

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur de la Recherche Scientifique
Université Constantine 1 des Frères Mentouri
Faculté des Sciences de la Nature et de la Vie
Département de Microbiologie

Matière : bioinformatique 1

TD N°3 : Interprétation des résultats de recherche dans Genbank (les différents types de fichiers)

Présenté par : KERKOUCHÉ Hibat errahmane (G3)

CHOUGUI Rihem (G1)

DORBANI Sara (G1)

CHAALA Amira Douaa (G3)

Année universitaire : 2025/2026

1. Objectif de TD

L'objectif de ce TD est de maîtriser la méthodologie de la recherche d'une séquence génétique stockée dans la base de données Genbank, par deux méthodes principales et d'apprendre l'interprétation des résultats de la séquence biologique tel que :la signification des résultats répétés et des pages Fasta, Genbank, Graphiques, et d'acquérir les approches techniques indispensables pour importer la séquence afin d'assurer leur exploitation dans des analyses subséquentes.

2. Recherche une séquence stockée dans Genbank

La recherche d'une séquence génétique stockée dans la base de données GenBank peut se faire de deux manières principales :

Premièrement, la recherche par le nom de l'espèce : accéder au NCBI et dans le menu déroulant sélectionner « Nucleotide » puis saisir le nom de l'espèce scientifique complète par exemple : *Tsukamurella paurometabola* 16S r-RNA ou *Tsukamurella sp.* 1534 dans la barre de recherche. Cette approche permet d'obtenir un grand nombre de résultats sous forme de liste de séquences déposé pour l'espèce donnée (Figure 1) et (Figure 2).

Figure 1 : Capture d'écran de l'interface Genbank lors de la recherche par nom d'espèce

Tsukamurella paurometabola 16S rRNA.

Figure 2 : Capture d'écran de l'interface Genbank lors de la recherche par nom d'espèce

Tsukamurella sp. 1534.

Deuxièmement, la recherche par le numéro d'accession : dans le menu déroulant de site NCBI sélectionner « Nucleotide », dans la barre de recherche saisir le numéro d'accession exemple : X53207 ou JQ407772, puis cliqué sur search. On accède immédiatement à la fiche exacte contenant la séquence recherchée. Cette méthode est la plus précise et la plus rapide lorsqu'on connaît déjà l'accession, car elle mène directement à la page Genbank, et à toutes les informations annotée (origine, fonctions, références, etc.) (Figure 3) et (Figure 4).

The screenshot shows a browser window with the URL [ncbi.nlm.nih.gov/nucleotide/X53207](https://www.ncbi.nlm.nih.gov/nucleotide/X53207). The page title is "Nucleotide" and the search term is "X53207". The main content displays the sequence information for "Tsukamurella paurometabolum (Rhodococcus aurantiacus) partial 16S rRNA". The sequence details include Locus X53207, Definition Tsukamurella paurometabolum (Rhodococcus aurantiacus) partial 16S rRNA, Accession X53207, Version X53207.1, Keywords 16S ribosomal RNA; ribosomal RNA, Source Tsukamurellaceae; Tsukamurella paurometabolism, Organism Bacteria; Actinomycetales; Actinomycetes; Mycobacteriales; Tsukamurellaceae; Tsukamurella, Reference 1 (bases 1 to 1342), Authors Collins,M.J., Smith,J., Dorsch,M. and Stackebrandt,E., Title Tsukamurellaceae nov., harboring Corynebacterium paurometabolum and Rhodococcus aurantiacus, Journal Int. J. Syst. Bacteriol. 38, 385-391 (1988), Reference 2 (bases 1 to 1342), Authors Stackebrandt,E., Title Direct Submission, Journal Submitted (09-MAY-1990) Stackebrandt E, Comment Rhodococcus aurantiacus is reclassified as Tsukamurella. On the right side, there are links for "Change region shown", "Customize view", "Analyze this sequence" (Run BLAST, Pick Primers, Highlight Sequence Features, Find in this Sequence), "Related information" (Taxonomy), and "LinkOut to external resources" (Tsukamurella paurometabolism BD-ID 17098, [BacDive]). The top right corner has a "Search" button and a "Help" link.

