



Université Mohamed Khider de Biskra
Faculté des Sciences Exactes et des Sciences de la Nature
et de la Vie
Département des Sciences de la Nature et de la Vie
Filière : Sciences Biologiques

Référence / 2023

MÉMOIRE DE MASTER

Spécialité : Microbiologie Appliquée

Présenté et soutenu par :
BENHAFID Malak et ELGHOUL Jihane

Le: jeudi 15 juin 2023

Utilisation d'outils bio-informatiques dans la caractérisation moléculaire des champignons « Saccharomyces »

Jury :

M.	Guemmaz Fateh	MAA	Université	Président
M.	Moussi Abdelhamid	Pr	Université	Rapporteur
M.	Rebai Redouane	MCA	Université	Examineur

Année universitaire : 2022/2023

Remerciements

*Nous rendons grâce à **Dieu** le tout puissant pour le courage et la patience qu'il nous a accordé pour mener à bien ce travail,*

*Nous tenons à remercier chaleureusement notre encadreur **Monsieur MOUSSI ABDELHAMID Professeur** à l'Université pour son implication, son soutien, sa disponibilité et surtout ses précieux conseils. Nous le remercions aussi pour la relecture de ce Mémoire,*

*Nous adressons nos sincères remerciements à **Monsieur le Président et les membres du jury** qui ont accepté d'examiner ce travail,*

*Nous remercions également **Madame MOKRANI Chef de Département** pour sa compétence,*

*Nous remercions tous **le personnel** du Département des Sciences de la Nature et de la Vie pour leur aide et compréhension,*

*Nous remercions également tous **les Professeurs** qui nous ont soutenus au cours de ces années d'étude,*

Nous tenons à exprimer nos sincères remerciements à toute personne qui a contribué de près ou de loin à notre mémoire de Master.

*Enfin Nous exprimons notre profonde gratitude à nos **mères** et à nos **pères** qui nous ont toujours encouragé et poussé à continuer nos études et réussir.*

Dédicace

Je dédie ce modeste travail,

A ma très chère Maman pour son affection, sa bienveillance et sa présence à côté de moi qui est ma source de force,

A mon très cher Papa, toujours à mes côtés pour me soutenir et m'encourager,

A mes très chers frères Mohamed Firas et Iyad Safouane, Puisse Dieu vous donne santé, bonheur, courage et surtout réussite,

A ma tante Djalila et son mari Rahmoune Saifi pour leur soutien, leur accueil chaleureux et surtout leur disponibilité et modestie,

A tous mes Amies, Souheila, Boubouti, Rayane, Jihane, Halima, Maroua, qui m'ont toujours encouragé, et à qui je souhaite plus de succès.

A notre Grand Père Jaloul, cet homme sérieux, ponctuel et toujours souriant qui nous à accompagner tout au long de ces années d'études dans le bus de l'université,

A toute ma famille.

Malak

Dédicace

Je dédie ce travail,

À ma mère Djamila et mon père Massoude pour leur soutien pendant toutes ces années,

À mes frères Aridj et Mouatez Dhiaed Dine,

À ma famille et surtout mes cousines, Iman, Soundos et Khouloud pour leur soutien moral à moi et mes amies,

À ma binôme Malak pour son effort dans ce projet,

Et sans oublier M. Moussi pour ces efforts avec nous dans ce mémoire.

Jihane

Table des matières

Liste des tableaux	I
Liste des figures	II
Liste des abréviations.....	III

Introduction	1
--------------------	---

Première partie : Synthèse bibliographique

Chapitre 1 : Généralité sur les *Saccharomyces*

1. Levure de <i>Saccharomyces</i>	2
2. Diversité du genre <i>Saccharomyces</i>	3
3. Importance en nutrition et santé	3
4. Barcode chez les champignons	4

Chapitre 2 : Etude phylogénétique

1. Définition de la phylogénie	6
2. Phylogénie moléculaire	6
3. Arbre phylogénétique	6
.....	7
4. Différents arbres phylogénétiques.....	7
4.1. Arbre non enraciné	7
4.2. Arbre enraciné	7
5. Construction d'arbres phylogénétiques	8

Chapitre 3 : Méthodes de délimitation

1. Délimitation.....	9
1.1. mPTP (multi-rate Poisson Tree Processes)	9
1.3. ASAP (Assemble Species by Automatic Partitioning).....	10
1.4. ABGD (Automatic <i>Barcoding</i> Gap Discovery).	10

Deuxième partie : Partie Expérimentale

Chapitre 4 : Méthodologie

1. Origine des séquences	11
1.1. GenBank	11
2. Construction de l'arbre phylogénétique	18
2.1. Méthodes de construction de l'arbre phylogénétique.....	18

3. Délimitation des espèces	20
3.1. mPTP (multi-rate Poisson Tree Processes)	21
3.2. bPTP (Bayesian Poisson Tree Process)	22
3.3. ASAP (Assemble Species by Automatic Partitioning).....	22
3.4. ABGD (Automatic <i>Barcoding</i> Gap Discovery).	24

Chapitre 5 : Résultats et Discussion

.1 Bilan sur les données	26
1.1. Arbre phylogénétique	26
1.2. Délimitation des espèces	26
Conclusion.....	33
Bibliographie.....	34
Références bibliographiques	34
Sites web.....	37
Annexes	38

Liste des tableaux

Tableau 1 : Espèces de Saccharomyces et leur numéro d'accèsion extraites de GenBank.....	15
Tableau 2 : Résultats de la méthode de délimitation bPTP.....	29
Tableau 3 : Résultats de la méthode de délimitation ASAP.	30
Tableau 4 : Résultats de méthode de délimitation ABGD.	31

Liste des figures

Figure 1 : Saccharomyces en microscopie DIC sous microscope optique Olympus BX61(Masur, 2010).....	2
Figure 2 : Structure de la cellule Saccharomyces sous microscope électronique.	2
Figure 3 : Structure d'ITS.....	4
Figure 4 : <i>Exemple d'un arbre phylogénétique</i>	7
Figure 5 : Différents arbre phylogénétique : (a) arbre non enraciné, (b) et (c) arbre enracines.	8
Figure 6 : mPTP.	9
Figure 7 : ASAP.	10
Figure 8 : ABGD.....	10
Figure 9 : Banques de données publiques GenBank	11
Figure 10 : Une recherche de « Saccharomyces » dans GenBank.	11
Figure 11 : BLAST.....	12
Figure 12 : BLAST suite.	12
Figure 13 : BLAST suite.	13
Figure 14 : BLAST suite.	13
Figure 15 : Séquences téléchargées avant nettoyage.....	14
Figure 16 : Séquences téléchargées après nettoyage.	14
Figure 17 : NG-phylogeny.	18
Figure 18 : Etapes NG-phylogeny.....	19
Figure 19 : Etapes NG-phylogeny cas ML.....	20
Figure 20 : Etapes NG-phylogeny cas Bayes.	20
Figure 21 : Plateforme mPTP.....	21
Figure 22 : Arbre phylogénétique format ML-Newick.	21
Figure 23 : Plateforme bPTP.....	22
Figure 24 : Arbre phylogénétique format Bayes-Newick.	22
Figure 25 : Plateforme ASAP.....	23
Figure 26 : Plateforme NG-phylogeny.	23
Figure 27 : Fichier format FASTA sans outgroupe.....	24
Figure 28 : Plateforme ABGD.....	24
Figure 29 : Fichier forma FASTA avec outgroup.	25
Figure 30 : Arbre phylogénétique ML enraciné de genre Saccharomyces.....	28

Liste des abréviations

NCBI : National Center for Biotechnology Information

MUSCLE : Multiple Sequence Comparison by Log- Expectation

PTP: Poisson tree processes

mPTP: Multi-rate Poisson tree processes

bPTP: Bayesian Poisson Tree Process

BLAST: Basic Local Alignment Search Tool

ABGD: Automatic Barcode Gap Discovery

ASAP: Assemble Species by Automatic Partitioning

ITS: Internal Transcribed Spacer

FASTA: Federal Acquisition Streamlining Act

Introduction

Introduction

Les champignons jouent un rôle essentiel dans notre environnement en tant que décomposeurs, symbiotes et agents pathogènes. Comprendre leur diversité et leur fonctionnement moléculaire est essentiel pour de nombreuses applications, allant de l'agriculture à la pharmacologie (Hussenet, 2017). Ces dernières décennies, les outils bio-informatiques ont révolutionné le domaine de la caractérisation moléculaire des organismes, en permettant une analyse approfondie des données génomiques et protéomiques.

Le genre *Saccharomyces* est composé de levures, dont certaines sont largement utilisées dans l'industrie alimentaire et de la biotechnologie, notamment *Saccharomyces cerevisiae*, une espèce clé pour la production de levure de boulangerie et pour la fermentation dans l'industrie brassicole et vinicole. Malgré l'importance économique de ces levures, leur diversité génétique et leur potentiel biotechnologique restent encore mal compris. Ainsi, la problématique de cette étude est de caractériser moléculairement les différentes espèces et souches de *Saccharomyces* en utilisant des outils bio-informatiques, afin de mieux comprendre leur diversité.

Les objectifs spécifiques de cette étude sont les suivants :

1. Réaliser une analyse phylogénétique des espèces de *Saccharomyces*, en utilisant de données génomiques disponibles dans les bases de données publiques, pour reconstruire les relations évolutives entre les différentes espèces.
2. Identifier les gènes spécifiques des différentes espèces de *Saccharomyces*, en comparant les génomes et en utilisant des approches de génomique comparative.
3. L'utilisation de diverses approches de délimitation des espèces permet d'établir des frontières claires entre les espèces existantes et de mettre en évidence de potentielles nouvelles espèces putatives.

Nous formulons les hypothèses de travail suivantes pour cette étude : Les différentes espèces de *Saccharomyces* présenteront des variations significatives au niveau de leur génome.

Pour atteindre nos objectifs, nous utiliserons une approche bio-informatique intégrée. Nous recueillerons les données génomiques des différentes espèces de *Saccharomyces* à partir des bases de données publiques, en mettant l'accent sur les séquences de génomes. Nous effectuerons une analyse phylogénétique en utilisant des méthodes d'alignement multiple et d'inférence phylogénétique pour reconstruire les relations évolutives entre les différentes espèces.

Première partie :
Synthèse bibliographique

Chapitre 1 : Généralité sur les *Saccharomyces*

1. Levure de *Saccharomyces*

Le nom de *Saccharomyces* désigne, champignon « *myces* » qui se nourrit du sucre saccharose « saccharo ». C'est un terme utilisé pour désigner le petit champignon microscopique qui compose les différentes sortes de levures qu'on utilise pour la fermentation (Larpent et Gourgaud, 1985).

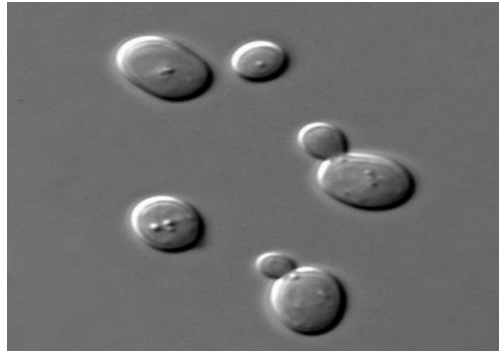


Figure 1 : *Saccharomyces* en microscopie DIC sous microscope optique Olympus BX61(Masur, 2010).

La cellule du champignon est minuscule, microscopique, ovale ou sphérique, souvent présentes sous forme de cellules isolées chaque cellule est délimitée par une paroi cellulaire composée de chitine. Dans le cytoplasme il y a des mitochondries, des ribosomes, un noyau vacuole etc. La paroi cellulaire étant perméable, les substances alimentaires se diffusent facilement dans la cellule. De même les déchets sont éliminés de la cellule par diffusion. Le protoplasme contient deux types d'enzymes, à savoir l'invertase extracellulaire qui sort de la cellule et hydrolyse le sucre de canne en dextrose. La zymase intracellulaire qui reste dans la cellule. Il agit sur le sucre et le convertit en alcool et en dioxyde de carbone. La levure est le premier organisme eucaryote dont le génome (environ 6000 gènes) a été complètement séquencé (Gofieau *et al*, 1996).

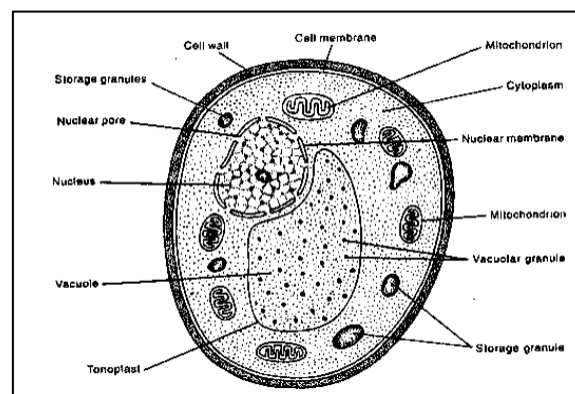


Figure 2 : Structure de la cellule *Saccharomyces* sous microscope électronique.

Les cellules sont des micro-organismes unicellulaires, classé comme étant anaérobie facultatif (Visser *et al*, 1990) avec une préférence pour le mode fermentatif (Barnett, 2003).

Classifications de l'espèce (Barnett, 1992) :

- **Régne** : *Fungi*.
- **Embranchement** : *Ascomycota*.
- **Sous-Embranchement** : *Saccharomycotina*.
- **Classe** : *Saccharomycetes*.
- **Ordre** : *Saccharomycetales*.
- **Famille** : *Saccharomycetaceae*.
- **Genre** : *Saccharomyces*.

2. Diversité du genre *Saccharomyces*

Parmi les alternatives possibles, les probiotiques (levures) sont utilisés pour soutenir la santé et les performances des animaux et les levures ont été largement utilisées pour améliorer la santé intestinale tant chez l'homme que chez l'animal (Geraldine et Gebreyohannes, 2014).

La diversité des espèces du genre *Saccharomyces* est très importante et la dernière classification (Kurtzman *et al*, 2011) a défini trois groupes dont celui regroupant toutes les souches de *Saccharomyces cerevisiae*, qui est donc considérée comme l'espèce fermentaire. Le genre *Saccharomyces* est également largement étudié en recherche biomédicale pour étudier la biologie cellulaire et la génétique, en raison de leur morphologie unicellulaire et de leur génome relativement simple. De nombreuses études ont analysé son évolution et sa dispersion géographique à travers le monde afin de mieux comprendre cette espèce. Elle est généralement associée aux boissons alcooliques (vin, bière), au pain ou encore à des souches utilisées pour la recherche (Börlin, 2015).

3. Importance en nutrition et santé

Les espèces de *Saccharomyces* sont d'importance économique, notamment *S. cerevisiae*, utilisée depuis des siècles dans la production de bière, de vin et de pain. *S. boulardii* est utilisée comme probiotique pour traiter certaines infections gastro-intestinales. Dans le secteur alimentaire, il s'agit de la production des levures et ferments pour la production du pain, des fromages et des boissons alcoolisées, mais aussi des arômes, de certains acides et de (Zaïd *et al*, 2020). Avec l'évolution des biotechnologies. Dans le domaine pharmaceutique et médical, elle est utilisée pour la production de vaccins, de probiotiques ou de protéines comme l'insuline (Celton, 2011).

Les levures deviennent des agents biologiques importants pour les biotransformations (éthanol, glycérol, acide ascorbique, lactique...). Le genre *Saccharomyces* est utilisé pour la fabrication de la bière depuis plus de 3000 ans, mais ce n'est qu'entre 1857 et 1863 que Louis Pasteur démontre son rôle dans la fermentation alcoolique (Hussenet, 2017).

4. Barcode chez les champignons

Le barcode chez les champignons consiste à utiliser une région d'ADN spécifique, généralement la région ITS (Internal Transcribed Spacer), pour identifier de manière fiable et rapide les espèces de champignons. Le gène ITS (Internal Transcribed Spacer) est une région non codante de l'ADN génomique des eucaryotes située entre les gènes codant pour l'ARNr 26S et 18S. Elle est composée de deux parties : l'ITS1, l'ITS2 et séparées par le gène ribosomal 5.8S (Voir figure ci-dessous). Il est utilisé en phylogénie des eucaryotes et en barcoding moléculaire des champignons pour mettre en évidence une différence génétique entre deux espèces (Sumida *et al*, 2004).

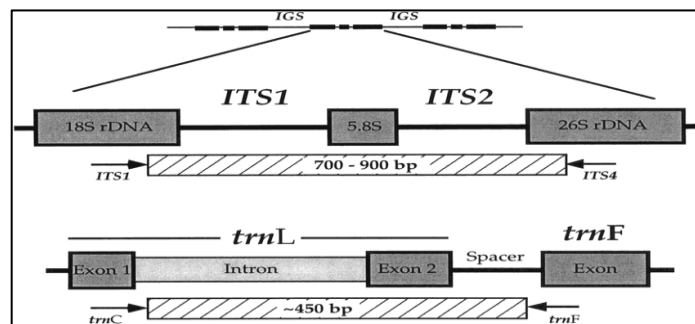


Figure 3 : Structure d'ITS.

