

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Ghardaia



Faculté des Sciences de la Nature et de Vie et Sciences de la Terre

Département de Biologie

Mémoire présenté en vue de l'obtention du diplôme de

MASTER

Filière : Sciences biologiques

Spécialité : Microbiologie appliquée

**Par : Bellouber Amoura
Bendouma Hadjer**

Thème

Analyse génomique de *Caldicoprobacter algeriensis* TH7C

Soutenu publiquement, le 23/09/2024, devant le jury composé de :

M ^{me} / Maidi L	Maitre-Assistant A	Univ-Ghardaia	Président
M./ Dif G	Maitre de conférences A	ENS-Laghouate	Directeur de mémoire
M./ Bakelli A	Maitre de conférences B	Univ-Ghardaia	Co-Directeur de mémoire
M ^{me} / Djemouai N	Maitre de conférences A	Univ-Ghardaia	Examineur

Année universitaire : 2023/2024

Dédicace

Je dédie ce travail à moi-même et mon époux, qui m'a permis de franchir cette étape et m'a soutenue énormément. Je lui présente également mes excuses pour sa patience et sa tolérance face à mes manquements durant cette année scolaire.

Je lui souhaite une bonne santé et du succès dans sa vie.

À ma famille, en commençant par mon père, que Dieu le garde, et ma mère, que Dieu ait son âme, qui m'ont élevés avec l'amour du succès dans la vie.

À mes enfants, je souhaite que Dieu les protège et les guide, et qu'ils réussissent dans leur vie.

À mes frères et sœurs, ainsi qu'à tous mes collègues.

À ma collègue binôme, Hadjer, je lui souhaite une vie pleine de succès.

À tous ceux qui m'ont aidé, ainsi qu'à ceux qui m'aiment de près ou de loin et soutiennent mon succès.

Amoura

Dédicace

Je dédie ce travail à mes parents qui m'ont soutenu tout au long de ma vie. J'espère que Dieu les protégera et prolongera leur vie. À mes enfants, que Dieu les garde et les protège, je leur souhaite du succès dans la vie. À mes frères et sœurs. À ma collègue, ma binôme Amoura, je lui souhaite une vie épanouie. À tous ceux qui m'ont aidé de près ou de loin, je souhaite à chacun le succès.

Hadjer

Remerciements

Tout d'abord, nous rendons grâce à Dieu pour Sa guidance et Son soutien tout au long de la réalisation de ce mémoire.

Nous tenons également à exprimer notre sincère gratitude à toutes les personnes qui ont contribué à ce travail. Nous remercions particulièrement Monsieur Dr: Dif Guendouz, le Directeur de ce mémoire, pour ses conseils précieux et son soutien constant.

Nos remerciements vont également à Monsieur DR: Bakessi Aissa, l'Assistant du Directeur, pour son aide précieuse.

Nous souhaitons également adresser nos remerciements à Madame Dr: Maidi Seïla, la Présidente du jury, ainsi qu'à Madame Dr: Djamouai nadjette, l'Examinatrice, pour leurs évaluations attentives et leurs remarques constructives.

Un grand merci à tous les enseignants qui nous ont accompagnés tout au long de notre parcours académique. Leur expertise et leur soutien ont été essentiels pour l'achèvement de ce travail.

Enfin, nous souhaitons exprimer notre reconnaissance à toutes les personnes qui, de près ou de loin, ont apporté leur soutien à ce projet.

Merci à tous pour votre aide et vos encouragements.

Bien cordialement,

[Amoura et Hadjer]

Résumé

Cette étude vise à analyser le génome complet d'une souche bactérienne, TH7C1. Une bactérie anaérobie thermophile (souche TH7C1) a été isolée de la source chaude hydrothermale de Guelma, dans le nord-est de l'Algérie, afin de détecter la présence de gènes intervenant dans la promotion de la croissance des plantes et ayant un impact sur les pertes importantes dans la sécurité alimentaire due à l'activité antagoniste contre les agents pathogènes des plantes. L'analyse taxonomique du génome a confirmé l'appartenance de la souche TH7C1 à une nouvelle espèce bactérienne, *Caldicoprobacter algeriensis*.

Dans cette démarche, le génome de la souche TH7C1 a été analysé à plusieurs niveaux à l'aide de différents outils informatiques pour étudier les gènes liés aux catégories fonctionnelles (COG) via une annotation génique à l'aide du serveur RAST, puis une identification génique par le programme Prokka. Cela a permis de découvrir et d'identifier les gènes liés aux mécanismes de promotion de la croissance des plantes et de lutte biologique. De plus, la recherche des groupes géniques responsables du métabolisme secondaire a été réalisée à l'aide de l'outil d'analyse bioinformatique antiSMASH.

Les résultats de l'analyse génétique de la souche TH7C1 ont révélé la présence de nombreux gènes contribuant à la tolérance au stress, tels que des gènes codant pour des enzymes et des composés antioxydants, ainsi que pour la production de radicaux libres du type superoxyde dismutase (SOD), peroxydases (POXs), et catalase (CAT). De plus, plusieurs gènes liés à des caractéristiques favorisant la croissance des plantes ont été identifiés, comme la solubilisation du phosphate organique, la production d'hormones végétales et l'absorption du fer.

En conclusion, les résultats obtenus soulignent l'importance de la souche TH7C1 en tant que candidat prometteur pour la promotion de la croissance des plantes et la contribution à la sécurité alimentaire.

Mots-clés : Outils de bioinformatique, sécurité alimentaire, gène, annotation du génome, amélioration de la croissance des plantes, résistance au stress.

Abstract :

This study aims to analyze the complete genome of a bacterial strain, TH7C1. A thermophilic anaerobic bacterium (strain TH7C1) was isolated from the hydrothermal hot spring of Guelma in the northeast of Algeria, with the purpose of detecting genes involved in promoting plant growth and impacting significant losses in food security due to antagonistic activity against plant pathogens. Taxonomic analysis of the genome confirmed the affiliation of strain TH7C1 with a new bacterial species, *Caldicoprobacter algeriensis*.

In this approach, the genome of strain TH7C1 was analyzed at multiple levels using various bioinformatics tools to study genes associated with functional categories (COG) through gene annotation using the RAST server, followed by gene identification using the Prokka program. This allowed for the discovery and identification of genes related to mechanisms promoting plant growth and biological control. Additionally, the search for gene clusters responsible for secondary metabolism was conducted using the antiSMASH bioinformatics analysis tool. The results of the genetic analysis of strain TH7C1 revealed the presence of numerous genes contributing to stress tolerance, such as genes encoding enzymes and antioxidant compounds, as well as the production of free radicals such as superoxide dismutase (SOD), peroxidases (POXs), and catalase (CAT). Furthermore, several genes related to traits promoting plant growth were identified, such as organic phosphate solubilization, production of plant hormones, and iron uptake.

In conclusion, the results underscore the importance of strain TH7C1 as a promising candidate for promoting plant growth and contributing to food security.

Keywords: Bioinformatic tools, food security, gene, genome annotation, enhancing plant-growth, stress resistance

المخلص:

تهدف هذه الدراسة إلى تحليل الجينوم الكامل لسلالة بكتيرية تُدعى TH7C1 تم عزل بكتيريا لا هوائية محبة للحرارة) السلالة (TH7C1T من الينابيع الساخنة في قالة بشمال شرق الجزائر، بغرض الكشف عن الجينات المتورطة في تعزيز نمو النباتات والتأثير على الخسائر الكبيرة في الأمن الغذائي بسبب نشاطها المضاد لمسببات الأمراض النباتية. أكدت التحاليل التصنيفية للجينوم انتساب السلالة TH7C1 إلى نوع بكتيري جديد، أُطلق عليه اسم *Caldicoprobacter algeriensis*.