Figure 3 : Capture d'écran de l'interface Genbank lors de la recherche par le numéro d'accésion (X53207).

→ Tsukamurella sp. 1534 16S ribosomal RNA gene, partial sequence

ncbi.nlm.nih.gov/nuccore/JQ407772

Nucleotide Nucleotide JQ407772 Advanced Search Help

GenBank ▾ Send to: ▾

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Go to: ▾

LOCUS JQ407772 1486 bp DNA linear BCT 07-MAR-2012

DEFINITION Tsukamurella sp. 1534 16S ribosomal RNA gene, partial sequence.

ACCESSION JQ407772

VERSION JQ407772.1

KEYWORDS

SOURCE Tsukamurella sp. 1534

ORGANISM Tsukamurella sp. 1534

Bacteria; Bacilli; Actinomycetota; Actinomycetes;

Mycobacteroles; Tsukamurellaceae; Tsukamurella.

REFERENCE 1 (bases 1 to 1486)

AUTHORS Oulmi,L. and Roux,V.

TITLE Tsukamurella algeriensis sp. nov. a novel actinomycete isolated from the sputum

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1486)

AUTHORS Oulmi,L. and Roux,V.

TITLE Direct Submission

JOURNAL Submitted (06-JAN-2012) Department of Biochemistry and Molecular Biology, Constantine University, Route Ain El Bey, Constantine

LinkOut to external resources

SILVA SSU Database [SILVA]

Activér Windows

Recent activity

Figure 4 : Capture d'écran de l'interface Genbank lors de la recherche par le numéro d'accésion (JO407772).

3. Comprendre les Résultats de la Recherche

La recherche d'une séquence dans Genbank à partir d'un numéro d'acquisition, comme X53207 ou JQ407772, permet d'accéder directement à une entrée unique, précise et validée, représentant une séquence spécifique déposée par un laboratoire donné. Chaque numéro d'acquisition renvoie à une fiche complète contenant l'ensemble des métadonnées essentielles : le nom de l'organisme, éventuellement son ancienne dénomination, la classification

systématique complète, la nature du gène (comme le 16S rRNA), la taille de l'amplicon en paire de bases, le type de la molécule, la date de soumission, les auteurs, les conditions ou l'origine du prélèvement, ainsi que des annotations fonctionnelles décrivant la localisation et les caractéristiques du gène. Ces fiches fournissent également la séquence nucléotidique brute, indispensable pour les analyses bioinformatiques telles que BLAST, l'alignement multiple ou la phylogénie.

La signification des résultats répétés :

La présence d'un nombre élevé de résultats répétés dans Genbank pour *Tsukamurella paurometabola* 16S rRNA (608 entrées) et pour *Tsukamurella sp.* 1534 (53 entrées) s'explique par le fait que Genbank est une base de données collaborative où de nombreux laboratoires du monde entier déposent leurs propres séquences, même lorsqu'il s'agit du même gène ou de la même espèce. Chaque séquence correspond à une souche différente, un échantillon différent ou un projet de séquençage distinct, ce qui crée une liste importante d'enregistrements. De plus, les gènes ribosomiques comme le 16S rRNA existent souvent en plusieurs copies légèrement différentes dans un même génome, ce qui contribue aussi à la diversité des dépôts. Les variations naturelles entre souches d'une même espèce produisent également des séquences légèrement divergentes, justifiant la multiplicité des résultats.

Ainsi, ces répétitions ne sont pas des erreurs : elles reflètent la diversité génétique de l'espèce, les différences entre échantillons et la richesse des données produites par les différents programmes de recherche, notamment les projets de séquençage partiels, complets ou shotgun qui génèrent plusieurs fragments ou assemblages du même génome.