La région ITS est maintenant largement utilisée comme « code-barre » génétique pour caractériser la diversité des champignons (Nguyen et Seifert, 2008). L'utilisation du barcode chez les champignons peut également être appliquée à des domaines de recherche tels que l'écologie, la biogéographie et la phylogénie des champignons. Il permet de comprendre la distribution géographique des espèces, leurs relations évolutives et leurs interactions avec d'autres espèces au sein de l'écosystème (Schoch *et al*, 2014).

Le barcode permet également de détecter les espèces de champignons qui peuvent présenter une menace pour la santé publique ou pour l'environnement, telles que les champignons toxiques ou les agents pathogènes des plantes. Cette méthode peut aider à prévenir la propagation des maladies des plantes et des animaux, et à protéger l'environnement contre les espèces invasives (Nilsson *et al*, 2014).

En somme, l'utilisation du barcode chez les champignons est un outil précieux qui contribue à une meilleure compréhension de la diversité fongique, à l'identification des espèces et à la protection de la santé publique et de l'environnement (Xu *et al*, 2014).

Chapitre 2 : Etude phylogénétique

1. Définition de la phylogénie

Le terme phylogénie ou phylogénèse provient du grec *phûlon* qui signifie « tribu, famille, clan, espèce » et *genesis* « création, origine ». Le terme de « phylogénie » a été inventé en 1866 par le biologiste allemand Ernst Haeckel pour désigner cette relation de filiation qui unit les êtres vivants (Haeckel, 1860).

En biologie, la phylogénie consiste ainsi en l'étude de l'évolution des relations entre des groupes d'organismes (par exemple, des espèces, des populations), et de ce qui est découvert par le biais des données de séquençage moléculaire et morphologique des matrices de données (site web1).

2. Phylogénie moléculaire

La phylogénie moléculaire permet d'étudier les espèces végétales, animales et microbiennes, sur les deux plans phénotypique et génotypique, afin de les classer en fonction de leurs ressemblances et en fonction de leurs structures géniques (liens de parenté). La phylogénie étudie, en fait, les relations de parenté entre les individus et représente sous forme d'arbre le résultat de ces relations (Luchetta *et al*, 2005).

Donc face à ce tas de données, il y aura besoin d'outils adéquats pour pouvoir traiter toutes ces informations et tirer un meilleur profit. La manipulation correcte des données initiales va permettre d'aboutir à des interprétations et des conclusions pertinentes : Grâce aux résultats de la phylogénie, le chercheur peut tirer des hypothèses sur les liens génétiques des espèces, les états ancestraux des caractères étudiés, la divergence ou la convergence des caractères. Il existe aujourd'hui une large gamme de marqueurs moléculaires qui peuvent être utilisés pour accomplir ce type de recherche. (Patwardhan *et al*, 2014).

3. Arbre phylogénétique

L'arbre phylogénétique présente les relations de parenté entre organismes vivants. Un arbre phylogénétique est une forme de classification des espèces. Cette classification traduit les relations de descendance des espèces avec modification de leurs caractères. Les caractères sont transmis d'une génération à l'autre à travers les mécanismes d'hérédité. Il montre qui est proche de qui, et non pas qui descend de qui (site web 2).

Un arbre est composé de quatre éléments principaux (Deriham, 2017).

- **La racine** : désignant l'ancêtre commun des espèces représentées dans l'arbre.
- **Les nœuds externes** : ou feuilles, qui représentent les unités taxonomiques (les espèces) dont les informations ont été utilisées lors de la construction de l'arbre.
- **Les nœuds internes** : représentant des ancêtres hypothétiques.
- **Les branches** : qui montrent les relations de descendance entre les nœuds de l'arbre.

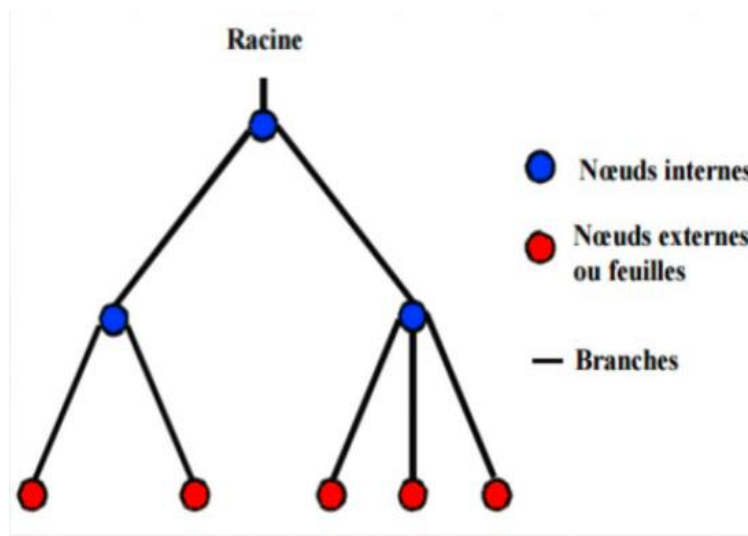


Figure 4 : Exemple d'un arbre phylogénétique

4. Différents arbres phylogénétiques

4.1. Arbre non enraciné

Un arbre non enraciné, il existe un seul et unique chemin menant d'un sommet à un autre. Les arrêtes représentent les liens de parenté entre les espèces. Dans un tel arbre, nous ne pouvons pas définir la relation entre un ancêtre et ses descendants (Rawlings, 2002).

4.2. Arbre enraciné

Un arbre enraciné comprend une racine qui représente l'ancêtre commun de toutes les espèces représentées. Lorsque l'ancêtre commun est identifié. Il est orienté dans le sens du temps d'évolution des espèces et présente une relation de descendance entre les nœuds. Souvent, il est impossible d'identifier l'origine de diversification des espèces. Il est impossible de retrouver la racine d'un arbre phylogénétique sans faire l'hypothèse de l'horloge moléculaire. Cette hypothèse suppose que les événements mutationnels se produisent à cadence régulière au cours du temps. Elle est peu réaliste en biologie, d'où l'intérêt accordé aux arbres

non enracinés. Les notions de temps et d'ancêtres se perdent avec ce type d'arbre. Il est souvent utilisé pour la classification des espèces (Rawlings, 2002).

La figure ci-dessous présente les deux types d'arbres pour quatre espèces a, b, c, d. Les figures (b) et (c) montrent la différence dans l'évolution entraînée par un changement de la position de la racine. Sur la figure (b) le sous arbre X regroupe les espèces a et b tandis que sur la figure (c), le sous arbre X est composé des espèces b, c et d.

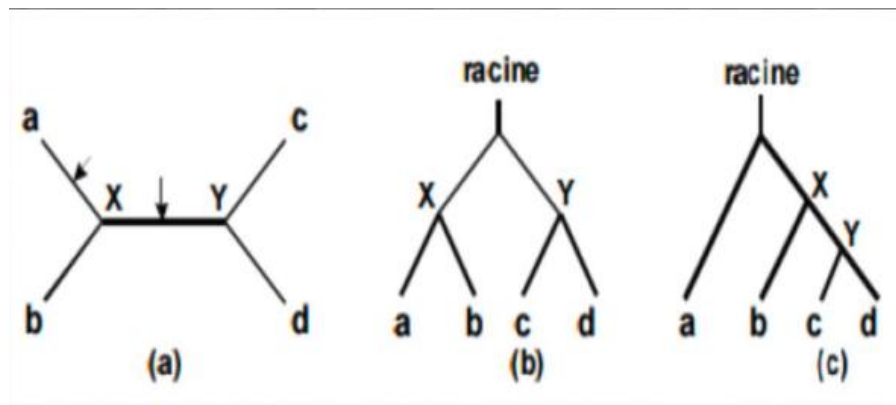


Figure 5: Différents arbre phylogénétique : (a) arbre non enraciné, (b) et (c) arbre enracinés.

5. Construction d'arbres phylogénétiques

La construction d'un arbre phylogénétique débute par ce que l'on appelle « alignement » qui consiste à mettre en correspondance les sites des séquences des espèces de manière à pouvoir les comparer les unes aux autres. Les séquences utilisées pour la reconstruction peuvent être de l'ADN ou de l'ARN. Les méthodes d'inférence phylogénétique présentées sont exposées en détails dans (Swofford, 1996).

Il existe quatre grandes approches pour inférer des arbres phylogénétiques :

L'approche de distances, le maximum de parcimonie, le maximum de vraisemblance et l'approche bayésienne (Felsenstein, 2004).

Chapitre 3 : Méthodes de délimitation

1. Délimitation

Les méthodes bio-informatiques de délimitation des espèces sont des méthodes analytiques qui permettent de séparer les séquences génétiques en groupes, qui sont biologiquement distincts, afin de délimiter des espèces. Ces méthodes sont de plus en plus utilisées pour l'identification des espèces, l'élucidation de leurs relations évolutives et l'étude de leur biologie.

Les méthodes de code à barres sont particulièrement importantes dans les enquêtes à grande échelle car elles favorisent la découverte rapide d'espèces et les estimations de la biodiversité parmi celles-ci, les méthodes basées sur la distance, sont le choix le plus courant (Kapli *et al*, 2017).

Nous avons appliqué quatre (4) méthodes : mPTP, bPTP, ABGD et ASAP.

1.1. mPTP (multi-rate Poisson Tree Processes)

Est une méthode améliorée qui atténué les lacunes théoriques et techniques du PTP. Il intègre différents niveaux de diversité génétique intraspécifique provenant de différences dans l'histoire évolutive ou l'échantillonnage de chaque espèce (Kapli *et al*, 2017).

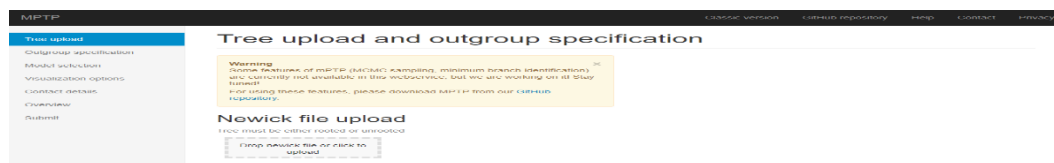


Figure 6: mPTP.

1.2. bPTP (Bayesian Poisson Tree Process)

bPTP est une implémentation bayésienne du modèle PTP pour la délimitation des espèces.



Figure 8 : bPTP.

1.3. ASAP (Assemble Species by Automatic Partitioning)

Méthode pour construire des partitions d'espèces à partir d'alignements de séquences de locus uniques (c'est-à-dire des ensembles de données de codes à barres) ASAP est la mise en œuvre d'un algorithme de clustering hiérarchique qui n'utilise que des distances génétiques par paires, évitant ainsi la charge de calcul de la reconstruction phylogénétique (Puillandre *et al*, 2021).

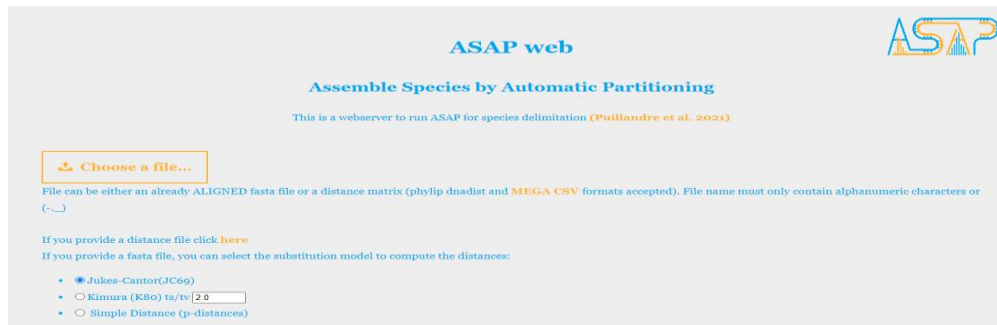


Figure 7: ASAP.

1.4. ABGD (Automatic Barcoding Gap Discovery).

La méthode ABGD (Automatic Barcode Gap Discovery) est un exemple de mise en œuvre largement utilisée de la délimitation phylogénétique des espèces. Le but du programme ABGD (Coissac *et al*, 2012) est d'identifier les seuils d'écart des codes-barres dans un processus automatisé.

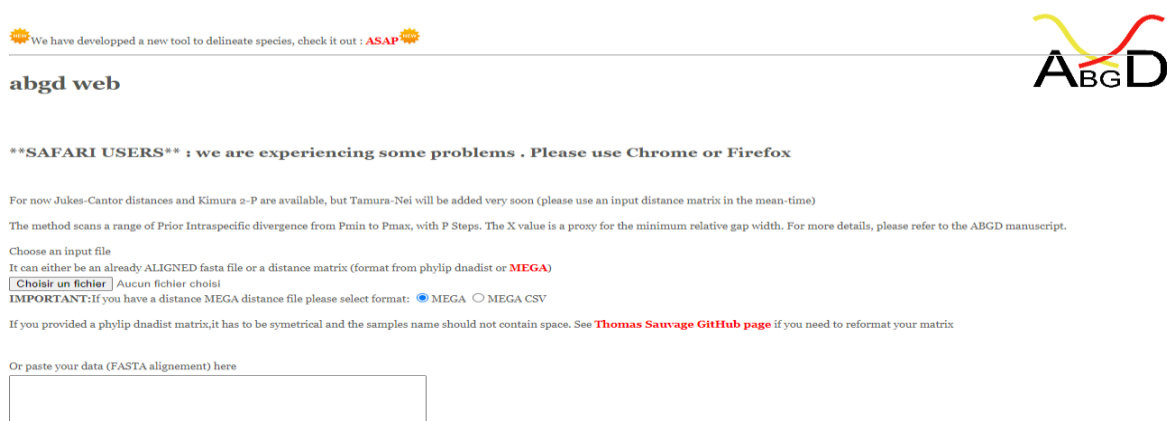


Figure 8 : ABGD.

Deuxième partie : Partie Expérimentale

Chapitre 4 : Méthodologie

1. Origine des séquences

La méthode de téléchargement et de curation des séquences de gènes ITS à partir de la base de données **GenBank** est importante pour l'étude et l'analyse du genre *Saccharomyces*. Elle commence par une recherche dans la base de données pour identifier les séquences d'ITS correspondant au genre *Saccharomyces*. Les séquences sont téléchargées et soumises à une analyse de qualité pour éliminer les séquences de mauvaise qualité ou les doublons. Les séquences sont ensuite alignées pour identifier les régions conservées et les régions variables d'ITS. Finalement, les séquences alignées peuvent être utilisées pour construire des arbres phylogénétiques afin d'étudier les relations évolutives entre les différentes espèces de *Saccharomyces*.

Notre analyse à débiter par une recherche de « *Saccharomyces* » dans la banque de données publiques **Genbank**.

1.1. GenBank

Banque de données publique complète de séquences nucléotides et d'annotations bibliographiques et biologiques. Il est construit et distribuée par le National Center for Biotechnology Information (NCBI) (Benson *et al*, 2018).

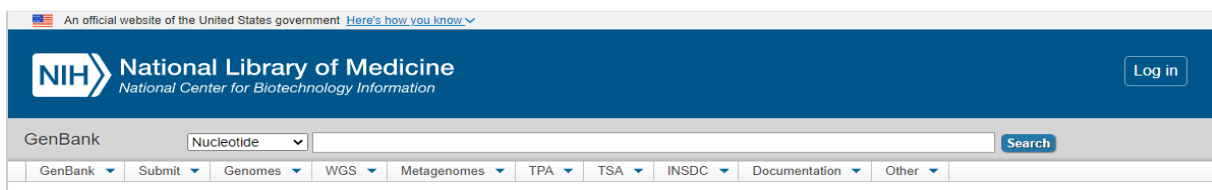


Figure 9: Banques de données publiques **GenBank**.

Nous avons obtenu à partir de recherche, des séquences combinés avec différents marqueurs. (Voir figure ci-dessous).

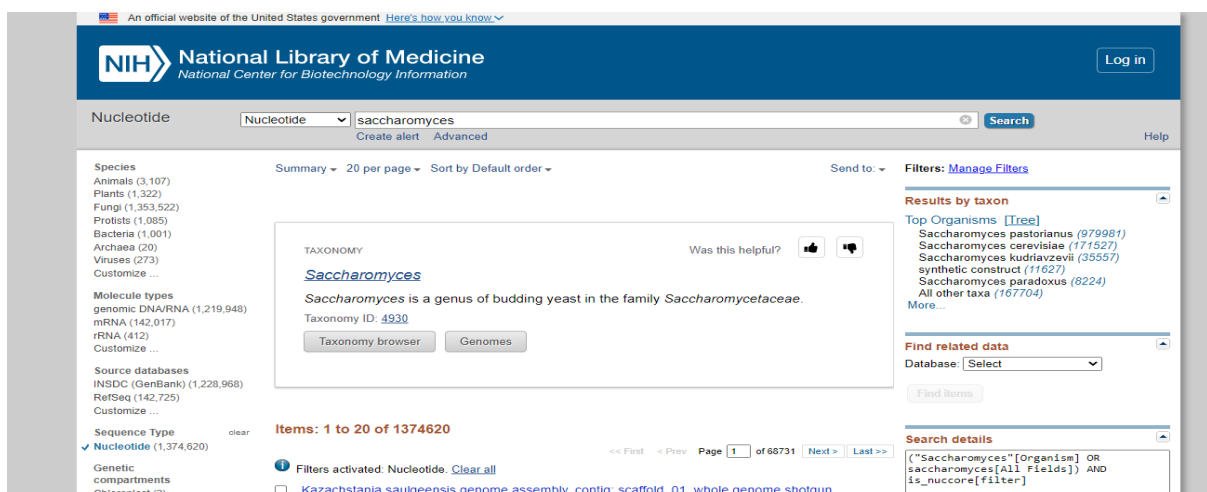


Figure 10 : Une recherche de « *Saccharomyces* » dans GenBank.