في هذه الدراسة، تم تحليل جينوم السلالة TH7C1 على مستويات متعددة باستخدام أدوات المعلوماتية الحيوية المختلفة لدراسة الجينات المرتبطة بالفئات الوظيفية (COG) من خلال تحديد الجينات باستخدام خادم RAST ، تلاه تحديد الجينات باستخدام برنامج Prokka. سمح هذا باكتشاف وتحديد الجينات المتعلقة بآليات تعزيز نمو النبات والمكافحة البيولوجية. بالإضافة إلى ذلك، تم البحث عن مجموعات الجينات المسؤولة عن الأيض الثانوي باستخدام أداة التحليل المعلوماتي الحيوي antiSMASH. كشفت نتائج التحليل الجيني للسلالة TH7C1 عن وجود العديد من الجينات التي تساهم في تحمل الإجهاد، مثل الجينات التي تشفر الإنزيمات والمركبات المضادة للأكسدة، وكذلك إنتاج الجذور الحرة مثل السوبر أكسيد ديسموتاز (SOD) ، البيروكسيداز (POXs)، والكاتالاز (CAT). علاوة على ذلك، تم تحديد العديد من الجينات المتعلقة بالصفات التي تعزز نمو النبات، مثل إذابة الفوسفات العضوي، إنتاج هرمونات النبات، وامتصاص الحديد. في الأخير، تؤكد النتائج على أهمية السلالة TH7C1 كمرشح واعد لتعزيز نمو النبات والمساهمة في الأمن الغذائي.

الكلمات المفتاحية: أدوات المعلوماتية الحيوية، الأمن الغذائي، الجين، تحديد الجينوم، تعزيز نمو النبات، مقاومة الإجهاد

A red decorative border with rounded corners and a double-line effect, enclosing the text.

Liste des legendes

Liste des figures

N° Figures	Titre	Page
Figure 1	<i>Reconstruction phylogénétique de la famille des Caldicoprobacteraceae</i>	5
Figure 2	<i>Position phylogénétique de l'espèce Acetomicrobium flavidum basée sur l'ARNr 16S</i>	6
Figure 3	<i>Morphologie de la souche TH7C1T</i>	8
Figure 4	<i>Présentation des trois niveaux d'annotation (Vannier, 2017).</i>	19
Figure 5	<i>présenter une capture d'ecran de site Web du NCBI</i>	22
Figure 6	<i>présenter une capture d'ecran de Plateforme TYGS</i>	23
Figure 7	<i>présenter une capture d'ecran de DDBJ DFAST</i>	24
Figure 8	<i>présenter une capture d'ecran de Plateforme BV-BRC</i>	25
Figure 9	<i>présenter une capture d'ecran de Programme Prokka Serveur RAST</i>	26
Figure 10	<i>présenter une capture d'ecran de Serveur antiSMASH</i>	27
Figure 11	<i>présenter une capture d'ecran de detection strictness</i>	28
Figure 12	<i>Arbre phylogénétique basé sur l'ADNr 16S généré par TYGS montrant la parenté de Caldicoprobacter algeriensis souche TH7C1 avec les souches types les plus proches,</i>	32
Figure 13	<i>Arbre phylogénomique basé sur les séquences génomiques dans l'arbre TYGS inféré avec FastME (Meier-Kolthof et Göker, 2019)</i>	33
Figure 14	<i>Vue circulaire du génome de la souche TH7C1</i>	35
Figure 15	<i>Vue d'ensemble des catégories de sous-systèmes de TH7C1. L'annotation du génome a été réalisée en utilisant le serveur RAST.</i>	36
Figure 16	<i>Les clusters de gènes de biosynthèse et les métabolites secondaires</i>	47

Liste des tableaux

N de tableau	Titre	Page
Tableau 1	Caractéristiques distinctives de <i>C. guelmensis</i> , <i>C. oshimai</i> , <i>C. algeriensis</i> et <i>Acetobacterium faecale</i> (Ojuederie, O. B., & Babalola,(2017)).	7
Tableau 2	Description des cinq principales familles de plate-formes de séquençage de nouvelle génération (Hodkinson & Grice, 2015; Pericard, 2017)	14
Tableau 3	Caractéristiques du génome de <i>Caldicoprobacter algeriensis</i> souche TH7C1. Les caractéristiques génétiques sont le résultat de la sortie combinée de l'annotation DFAST	34
Tableau 4	Gènes impliqués dans la tolérance au stress osmotique.	37
Tableau 5	Gènes impliqués dans la résistance au stress thermique	38
Tableau 6	Gènes impliqués dans la formation et la germination des spores	39
Tableau 7	Gènes impliqués dans le transport et la résistance aux métaux lourds	41
Tableau 8	Gènes impliqués dans la protection contre le stress oxydatif.	42
Tableau 9	Gènes impliqués dans le métabolisme de l'azote, du soufre et du phosphore	43
Tableau 10	Gènes impliqués dans la chimiotaxie et la motilité	44
Tableau 11	Gènes impliqués dans la production/modulation des phytohormones.	46

A red decorative border with rounded corners and a double-line effect, enclosing the title text.

Liste des abréviations

Liste des symboles et abréviations

- **ADN** : Acide désoxyribonucléique
- **ARN** : Acide ribonucléique
- **ARNr**: ARN Ribosomal
- **ATP**: Adénosine Triphosphate
- **C** : Degré Celsius
- **CDS** : Coding DNA Sequence
- **FAO**: Food and Agriculture Organization
- **GC** : Guanine-Cytosine
- **PGPR** : Plant Growth-promoting rhizobacteria
- **WGS** : Whole Genome Sequencing
- **NGS** : Next Generation Sequencing
- **PCR** : Polymerase Chain Reaction

SOMMAIRE

Titre	Page
Introduction	1
I. Synthèse bibliographique	4
1. Description de la famille	4
2. Structure phylogénétique de la famille et ses genres	4
3 Généralités sur la souche Caldicoprobacter algriensis TH7C1	7
4- Morphologie de la souche TH7C1T	8
5- Description modifiée du gesdfHJNFD Caldicoprobacter	8
6- Conditions de croissance optimales	9
7- Propriétés métaboliques	9
Étude génomique des Bactéries favorisant la croissance des plantes	12
1. Séquençage de génomes	12
2. Séquençage de nouvelle génération	13
3. Principales technologies de NGS	14
4. Analyse génomique	15
5. Reconstruction de séquences (assemblage)	17
6. Annotation du génome	18
6.1. Annotation structurale	19
6.2. Annotation fonctionnelle	20
Matériel et méthodes	22
1-Outils utilisés (Materiels)	24
2-Méthode	29
3-Résultats	32
4-Discussion	50
Conclusion	54

Introduction

Introduction

Le système agricole contemporain est confronté à une série de défis majeurs, parmi lesquels la diminution de la productivité et la dégradation de la durabilité des écosystèmes agricoles, engendrées par la pollution environnementale, la sécheresse et la raréfaction des terres agricoles fertiles. Les projections des Nations unies suggèrent que la population mondiale pourrait dépasser les 9 milliards d'habitants d'ici 2050, entraînant ainsi une hausse constante de la demande alimentaire, allant jusqu'à 60 à 46 % (**Alexandratos, 2012**).

La salinisation des terres représente un défi majeur à l'échelle mondiale. Selon la FAO et les estimations les plus récentes, environ 800 millions d'hectares de terres sont déjà touchés par la salinisation, ce qui équivaut à environ 6% de la surface terrestre totale et 20% de la superficie cultivée dans le monde, menaçant sérieusement une superficie équivalente (**Jia et al., 2011**)/

Ce chiffre continue d'augmenter d'année en année en raison de l'utilisation excessive d'engrais chimiques et de l'irrigation mal maîtrisée, contribuant ainsi à l'aggravation du problème de la salinisation des sols (**Messedi et Abdelly, 2004**).