Les résultats de la recherche dans GenBank peuvent être consultés à travers trois pages complémentaires : Genbank, FASTA et Graphics. La page GenBank fournit une fiche complète détaillant toutes les métadonnées associées à la séquence, notamment le nom de l'organisme, le locus génétique, la taille de la séquence, les auteurs et les domaines fonctionnels et la séquence complète écrit en blocs de 10 bases, terminé par le symbole « // ». La page FASTA permet d'extraire la séquence pour des analyses bioinformatiques, structurée selon une ligne d'en-tête débutant par le symbole « > » contenant les métadonnées, suivie de la séquence nucléotidique (A, T, C, G) repartie sur plusieurs lignes; ce format est essentiel pour réaliser des alignements de séquences ou construire des arbres phylogénétiques.

Enfin, la page Graphics offre une représentation visuelle du gène au sein du génome, indiquant son emplacement exact, son orientation et les gènes voisins, ce qui facilite l'interprétation du contexte génomique et la compréhension des relations fonctionnelles avec

d'autres gènes. Ensemble, ces trois pages permettent d'analyser de manière complète et structurée les données obtenues lors d'une recherche dans Genbank.

4. Importer une séquence stockée dans une base de données

L'importation d'une séquence génétique à partir d'une base de données telle que Genbank (NCBI) repose sur une procédure précise permettant d'obtenir la séquence brute au format FASTA.

Accéder au site NCBI et dans le menu déroulant sélectionné « Nucleotide », puis saisir le numéro d'accès : X53207, Cliquer sur search.

On accède directement à la fiche exacte contenant la séquence recherchée (Figur 5).

La capture d'écran montre une interface web de recherche sur Genbank. Le champ de recherche indique "Nucleotide" et la valeur "x53207". La page affiche les détails de la séquence :

- DEFINITION:** Tsukamurella paurometabolum (Rhodococcus aurantiacus) partial 16S rRNA
- VERSION:** X53207.1
- KEYWORDS:** 16S ribosomal RNA; ribosomal RNA.
- ORGANISM:** Tsukamurella paurometabolum
- REFERENCE:** 1 (bases 1 to 1342)
- AUTHORS:** Collins,M.D., Smida,J., Dorsch,M. and Stackebrandt,E.
- TITLE:** Tsukamurella gen. nov. harboring Corynebacterium paurometabolum and Rhodococcus aurantiacus
- JOURNAL:** Int J Syst Bactriol. 38, 385-391 (1988)
- REFERENCE:** 2 (base 1 to 1342)
- AUTHORS:** Stackebrandt,E.
- TITLE:** Direct Submission
- JOURNAL:** Submitted (09-1990) Stackebrandt E
- COMMENT:** Rhodococcus aurantiacus is reclassified as Tsukamurella

Figure 5 : Capture d'écran de l'interface Genbank afficher la séquence recherchée

Puis cliquez sur « FASTA » en haut de page, la séquence s'affiche au format FASTA (Figure 6).