Nous avons par la suite choisi une séquence *S. cerevisiae* avec marqueur ITS dont le numéro d'accèsion est comme suit : MW057253.1 (annexe1), pour la blaster dans la plateforme BLAST, puis on a saisi le genre *Saccharomyces* dans la case **ADD organism** et sélectionner le nombre maximum de séquences alignées 5000 dans la case **Max target sequences** et appuyer sur la touche **BLAST**.

The screenshot shows the BLAST web interface. At the top, there is a navigation bar with the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". Below this, the page title is "BLAST® » blastn suite". The main content area is titled "Standard Nucleotide BLAST".

The "Enter Query Sequence" section contains a text area with the following sequence:


```

  ATACTGAGCGTATTGGAACGTTATCGATAAGAGAGAGCGCTAGGCGAA
  CAATGTTCTTAAAGTTGACCTCAAATCAG
  GTAGGAGTACCCGCTGAACCTAAGCATATCAATAAGCGGAGGA
  
```

 Below the text area, there are options to "Or, upload file" and "Job Title" (MW057253.1 S cerevisiae). There is also a checkbox for "Align two or more sequences".

The "Choose Search Set" section shows the "Database" set to "Standard databases (nr etc.)". There is a link to "Try experimental taxonomic nt databases". The "Organism" field is set to "Saccharomyces (taxid:4930)".

Red arrows point to the "Clear" button next to the "Enter accession number(s), gi(s), or FASTA sequence(s)" label, the "Job Title" field, and the "Saccharomyces (taxid:4930)" field.

Figure 11: BLAST.

The screenshot shows the "Algorithm parameters" section of the BLAST suite. It is divided into three main sections: "General Parameters", "Scoring Parameters", and "Filters and Masking".

The "General Parameters" section includes:

- Max target sequences:** Set to 5000 (highlighted in yellow with a red arrow).
- Short queries:** Checked "Automatically adjust parameters for short input sequences".
- Expect threshold:** Set to 0.05.
- Word size:** Set to 28.
- Max matches in a query range:** Set to 0.

The "Scoring Parameters" section includes:

- Match/Mismatch Scores:** Set to 1,-2.
- Gap Costs:** Set to Linear.

The "Filters and Masking" section includes:

- Filter:** Checked "Low complexity regions".
- Mask:** Checked "Mask for lookup table only".

At the bottom, there is a large blue "BLAST" button. To its right, there is a summary of the search: "Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)" and a checked checkbox for "Show results in a new window".

Red arrows point to the "BLAST" button and the "Show results in a new window" checkbox.

Figure 12: BLAST suite.

Parmi les 5000 séquences alignées nous avons obtenus 3500 séquences combinés.

The screenshot shows the BLAST suite interface. At the top, it displays 'BLAST® » blastn suite » results for RID-73ZG3SM6013'. Below this, there are navigation links like 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. A search bar contains 'MW057253.1 S cerevisiae'. The 'Filter Results' panel is visible, with 'Organism' set to 'S cerevisiae'. The 'Sequences producing significant alignments' section shows a dropdown menu set to '5000' sequences. A red arrow points to this dropdown menu.

Figure 13 : BLAST suite.

En se basant sur les valeurs de (**Query coverage, E-Value et Percent Identity**), nous avons pu faire un curage manuel de ces 3500 séquences combinés. On a procédé à la suppression des séquences identiques similaires, des séquences contenant des bases ambiguës et des séquences complètes et enfin éliminer les doublons ce qui nous a permis de réduire le nombre de séquences à 136 séquences.

The screenshot shows a detailed view of the BLAST results. A table lists sequences with columns: Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, and Acc. Len. A dropdown menu is open over the 'E value' column, showing options like 'FASTA (complete sequence)', 'FASTA (aligned sequences)', 'GenBank (complete sequence)', 'Hit Table (text)', 'Hit Table (CSV)', 'Text', 'Descriptions Table (CSV)', 'XML', and 'ASN.1'. Red circles highlight the 'E value' and 'Per. Ident' columns in the table. The 'Sequences producing significant alignments' section shows '136 sequences selected'.

Figure 14 : BLAST suite.

Cet ensemble de données vont nous permettre de les télécharger sous format **FASTA**, pour notre analyse phylogénétique visant à mieux comprendre la diversité et la taxonomie du genre *Saccharomyces*.

```
*Sequences_Saccharomyces 21 - Bloc-notes
Fichier Edition Format Affichage Aide
>KY105144.1 Saccharomyces cf. cerevisiae/paradoxus culture CBS:6751 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA
GTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGAAATTAATAATTTTGATATTGGATTTTTTGTGTTT
GGCAAGAGCGTGAGAGCTTTTACTGGCAAGAAGACAAGAGATGGAGAGTCCAGCCGGGCTGCGCTTAAGTGC GCGGTC
TTGCTAGGCTTGCAAAATTTCTTCTGCTATTCCAACAGTGAGACTTCTCTGTTTTGTTATAGGACAAATAAAACCGT
TTCAATACAACACACTGTGGAGTTTTATACCTTTGCAACTTTTCTTTGGGCTTCGAGCAATCGAGCCAGAGGTAACA
AACACAACAATTTTATTTATTTATTTAAATTTTGTCAAAAACAAGAATTTTCGTAACGGAAATTTAAAAATAATTA
AAACTTTCAACAACGGATCTCTGGTTCTCGCATCGATGAAGAACCAGCGAAATGCGATACGTAATGTAATGCAGAA
TTCCGTAATCATCGAATCTTGAACGCACATTGCGCCCTTGGTATTCCAGGGGGCATGCTGTTGAGCGTCAATTTCC
TTCTCAAACTTCTGTTGGTAGTGAGTGATACTCTGGAGTTAACTTGAATGCTGGCCTTTTCATTGGATGTTTTT
TTTTCAAAGAGAGGTTTCTCGCGTCTTGAGGTATAATGCAAGTACGGTCTTTAGGTTTACCAACTGCGGCTAATC
TTTTTTGACTGAGCGTATTGAAACGTTATCGATAAGAAGAGAGCGTCTAGGCGAACAAATGTTCTTAAAGTTGACCTCA
AATCAGGTAGGAGTACCCGCTGAACCTAAGCATATCAA
>KY105171.1 Saccharomyces cf. cerevisiae/paradoxus culture CBS:2834 small subunit ribosomal RNA
TGTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGAAATTAATAATTTTGATATTGGATTTTTTGTGTT
TGGCAAGAGCGTGAGAGCTTTTACTGGCAAGAAGACAAGAGATGGAGAGTCCAGCCGGGCTGCGCTTAAGTGC GCGGTC
CTTGTAGGCTTGCAAAATTTCTTCTGCTATTCCAACAGTGAGACTTCTCTGTTTTGTTATAGGACAAATAAAACCG
TTTCAATACAACACACTGTGGAGTTTTATACCTTTGCAACTTTTCTTTGGGCTTCGAGCAATCGAGCCAGAGGTAACA
AACACAACAATTTTATTTATTTCAATTTAAATTTTGTCAAAAACAAGAATTTTCGTAACGGAAATTTAAAAATAATTA
AAACTTTCAACAACGGATCTCTGGTTCTCGCATCGATGAAGAACCAGCGAAATGCGATACGTAATGTAATGCAGAA
ATTCCGTAATCATCGAATCTTGAACGCACATTGCGCCCTTGGTATTCCAGGGGGCATGCTGTTGAGCGTCAATTTCC
TTCTCAAACTTCTGTTGGTAGTGAGTGATACTCTGGAGTTAACTTGAATGCTGGCCTTTTCATTGGATGTTTTT
TTTTCAAAGAGAGGTTTCTCGCGTCTTGAGGTATAATGCAAGTACGGTCTTTAGGTTTACCAACTGCGGCTAATC
TTTTTTGACTGAGCGTATTGAAACGTTATCGATAAGAAGAGAGCGTCTAGGCGAACAAATGTTCTTAAAGTTGACCTCA
AATCAGGTAGGAGTACCCGCTGAACCTAAGCATATCAA
>KY104983.1 Saccharomyces cf. bayanus/pastorianus culture CBS:375 small subunit ribosomal RNA
GTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGAAATTAATAATTTTGATATTGGATTTTTTGTGTTT
GGCAAGAGCGTGAGAGCTTTTACTGGCAAGAAGACAAGAGATGGAGAGTCCAGCCGGGCTGCGCTTAAGTGC GCGGTC
TTGCTAGGCTTGCAAAATTTCTTCTGCTATTCCAACAGTGAGACTTCTCTGTTTTGTTATAGGACAAATAAAACCG
TTCAATACAACACACTGTGGAGTTTTATACCTTTGCAACTTTTCTTTGGGCTTCGAGCAATCGAGCCAGAGGTAACA
AACACAACAATTTTATTTATTTCAATTTAAATTTTGTCAAAAACAAGAATTTTCGTAACGGAAATTTAAAAATAATTA
AAACTTTCAACAACGGATCTCTGGTTCTCGCATCGATGAAGAACCAGCGAAATGCGATACGTAATGTAATGCAGAA
TTCCGTAATCATCGAATCTTGAACGCACATTGCGCCCTTGGTATTCCAGGGGGCATGCTGTTGAGCGTCAATTTCC
TTCTCAAACTTCTGTTGGTAGTGAGTGATACTCTGGAGTTAACTTGAATGCTGGCCTTTTCATTGGATGTTTTT
TTTTCAAAGAGAGGTTTCTCGCGTCTTGAGGTATAATGCAAGTACGGTCTTTAGGTTTACCAACTGCGGCTAATC
TTTTTTGACTGAGCGTATTGAAACGTTATCGATAAGAAGAGAGCGTCTAGGCGAACAAATGTTCTTAAAGTTGACCTCA
AATCAGGTAGGAGTACCCGCTGAACCTAAGCATATCAA
```

Figure 15: Séquences téléchargées avant nettoyage.

```
*Sequences_Saccharomyces 21 - Bloc-notes
Fichier Edition Format Affichage Aide
>KY105144.1 S cf. cerevisiae/paradoxus
GTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGAAATTAATAATTTTGATATTGGATTTTTTGTGTTT
GGCAAGAGCGTGAGAGCTTTTACTGGCAAGAAGACAAGAGATGGAGAGTCCAGCCGGGCTGCGCTTAAGTGC GCGGTC
TTGCTAGGCTTGCAAAATTTCTTCTGCTATTCCAACAGTGAGACTTCTCTGTTTTGTTATAGGACAAATAAAACCGT
TTCAATACAACACACTGTGGAGTTTTATACCTTTGCAACTTTTCTTTGGGCTTCGAGCAATCGAGCCAGAGGTAACA
AACACAACAATTTTATTTATTTCAATTTAAATTTTGTCAAAAACAAGAATTTTCGTAACGGAAATTTAAAAATAATTA
AAACTTTCAACAACGGATCTCTGGTTCTCGCATCGATGAAGAACCAGCGAAATGCGATACGTAATGTAATGCAGAA
TTCCGTAATCATCGAATCTTGAACGCACATTGCGCCCTTGGTATTCCAGGGGGCATGCTGTTGAGCGTCAATTTCC
TTCTCAAACTTCTGTTGGTAGTGAGTGATACTCTGGAGTTAACTTGAATGCTGGCCTTTTCATTGGATGTTTTT
TTTTCAAAGAGAGGTTTCTCGCGTCTTGAGGTATAATGCAAGTACGGTCTTTAGGTTTACCAACTGCGGCTAATC
TTTTTTGACTGAGCGTATTGAAACGTTATCGATAAGAAGAGAGCGTCTAGGCGAACAAATGTTCTTAAAGTTGACCTCA
AATCAGGTAGGAGTACCCGCTGAACCTAAGCATATCAA
>KY105171.1 S cf. cerevisiae/paradoxus
TGTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGAAATTAATAATTTTGATATTGGATTTTTTGTGTT
TGGCAAGAGCGTGAGAGCTTTTACTGGCAAGAAGACAAGAGATGGAGAGTCCAGCCGGGCTGCGCTTAAGTGC GCGGTC
CTTGTAGGCTTGCAAAATTTCTTCTGCTATTCCAACAGTGAGACTTCTCTGTTTTGTTATAGGACAAATAAAACCG
TTTCAATACAACACACTGTGGAGTTTTATACCTTTGCAACTTTTCTTTGGGCTTCGAGCAATCGAGCCAGAGGTAACA
AACACAACAATTTTATTTATTTCAATTTAAATTTTGTCAAAAACAAGAATTTTCGTAACGGAAATTTAAAAATAATTA
AAACTTTCAACAACGGATCTCTGGTTCTCGCATCGATGAAGAACCAGCGAAATGCGATACGTAATGTAATGCAGAA
ATTCCGTAATCATCGAATCTTGAACGCACATTGCGCCCTTGGTATTCCAGGGGGCATGCTGTTGAGCGTCAATTTCC
TTCTCAAACTTCTGTTGGTAGTGAGTGATACTCTGGAGTTAACTTGAATGCTGGCCTTTTCATTGGATGTTTTT
TTTTCAAAGAGAGGTTTCTCGCGTCTTGAGGTATAATGCAAGTACGGTCTTTAGGTTTACCAACTGCGGCTAATC
TTTTTTGACTGAGCGTATTGAAACGTTATCGATAAGAAGAGAGCGTCTAGGCGAACAAATGTTCTTAAAGTTGACCTCA
AATCAGGTAGGAGTACCCGCTGAACCTAAGCATATCAA
>KY104983.1 S cf. bayanus/pastorianus
GTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGAAATTAATAATTTTGATATTGGATTTTTTGTGTTT
GGCAAGAGCGTGAGAGCTTTTACTGGCAAGAAGACAAGAGATGGAGAGTCCAGCCGGGCTGCGCTTAAGTGC GCGGTC
TTGCTAGGCTTGCAAAATTTCTTCTGCTATTCCAACAGTGAGACTTCTCTGTTTTGTTATAGGACAAATAAAACCGT
TTCAATACAACACACTGTGGAGTTTTATACCTTTGCAACTTTTCTTTGGGCTTCGAGCAATCGAGCCAGAGGTAACA
AACACAACAATTTTATTTATTTCAATTTAAATTTTGTCAAAAACAAGAATTTTCGTAACGGAAATTTAAAAATAATTA
AAACTTTCAACAACGGATCTCTGGTTCTCGCATCGATGAAGAACCAGCGAAATGCGATACGTAATGTAATGCAGAA
TTCCGTAATCATCGAATCTTGAACGCACATTGCGCCCTTGGTATTCCAGGGGGCATGCTGTTGAGCGTCAATTTCC
TTCTCAAACTTCTGTTGGTAGTGAGTGATACTCTGGAGTTAACTTGAATGCTGGCCTTTTCATTGGATGTTTTT
TTTTCAAAGAGAGGTTTCTCGCGTCTTGAGGTATAATGCAAGTACGGTCTTTAGGTTTACCAACTGCGGCTAATC
TTTTTTGACTGAGCGTATTGAAACGTTATCGATAAGAAGAGAGCGTCTAGGCGAACAAATGTTCTTAAAGTTGACCTCA
ATCAGGTAGGAGTACCCGCTGAACCTAAGCATATCAA
```

Figure 16: Séquences téléchargées après nettoyage.

Les séquences nucléotidiques utilisées dans notre analyse et leur numéro d'accèsion sont présentés dans le tableau suivant :

Tableau 1: Espèces de *Saccharomyces* et leur numéro d'accèsion extraites de GenBank.