Dans les régions semi-arides et arides, la salinité représente un obstacle majeur à l'amélioration de la productivité agricole et au développement (**Abdellatef, 2010 ; UNU-INWEH, 2014 ; Brahimi et Lamiche, 2017**). Ces zones, couvrant près de 95% du territoire national en Algérie (Halitim, 2011), sont caractérisées par des sécheresses rigoureuses et fréquentes, où la salinisation des sols se présente comme l'un des principaux facteurs limitant le développement des plantes (**Farissi et al., 2014**).

Par conséquent, l'urgence se fait sentir pour mettre en place des stratégies novatrices et durables afin de répondre à cette demande croissante en alimentation et de prévenir la détérioration continue de notre écosystème. Parmi ces stratégies figure l'utilisation de micro-organismes, notamment les bactéries connues sous le nom de Rhizobactéries Promotrices de la Croissance des Plantes (PGPR), qui favorisent le développement racinaire et améliorent la santé des plantes sans pollution environnementale (**Dif et al., 2021, 2022**).

Les PGPR peuvent avoir un impact positif sur la croissance des plantes (**Guapta et al., 2002**) en facilitant l'absorption des éléments nutritifs présents dans le sol (tels que le phosphore et le fer), en produisant des hormones et des régulateurs de croissance, en améliorant la structure du sol et en détoxifiant les sols contaminés en absorbant les métaux lourds et les composés chimiques nocifs

(Ojuederie et Babalola, 2017). De plus, ils peuvent aider les plantes à tolérer le stress salin (Tank et Saraf, 2010).

La lutte biologique et la protection des plantes contre les agents pathogènes végétaux en contrôlant, inhibant ou renforçant les propriétés de résistance des plantes sont également des domaines d'application des PGPR (Toumatia et al., 2016).

Il existe plusieurs genres de bactéries, qu'elles soient symbiotiques comme *Bradyrhizobium*, *Mesorhizobium* et *Rhizobium*, comme mentionné par (Pereira-Gomez et al., 2020), Ou non-symbiotiques comme *Azospirillum*, *Klebsiella*, *Bacillus* et *Pseudomonas*, comme indiqué par (Fitriatin et Nurmala, 2019; Prasad et al., 2015) pour *Azotobacter*. Ces bactéries sont utilisées dans les domaines de la recherche et sur le terrain en tant que vaccins biologiques pour favoriser la croissance des cultures et les protéger contre les agents pathogènes des maladies végétales.

En ce qui concerne les caractéristiques phénotypiques et fonctionnelles des bactéries, elles dépendent principalement de leur génome, et il est possible de prédire de nombreuses caractéristiques en identifiant leur séquence génomique (Kumar et al., 2015; Hil et al., 2013). La technologie de séquençage de nouvelle génération (NGS) a révolutionné les sciences biologiques en raison de sa rapidité et de sa précision inégalées (Vincent et al., 2017). En combinant le séquençage NGS avec les progrès de l'informatique et des algorithmes, il est possible d'analyser rapidement et efficacement les séquences génomiques et de prédire les gènes et les caractéristiques des souches bactériennes (Drouin, 2018), offrant ainsi une approche prometteuse pour économiser du temps et des efforts dans divers domaines d'application (Thakur, 2018).

Dans le cadre de cette étude, nous cherchons à mener une analyse génomique approfondie à l'aide d'outils et de logiciels informatiques de pointe sur une souche bactérienne spécifique *C. algeriensis* TH7C1 qui n'a pas encore fait l'objet d'une étude approfondie, dans le but d'identifier des gènes ayant un impact positif sur la stimulation et le soutien de la croissance des plantes. Cela permettra de déterminer l'importance de cette souche en tant que vaccin biologique amélioré pour la croissance des cultures et la protection contre les maladies des plantes.

Ce travail se conclura par une synthèse des principaux résultats obtenus au terme des recherches menées, en soulignant les perspectives prometteuses pour l'application des PGPR (*C. algeriensis* TH7C1).



Chapitre I

Synthèse Bibliographique

I. Synthèse Bibliographique :

1. Description de la famille

Le nom *Caldicoprobacter algeriensis* TH7C1 est dérivé du genre *Caldicoprobacter*, qui est un genre de la famille des *Caldicoprobacteraceae*. Le nom du genre "*Caldicoprobacter*" vient du latin "calidus" signifiant "chaud", "probare" signifiant "tester" ou "éprouver", et "bacter" signifiant "bactérie", faisant référence à sa capacité à croître dans des conditions thermophiles. Le terme "algeriensis" fait référence à l'origine géographique de la souche, l'Algérie. Et "TH7C1" peut être un code spécifique utilisé pour identifier cette souche dans des contextes de recherche ou de laboratoire. (Yokoyama et al, 2010)

Coloration de Gram positive Comprend les espèces thermophiles sporulées et non sporulées cellules en forme de bâtonnet à chimioorganotrophie anaérobie métabolisme. Immobile. Les acides gras cellulaires sont composés principalement de 17 et 15 carbones saturés, ramifiés acides gras (iso-C17 :0, iso-C15 :0 et anteiso-C17 :0). Le peptidoglycane le type est A1g. Les lipides polaires n'ont pas été analysés. Valeurs G+C de L'ADN varie entre 41 et 45% en moles. Isolé des excréments de mouton et des sources chaudes. (Yokoyama et al, 2010)

2. Structure phylogénétique de la famille et ses genres

Analyse phylogénétique de *Caldicoprobacter* spp. en utilisant le 16S Les séquences des gènes d'ARNr ont révélé qu'elles forment une branche distincte dans l'ordre des Clostridiales et comprend également *Acetomicrobium* matières fécales isolées des boues d'épuration incubées à 72 C° (Winter et al., 1987)

La séquence du gène de l'ARNr 16S d'A. faecale a été établie en décembre 2011(Figure 01)

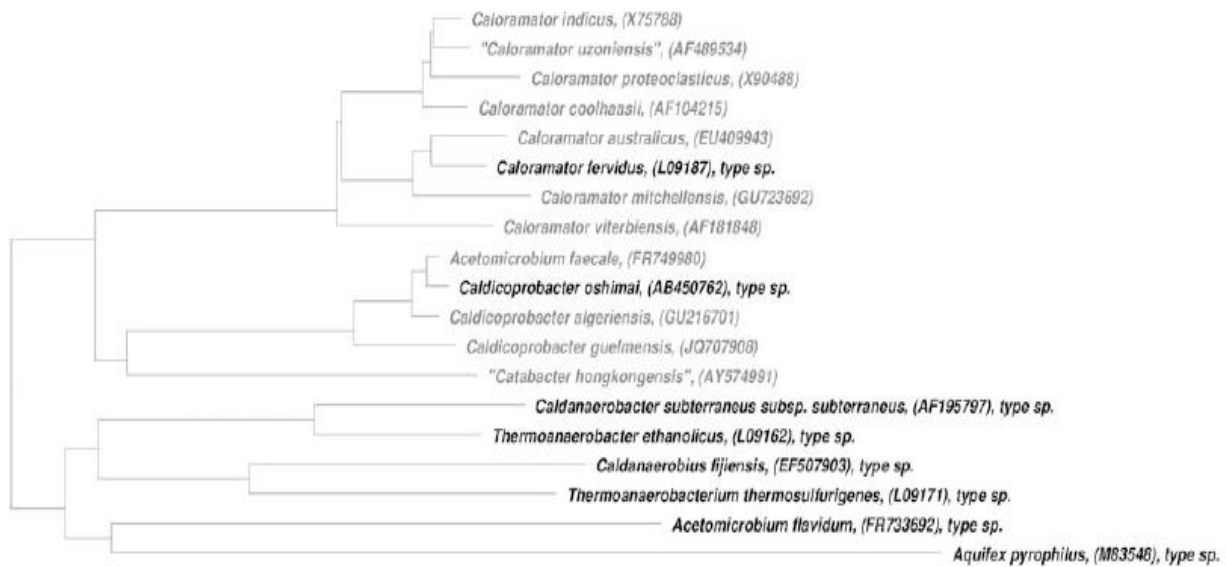


Figure 1: Reconstruction phylogénétique de la famille des *Caldicoprobacteraceae*

après la description du famille des *Caldicoprobacteraceae*. Cela démontre qu'*A. faecale* doit être reclassé dans le genre *Caldicoprobacter* comme *C. fécale*, peigne. nov. (**Bouanane-Darenfed et al, 2011**),

Basé principalement sur ses caractéristiques phénotypiques et génétiques, *A. faecale* (**Winter et al., 1987**) a été reconnu pour la première fois comme membre du genre *Acetomicrobium* (**Goodfellow et al, 2011**). de la famille des Bacteroidaceae, de l'ordre des Bacteroidales, du phylum des Bacteroidetes, *A. flavidum* étant l'espèce type de ce genre. Par ailleurs, la séquence du gène de l'ARNr 16S d'*A. flavidum*, établie en janvier 2012, a démontré que *A. flavidum* et *A. faecale* ne doivent pas être considérées au même niveau de genre et a confirmé l'affiliation d' *A. faecale* au genre *Caldicoprobacter* (98 % et 98,9 % de similarité avec *C. algeriensis* et *C. oshimai*, respectivement) comme déjà proposé par (**Bouanane-Darenfed et al, 2011**).