La capture d'écran montre la même interface NCBI que la figure 5, mais avec le filtre "FASTA" appliqué. La page affiche la séquence de la manière suivante :

```

>X53207.1 Tsukamurella paurometabolum (Rhodococcus aurantiacus) partial 16S rRNA
TCATGGAGAGTTNGATCCTGGCTCAAGGACGAACGCTNGCGCGTCTAACACATGCAAGTCGAACGGT
AAGGATCAAAGCTNGCTNNCTGTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GGATAAGCCCTGGAAACTTGTCTTAATACGGGATATGACCTCTCCCTGNATGGGGGTGGTGGAAAGCTT
TTGGTGTACAGAGTGGCCCGGGCTTACAGCTTGGGTGGGTGATGGCTACCAAGGGCAGACGGGG
TAGCNGNCCCTGAGAGGGTGACCGGCCACACTGGGACTTGAGANACGGCCAGACTCTACGGGAGGCAGCA
GTGGGGAAATTGCAACATGGGGCAAGCTNATGYAGCGACGCCCGCGTGAAGGGATGACGCCCTTCGGGT
TGTAACCTNNTTCAGTAGGGAGCGAAAGCAGCGAACGTTGACGGGACTTACAGAAAGNAACGCCCAACTACAGT
GCCAGCAGCCGGTAAATACGTAGGGTGGCAGGGTTGTCGGGATTTACTGGGCGTAAAGNGGTCGTAGGC
NGTTTGTGGCGTGTGTGAAAAAACCGAGGCTTAACCTCGGGGCTGCAAGGGATACGGGGCAGACTGGAG
TACTGTAGGGGAGATTGAAATTCTGTGTAGGGTGGTNNATGCCAGATATCAGGGAGGAACACGGGTTGG
GAAGGGCGGGTCTGGGGCAGTAACTGNCGCTGAAGGAGGAAGGGATGGGGTAGCGAACAGGATTAGATACC
CTGGTATCCATGCCGTAAACGGTGGGTACTAGGTGTTGGTTCTCCACGGGATCCTGGCGTGAACGCTA
ACGCATTAAGTACCCCGCCCTGGGGAGTACGGCGCAAGGCTAAACACTCAAAGGAATTGACGGGGGCCCGC
ACAAGGGCGGGAGATGTGGAAATTCTGAGGCAACGGCGAAGAAACCTTACCTGGGTTGACATATAGAGG
ATCGGAGGAGATGTGTGAAATTCTCATGGCGCTATAGGTGGGTGAGGAGGAGGAGGAGGAGGAGGAGG
GGAGAGATGTGGTTAAAGGCCCGAGAGGGCGGCCCTGTGCAAGTGTGGGGAGGAGGAGGAGGAGGAGG
ACTCGTGGAGGAGACTGCCGGGGTNAACTCGGGAGGGTGGNGATGACGTCAGTCAAGTCAATNATGCCCTTATG
TCCAGGGCTTCACACATGCTNCHATGGCGCGTACAGNGGGCTNCGATAACCGTNAGGGTGGAGCGAATCCCT
NAAAGGGCGTCTCAGTCGGATTGGGTTCTGNAACTNGACCCNATGAAGTCGGAGTCGAGTAGTAATCGCA
GNTCAGNAACGC

```

Figure 6 : Capture d'écran de Séquence affichée au format FASTA sur NCBI

Sélectionnez tout la séquence à partir du symbole « > » jusqu'à la fin de la séquence (Figure 7).

```
>X53207.1 Tsukamurella paurometabolum (Rhodococcus aurantiacus) partial 16S rRNA
TCAATGGAGAGTTNGATCCTGGCTCAGGACGAACGCTNGCGCGTGTAAACACATGCAAGTCGAACGGT
AAGGATCAAGCTNGCTNNGTCTGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GGATAAGCCTGGAAACTTGGTCTAAATACCGGATATGACCTNCTCCTGNATGGGGTTGGTGGAAAGCTT
TTGTGGTACAGGATGGCCCGGCCCTATCAGCTGGTGGGTGGGTNATGCCCTACCAAGGGCAGACGGG
TAGCNGNCCTGAGAGGGTGACCGGCCACACTGGGACTGAGANACGGCCAGACTCCTACGGGAGGCAGCA
GTGGGGAAATTGCAAAATGGGCGCAAGCCTNATGYAGCGACGCCGCGTGGGGATGACGGCCTTCGGGT
TGTAACCTNNTTCAGTAGGGACGAAGCGCAAGGTGACGGTACCTACAGAAGNAGCACCGCCAACTACGT
GCCAGCAGCCCGGTAATACGTAGGGTGGCAGGCTTACTGGCGTAAGNGCTGTAGGC
NGTTGTCGGCTGTGAAACCCGAGGCTTAACCTGGGCTGAGGGTGGGTAGCGAACAGGATTAGATACC
TACTGTAGGGAGACTTGGGTTNATGGCTAGGGGTTGGGTAGCGAACAGGATTAGATACC
CTGGTAGTCATGGCTAAACGGTGGGTACTAGGTGGTTCTCCACGGGATCGCTGGTAGCTA
ACGGCTTAAGTACCCCGCTGGGGAGTACGGGCCAAAGCTAAACCTGGGCTGAGGGTGGGTAGCGAACAGGATTAGATACC
ACAAGCGGGAGCATGGGATTCAGTGGGTTGGGTAGCGAACAGGATTAGATACC
ATCGCCCGAGAGATGTGGTATGGCTAGTGGCTCTATACAGGTGGTGTGACGGCTGTAGCTGTC
GTGAGATGTGGGTTAAGTCCCGCAAGCGTAACGGCTTACAGGGCTNCGATACCGTNAGGGGGAGCGAACCCCT
TCCAGGGCTTCACACATGCTNCATGGCGCTACAGNGGGCTNCGATACCGTNAGGGGGAGCGAACCCCT
NAAAGCGCGTCTCAGTTGGATTGGGTCTGNAACTNGACCCNATGAAGTCGGAGTCGCTAGTAATCGCA
GNTCAGNAACGC
```