Taxon (Genre <i>Saccharomyces</i>)	Numéro D'accèsion (GenBank)
<i>Saccharomyces sp.</i>	MK267682.1
<i>Saccharomyces cf. cerevisiae/paradoxus</i>	KY105144.1
<i>Saccharomyces cf. cerevisiae/paradoxus</i>	KY105171.1
<i>Saccharomyces cf. bayanus/pastorianus</i>	KY104983.1
<i>Saccharomyces cf. bayanus/pastorianus</i>	KY104974.1
<i>Saccharomyces sp. Boulardii</i>	MZ377294.1
<i>Saccharomyces sp.</i>	MK271335.1
<i>Saccharomyces sp.</i>	MT903454.1
<i>Saccharomyces arboricola</i>	EF580917.1
<i>Saccharomyces sp. Lalvin</i>	AY942706.1
<i>Saccharomyces sp.</i>	OP919466.1
<i>Saccharomyces sp.</i>	KU350338.1
<i>Saccharomyces sp.</i>	OP934234.1
<i>Saccharomyces sp.</i>	KU350336.1
<i>Saccharomyces sp.</i>	MF438280.1
<i>Saccharomyces sp. Lalvin</i>	AY942702.1
<i>Saccharomyces sp.</i>	HG764814.1
<i>Saccharomyces sp. Assmannshausen</i>	DQ153232.1
<i>Saccharomyces cariocanus</i>	KY104991.1
<i>Saccharomyces cariocanus</i>	NR_144772.1
<i>Saccharomyces sp. Zymaflore</i>	AY942703.1
<i>Saccharomyces cariocanus</i>	AJ229061.1
<i>Saccharomyces sp. Anchor</i>	DQ153227.1
<i>Saccharomyces cariocanus</i>	MH595331.1
<i>Saccharomyces cariocanus</i>	KY104988.1
<i>Saccharomyces arboricola</i>	KY104943.1
<i>Saccharomyces arboricola</i>	MZ068118.1
<i>Saccharomyces chevalieri</i>	AF005714.1
<i>Saccharomyces ellipsoideus</i>	AF005716.1
<i>Saccharomyces sp. isolate cerevisiae</i>	MT163713.1
<i>Saccharomyces cf. cerevisiae/paradoxus</i>	KY109343.1
<i>Saccharomyces cf. cerevisiae/paradoxus</i>	KY109359.1
<i>Saccharomyces cf. cerevisiae/paradoxus</i>	KY109332.1
<i>Saccharomyces chevalieri</i>	AF005707.1
<i>Saccharomyces ellipsoideus</i>	AF005708.1
<i>Saccharomyces eubayanus</i>	MW710882.1
<i>Saccharomyces eubayanus</i>	OK051019.1
<i>Saccharomyces eubayanus</i>	OK051021.1
<i>Saccharomyces eubayanus</i>	KM384476.1

<i>Saccharomyces eubayanus</i>	KR871556.1
<i>Saccharomyces eubayanus</i>	KR871555.1
<i>Saccharomyces eubayanus</i>	KR871554.1
<i>Saccharomyces eubayanus</i>	NR_137586.1
<i>Saccharomyces eubayanus</i>	OK051441.1
<i>Saccharomyces eubayanus</i>	OK051321.1
<i>Saccharomyces pastorianus</i>	KJ507664.1
<i>Saccharomyces pastorianus</i>	NR_165985.1
<i>Saccharomyces pastorianus</i>	MW710905.1
<i>Saccharomyces pastorianus</i>	MW710883.1
<i>Saccharomyces pastorianus</i>	MW710918.1
<i>Saccharomyces pastorianus</i>	MW710897.1
<i>Saccharomyces pastorianus</i>	KY105221.1
<i>Saccharomyces pastorianus</i>	AB279757.1
<i>Saccharomyces pastorianus</i>	MW980898.1
<i>Saccharomyces pastorianus</i>	KY109458.1
<i>Saccharomyces bayanus</i>	MK267707.1
<i>Saccharomyces bayanus</i>	NR_165984.1
<i>Saccharomyces bayanus</i>	MH023214.1
<i>Saccharomyces bayanus</i>	KJ706988.1
<i>Saccharomyces bayanus</i>	D89887.1
<i>Saccharomyces bayanus</i>	KY104950.1
<i>Saccharomyces bayanus</i>	KY104956.1
<i>Saccharomyces bayanus</i>	Z95947.1
<i>Saccharomyces bayanus</i>	Z95946.1
<i>Saccharomyces bayanus</i>	KY104965.1
<i>Saccharomyces paradoxus</i>	KY105212.1
<i>Saccharomyces paradoxus</i>	MH013273.1
<i>Saccharomyces paradoxus</i>	KY105208.1
<i>Saccharomyces paradoxus</i>	KY105213.1
<i>Saccharomyces paradoxus</i>	KY105211.1
<i>Saccharomyces paradoxus</i>	OK135369.1
<i>Saccharomyces paradoxus</i>	OP899938.1
<i>Saccharomyces paradoxus</i>	MZ185342.1
<i>Saccharomyces paradoxus</i>	KY105219.1
<i>Saccharomyces paradoxus</i>	KF057503.1
<i>Saccharomyces paradoxus</i>	KT207224.1
<i>Saccharomyces paradoxus</i>	KP250839.1
<i>Saccharomyces paradoxus</i>	OK052373.1
<i>Saccharomyces paradoxus</i>	OK051055.1
<i>Saccharomyces paradoxus</i>	NR_138272.1
<i>Saccharomyces boulardii</i>	AY428861.1
<i>Saccharomyces boulardii</i>	OP925906.1
<i>Saccharomyces boulardii</i>	MH266045.1
<i>Saccharomyces boulardii</i>	MK672872.1
<i>Saccharomyces boulardii</i>	MK672873.1
<i>Saccharomyces boulardii</i>	MW959788.1
<i>Saccharomyces boulardii</i>	OM722057.1

<i>Saccharomyces boulardii</i>	MW596413.1
<i>Saccharomyces boulardii</i>	MZ148310.1
<i>Saccharomyces cerevisiae</i>	KY105007.1
<i>Saccharomyces cerevisiae</i>	MN158119.1
<i>Saccharomyces cerevisiae</i>	KC183726.1
<i>Saccharomyces cerevisiae</i>	MZ361705.1
<i>Saccharomyces cerevisiae</i>	MT322849.1
<i>Saccharomyces cerevisiae</i>	KU131578.1
<i>Saccharomyces cerevisiae</i>	KJ502662.1
<i>Saccharomyces cerevisiae</i>	AM262830.1
<i>Saccharomyces cerevisiae</i>	AM262827.1
<i>Saccharomyces cerevisiae</i>	KT175188.1
<i>Saccharomyces cerevisiae</i>	KJ502661.1
<i>Saccharomyces cerevisiae</i>	MK439496.1
<i>Saccharomyces cerevisiae</i>	MH453895.1
<i>Saccharomyces cerevisiae</i>	KP132589.1
<i>Saccharomyces cerevisiae</i>	MZ098648.1
<i>Saccharomyces cerevisiae</i>	KC254082.1
<i>Saccharomyces cerevisiae</i>	FN393995.1
<i>Saccharomyces cerevisiae</i>	GQ376091.1
<i>Saccharomyces cerevisiae</i>	FJ793809.1
<i>Saccharomyces cerevisiae</i>	AM262826.1
<i>Saccharomyces cerevisiae</i>	KC588952.1
<i>Saccharomyces cerevisiae</i>	MW057253.1
<i>Saccharomyces cerevisiae</i>	OP252624.1
<i>Saccharomyces cerevisiae</i>	KP132590.1
<i>Saccharomyces cerevisiae</i>	KC254075.1
<i>Saccharomyces cerevisiae</i>	OQ711762.1
<i>Saccharomyces cerevisiae</i>	LC413776.1
<i>Saccharomyces cerevisiae</i>	MG015999.1
<i>Saccharomyces cerevisiae</i>	KU729071.1
<i>Saccharomyces cerevisiae</i>	OM348909.1
<i>Saccharomyces cerevisiae</i>	KT726924.1
<i>Saccharomyces cerevisiae</i>	OL457918.1
<i>Saccharomyces cerevisiae</i>	KY105021.1
<i>Saccharomyces cerevisiae</i>	AM900400.1
<i>Saccharomyces cerevisiae</i>	MG720174.1
<i>Saccharomyces uvarum</i>	NR_153310.1
<i>Saccharomyces uvarum</i>	EU145771.1
<i>Saccharomyces uvarum</i>	KJ507662.1
<i>Saccharomyces uvarum</i>	KY105228.1
<i>Saccharomyces uvarum</i>	KY105229.1
<i>Saccharomyces uvarum</i>	MW710375.1
<i>Saccharomyces uvarum</i>	OQ747985.1
<i>Saccharomyces uvarum</i>	MW980899.1
<i>Saccharomyces uvarum</i>	MW980894.1
<i>Saccharomyces uvarum</i>	MW980897.1

2. Construction de l'arbre phylogénétique

L'arbre phylogénétique, également appelé phylogénie, est un diagramme qui décrit les lignes d'évolution de différentes espèces, organismes ou gènes à partir d'un ancêtre commun. Les phylogénies sont utiles pour organiser la connaissance de la diversité biologique, pour structurer les classifications et pour donner un aperçu des événements qui se sont produits au cours de l'évolution (Baum, 2008).

La construction phylogénétique, consiste à représenter sous forme d'un arbre les relations évolutives entre nos séquences.

2.1. Méthodes de construction de l'arbre phylogénétique

Pour l'étude des relations phylogénétiques entre les champignons *Saccharomyces*, nous avons utilisés la plateforme en ligne NG-phylogeny qui est disponible sur le lien suivant : <https://ngphylogeny.fr/>.



Figure 17: NG-phylogeny.

Une fois la fenêtre apparaît à l'écran nous avons mis le curseur sur **phylogeny analysis** puis cliquez l'option **à la Carte** une fenêtre apparaît de nouveau qui rassemble les quatre étapes de la construction de l'arbre (**Multiple Alignment, Alignment Curation, Tree Inference, Tree Rendering**).

Chaque étape présente différents logiciels qu'on peut choisir pour notre traitement.

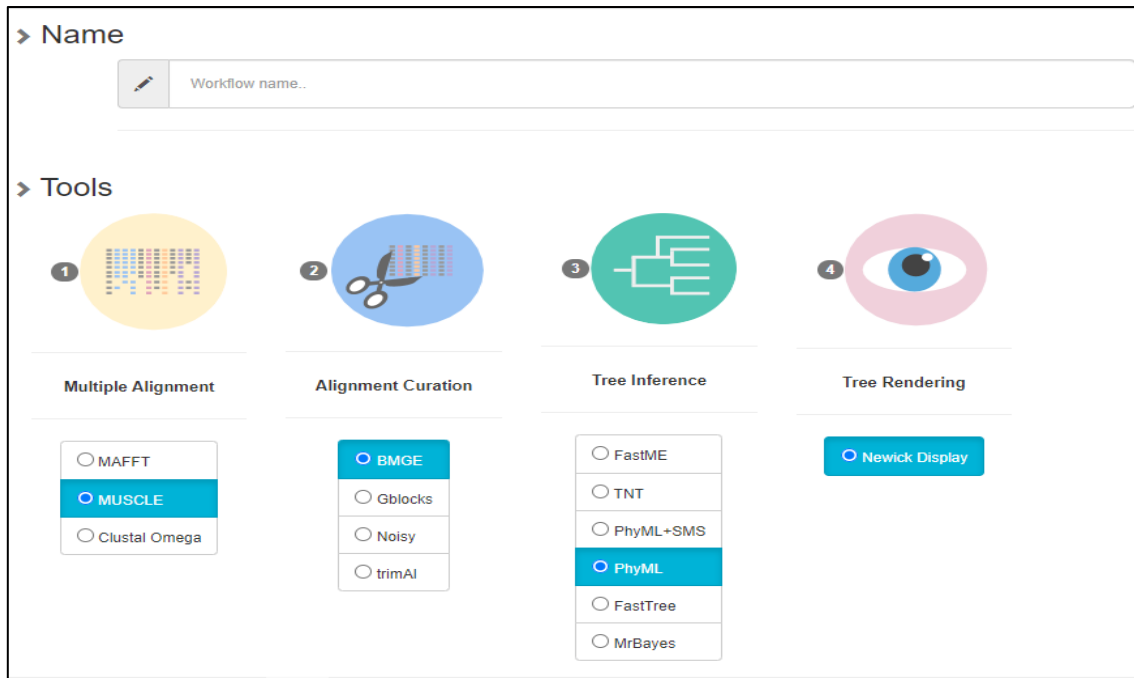


Figure 18 : Etapes NG-phylogeny.

Concernant la première étape qui est **l'alignement multiple des séquences** nous avons choisis **MUSCLE** (Multiple Sequence Comparison by Log-Expectation) est un logiciel informatique pour l'alignement de séquences multiples de protéines et de nucléotides (Edgar, 2004).

Quant à la deuxième étape celle du **nettoyage des séquences nucléiques**, nous avons sélectionné la méthode **BMGE** (Block Mapping and Gathering with Entropy), qui est conçu pour sélectionner des régions dans un alignement de séquences multiples qui conviennent à l'inférence phylogénétique. BMGE met également en œuvre des méthodes d'élagage et de recodage visant à minimiser les artefacts de reconstruction de la phylogénie dus à l'hétérogénéité de la composition (Crisuolo et Gribaldo, 2010).

Cependant, la troisième étape qui consiste à **la construction d'arbre cas ML**, nous avons sélectionné la méthode **PhyML** (logiciel de phylogénie basé sur le principe du maximum de vraisemblance). Ce logiciel a été largement utilisé en raison de sa simplicité et d'un bon compromis entre précision et rapidité (Guindon et Gascuel, 2003).

Pour **la construction d'arbre cas Bayes** les étapes déjà citer sont les même pour les deux arbres seulement pour ce cas nous avons sélectionné la méthode **MrBayes** (programme d'inférence bayésienne de la phylogénie doté d'une interface en ligne de commande et devrait fonctionner sur diverses plates-formes informatiques) (Huelsenbeck et Ronquist, 2001).

Enfin pour la quatrième ou dernière étape il s'agit de **la visualisation de l'arbre** nous avons sélectionné directement Newick **Display** (format simple, strictement symbolisé, représentant les arbres phylogénétiques) (Cardona *et al*, 2008).

Par la suite nous avons cliqué sur **Create Workflow**, une fenêtre est apparue on a ajouté le fichier contenant nos données (séquences nettoyées) puis on a cliqué sur **Submit** pour obtenir des résultats qui nous permettent d'établir les méthodes de délimitation.

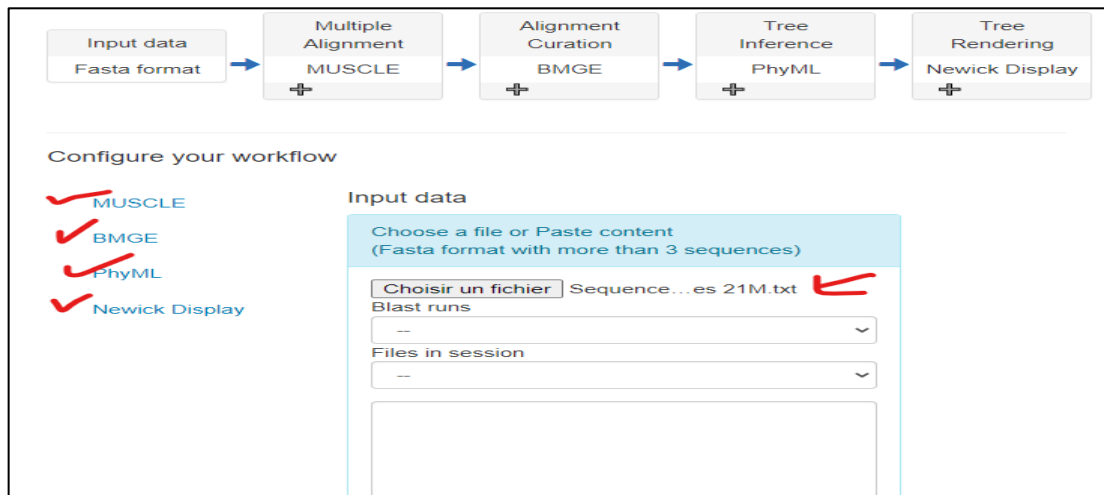


Figure 19 : Etapes NG-phylogeny cas ML.

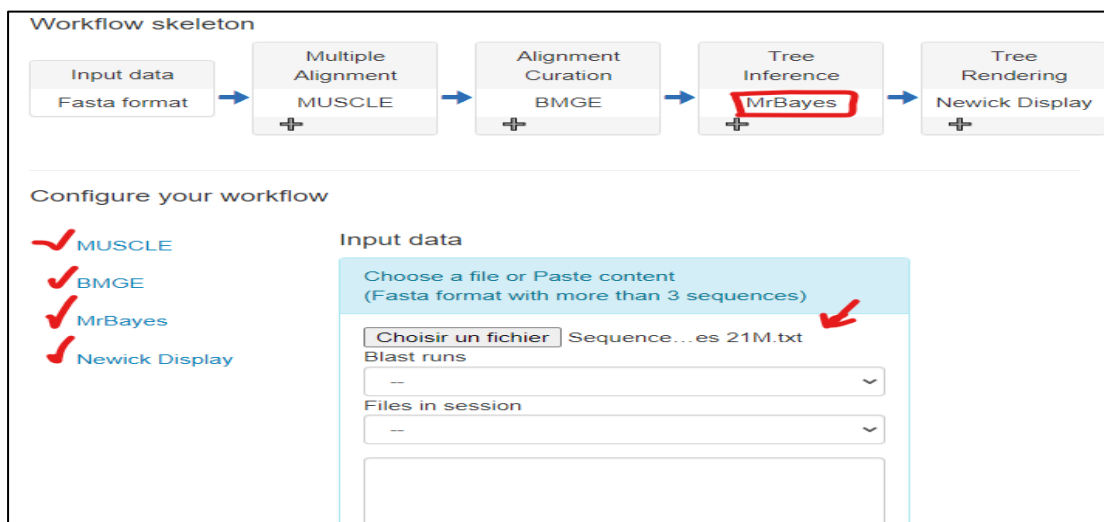


Figure 20: Etapes NG-phylogeny cas Bayes.

3. Délimitation des espèces

Méthodes de plus en plus utilisées pour séparer les séquences génétiques en groupes, l'identification des espèces et l'élucidation de leurs relations évolutives et l'étude de leur biologie.

Pour la réalisation de la délimitation des espèces nous avons appliqué quatre (4) méthodes mPTP, bPTP, ASAP et ABGD.

