible l'une des disciplines les plus actives de la bioinformatique : la reconstruction phylogénétique. Celle-ci a pour objet de reconstruire les relations de parenté entre des espèces, en se basant sur des caractères morphologiques, des séquences de nucléotides ou d'acides aminés. Dans un premier temps, nous avons abordé la méthodologie de la reconstruction phylogénétique, depuis l'acquisition des données, jusqu'à l'inférence et l'évaluation des arbres phylogénétiques. Ensuite, nous avons utilisé les quatre différentes méthodes (UPGMA, Neighbor-Joining, Parcimonie et Maximum de Vraisemblance) pour inférer la phylogénie de 34 espèces du genre Genista sur la base des séquences ITS1 et ITS2 d'ADN ribosomique. Nos résultats ont clairement montré une forte congruence entre les groupes taxonomiques générés par les deux méthodes: Neighbor-joining et maximum de vraisemblance, et aussi, la robustesse de ces deux méthodes qui ont montré les valeurs booststrap les plus élevées. Nos résultats ne sont pas très congruents avec la classification traditionnelle basée sur des caractères morphologiques proposée par Gibbs (1966), en effet, au niveau des sections et hormis pour les deux sections: Genista et Volegera qui se sont montrées monophylétiques, toutes les autres sections se sont montrées polyphylétiques et enfin, nos résultats ont également démontré le caractère atypique de Genista saharae.

Mots clés : reconstruction phylogénétique, ADN ribosomique, Genista,

Abstract

The aim of this work is to expose as fully as possible one of the most active disciplines of bioinformatics: phylogenetic reconstruction. It aims to reconstruct relationships between species, based on morphological characteristics, nucleotide sequences or amino acids. Initially, we discussed the methodology of phylogenetic reconstruction, from data acquisition to the inference and evaluation of phyllogenetic trees. Then we used the four different methods (UPGMA, Neighbor-Joining, Parsimony and Maximum Probability) to infer the phylogenicity of 34 Genista species based on the ITS1 and ITS2 sequences of ribosomal DNA. Our results clearly showed a strong congruence between the taxonomic groups generated by both methods: Neighbor-joining and maximum probability, and also, the robustness of these two methods which showed the highest booststrap values. Our results are not very consistent with the traditional classification based on morphological characteristics proposed by Gibbs (1966), in fact, at the level of the sections and except for the two sections: Genista and Volegera which proved to be monophyletic, all the other sections were polyphyletic and finally, our results also demonstrated the atypical character of Genista sahara

Keywords: phylogenetic reconstruction, ribosomal DNA, Genista,

ملخص

تهدف هذه المبادرة إلى إظهار على نطاق أقصى قدر ممكن واحدة من المراحل الأكثر نشاطا في علم البيولوجيا: إعادة بناء الفيتولوجيا. يهدف هذا المنهج إلى إعادة بناء العلاقة بين الأنواع، على أساس العناصر الماركسية، أو سلسلة النيوتيدات أو الأحماض الأمينية .في البداية ، قمنا بتنفيذ أساليب إعادة بناء الفيلوجينات ، من الحصول على البيانات إلى التخطيط والتقييم الفيلوبينات. ثم استخدمنا أربعة أساليب مختلفة (UPGMA ، Neighbor-Joining، Parcimonia، Maximum of Probability)لإدخال الفيلولوجيا من 34 أنواع من النوع Genista على أساس سلسلة ITS1 و ITS2 من DNA الروبووسوم. أظهرت النتائج لدينا بوضوح صلة قوية بين المجموعات التمويلية التي تم إنشاؤها من قبل الطرق اثنين: الجانب المشترك والحد الأقصى من المرجح، وأيضًا، قوة هذه الطرق التي أثبتت أعلى قيمة .booststrap لم تكن نتائجنا متوافقة تماماً مع التصنيف التقليدي المعتمد على الخصائص المنطقية التي تم طرحها من قبل جيبس (1966)، في الواقع، على مستوى الأقسام، وباستثناء كلا القطاعات Genista :و Volegera التي أثبتت نفسهاmonophyletic ، جميع الأقطاعات الأخرى أثبت نفسها polyphyletic وأخيرا، أظهرت نتائجنا أيضًا طبيعة Genista saharae غير عادية.

الكلمات المفتاحية: إعادة بناء الفيلوجينات، DNAالروبوسوم، Genista