Compte tenu de leurs séquences génétiques d'ARNr 16S, il semble qu' *A. flavidum*, les trois espèces d' *Anaerobaculum* (Ab) décrites jusqu'à présent, *Ab. mobile* (**Javier-Menes R, Muxi L, 2002**), *Ab. thermoterrenum* (**1997, p. Rees et al**), et Un B. hydrogéniformans (**Maune MW, Tanner RS , 2012**) et *A. flavidum* appartiennent au même clade phylogénétique avec des niveaux élevés (> 95 %) de similarité (**Figure 02**).

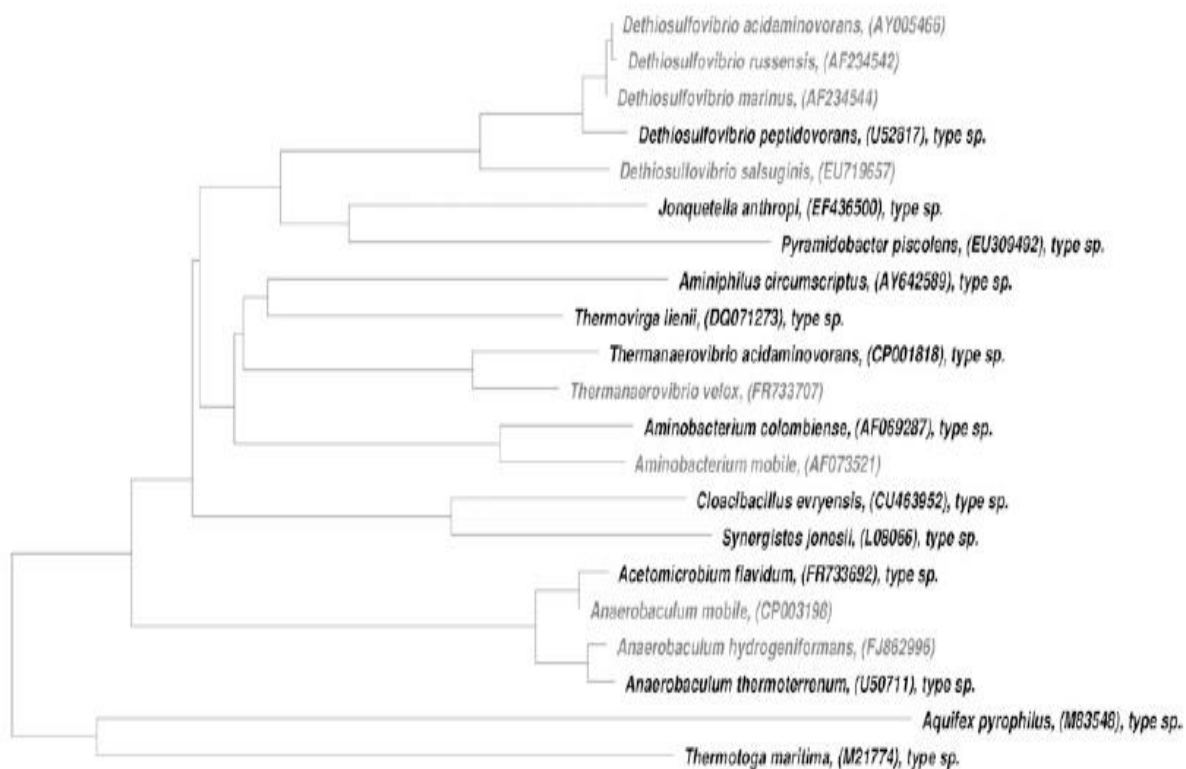


Figure 2: Position phylogénétique de l'espèce *Acetomicrobium flavidum* basée sur l'ARNr 16S

A cet égard, ces trois espèces d'*Anaerobaculum* devraient être reclassé au sein du genre *Acetomicrobium* qui a la priorité sur le genre *Anaerobaculum* (1997, p. Rees et al) depuis son validé ultérieurement après le genre *Acetomicrobium* (liste de validation, 1985, Int. J. Syst. Bactérie. 35, 223-225).

Par conséquent, *Acetomicrobium* spp. devrait appartenir à la famille des Synergistacées, ordre des Synergistales, phylum Synergistetes. À ce jour, le genre *Caldicoprobacter* comprend trois genres, *C. oshimai*, *C. algeriensis* et *C. guelmensis*, isolé du mésothermique (Yokoyama et al, 2010) et environnements chauds (Bouanane-Darenfed et al, 2011)

Tableau 2: Caractéristiques distinctives de *C. guelmensis*, *C. oshimai*, *C. algeriensis* et *Acetobacterium faecale* (Ojuederie, O. B., & Babalola,(2017)).

Caractéristiques	1	2	3	4
Isolation	Source chaude	Sludge samples	Échantillons de boues	Source chaude
Motilité	-	+	-	-
Morphologie	Tige	Tige	Bâtonnet sporulé	Tige
Gramme	Positive	Négative	Positive	Positive
Température optimale(c)°	65(45-85)	73-70	70(77-44)	65(55-75)
Plage de pH optimale	6.8(5-9)	6.5(5.5-9)	7.2(5.9-8.6)	6.9(6.2-8.3)
Nac optimal%	0-2	0-3	0-2	>5
Substrat de croissance				
Xylane	+	-	+	-
Raffinose	-	ND	+	+
Melibiose	-	+	ND	+
Mannitol	-	-	-	+
Pyruvate	+	-	ND	+
Produits finaux de la fermentation du glucose	Acétate, lactate, H ₂ ·CO ₂	Acétate, lactate, Éthanol·H ₂ ·CO ₂	Acétate, lactate, Éthanol·H ₂ ·CO ₂	Acétate, lactate, Éthanol·H ₂ ·CO ₂
G+C%	41.6	45	45.4	44.7

3- Généralités sur la souche *Caldicoprobacter algeriensis* TH7C1

Caldicoprobacter algeriensis sp. nov. strain TH7C1 est une nouvelle espèce de bactérie thermophile anaérobie isolée à partir d'une source hydrothermale chaude de Guelma en Algérie. Voici les principales caractéristiques de cette souche bactérienne :

- C'est une bactérie thermophile anaérobie, ce qui signifie qu'elle se développe à des températures élevées en absence d'oxygène.
- Elle a été isolée à partir d'un échantillon prélevé dans une source hydrothermale chaude située à Guelma, en Algérie.
- Les analyses phylogénétiques basées sur la séquence de l'ARNr 16S ont montré que *C. algeriensis* représente une nouvelle espèce au sein du genre *Caldicoprobacter*.
- La souche type TH7C1 a été déposée dans deux collections de cultures, la Collection de l'Institut Pasteur (CIP) et la Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ).
- C'est la première fois qu'une nouvelle espèce du genre *Caldicoprobacter* est décrite, élargissant ainsi la diversité connue de ce genre bactérien.