3.1. mPTP (multi-rate Poisson Tree Processes)

Nous avons au départ accédé à la plateforme **mPTP** à partir du lien suivant : <https://mptp.h-its.org/#/tree>

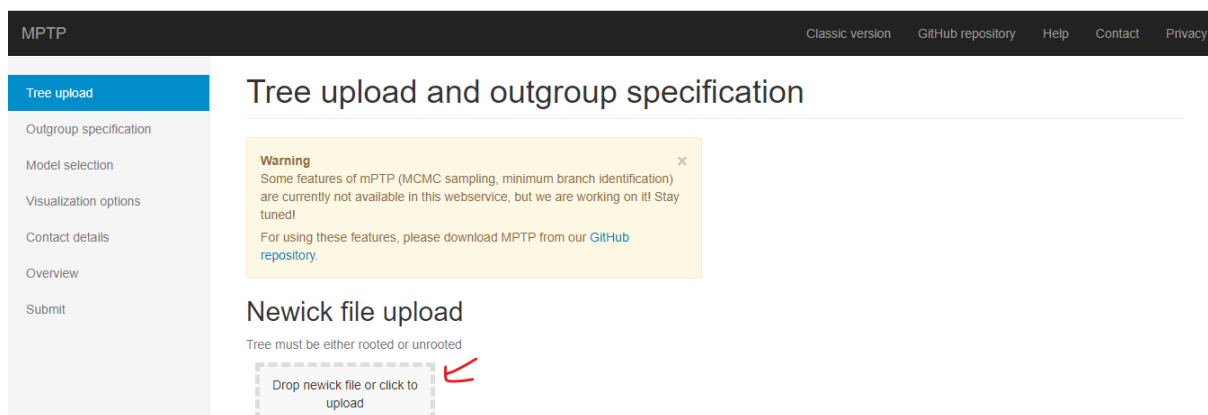


Figure 21: Plateforme mPTP.

Nous avons introduit notre fichier de L'arbre phylogénétique sous format **ML-Newick** obtenu à partir des étapes précédentes.

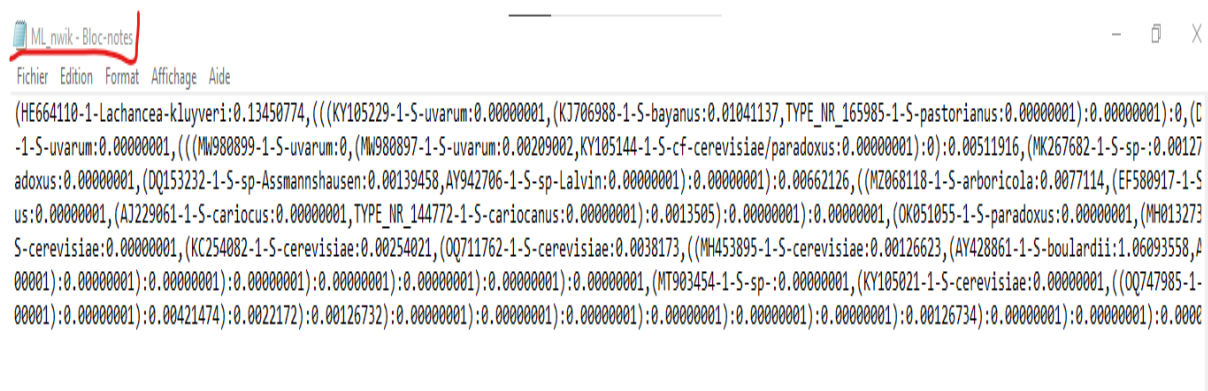


Figure 22 : Arbre phylogénétique format ML-Newick.

Nous avons ensuite sélectionné l'**outgroup** (exo groupe), espèce *Lachancea kluyveri* (numéro d'accension de GenBank HE664110.1) puis appuyer sur **Model selection** et **Visualization options** ensuite **Contact details** et enfin on introduit l'email et cliquer sur **Overview** pour démarrer l'opération de délimitation et obtenir les résultats.

3.2. bPTP (Bayesian Poisson Tree Process)

Nous avons au départ accédé à la plateforme **bPTP** à partir du lien suivant : <https://species.h-its.org/ptp/>

bPTP server: a Bayesian implementation of the PTP model for species delimitation
with *PhyloMap* for visualization

Please cite: **A General Species Delimitation Method with Applications to Phylogenetic Placements.** Zhang, Jiajie, Kapli, P., Pavlidis, P., and Stamatakis, A. *Bioinformatics (Oxford, England)*(2013), 29 (22): 2869-2876

bPTP is an updated version of the original maximum likelihood PTP (maximum likelihood PTP search result is part of the bPTP results), it adds Bayesian support (BS) values to delimited species on the input tree. Higher BS value on a node indicates all descendants from this node are more likely to be from one species. Note this web server only allows a single phylogenetic tree as input - to run on multiple trees (from Bayesian analysis or bootstrap), please download the **bPTP and PTP standalone**.

If you are not familiar with Bayesian analysis, please [read this first](#).

The server will only accept Newick format or NEXUS format with no annotations on the tree.
Here is an [example of Newick format](#) and an [example of NEXUS format](#).

Special note on taxa name: A valid taxa name should only contain letters, numbers, _ and -; and should NOT be pure numbers. Any other characters, in particular space, \$, #, %, (and), will cause errors!

My phylogenetic input tree, if input file contains multiple trees, only the first tree will be used:
 Bays Nwk.txt

My tree is:
Rooted If unrooted, the tree will be rooted at the longest branch.

No. MCMC generations:
100000 For small trees (<50 taxa), 100000 is usually enough, but you should always check for convergence.

Figure 23 : Plateforme bPTP.

Nous avons introduit notre fichier de l'arbre phylogénétique sous format Bayes-Newick obtenu à partir des étapes précédentes.

```
(AY428861-1-S-boulardii:0.4970748,((AF005708-1-S-ellipsoideus:0.0005732778,AF005707-1-S-chevalieri:0.002126118):0.0827782[1],(KY109458-1-S-pastorianus:0.06535966,(KY109456-1-S-cerevisiae:0.0009860614,KC588952-1-S-cerevisiae:0.0008449096,FJ793809-1-S-cerevisiae:0.0005515394,GQ376091-1-S-cerevisiae:0.0006238592,FH393995-1-S-cerevisiae:0.0004511604,TYPE_AM262830-1-S-cerevisiae:0.0006913062,KT175188-1-S-cerevisiae:0.0007567769,OP899938-1-S-paradoxus:0.001639614,OM348909-1-S-cerevisiae:0.001124542,(OR100000-1-S-paradoxus:0.001146639,KF057503-1-S-paradoxus:0.001639685,KY104988-1-S-cariocanus:0.0007809862,KY105219-1-S-paradoxus:0.002144744,KY104991-1-S-cariocanus:0.005895175,(TYPE_AM100000-1-S-pastorianus:0.0007464371,MW710897-1-S-pastorianus:0.0005528177):0.003093892[0.986754967]),(TYPE_IR_137586-1-S-eubayanus:0.001074182,KR871554-1-S-eubayanus:0.00092627,KR871555-1-S-eubayanus:0.0009315498,MH023214-1-S-bayanus:0.0007445684,MW710375-1-S-uvorum:0.00192903,OK051021-1-S-eubayanus:0.001035913,MN980894-1-S-uvorum:0.004908449,(MW710883-1-S-pastorianus:0.0009315498,MH023214-1-S-bayanus:0.0007445684,MW710375-1-S-uvorum:0.00192903,OK051021-1-S-eubayanus:0.001035913,MN980894-1-S-uvorum:0.004908449,(MW710883-1-S-pastorianus:0.0009315498,...
```

Figure 24 : Arbre phylogénétique format Bayes-Newick.

Après l'introduction du fichier nous avons choisi l'option arbre avec racine (**Rooted**) puis cliquez sur **Submit** pour démarrer la délimitation et recevoir par la suite les résultats.

3.3. ASAP (Assemble Species by Automatic Partitioning)

Nous avons au départ accédé à la plateforme **ASAP** à partir du lien suivant : <https://bioinfo.mnhn.fr/abi/public/asap/>



Figure 25 : Plateforme ASAP.

On a ensuite téléchargé le fichier sous format **FASTA** qui comporte toutes les séquences du genre *Saccharomyces* qui sont déjà alignées et nettoyé à partir de la plateforme NG-phylogeny.

Newick Display	13.	All tree images	✓	+ ⚙️ ▶️ ⬇️ 📄 tar
	12.	Tree image	✓	+ ⚙️ ▶️ ⬇️ 📄 svg
PhyML	11.	Mapping between short sequence id and names (useful to interpret some bootstrap log files if any)	✓	+ ⚙️ ▶️ ⬇️ 📄 bxt
	10.	PhyML Newick tree	✓	+ ⚙️ ▶️ ⬇️ 📄 nhx Viewer ITol
	9.	PhyML Statistics	✓	+ ⚙️ ▶️ ⬇️ 📄 bxt
	8.	PhyML log	✓	+ ⚙️ ▶️ ⬇️ 📄 bxt
BMGE	7.	BMGE Cleaned sequences Html	✓	+ ⚙️ ▶️ ⬇️ 📄 html
	6.	BMGE Cleaned sequences Nexus	✓	+ ⚙️ ▶️ ⬇️ 📄 nex
	5.	BMGE Cleaned sequences Fasta	✓	+ ⚙️ ▶️ ⬇️ 📄 fasta MSAViewer
MUSCLE	4.	BMGE Cleaned sequences Phylip	✓	+ ⚙️ ▶️ ⬇️ 📄 phylip
	3.	Muscle alignment (html)	✓	+ ⚙️ ▶️ ⬇️ 📄 fasta MSAViewer
Upload File	2.	Muscle alignment	✓	+ ⚙️ ▶️ ⬇️ 📄 fasta MSAViewer
	1.	Sequences_Saccharomyces 21M.txt	✓	+ ⚙️ ▶️ ⬇️ 📄 fasta MSAViewer

Figure 26 : Plateforme NG-phylogeny.

Nous avons éliminé l'outgroup (*Lachancea kluyveri* numéro d'accension de GenBank HE664110.1) du fichier déjà télécharger et enfin cliquez sur GO pour démarrer la délimitation et obtenir des résultats.

```

BMGECleanog - Bloc-notes
Fichier Edition Format Affichage Aide
>AY428861.1 S boulardii
-----TCCTCCGCTA---TTGATATGCTTAAGTTCCTACCTGATTGGTCAAACCT---GAACATTAACGTCCAATACGCTCAGTATAAAAAAATAGCCGAGTTGGACTAAAA---CGACGTACTTGCATTATA---CCT-----CAGCACG
>AF085707.1 S chevalieri
GGTAGGAGTACCCTCGCTGAACCTAAGCATATCAATA-----GGAAAGAAACCAACCGGATGGCT-----TAGTAACGGCGAGTGAAGC---GGCAAAGCTCAAATTTGAAATCTTGCCTGAGTTGTAATTTGCTTTGGGGCGTTCC----T
>KY189359.1 S cf. cerevisiae/paradoxus
GAAGAGAGCGTTTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCTTAAGCATATCAATAAGCGGAGGAAAAGAAAC--CAACCGGATTGGCTCAAATTTGAAATCTTGCCTGAGTTGTAATTTGCTTTGGGGCGTTCC----T
>KY189343.1 S cf. cerevisiae/paradoxus
GAAGAGAGCGCTTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCTT---AAGCATATCAATAAGGAGGAAAAGAAACCAACCGGATTGGCTCAAATTTGAAATCTTGCCTGAGTTGTAATTTGCTTTGGGGCGTTCC----T
>AF085708.1 S ellipsoideus
GGTAGGAGTACCCTCGCTGAACCTAAGCATATCAATA-----GGAAAGAAACCAACCGGATGGCT-----TAGTAACGGCGAGTGAAGC---GGCAAAGCTCAAATTTGAAATCTTGCCTGAGTTGTAATTTGCTTTGGGGCGTTCC----T
>KY189332.1 S cf. cerevisiae/paradoxus
GAAGAGAGCGCTTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCTTAAGCATATCAATAAGCGGAGGAAAAGAAAC--CAACCGGATTGGCTCAAATTTGAAATCTTGCCTGAGTTGTAATTTGCTTTGGGGCGTTCC----T
>KY189458.1 S pastorianus
GAAGAGAGCGCTTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCTTAAGCATATCAATAAGCGGAGGAAAAGAAAC--CAACCGGATTGGCTCAAATTTGAAATCTTGCCTGAGTTGTAATTTGCTTTGGGGCGTTCC----T
>MT163713.1 S sp.cerevisiae
-----
>MZ068118.1 S arboricola
-----CTGGCAAGAGATAAAGAGATGGAGAGTCCAG--GGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>JQ914752.1 S arboricola
-----AAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCGTGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGCGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>EF580917.1 S arboricola
GGAAAGATCATTAAAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCGTGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGCGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>KY184943.1 S arboricola
-----ATTTTGAAATTTTTTTGTTTTGGCAAGAGCGTGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGCGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>HG764814.1 S sp.
---AGGATCATTAAAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCATG--AGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGTGGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>KT287224.1 S paradoxus
GGAAAGATCATTAAAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCATGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGTGGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>KY184991.1 S cariocanus
---AGTTCATTAAAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCATGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGTGGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>MZ185342.1 S paradoxus
-----AAATTAATAATTTTGA--ATTTTGGTGGCAAGAGCATGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGTGGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>KY185211.1 S paradoxus
GGAAAGATCATTAAAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCATGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGTGGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG
>KY185213.1 S paradoxus
GGAAAGATCATTAAAGAAATTAATAATTTTGAAATTTTTTTGTTTTGGCAAGAGCATGAGAGCTTTTACTGGGCAAGAGATAAAGAGATGGAGAGTCCAGTGGGGCTGCGCTTAAAGTGGCGGTCTTTCTAGACTTGAAGTTCTTCTTCTGATTCCAAACAG

```

Figure 27 : Fichier format FASTA sans outgroupe.

3.4. ABGD (Automatic *Barcoding* Gap Discovery).

Nous tout d’abord accédé à la plateforme **ABGD** à partir du lien suivant :

<https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>

The screenshot shows the ABGD web interface. At the top, it says "It can either be an already ALIGNED fasta file or a distance matrix (format from phylip dnadist or MEGA)". Below this, there is a "Choisir un fichier" button and a "BMGECleanog fasta" file name. An "IMPORTANT" note follows: "If you have a distance MEGA distance file please select format: MEGA MEGA CSV". A note below that says: "If you provided a phylip dnadist matrix, it has to be symmetrical and the samples name should not contain space. See Thomas Sauvage GitHub page if you need to reformat your matrix".

There is a section "Or paste your data (FASTA alignment) here" with a red arrow pointing to a large empty text box. Below this, it says "if you select a file AND paste data, only the pasted data will be processed".

Parameters are set as follows: Pmin: 0.001, Pmax: 0.1, Steps: 10, X (relative gap width): 0.5. Under "If you enter a fasta file you can select your distance:", Jukes-Cantor(JC69) is selected, Kimura (K80) is selected, and TS/TV is selected with a value of 2.0. There is also an option for "Simple Distance". A checkbox for "save multipartion Spart file" is unchecked.

The "Nb bins (for distance distribution)" is set to 20. At the bottom, there is a "Go" button with a red arrow pointing to it.

Figure 28 : Plateforme ABGD.