Figure 7 : Capture d'écran de Séquence en format FASTA sélectionnée.

Copier la séquence FASTA et la coller dans un éditeur de texte simple, tel que Bloc-notes afin de créer un fichier texte local, pour faciliter l'utilisation de la séquence pour l'analyse bioinformatique (Figure8).

```
>X53207.1 Tsukamurella paurometabolum (Rhodococcus aurantiacus) partial 16S rRNA
TCAATGGAGAGTTNGATCCTGGCTCAGGACGAACGCTNGCGCGTGTAAACACATGCAAGTCGAACGGT
AAGGATCAAGCTNGCTNNGTCTGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GGATAAGCCTGGAAACTTGGTCTAAATACCGGATATGACCTNCTCCTGNATGGGGTTGGTGGAAAGCTT
TTGTGGTACAGGATGGCCCGGCCCTATCAGCTGGTGGGTGGGTNATGCCCTACCAAGGGCAGACGGG
TAGCNGNCCTGAGAGGGTGACCGGCCACACTGGGACTGAGANACGGCCAGACTCCTACGGGAGGCAGCA
GTGGGGAAATTGCAAAATGGGCGCAAGCCTNATGYAGCGACGCCGCGTGGGGATGACGGCCTTCGGGT
TGTAACCTNNTTCAGTAGGGACGAAGCGCAAGGTGACGGTACCTACAGAAGNAGCACCGCCAACTACGT
GCCAGCAGCCCGGTAATACGTAGGGTGGCAGCGTGTCCGGATTTACTGGCGTAAGNGCTGTAGGC
NGTTGTCGGCTGTGAAACCCGAGGCTTAACCTGGGCTGAGGGTGGGTAGCGAACAGGATTAGATACC
TACTGTAGGGAGACTTGGGTTNATGGCTAGGGGTTGGGTAGCGAACAGGATTAGATACC
CTGGTAGTCATGGCTAAACGGTGGGTACTAGGTGGTTCTCCACGGGATCGCTGGTAGCTA
ACGGCTTAAGTACCCCGCTGGGGAGTACGGGCCAAAGCTAAACCTGGGCTGAGGGTGGGTAGCGAACAGGATTAGATACC
ACAAGCGGGAGCATGGGATTCAGTGGGTTGGGTAGCGAACAGGATTAGATACC
ATCGCCCGAGAGATGTGGTATGGCTAGTGGCTCTATACAGGTGGTGTGACGGCTGTAGCTGTC
GTGAGATGTGGGTTAAGTCCCGCAAGCGTAACGGCTTACAGGGCTNCGATACCGTNAGGGGGAGCGAACCCCT
TCCAGGGCTTCACACATGCTNCATGGCGCTACAGNGGGCTNCGATACCGTNAGGGGGAGCGAACCCCT
NAAAGCGCGTCTCAGTTGGATTGGGTCTGNAACTNGACCCNATGAAGTCGGAGTCGCTAGTAATCGCA
GNTCAGNAACGC
```

Figure 8 : Capture d'écran de Séquence en format FASTA collée dans un éditeur de texte (Blok-notes)

5. Conclusion

La recherche dans GenBank par numéro d'acquisition offre une information fiable, non ambiguë et parfaitement traçable, contrairement à la recherche par nom d'espèce qui renvoie souvent de multiples résultats provenant de différentes souches, laboratoires ou variantes génétiques. Les formats GenBank, FASTA et Graphics offrent les informations nécessaires pour les analyser, et le format FASTA est essentiel pour les outils bioinformatiques comme BLAST.