4- Morphologie de la souche TH7C1

Les colonies obtenues dans des tubes roulés étaient rondes, crémeuses, et jaune pâle. Ils mesuraient 2 à 3 mm de diamètre après 3 jours d'incubation à 70 C. La souche TH7C1 était immobile bactérie en forme de bâtonnet se produisant seule, en paires ou parfois sous forme de longues chaînes d'une taille allant de 0,7 à 1,0 à 2,0–6,0 µl (Fig. 3A). Aucune spore n'a été observée.

Cellules colorées à Gram positif. La structure de la paroi cellulaire de la souche TH7C1 était une seule couche (type Gram positif) (Fig. 3b).

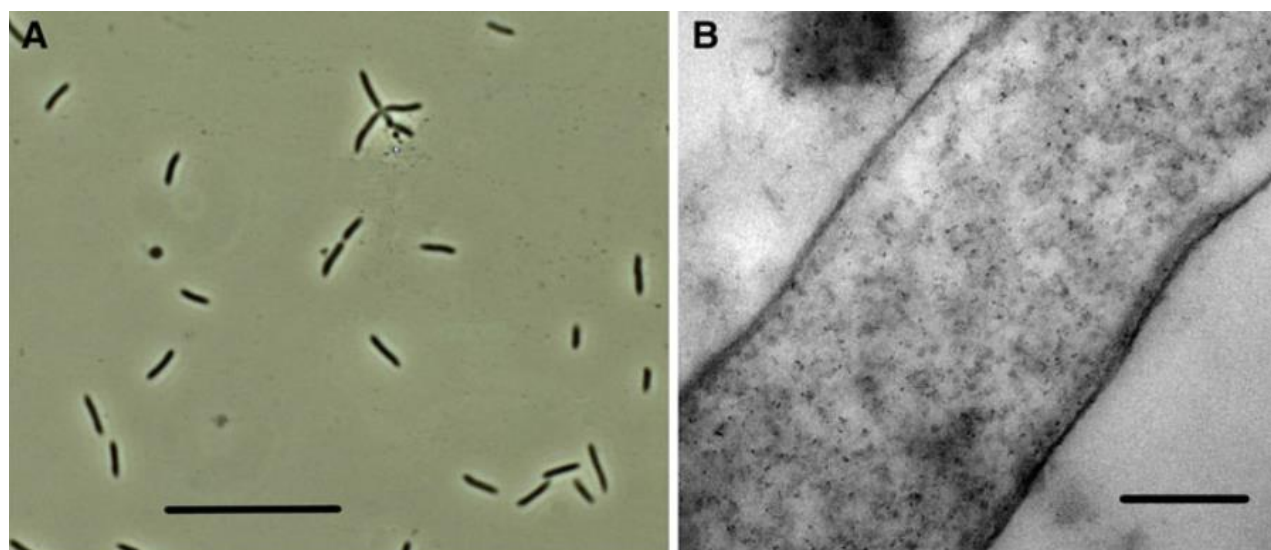


Figure 3: Morphologie de la souche TH7C1T. a. contraste de phase micrographie de cellules en forme de bâtonnet. Barre 20 µl. b. Électron micrographie de coupes fines. Bar 0,2 µl (Ojuederie, O. B., & Babalola, (2017).

5- Description modifiée du *Caldicoprobacter* :

La description du genre est la même que celle donnée par (Yokoyama et al.) sauf que la capacité à former des spores et réduire le mannitol sont variables (Benson DA, Karsch-Mizrachi I, Lipman DJ, 2008)

Description de *Caldicoprobacter algeriensis* sp. nov. (al.ge.ri.en'sis. M.L. fem. N. faisant référence à l'Algérie, le pays où la bactérie a été découverte pour la première fois). Cellules coloration Gram positif. Ils ont un type de Gram positif paroi cellulaire. (Slobodkina GB, Reysenbach A-L, Panteleeva AN. et al., 2012)

Ce sont des bâtonnets immobiles apparaissant seuls (0,7–1 à 2–6 ml), par paires, ou occasionnellement en longues chaînes. Spores ne sont pas observées. Sucres strictement anaérobies et fermentants utilise du glucose, du galactose, du lactose, de la raffinose, du fructose, du ribose, xylose, arabinose, maltose, mannitol, cellobiose, mannose, le mélibiose, le saccharose, le xylane et le pyruvate, mais pas les casaminoacides, peptone, gélatine, amidon, éthanol, succinate, le formiate et le fumarate. Sulfate, thiosulfate, soufre élémentaire, le sulfite, le nitrate et le nitrite ne sont pas utilisés comme terminaux accepteurs d'électrons. La température optimale est de 65 °C (plage 55 à 75 °C ; pas de croissance à 50 et 80 °C). pH optimal à 6,9 (pas de croissance à 6,3 et 8,6).

La croissance se produit à la concentration de NaCl allant de 0 à 0,5 % (p/v). Sous optimal conditions de croissance, le temps de génération est de 2 h.

Le principal les acides gras membranaires sont C17:0 (iso et anteiso), C16:0, et acides C15:0.

La teneur en G+C de l'ADN génomique est de 44,7 %. Le type la souche est la souche TH7C1^T (= DSM 22661^T = JCM 16184^T) qui a été isolée d'une source thermale située à A 20 km de Guelma au nord-est de l'Algérie (70 250 Est, 36 270 Nord), à 320 m d'altitude. (**Benson DA, Karsch-Mizrachi I, Lipman DJ , 2008**)

6- Conditions de croissance optimales

La souche était strictement anaérobie et thermophile. L'optimal la température de croissance était de 65 °C (plage de 55 à 75 °C ; aucune croissance à 50 et 80 °C). L'isolat n'a pas nécessité de NaCl pour croissance. La plage de pH optimale pour la croissance était de 6,9 (plage 6.2 à 8.3).

7- Propriétés métaboliques

L'isolat nécessitait de l'extrait de levure (0,1 %) pour sa croissance, ce qui ne pouvait pas être remplacé par des peptides, de la biotrypcase ou des vitamines. Sulfate, thiosulfate, soufre élémentaire, sulfite, nitrate et les nitrites n'ont pas été utilisés comme accepteurs d'électrons terminaux. Souche TH7C1 s'est développée sur du glucose, du galactose, du lactose, du raffinose, fructose, ribose, xylose, arabinose, maltose, mannitol, cellobiose, mannose, mélibiose, saccharose, xylane et pyruvate, mais pas sur les casaminoacides, la peptone, la gélatine, l'amidon, éthanol, succinate, formiate et fumarate.

les produits finaux résultant de la fermentation du glucose ont été acétate, lactate, éthanol, CO₂ et H₂. En croissance optimale conditions, la durée de génération était d'environ 2 heures.

Souche TH7C1 s'est développée avec le xylane comme source d'énergie. Xylanolytique l'activité a été trouvée extracellulaire. La souche TH7C1 était sensible à l'ampicilline (150 µg ml⁻¹) et au chloramphénicol (50 µg ml⁻¹). Morphologie de la souche TH7C1. un contraste de phase micrographie de cellules en forme de bâtonnet. Barre 20 µm. b Électron micrographie de coupes fines. Bar 0,2 µm (Joseph Felsenstein, 1985).

Chapitre II

Étude génomique des Bactéries favorisant la croissance des plantes

1. Séquençage de génomes

Le séquençage de l'ADN est une technique qui a révolutionné la biologie moléculaire dans les années 1970, c'est un procédé chimique qui permet d'établir l'ordre dans lequel les quatre nucléotides sont enchainés dans un brin d'ADN donné. Ce procédé, reposant sur la synthèse d'un brin d'ADN par une ADN-polymérase, a été mis au point pour la première fois en 1977 (**Sanger et al., 1977**).