On a ensuite introduit notre fichier format **FASTA** avec outgroupe (*Lachancea kluyveri* numéro d’accession de GenBank HE664110.1), et cliquez sur **GO** pour lancer la délimitation et obtenir des résultats.


```

BMGEcleanog - Bloc-notes
Fichier Edition Format Affichage Aide
>AY428861.1 S boulardii
-----TCCCTCCGTA--TTGATATGCTTAAGTTCTACCTGATTTGGTCAAACCT---GAACATTAACTTCCAATACGCTCAGTATAAAAAAATAGCCGCAGTTGGACCTAAAA---CGACCGTACTTGCATTATA-----CCT-----CAGCACG
>AF085707.1 S chevalieri
GGTAGGAGTACCCTGAACTTAAGCATATCAATAA-----GGAAAGAAACCAACCGATTGCT-----TAGTAACGGCGAGTGAAGC--GGCAAAGCTCAAATTTGAAATCTTGCCCGAGTTGTAATTTGCTTTGGGGCCGTTCC-----T
>KY109359.1 S cf. cerevisiae/paradoxus
GAAGAGAGCGTTTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCTAAGCATATCAATAAGCGGAGGAAAAGAAAC--CAACCGGGATTGCGCTCAAATTTGAAATCTTGCCCGAGTTGTAATTTGCTTTGGGGCCGTTCC-----T
>KY109343.1 S cf. cerevisiae/paradoxus
GAAGAGAGCGTCTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCT--AAGCATATCAATAAGGGAGGAAAAGAAACCAACCGGGATTGCGCTCAAATTTGAAATCTTGCCCGAGTTGTAATTTGCTTTGGGGCCGTTCC-----T
>AF085708.1 S ellipsoideus
GGTAGGAGTACCCTGAACTTAAGCATATCAATAA-----GGAAAGAAACCAACCGATTGCT-----TAGTAACGGCGAGTGAAGC--GGCAAAGCTCAAATTTGAAATCTTGCCCGAGTTGTAATTTGCTTTGGGGCCGTTCC-----T
>KY109332.1 S cf. cerevisiae/paradoxus
GAAGAGAGCGTCTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTGAACCTAAGCATATCAATAAGCGGAGGAAAAGAAAC--CAACCGGGATTGCGCTCAAATTTGAAATCTTGCCCGAGTTGTAATTTGCTTTGGGGCCGTTCC-----T
>KY109458.1 S pastorianus
GAAGAGAGCGTCTAGGGAA--CAATGTTCTTAAAGT--CCTCAAATCAGGTAGGAGTAGCTAAGCATATCAATAAGCGGAGGAAAAGAA-----ACCAACCGGGATTGCGCTCAAATTTGAAATCTTGCCCGAGTTGTAATTTGCTTTGGGGCCGTTCC-----T
>MT163713.1 S sp.cerevisiae
-----
>MZ068118.1 S arboricola
-----CTGGGCAAGAGTATAAGAGATGGAGAGTCCAG--GGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>JQ914752.1 S arboricola
-----AAGAAATTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGTATAAGAGATGGAGAGTCCAGGCGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>EF580917.1 S arboricola
GGAGGATCATTAAAGAAATTTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGTATAAGAGATGGAGAGTCCAGGCGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>KY104943.1 S arboricola
-----ATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGTATAAGAGATGGAGAGTCCAGGCGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>HG764814.1 S sp.
---AGGATCATTAAAGAAATTTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATG--AGCTTTTCTGGGCAAGAAATACAAGAGATGGAGAGTCCAGTGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>KT207224.1 S paradoxus
GGANGATNATTAAAGAAATTTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGACAAGAGATGGAGAGTCCAGTGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>KY104991.1 S cariocanus
---AGTTCATTAAAGAAATTTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAAAGCTTTTCTGGGCAAGAGACAAGAGATGGAGAGTCCAGTGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>MZ185342.1 S paradoxus
-----AAAATTAATAAATTTGA--ATTTTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGACAAGAGATGGAGAGTCCAGTGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>KY105211.1 S paradoxus
GGAGGATCATTAAAGAAATTTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGACAAGAGATGGAGAGTCCAGTGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG
>KY105213.1 S paradoxus
GGAGGATCATTAAAGAAATTTAATAAATTTTGAATAATTTTGTGTTTGGCAAGAGCATGAGAGCTTTTCTGGGCAAGAGACAAGAGATGGAGAGTCCAGTGGGCTGCGCTTAAGTGCCTTTCTAGACTTGAAGTTCTTTCTTGTGCTATTCCAAACAG

```

Figure 29 : Fichier forma FASTA avec outgroup.

Chapitre 5 : Résultats et Discussion

Dans ce chapitre, nous présenterons une analyse complète des résultats obtenus, accompagnée d'une discussion. Nous aborderons le processus de téléchargement des séquences, leur identification, la construction de l'arbre phylogénétique et la délimitation des espèces.

1. Bilan sur les données

Grâce à l'utilisation de la plateforme GenBank, nous avons trouvé un total de 20 espèces appartenant au genre *Saccharomyces*. Ces espèces comprennent *Saccharomyces cf. cerevisiae/paradoxus*, *Saccharomyces cf. bayanus/pastorianus*, *Boulardii*, *Saccharomyces arboricola*, *Saccharomyces sp. Lalvin*, *Saccharomyces sp. Assmannshausen*, *Saccharomyces cariocanus*, *Saccharomyces sp. Zymaflore*, *Anchor*, *Saccharomyces chevalieri*, *Saccharomyces ellipsoideus*, *isolate cerevisiae*, *Saccharomyces eubayanus*, *Saccharomyces pastorianus*, *Saccharomyces bayanus*, *Saccharomyces paradoxus*, *Saccharomyces boulardii*, *Saccharomyces cerevisiae* et *Saccharomyces uvarum*, *Saccharomyces sp1.*, *Saccharomyces sp2*, *Saccharomyces sp3* et *Saccharomyces sp4*. L'ensemble de ces espèces est représenté par un total de 136 séquences, ce qui correspond à 136 enregistrements publiés. Chaque séquence possède un numéro d'accès unique.

1.1. Arbre phylogénétique

Les analyses phylogénétiques des séquences des espèces impliquées dans cette étude ont été reconstruites dans les cadres du maximum de vraisemblance (ML) en utilisant des ensembles de données ITS, avec *Lachancea kluyveri* (HE664110.1) comme outgroup (Annexe2).

1.2. Délimitation des espèces

Bien que les méthodes de délimitation des espèces basées sur le marqueur moléculaire ITS aient généralement abouti à des unités taxonomiques opérationnelles (MOTU) plus ou moins cohérentes avec les morphospécies, il y avait également quelques différences et incohérences entre elles (Figure 31 ; tab 2, 3 et 4 ; annexes 3, 4 et 5).

Si l'on compare les quatre méthodes de délimitation étudiées pour notre cas, nous obtenons différents résultats qui sont les suivants :

mPTP (multi-rate Poisson Tree Processes)

Le résultat présenté dans la figure 31 de la délimitation obtenue par l'analyse mPTP était cohérent, 2 espèces supplémentaires ont été reconnues.

Le premier MOTU qui débute à partir de l'outgroup (*Lachancea kluyveri*) est composé de 3 espèces différentes *S. uvarum*, *S. bayanus*, **TYPE_NR_S. pastorianus**, sachant bien que nous avons une espèce type qui ne ressemble pas aux autres espèces, cette fusion est due soit à une erreur dans leurs noms ou parce qu'elles ont été mal identifiées.

La deuxième et les troisièmes MOTU sont composés de l'espèce **TYPE_NR_S. bayanus** et *S. bayanus* il se peut qu'il y ait un haut degré de similarité.

Le quatrième MOTU présente deux espèces, **TYPE_NR_S. eubayanus** et **TYPE_NR_S. uvarum** alors que les autres espèces sont totalement différents et présentent une fusion énorme ceci est due soit à une erreur dans leur nom soit à une mauvaise identification.

Le cinquième MOTU est composé de différents espèces non identiques cela signifie qu'il y ait une mauvaise identification tandis que le sixième clade présente une nette similarité entre les 4 espèces *S. arboricola*.

Le septième MOTU est composé de quelques espèces totalement différent cela peut être dû à une erreur ou autre.

Le huitième MOTU présente deux espèces type regroupés ensemble cela est probablement due à un haut degré de similarité. **TYPE_NR_S. paradoxus** et **TYPE_NR_S. cariocanus**.

La neuvième jusqu'au dix-neuvième MOTU les espaces sont complètement différentes ceci est aussi due à une erreur ou une mauvaise identification.

En fin pour le douzième MOTU il existe une forte similarité entre les espèces du **TYPE_AM_ S. cerevisiae**.

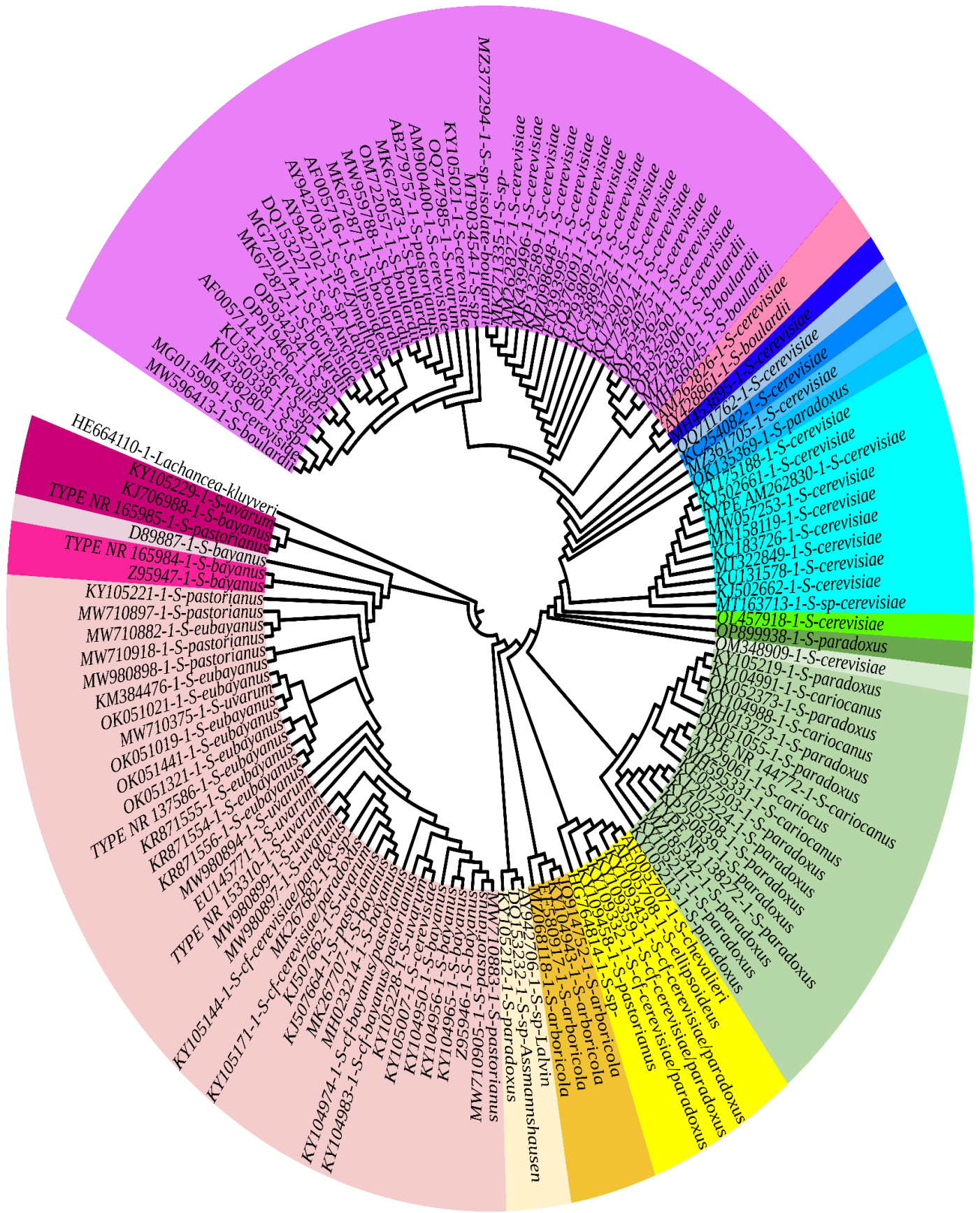


Figure 30: Arbre phylogénétique ML enraciné de genre *Saccharomyces*.

Pour les trois autres méthodes de délimitations, nous avons enregistré :

bPTP (Bayesian Poisson Tree Process), composé de 7 clades (Annexe3 ; tab 2)

Tableau 2 : Résultats de la méthode de délimitation bPTP.

MOTUs	Espèces
MOTU	<i>S.boulandii</i>
MOTU	<p><i>S.cerevisiae/S.cerevisiae/S.sp.cerevisiae/S.ellipsoideus/S.boulandii/S.cerevisiae/S.pastorianus/S.cerevisiae/S.uvarum/S.boulandii/S.boulandii/S.sp.Zymaflore/S.sp/S.sp/S.sp.Lalvin/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.boulandii/S.sp/S.sp.isolate.boulandii/S.sp/S.sp/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.sp.Anchor/S.chevalieri/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.paradoxus/S.cerevisiae/S.cerevisiae/TYPE_S.cerevisiae/S.cerevisiae/S.paradoxus/S.cerevisiae/S.boulandii/S.boulandii/S.sp/S.sp/S.boulandii/S.boulandii/S.cerevisiae/S.cerevisiae/S.cerevisiae/S.boulandii/S.paradoxus/S.paradoxus/TYPE_S.paradoxus/S.paradoxus/S.paradoxus/S.paradoxus/S.paradoxus/S.paradoxus/S.paradoxus/S.cariocanus/S.cariocanus/S.cariocanus/S.cariocanus/S.cariocanus/S.cariocanus/TYPE_S.cariocanus/S.cariocanus/S.sp/S.arboricola/S.arboricola/S.arboricola/S.arboricola/S.sp.Assmannshausen/S.paradoxus/S.sp.Lalvin/S.eubayanus/TYPE_S.pastorianus/S.pastorianus/S.bayanus/TYPE_S.bayanus/S.uvarum/S.bayanus/S.pastorianus/S.bayanus/Lachancea-kluuyveri/S.pastorianus/S.pastorianus/TYPE_S.eubayanus/S.eubayanus/S.eubayanus/S.eubayanus/S.eubayanus/S.eubayanus/S.eubayanus/S.eubayanus/S.bayanus/S.uvarum/S.uvarum/S.bayanus/S.bayanus/S.bayanus/S.sp/S.cf.bayanus/pastorianus/S.cf.cerevisiae/paradoxus/TYPE_S.uvarum/S.bayanus/S.uvarum/S.cf.cerevisiae/paradoxus/S.uvarum/S.cf.bayanus/pastorianus/S.cerevisiae/S.uvarum/S.pastorianus/S.bayanus/S.uvarum/S.eubayanus/S.uvarum/S.pastorianus/S.pastorianus</i></p>
MOTU	<i>S.ellipsoideus/S.chevalieri</i>
MOTU	<i>S.pastorianus</i>
MOTU	<i>S.cf.cerevisiae/paradoxus</i>
MOTU	<i>S.cf.cerevisiae/paradoxus</i>
MOTU	<i>S.cf.cerevisiae/paradoxus</i>

Les résultats de l'analyse **ASAP (Assemble Species by Automatic Partitioning)** ont permis d'obtenir une partition avec le meilleur score ASAP. Plus le score ASAP est faible (dans ce cas, il est égal à 3), meilleure est la partition. La meilleure partition ASAP comprend 3 MOTU (Tab 3 ; annexe 4).

Tableau 3 : Résultats de la méthode de délimitation ASAP.

MOTUs	Espèces
MOTU 1	<i>S. boulardii</i>
MOTU 2	<i>S. chevalieri</i> / <i>S. ellipsoideus</i> / <i>S.cf. cerevisiae/paradoxus</i> / <i>S.cf. cerevisiae/paradoxus</i> / <i>S.cf. cerevisiae/paradoxus</i> / <i>S. pastorianus</i>
MOTU 3	<i>S.sp. cerevisiae</i> / <i>S. arboricola</i> / <i>S. arboricola</i> / <i>S. arboricola</i> / <i>S. arboricola</i> / <i>S. paradoxus</i> / TYPE_S paradoxus / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. cariocanus</i> / <i>S. cariocanus</i> / <i>S. paradoxus</i> / TYPE_S. cariocanus / <i>S. paradoxus</i> / <i>S. cariocanus</i> / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. paradoxus</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / TYPE_S. cerevisiae / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. paradoxus</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. chevalieri</i> / <i>S. sp.</i> / <i>S.sp.</i> / <i>S. boulardii</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. boulardii</i> / <i>S. boulardii</i> / <i>S. ellipsoideus</i> / <i>S. cerevisiae</i> / <i>S. sp.</i> / <i>S. sp.</i> / <i>S.sp. isolate boulardii</i> / <i>S.sp.</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S.sp.</i> / <i>Lalvin</i> / <i>S.sp.</i> / <i>S.sp.</i> / <i>S.sp. Zymaflore</i> / <i>S. boulardii</i> / <i>S. boulardii</i> / <i>S. boulardii</i> / <i>S. pastorianus</i> / <i>S. cerevisiae</i> / <i>S. boulardii</i> / <i>S. cerevisiae</i> / <i>S. boulardii</i> / <i>S. boulardii</i> / <i>S. uvarum</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. cerevisiae</i> / <i>S. paradoxus</i> / <i>S.sp. Anchor</i> / <i>S.sp.</i> / <i>S.sp. Lalvin</i> / <i>S. paradoxus</i> / <i>S.sp. Assmannshausen</i> / <i>S. bayanus</i> / <i>S. pastorianus</i> / <i>S. bayanus</i> / <i>S. uvarum</i> / <i>S. bayanus</i> / TYPE_S. pastorianus / TYPE_S bayanus / <i>S. eubayanus</i> / <i>S. eubayanus</i> / <i>S. eubayanus</i> / <i>S. eubayanus</i> / <i>S. eubayanus</i> / <i>S. eubayanus</i> / TYPE_S. eubayanus / <i>S. eubayanus</i> / <i>S. eubayanus</i> / <i>S. bayanus</i> / <i>S. pastorianus</i> / <i>S. uvarum</i> / <i>S. cerevisiae</i> / <i>S.cf. bayanus/pastorianus</i> / <i>S. uvarum</i> / <i>S.cf. cerevisiae/paradoxus</i> / <i>S. uvarum</i> / <i>S. bayanus</i> / TYPE_S. uvarum / <i>S.cf. bayanus/pastorianus</i> / <i>S. bayanus</i> /

	<i>Zymaflore /S. boulardii /S. boulardii /S. boulardii /S. uvarum /S. cerevisiae / S. boulardii /S. pastorianus / S. cerevisiae / S. boulardii /S. ellipsoideus</i>
MOTU 12	<i>S. sp. Lalvin /S. paradoxus /S. sp. Assmannshausen</i>
MOTU 13	<i>S. bayanus</i>
MOTU 14	<i>S. pastorianus /S. uvarum /S. bayanus /S. eubayanus / S. pastorianus / S. pastorianus /S. uvarum /S. uvarum /TYPE_ S. bayanus /S. bayanus / S. pastorianus /TYPE_ S. pastorianus /S. bayanus / S. pastorianus /S. uvarum / S. cerevisiae / S. cf. bayanus/pastorianus /S. uvarum /S.cf. cerevisiae/paradoxus /S. uvarum /S. bayanus /TYPE_ S. uvarum /S.cf. cerevisiae/paradoxus / S. cf. bayanus/pastorianus /S. sp. /S bayanus / S. bayanus /S. bayanus /S. pastorianus /S. uvarum /S. uvarum / S. bayanus / S. eubayanus /S. pastorianus / S. eubayanus / S. eubayanus /S. eubayanus / S. eubayanus /S. eubayanus / S. eubayanus /S. eubayanus / TYPE_ S. eubayanus</i>

Le barcoding permet généralement l'identification de plusieurs espèces mais pour notre cas il n'a pas pu les identifier.

Le nombre de groupes enregistré est très réduit, les résultats ne seront donc pas fiables par rapport à la méthode **mPTP (multi-rate Poisson Tree Processes)**, dont les résultats obtenus sont plus proches de la réalité.

Nos résultats ont révélé une relation phylogénétique ambiguë entre les différentes méthodes de délimitations du genre *Saccharomyces*. Avec un regroupement plus ou moins fiable des espèces (*S. pastorianus*, *S. bayanus*, *S. eubayanus*, *S. uvarum*, *S. paradoxus*, *S. cariocanus* et *S-cerevisiae*) pour le cas de la méthode **mPTP (multi-rate Poisson Tree Processes)**.

Conclusion

Conclusion

Notre travail basé sur l'utilisation d'outils bio-informatique combinée à des méthodes de délimitation des espèces nous a permis de réaliser un arbre phyllogénétique du genre *Saccharomyces* et ce pour mieux comprendre la diversité et les relations évolutives du champignon.

Selon les données recueillis de la base de données GenBank, le champignon du genre *Saccharomyces* contient 20 espèces représentés par 136 séquences ces dernières, ont été nettoyés (supprimées) manuellement puis introduites directement dans la plateforme NG_phylogeny afin de construire l'arbre phylogénétique en se basant sur les deux méthodes Maximum likelihood (ML) et Mr Bayes.

Quant à la délimitation des espèces nous avons utilisés quatre méthodes, mPTP, bPTP, ASAP et ABGD. Les résultats obtenus sont complètement différents les uns des autres seule la méthode de mPTP est plus ou moins fiable voir une meilleure distribution et classification des séquences dans l'arbre phylogénétique.

Le Barcoding peut également être utile pour décrire les groupes d'espèces du genre *Saccharomyces* avec une taxonomie mal renseignée, ou présentant de fortes difficultés pour l'identification.

Malgré les lacunes obtenues à partir des bases de données publiques GenBank, séquences, répétées ou doublant, séquences ambiguës, et spécimen mal identifié tous ces informations peuvent perturber les résultats de l'étude phylogénétique. Sachant bien que le rôle le plus important de ces bases de données est de préserver la biodiversité.

Enfin, l'utilisation d'outils bio-informatiques, combinée à des méthodes de délimitation des espèces, a permis de mieux comprendre la diversité et les relations évolutives des champignons *Saccharomyces*. Cette recherche pourrait conduire au développement de stratégies plus efficaces pour l'identification et le contrôle des pathogènes fongiques et d'autres champignons bénéfiques. D'autres recherches sont nécessaires pour explorer d'autres techniques moléculaires et outils bio-informatiques dans la délimitation des espèces.

Bibliographie

Références bibliographiques

1. **Barnett, J. A. (1992).** The taxonomy of the genus *Saccharomyces* Meyen ex Reess: a short review for non-taxonomists. *Yeast*, 8(1), 1-23.
2. **Barnett, P. G. (2003).** Determination of VA health care costs. *Medical Care Research and Review*, 60(3_suppl), 124S-141S.
3. **Baum, D. 2008.** Reading a phylogenetic tree: the meaning of monophyletic groups. *Nature Education*, 1(1), 190.
4. **Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018).** GenBank. *Nucleic acids research*, 46(Database issue), D41.
5. **Börlin, M. (2015).** Diversité et structure de population des levures *Saccharomyces cerevisiae* à l'échelle du vignoble bordelais : Impact de différents facteurs sur la diversité (Doctoral dissertation, Université de Bordeaux).
6. **Cardona, G., Rosselló, F., & Valiente, G. (2008).** Extended Newick: it is time for a standard representation of phylogenetic networks. *BMC bioinformatics*, 9(1), 1-8.
7. **Celton, M. (2011).** Etude de la réponse de *Saccharomyces cerevisiae* à une perturbation NADPH par une approche de biologie des systèmes (Doctoral dissertation, Montpellier, SupAgro).
8. **Criscuolo, A., & Gribaldo, S. (2010).** BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC evolutionary biology*, 10, 1-21.
9. **Coissac, E., Riaz, T., & Puillandre, N. (2012).** Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular ecology*, 21(8), 1834-1847.
10. **Deriham T. E. (2017).** Utilisation des arbres phylogénétiques dans l'alignement de séquence (Doctoral dissertation, FACULTE : Mathématiques et Informatique-- université MOHAMED BOUDIAF-M'SILA).
11. **Edgar, R. C. (2004).** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792-1797.
12. **Felsenstein, J., & Felsenstein, J. (2004).** *Inferring phylogenies* (Vol. 2, p. 664). Sunderland, MA : Sinauer associates.
13. **Géraldine, K. U. H. N., & GEBREYOHANNES, T. K. (2014).** Effet de la levure probiotique *Saccharomyces cerevisiae* 47 sur la santé, le microbiote et les performances zootechniques de porcelets sevrés.

14. **Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., ... & Oliver, S. G. (1996).** Life with 6000 genes. *Science*, 274(5287), 546-567.
15. **Guillaume, L. (2007).** Qu'est-ce qu'un arbre phylogénétique ?
<http://acces.ens-lyon.fr/biotic/evolut/parente/html/arbphyl.htm> , (consulté le 25 Avril 2023).
16. **Guindon, S., & Gascuel, O. (2003).** A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, 52(5), 696-704.
17. **Haeckel, E. H. P. A. (1860).** Fernere Abbildungen und Diagnosen neuer Gattungen und Arten von lebenden Radiolarien des Mittelmeeres. *Monatsberichte der Koniglichen Preuss. Akademie der Wissenschaften zu Berlin*, 835-845.
18. **Huelsenbeck, J. P., & Ronquist, F. (2001).** MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755.
19. **Hussenet, C. (2017).** Instrumentation, modélisation et automatisation de fermenteurs levuriers à destination oenologique (Doctoral dissertation, Université Paris-Saclay).
20. **Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017).** Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, 33(11), 1630-1638.
21. **Kurtzman, C. P., Fell, J. W., & Boekhout, T. (Eds.). (2011).** The yeasts: a taxonomic study. Elsevier.
22. **Larpent, J. P., & Larpent-Gourgaud, M. (1985).** *Eléments de microbiologie*. Hermann, France, 464p.
23. **Luchetta, P., Maurel, M. C., Higué, D., & Vervoort, M. (2005).** *Évolution moléculaire*.
24. **Nguyen, H. D. T., & Seifert, K. A. (2008).** Description and DNA barcoding of three new species of *Leohumicola* from South Africa and the United States. *Persoonia-Molecular Phylogeny and Evolution of Fungi*, 21(1), 57-69.
25. **Nilsson, R. H., Hyde, K. D., Pawłowska, J., Ryberg, M., Tedersoo, L., Aas, A. B., ... & Abarenkov, K. (2014).** Improving ITS sequence data for identification of plant pathogenic fungi. *Fungal Diversity*, 67, 11-19.
26. **Patwardhan, A., Ray, S., Roy, A. (2014).** Molecular markers in phylogenetic studies-a review. *Journal of Phylogenetics & Evolutionary Biology*, 2014.

27. Phylogénie, 2007.

<https://www.aquaportail.com/definition-1802-phylogenie.html>, (consulté le 22 Avril 2023).

- 28. Puillandre, N., Brouillet, S., & Achaz, G. (2021).** ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources*, 21(2), 609-620.
- 29. Rawlings, D. E. (2002).** Heavy metal mining using microbes. *Annual Reviews in Microbiology*, 56(1), 65-91.
- 30. Schoch, C. L., Robbertse, B., Robert, V., Vu, D., Cardinali, G., Irinyi, L., ... & Federhen, S. (2014).** Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. *Database*, 2014, bau061.
- 31. Sumida, M., Kato, Y., & Kurabayashi, A. (2004).** Sequencing and analysis of the internal transcribed spacers (ITSs) and coding regions in the EcoR I fragment of the ribosomal DNA of the Japanese pond frog *Rana nigromaculata*. *Genes & genetic systems*, 79(2), 105-118.
- 32. Swofford, D. L. (1996).** Phylogenetic inference. *Molecular systematics*.
- 33. Visser, W., Scheffers, W. A., Batenburg-van der Vegte, W. H., & van Dijken, J. P. (1990).** Oxygen requirements of yeasts. *Applied and environmental microbiology*, 56(12), 3785-3792.
- 34. Xu, J., Feng, Y., Xu, Z., Bailey, B. A., Liu, Y., & Zeng, Q. (2014).** DNA barcoding reveals the diversity of Chinese medicinal Fungi including medicinal Ascomycota (Phylum) in China. *PloS One*, 9(4), e92077.
- 35. Zaïd, S. M., Mustapha, R., El Mire, S., Mohammed, L., & Sarra, A. (2020).** *Saccharomyces cerevisiae*, une levure de plus en plus impliquée dans les infections urinaires : à propos de 3 cas. *The Pan African Medical Journal*, 35.

Sites web

<https://species.h-its.org/ptp/>

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

<https://ngphylogeny.fr/>

<https://www.ncbi.nlm.nih.gov/genbank/>

<https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>

<https://bioinfo.mnhn.fr/abi/public/asap/asapweb.html>

<http://acces.ens-lyon.fr/biotic/evolut/parente/html/arbphyl.htm> (consulté le 25 Avril 2023).

<https://www.aquaportail.com/definition-1802-phylogenie.html> (consulté le 22 Avril 2023).

Annexes

Annexe 1 : Exemple d'une séquence nucléotidique dans la base de données GenBank

(*S. cerevisiae* MW057253.1).

National Library of Medicine
National Center for Biotechnology Information
Log in

Nucleotide Search

GenBank
Send to: ▾
Change region shown ▾

Saccharomyces cerevisiae strain 2019_SD56 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

GenBank: MW057253.1
[FASTA](#) [Graphics](#)

[Go to: ☺](#)

LOCUS	MW057253	843 bp	DNA	linear	PLN 08-OCT-2020
DEFINITION	Saccharomyces cerevisiae strain 2019_SD56 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence.				
ACCESSION	MW057253				
VERSION	MW057253.1				
KEYWORDS	-				
SOURCE	Saccharomyces cerevisiae (baker's yeast)				
ORGANISM	Saccharomyces cerevisiae Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
REFERENCE	1 (bases 1 to 843)				

REFERENCE	2 (bases 1 to 843)				
AUTHORS	Do Ngoc Anh,D., Le Tran Anh,D., Nguyen Van Tuan,B., Nguyen T Nhu Quynh,B., Le Quoc Tuan,M. and Nguyen Khac Luc,D.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-OCT-2020) Parasitology, Vietnam military medical university, Phung Hung road, Hadong district, Hanoi, Hanoi 100000, Viet Nam				
COMMENT	##Assembly-Data-START## Sequencing Technology :: Sanger dideoxy sequencing ##Assembly-Data-END##				
FEATURES	Location/Qualifiers				
source	1..843 /organism="Saccharomyces cerevisiae" /mol_type="genomic DNA" /strain="2019_SD56" /isolation_source="vaginal discharge" /host="Homo sapiens" /db_xref="taxon:4932" /country="Viet Nam: Hanoi"				
misc_RNA	<1..843 /note="contains small subunit ribosomal RNA, internal transcribed spacer 1, 5.8S ribosomal RNA, internal transcribed spacer 2, and large subunit ribosomal RNA"				
ORIGIN	<pre> 1 tccgtagggt aacctcggg aggatcatta aagaaattta ataatttga aaatggatt 61 tttgttttg gcaagagcat gagagctttt actgggcaag aagacaagag atggagagtc 121 cagccggggc tgcgcttaag tgcgcggctt tgctaggcct gtaagtttct ttctgctat 181 tccaaacggt gagagatttc tgtccttttg ttataggaca attaaaaccg ttccaataca 241 acacactgtg gagttttcat tatccttgca actttttcct tgggcattcg agcaatcggg 301 gccccagagg acaaaacaca aacaatttta ttatttcatt aaattttgt caaaaacaag 361 aattttcgta actggaattt ttaaaattt aaaaactttc aacaacggat ccttggttc 421 tgcacatgat gaagaacgca gcgaaatgcg atacgtaatg tgaattgcag aattcctgta 481 atcatgcaat cttgaaacgc acattgcgcc ccttggtatt ccaggggcca tgcctgttg 541 agcgtcattt ccttctcaaa cattctgttt gtagtgagt gatactttt ggagttaact 601 tgaattgctt ggccttttca ttggatgttt ttttttcca aagagaggtt tcttcgctg 661 cttgaggtat aatgcaagta cggctgtttt aggttttacc aactcggct aactttttt 721 atactgagcg tatggaaacg ttatcgataa gaagagagcg tctagcgaa caatgttct 781 aagtttgac ctcaaatcag gtaggagtac ccgctgaact taagcatatc aataagcggg 841 gga </pre>				