Au cours des dernières années, la méthode de Sanger a été largement développée grâce à plusieurs avancées technologiques importantes telles que le développement de la synthèse chimique automatisée des oligonucléotides qui sont utilisés comme amorces dans la synthèse, l'introduction de traceurs fluorescents à la place des marqueurs radioactifs, la mise au point d'une méthode d'amplification linéaire utilisant peu de matrices et l'utilisation de séquenceurs automatiques ainsi que de l'électrophorèse capillaire (**El Rhaffouli et al., 2014**).

Les premiers génomes complètement séquencés ont tout d'abord été ceux de virus puis de micro-organismes. Le séquençage complet des génomes de trois microorganismes modèles qui sont la levure *Saccharomyces cerevisiae* et les bactéries *Bacillus subtilis* et *Escherichia coli*, a été effectué grâce au travail concerté, durant plusieurs années, de laboratoires de recherche regroupés en consortiums internationaux.

Les séquences de ces espèces ont été obtenues en totalité entre 1996 et 1997. Parallèlement, des entreprises privées se sont lancées de manière efficace dans le séquençage. Ainsi, le premier séquençage complet de génome bactérien a été celui de la bactérie pathogène *Haemophilus influenzae*, une bactérie responsable d'infections broncho-pulmonaires chez les jeunes enfants, dont le génome est composé d'un seul chromosome circulaire long de 1830140 paires de bases et comportant 1738 gènes (**Fleischmann et al., 1995**).

Pour la première fois, la base génétique d'un isolat bactérien a été complètement caractérisée. À mesure que de nouvelles technologies de séquençage seront mises au point pour faciliter l'accès aux informations sur les séquences génomiques, la révolution génomique en cours devrait continuer à avoir un effet transformateur et avoir un impact sur de nombreux aspects de notre vie. Primitivement, le séquençage manuel ne permettait de séquencer que 800 bases en environ 5 jours et nécessitait l'utilisation de radioéléments. Avec les séquenceurs capillaires automatiques, la

cadence est passée dans les années deux mille à environ 500 kilos bases (kb) par jour. Puis, à partir de 2006, une convergence entre les techniques de microfluidique, les nanotechnologies et les progrès de l'informatique, a permis l'émergence de nouvelles machines capables de réaliser plusieurs millions de réactions de séquence en parallèle. Ces automates sont appelés séquenceurs à très haut débit ou séquenceurs de nouvelle génération (**Sanger et al., 1977; Therrien-Laperrière, 2018**).

2. Séquençage de nouvelle génération

Les méthodes de séquençage de nouvelle génération permettent de séquencer un grand nombre de molécules d'ADN en parallèle (**Metzger et al., 2020**)

En 2007, la deuxième génération des appareils Next Generation Sequencing (NGS) a été développée. Les termes « deuxième » et « troisième génération » sont synonymes de l'évolution des technologies depuis la méthode de séquençage direct de Sanger dite de « première génération » (**Meldrum et al., 2017**)

Aujourd'hui, une seule machine peut séquencer 1 Gigabase (Gb) en quelques heures et les développements technologiques annoncés sont plus vertigineux encore (**Lacoste et al., 2017**).

Toutefois, avec la possibilité de séquençage du génome entier (« whole genome sequencing », WGS) de bactéries, virus et champignons, nous assistons probablement à une véritable révolution (**Deng et al., 2016**).

La technologie du WGS permet de réaliser des analyses avec une précision jusqu'alors inégalée et fournissant une quantité d'informations aussi inégalée.

Alors qu'il y a encore quelques années, le WGS était limité par les coûts élevés liés au matériel et aux réactifs, cet obstacle a été surmonté grâce à la disponibilité croissante et au développement de la technologie (**Dif, 2022**).

En parallèle, la bio-informatique pour l'analyse des grandes quantités de données complexes a accompli d'énormes progrès. Bien que pour la plupart des génomes bactériens, on compte « seulement » entre deux et neuf mégabases, le volume des données à analyser peut être démultiplié en raison d'une couverture « coverage » multiple lors du séquençage. Des ordinateurs très performants doivent dès lors être utilisés pour l'analyse bio-informatique. (**Dif, 2022**).

La définition de processus analytiques automatiques (« analytical pipelines ») au moyen de logiciels commerciaux permet un traitement relativement rapide des données en l'espace de quelques heures (Meinel et al., 2017).

3. Principales technologies de NGS

Cinq technologies NGS à haut débit de « deuxième et troisième génération » sont apparues de façon quasi simultanée : le pyroséquençage (454 Life Sciences, Roche Diagnostics, Bale, Suisse), le séquençage avec des terminateurs réversibles (HiSeq2000, Illumina/Solexa, Inc., San Diego, CA, USA), le séquençage par ligation (SOLiD™, Life Technologies, Carlsbad, CA, USA), le séquençage avec la technologie "Ion Torrent" et le séquençage Pacific Biosciences (Hodkinson & Grice, 2015). Dans le tableau suivant (Tableau 2) on présente les cinq principaux types de plates-formes qui ont été utilisés pour les études sur le microbiome :

Tableau 3 Description des cinq principales familles de plate-formes de séquençage de nouvelle génération (Hodkinson & Grice, 2015; Pericard, 2017).

Technologie	Longueur lectures	Précision	Nb de lectures par run	Coût (\$US/Mbp)
Sanger	400-900 bp	99,9%	N/A	2400 \$
Illumina	MiSeq : 50-600 bp HiSeq : 50-500 bp	99,9%	MiSeq : 1-25M HiSeq : 0,3-2 G	0,05-0,15 \$
Roche-454	700 bp	99,9%	1 M	10 \$
Ion Torrent	jusqu'à 400 bp	98%	jusqu'à 80 M	1 \$
Pacific Biosciences	10-15 kbp moy max > 40 kbp	87%	500-1000 M	0,13-0,60 \$
Nanopore	jusqu'à 500 kbp	~92-97%	variable	~0,5 \$

Illumina produit la famille de plateformes les plus utilisées, cette technologie a été introduite en 2006 et a été rapidement adoptée par de nombreux chercheurs, car une plus grande quantité de données pourrait être générée de manière plus rentable. Pour cette technique, l'amplification de l'échantillon à analyser ne s'effectue pas en solution, mais sur un support solide.

La réaction de séquençage est alors réalisée directement sur le support où l'ADN a été amplifié. Elle se déroule position après position et après ajout d'un mélange contenant toutes les bases associées chacune à un fluorophore différent. L'extrémité de ces bases est protégée pour empêcher l'addition de bases supplémentaires à chaque cycle d'incorporation. Un laser détecte les bases incorporées au fur et à mesure. Le clivage des fluorophores permet ensuite l'incorporation de la base suivante (**Ilie et al., 2014**) Les lectures Illumina sont considérées comme de «courtes» lectures bien que certaines avancées dans la technologie de séquençage leur permettent maintenant de rivaliser avec les lectures 454 en terme de taille (présentement 300 nucléotides) (**Ilie et al., 2014**), la lecture est effectuée ainsi cycle après cycle. Cette méthode permet l'acquisition en parallèle de plus de 3 milliards de séquences de 150 paires de bases.

Chaque position étant lue l'une après l'autre, les erreurs principales de cette technologie sont des erreurs de substitution d'une base par une autre (**Desai & Jere, 2012; Loman et al., 2012; Ilie et al., 2014**). Les erreurs de type déphasage affligent $\sim 0,015\%$ des lectures Illumina (**Luo et al., 2012**).

La formation de lectures chimériques et la présence de nucléotides provenant d'amorces et d'adaptateurs de séquençage limitent également l'utilisabilité des lectures de séquençage (**Tsai et al., 2014**).

4. Analyse génomique

La génomique est une discipline de la biologie moderne. L'évolution des technologies de séquençage a permis le séquençage et l'étude de nombreux génomes issus de divers organismes pris individuellement. Le génome est l'ensemble du matériel génétique d'un individu ou d'une espèce codée dans son ADN (**Vannier, 2017**). La connaissance des gènes est une étape indispensable à la compréhension des phénomènes biologiques d'un organisme. La génomique est l'étude des génomes. Elle concerne le séquençage du code génétique complet (ADN) des organismes vivants. C'est une science qui examine comment l'ensemble du génome interagit avec l'environnement.