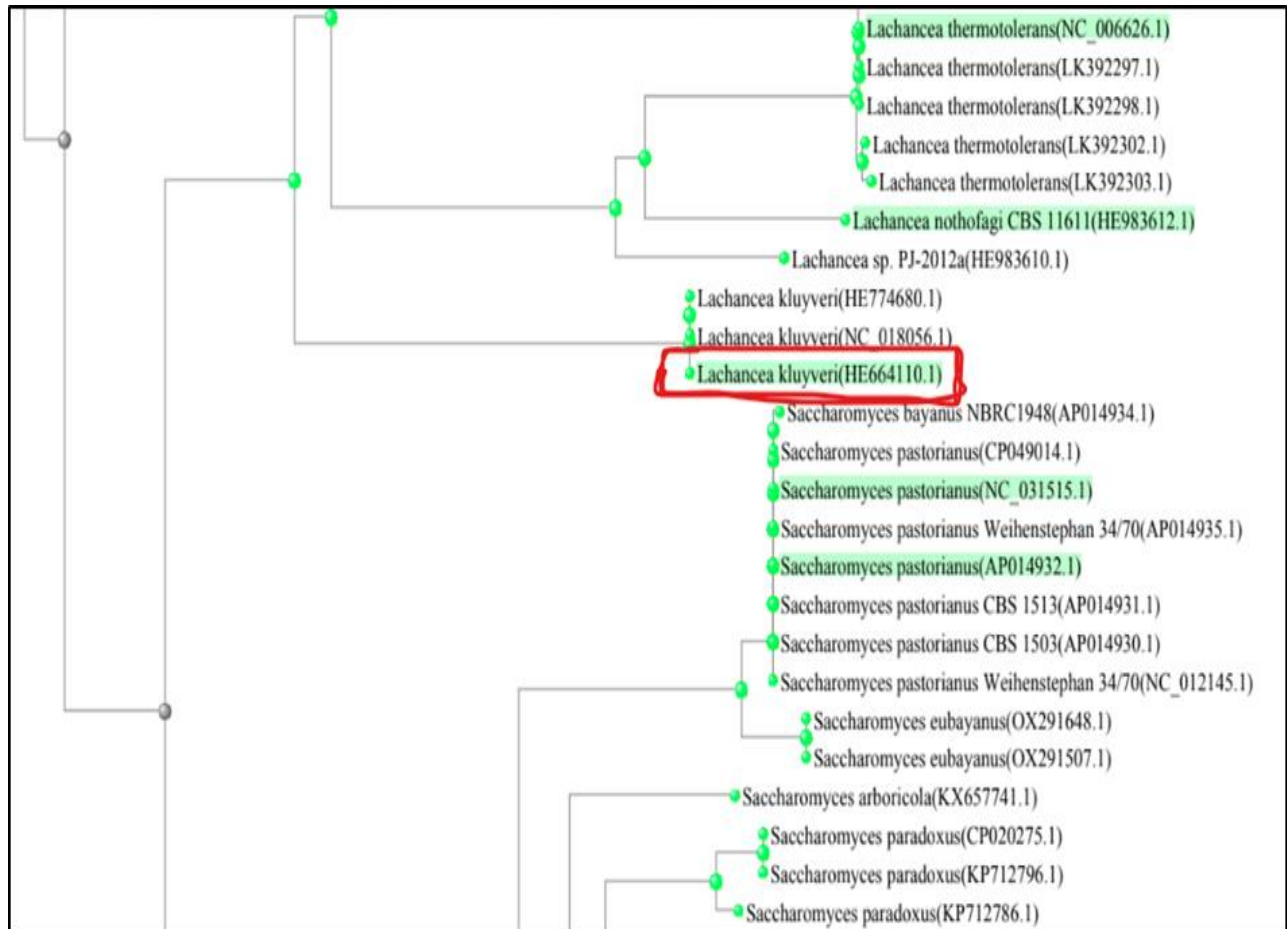
Saccharomyces (1374620) Nucleotide

saccharomyces AND (is_nucore[filter]) (1374620) Nucleotide

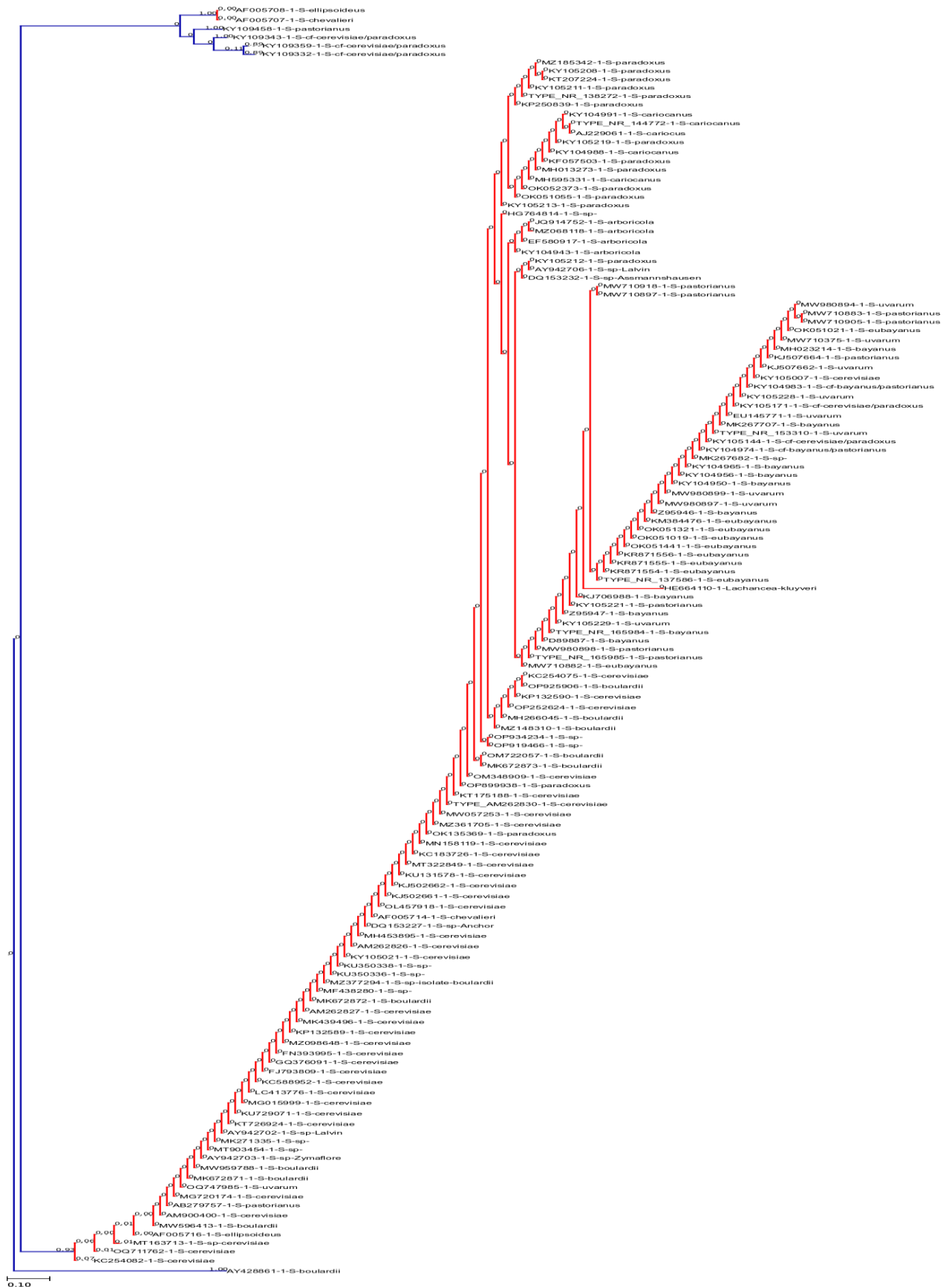
Saccharomyces paradoxus strain LL12_025 internal transcribed spacer 1, partial Nucleotide

[See more...](#)

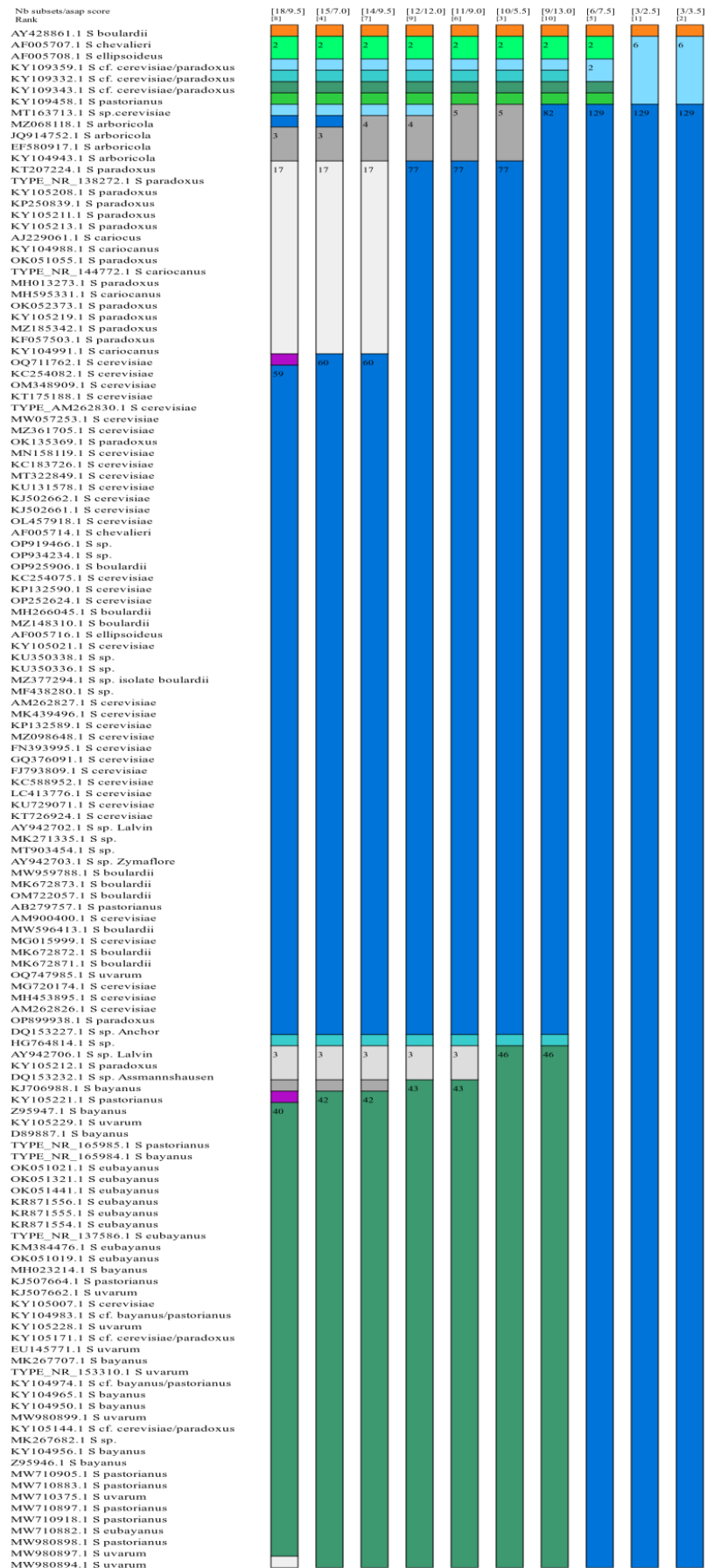
Annexe 2 : L'outgroup de l'arbre phylogénétique.



Annexe 3 : Arbre phylogénétique rectangulaire de la méthode de délimitation bPTP.



Annexe 4 : Résultats de la méthode de délimitation ASAP format graphe.



Annexe 5 : Résultats de la méthode de délimitation ABGD.**Group[1] n: 1 ;id:**AY428861.1 *S boulardii***Group[2] n: 2 ;id:**AF005707.1 *S chevalieri*AF005708.1 *S ellipsoideus***Group[3] n: 1 ;id:**KY109359.1 *S cf. cerevisiae/paradoxus***Group[4] n: 1 ;id:**KY109343.1 *S cf. cerevisiae/paradoxus***Group[5] n: 1 ;id:**KY109332.1 *S cf. cerevisiae/paradoxus***Group[6] n: 1 ;id:**KY109458.1 *S pastorianus***Group[7] n: 1 ;id:**MT163713.1 *S sp. cerevisiae***Group[8] n: 4 ;id:**MZ068118.1 *S arboricola*JQ914752.1 *S arboricola*EF580917.1 *S arboricola*KY104943.1 *S arboricola***Group[9] n: 1 ;id:**HG764814.1 *S sp.***Group[10] n: 17 ;id:**

KT207224.1 *S paradoxus* / KY104991.1 *S cariocanus* /MZ185342.1 *S paradoxus*
 /KY105211.1 *S paradoxus* /KY105213.1 *S paradoxus* /TYPE_NR_138272.1 *S paradoxus*/
 KY105208.1 *S paradoxus* /KP250839.1 *S paradoxus* /AJ229061.1 *S cariocanus*
 /KY105219.1 *S paradoxus* /KY104988.1 *S cariocanus* /KF057503.1 *S paradoxus*
 /MH013273.1 *S paradoxus* /MH595331.1 *S cariocanus* /OK052373.1 *S paradoxus*
 /OK051055.1 *S paradoxus* /TYPE_NR_144772.1 *S cariocanus*

Group[11] n: 60 ;id:

OQ711762.1 *S cerevisiae* /KC254082.1 *S cerevisiae* /OM348909.1 *S cerevisiae* /OP899938.1
S paradoxus /KT175188.1 *S cerevisiae* /OP919466.1 *S sp.* /OP934234.1 *S sp.*
 /TYPE_AM262830.1 *S cerevisiae* /MW057253.1 *S cerevisiae* /MZ361705.1 *S cerevisiae*
 /OK135369.1 *S paradoxus* /MN158119.1 *S cerevisiae* /KC183726.1 *S cerevisiae*
 /MT322849.1 *S cerevisiae* /KU131578.1 *S cerevisiae* /KJ502662.1 *S cerevisiae* /KJ502661.1
S cerevisiae /OL457918.1 *S cerevisiae* /AF005714.1 *S chevalieri* /DQ153227.1 *S sp.* Anchor
 /MH453895.1 *S cerevisiae* /AM262826.1 *S cerevisiae* /OP925906.1 *S bouldarii* /KC254075.1
S cerevisiae /KP132590.1 *S cerevisiae* /OP252624.1 *S cerevisiae* /MH266045.1 *S bouldarii*
 /MZ148310.1 *S bouldarii* /KY105021.1 *S cerevisiae* /KU350338.1 *S sp.* /KU350336.1 *S sp.*
 /MZ377294.1 *S sp. isolate bouldarii* /MF438280.1 *S sp.* /MK672872.1 *S bouldarii*
 /AM262827.1 *S cerevisiae* /MK439496.1 *S cerevisiae* /KP132589.1 *S cerevisiae* /
 MZ098648.1 *S cerevisiae* /FN393995.1 *S cerevisiae* /GQ376091.1 *S cerevisiae* /FJ793809.1
S cerevisiae /KC588952.1 *S cerevisiae* /LC413776.1 *S cerevisiae* /MG015999.1 *S cerevisiae*
 /KU729071.1 *S cerevisiae* /KT726924.1 *S cerevisiae* /AY942702.1 *S sp. Lalvin* /MK271335.1
S sp. /MT903454.1 *S sp.* /AY942703.1 *S sp. Zymaflore* /MW959788.1 *S bouldarii* /
 MK672873.1 *S bouldarii* /MK672871.1 *S bouldarii* /OQ747985.1 *S uvarum* /MG720174.1 *S*
cerevisiae /OM722057.1 *S bouldarii* /AB279757.1 *S pastorianus* /AM900400.1 *S cerevisiae*
 /MW596413.1 *S bouldarii* /AF005716.1 *S ellipsoideus*

Group[12] n: 3 ;id:

AY942706.1 *S sp. Lalvin* /KY105212.1 *S paradoxus* /DQ153232.1 *S sp. Assmannshausen*

Group[13] n: 1 ;id:

KJ706988.1 *S bayanus*

Group[14] n: 42 ;id:

KY105221.1 *S pastorianus* /MW980894.1 *S uvarum* /Z95947.1 *S bayanus* /OK051021.1 *S*
eubayanus /MW710897.1 *S pastorianus* /MW710918.1 *S pastorianus* /MW710375.1 *S uvarum*
 /KY105229.1 *S uvarum* /TYPE_NR_165984.1 *S bayanus* /D89887.1 *S bayanus* /
 MW980898.1 *S pastorianus* /TYPE_NR_165985.1 *S pastorianus* /MH023214.1 *S bayanus* /
 KJ507664.1 *S pastorianus* /KJ507662.1 *S uvarum* /KY105007.1 *S cerevisiae* /KY104983.1 *S*
cf. bayanus/pastorianus /KY105228.1 *S uvarum* /KY105171.1 *S cf. cerevisiae/paradoxus* /

EU145771.1 *S uvarum*/ MK267707.1 *S bayanus*/ TYPE_NR_153310.1 *S uvarum*/
KY105144.1 *S cf. cerevisiae/paradoxus*/ KY104974.1 *S cf. bayanus/pastorianus*
/MK267682.1 *S sp.* /KY104965.1 *S bayanus* /KY104956.1 *S bayanus* /KY104950.1 *S*
bayanus/ MW710905.1 *S pastorianus* /MW980899.1 *S uvarum* /MW980897.1 *S uvarum*/
Z95946.1 *S bayanus*/ MW710882.1 *S eubayanus* /MW710883.1 *S pastorianus* /KM384476.1
S eubayanus /OK051321.1 *S eubayanus* /OK051019.1 *S eubayanus* /OK051441.1 *S*
eubayanus /KR871556.1 *S eubayanus*/ KR871555.1 *S eubayanus* /KR871554.1 *S eubayanus*/
TYPE_NR_137586.1 *S eubayanus*

المخلص

كان الهدف من هذه الدراسة هو استخدام أدوات المعلوماتية الحيوية لوصف جزيئي وتحديد أنواع الفطريات *Saccharomyces*. يعد فهم السلالة وتنوع الفطريات أمرًا بالغ الأهمية لفهم خصائصها البيولوجية، وقد وجد أن أدوات المعلوماتية الحيوية مفيدة بشكل خاص في تحليل كميات كبيرة من بيانات التسلسل في أبحاث الفطر. تسلسل تم تنزيله من قاعدة بيانات GenBank العامة، وتنظيفه يدويًا، ثم أدخلتها في منصة NG_phylogeny لبناء شجرة وراثية باستخدام طرق الاحتمال القصوى (ML) والسيد Bayes. أكدت النتائج التي تم الحصول عليها توزيع التسلسلات في تسعة عشر MOUTUs بفضل طريقة تعيين mPTP. وهكذا، توضح هذه الدراسة فائدة أدوات المعلوماتية الحيوية للتوصيف الجزيئي وتحديد أنواع فطر *Saccharomyces*، مما يساهم في فهم أفضل لنباتهم وتنوعهم البيولوجي.

الكلمات المفتاحية: *Saccharomyces*، GenBank، NG_phylogeny، تحديد الأنواع، Phylogeny.

Résumé

L'objectif de cette étude était d'utiliser des outils bio-informatiques pour caractériser moléculairement et délimiter les espèces de champignons *Saccharomyces*. La compréhension de la phylogénie et de la diversité des champignons est cruciale pour comprendre leurs propriétés biologiques, et les outils bio-informatiques se sont révélés particulièrement utiles pour analyser de grandes quantités de données de séquence dans la recherche sur les champignons. Les séquences ont été téléchargées à partir de la base de données publique GenBank, les ont nettoyées manuellement, puis les ont introduites dans la plateforme NG_phylogeny pour construire un arbre phylogénétique en utilisant les méthodes du maximum de vraisemblance (ML) et Mr Bayes. Les résultats obtenus ont confirmé la répartition des séquences en dix-neuf MOUTUs grâce à la méthode de délimitation mPTP. Ainsi, cette étude démontre l'utilité des outils bio-informatiques pour la caractérisation moléculaire et la délimitation des espèces de champignons *Saccharomyces*, ce qui contribue à une meilleure compréhension de leur phylogénie et de leur diversité biologique.

Mots clés : *Saccharomyces*, GenBank, NG_phylogeny, Délimitation des espèces, Phylogénie.

Abstract

The objective of this study was to use bioinformatics tools to molecularly characterize and delimit *Saccharomyces* fungal species. Understanding the phylogeny and diversity of fungi is crucial to understanding their biological properties, and bioinformatics tools have proved particularly useful for analyzing large amounts of sequence data in fungal research. The sequences were downloaded from the public GenBank database were manually cleaned, then fed into the NG_phylogeny platform to construct a phylogenetic tree using maximum likelihood (ML) and Mr. Bayes methods. The results confirmed the division of sequences into nineteen MOUTUs using the mPTP delimitation method. This study demonstrates the usefulness of bioinformatics tools for the molecular characterization and delimitation of *Saccharomyces* mushroom species, contributing to a better understanding of their phylogeny and biological diversity.

Keywords: *Saccharomyces*, GenBank, NG_phylogeny, Species delimitation, Phylogeny.