La génomique implique le séquençage de grandes quantités d'ADN et ainsi, produit une énorme quantité de données qui peuvent être conservées, organisées et consultées L'arrivée des méthodes de séquençage de nouvelle génération amorça une autre révolution. Les avancées en génomique y sont pour beaucoup dans la croissance du domaine technique de la bioinformatique, qui consiste

en l'application de la science et de la technologie informatiques à la gestion de l'information biologique .

La génomique se divise en deux branches :

- La génomique structurale, qui se charge du séquençage du génome entier ;
- La génomique fonctionnelle, qui vise à déterminer la fonction et l'expression des gènes séquencés en caractérisant le transcriptome et le protéome **(Ponsart et al., 2014)**.

4.1. Génomique structurale

La génomique structurale est définie comme l'analyse de la séquence et de la structure des éléments du génome : les gènes, les régulateurs et les éléments mobiles **(Figueira Aburjaile et al., 2021)**.

La génomique structurale vise à accroître nos connaissances de base des macromolécules biologiques tout en diminuant les coûts moyens de la détermination de la structure. Parallèlement, la bio-informatique structurale est liée aux analyses et à la prédiction de la structure tridimensionnelle (3D) des macromolécules biologiques, telles que : les protéines, l'ARN et l'ADN. Le terme structurel est le même qu'en biologie structurale, et la bioinformatique structurale est considérée comme une partie de la biologie structurale computationnelle, alors que la génomique structurale tend à décrire la structure 3D de chaque protéine codée par un génome donné. Cette approche fournit une méthode à haut débit pour la détermination de la structure des approches expérimentales et de modélisation **(de Sousa et al., 2018)**.

4.2. Génomique fonctionnelle

La génomique fonctionnelle est une branche qui intègre des études de biologie moléculaire et de biologie cellulaire et traite l'ensemble de la structure, la fonction et la régulation d'un gène **(Le Mignon et al., 2010)**.

La génomique fonctionnelle est l'étude à l'échelle du génome de la fonction de l'ADN (y compris les gènes), alors elle utilise des données du génome pour étudier l'expression génique et protéique à l'échelle du génome et leurs fonctions. Pour cela, des méthodes à haut débit sont utilisées pour comprendre la transcription et la traduction des gènes et l'Interaction Protéine-Protéine (IPP).

Les études de génomique fonctionnelle sont liées à la fonction des gènes dans le génome **(De Sousa et al., 2018)**.

Le rôle du gène et sa modulation dans une condition donnée est crucial pour mieux

comprendre la biologie d'un organisme. Cette étude implique des études des transcrits et des protéines qui jouent un rôle essentiel dans le fonctionnement des processus biochimiques des cellules, associés avec la régulation (**De Sousa et al., 2018**).

5. Reconstruction de séquences (assemblage)

L'assemblage est le processus de reconstruction du génome qui a été fragmenté pour être séquencé et qui doit être reconstitué pour permettre l'étude de sa structure et de son contenu (**Amador, 2019**)

Le séquençage à haut débit fournit un ensemble de courtes lectures (reads) de séquençage. Il est souvent nécessaire de pouvoir reconstruire la séquence initiale à partir de ces fragments en l'absence d'un génome de référence. C'est ce qu'on appelle le problème de l'assemblage, qui est central en bio-informatique, au même titre que l'alignement de séquences.

Nous présentons tout d'abord l'assemblage dans le contexte classique de la génomique, avec la reconstruction d'un génome. Un assemblage consiste à aligner et/ou fusionner des petits fragments d'ADN ou d'ARN issus d'une plus longue séquence afin de reconstruire cette séquence originale. Il s'agit d'une étape d'analyse *in silico* qui succède au séquençage de l'ADN ou de l'ARN d'un organisme unique, d'une colonie de clones (bactériens par exemple), ou encore d'un mélange complexe d'organismes. Le problème de l'assemblage peut être comparé de manière imagée à celui de la reconstruction du texte d'un livre à partir de plusieurs copies de celui-ci, préalablement déchiquetées en petits morceaux. Son application principale en bioinformatique est la reconstruction de génomes, ou de gènes, à partir d'un ensemble de lectures de séquençage (**Pericard, 2017**).

Puisque différentes méthodes existent pour reconstruire *in silico* un génome fragmenté, il est important de pouvoir les comparer en évaluant la qualité de différents assemblages proposés (**Amador, 2017**).

Le produit de l'assemblage est, en général, plusieurs séquences contiguës appelées contigs et l'évaluation des assembleurs consiste généralement à comparer ces contigs (**Salzberg et al., 2012**)

Quand le séquençage et l'assemblage des données nucléotidiques sont disponibles, nous pouvons commencer la phase d'annotation de notre génome d'intérêt (**Ortet, 2018**).

6. Annotation du génome

Afin de documenter le génome, le transcriptome et le protéome, il est nécessaire de passer par le processus d'annotation (**Mercier, 2017**) .

L'annotation d'un génome correspond à prédire et localiser l'ensemble des séquences fonctionnelles biologiquement (gènes) du génome et à déterminer et identifier leur structure, leur fonction ainsi que les relations entre les entités biologiques relatives au génome (**Amador, 2017**). L'annotation d'une séquence génomique peut être abordée à différents niveaux d'analyse, une première phase incontournable consistant à identifier les gènes de l'organisme, c'est-à-dire à trouver leur localisation précise sur la *Revue bibliographique* 33 séquence du génome. Dans une seconde étape, on cherche ensuite à assigner une (ou plusieurs) fonction biologique à chacun de ces gènes hypothétiques. Ces étapes reposent initialement sur l'utilisation d'outils algorithmiques sophistiqués, dont le développement constitue l'un des champs de la bio-informatique. Cependant, cette annotation expertisée nécessite un travail manuel de vérification, long et fastidieux (**Poux S & Gaudet P, 2017**).

L'objectif de l'annotation des fonctions géniques est de fournir un catalogue des capacités moléculaires et/ou biochimiques dont est pourvu un organisme (**Mercier, 2017**) (**Figure 5**).

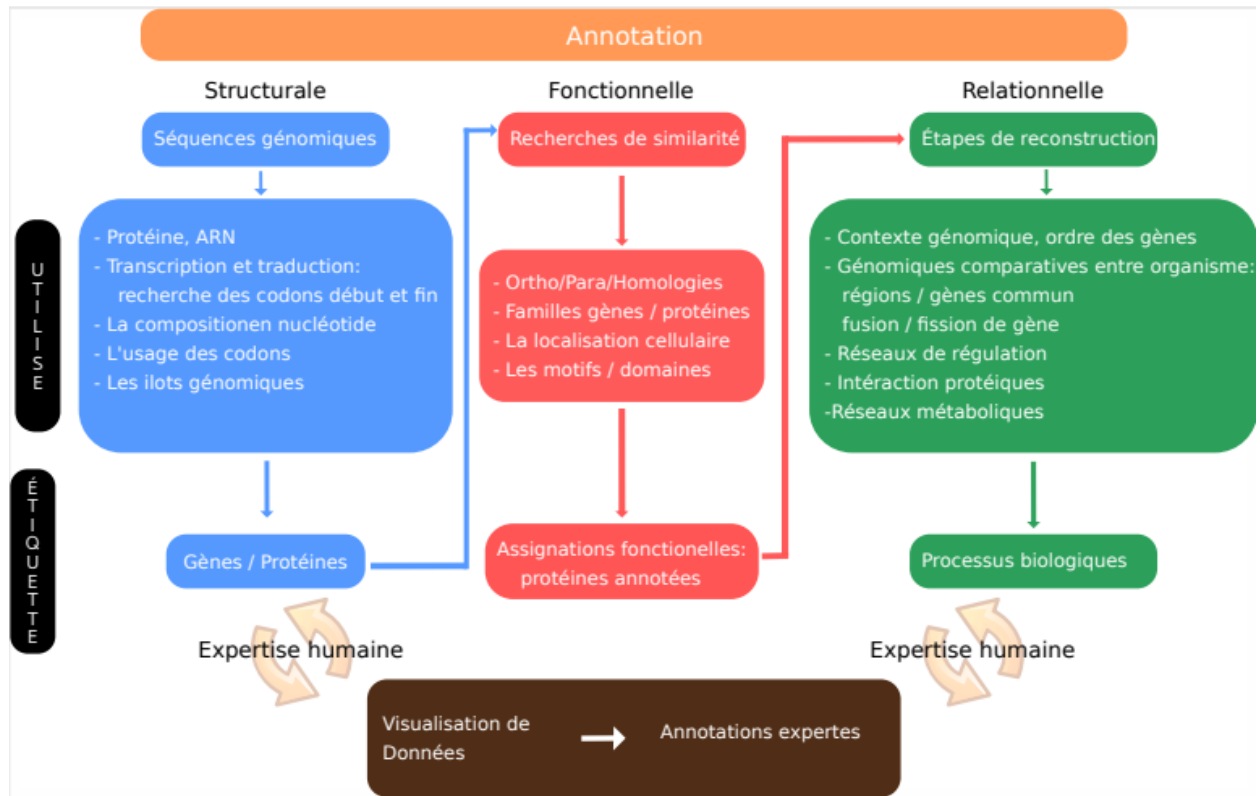


Figure 4: Présentation des trois niveaux d'annotation (Vannier, 2017).

Annotation structurale : identification d'objets d'intérêt biologique sur la séquence d'ADN (gènes de protéine, gènes d'ARN, signaux, etc.),

Annotation fonctionnelle : attribution des objectifs biologiques réalisés par les séquences précédemment identifiées (usuellement par homologie avec une séquence de fonction connue),

Annotation relationnelle : regroupe différentes relations établies entre les séquences pour décrire des objectifs ou modules biologiques par exemple une voie métabolique.

6.1. Annotation structurale

L'objectif de l'annotation syntaxique est de définir les bornes des gènes (CDS Coding Sequence et différents ARN), les alternances introns/ exons , mais également des pseudogènes (gènes inactifs ou non fonctionnels du génome), les régions non traduites (UTR, Untranslated region) et les séquences répétées (transposons, microsatellites, minisatellites, virus intégrés dans le génome, etc.). Cette annotation peut être approfondie pour identifier les régions promotrices, régulatrices, les sites de début de transcription (TSS, transcription start site), les sites de fixation d'histones, de

réplication, etc. Pour des séquences de transcrits, il s'agit essentiellement de localiser la région codante (**Ouyang et al., 2007**)

6.2. Annotation fonctionnelle

L'annotation fonctionnelle a pour principal but d'attribuer les fonctions biologiques de produit des gènes identifiés lors de l'annotation structurale (**Ortet, 2018; Therrien-Laperrière, 2018**).

L'annotation fonctionnelle de gènes en étudiant les gènes différentiellement exprimés dans des environnements contrastés (**Vannier, 2017**).

Elle consiste à attribuer une fonction biologique à chacun des gènes ou produits de gènes d'un génome ou d'une collection de transcrits. Alors elle permet de se forger rapidement une idée globale de cette fonction biologique réelle (**Bright et al., 2009**).

Il y'a deux façons d'annotation de génomes:

- **L'annotation automatique:** s'appuie (essentiellement) sur des comparaisons des séquences à annoter avec les séquences présentes dans les banques de données.

S'appuyer uniquement sur des algorithmes recherchant des similarités (de séquence, de structure, de motifs, etc.), permettant de prédire (en fait deviner) la fonction d'un gène. Elle aboutit au transfert « automatique » de l'information figurant dans l'étiquette d'un gène « similaire » d'un génome déjà annoté au génome en cours d'annotation (**Amador, 2017 ; Ortet, 2018**).

- **L'annotation manuelle :** basé sur l'expérience par des experts qui valident ou invalident la prédiction en fonction de leurs connaissances ou de résultats expérimentaux. Celle-ci peut ainsi éviter le transfert automatique d'erreurs et donc leur propagation, ce qui peut devenir le grand problème que devra affronter la génomique, compte tenu de l'afflux massif de données issues en particulier, des nouvelles techniques de séquençage (**Ortet, 2018**).

A red decorative border with a double-line effect and rounded corners, enclosing the title text.

Materiels et Méthodes

Matériels

Outils utilisés (Materiels)

Site web du NCBI (National Center for Biotechnology Information)

Le site web du NCBI (<https://www.ncbi.nlm.nih.gov>) sert de base de données complète contenant des informations sur les gènes provenant de multiples sources, accessibles aux professionnels et aux chercheurs en biologie moléculaire (Brown et al., 2015).

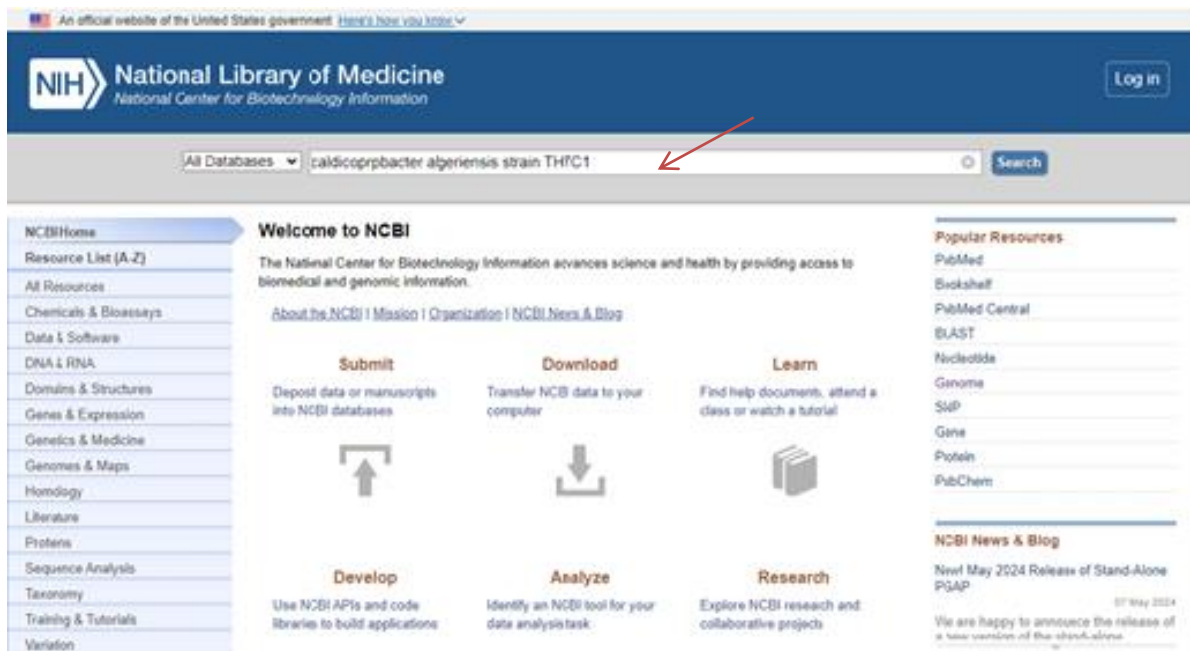


Figure 5: présenter une capture d'ecran de site Web du NCBI

TYGS (Type Strain Genome Server) Plateforme

TYGS (<http://tygs.dsmz.de>) est un serveur web à haut débit facilitant la classification des bactéries basée sur les séquences génomiques. Il est lié à une vaste base de données d'informations

génomiques et taxonomiques, aidant à la classification et à la taxonomie microbienne (**Göker et Meier-Kolthoff, 2019**).



Figure 6: présenter une capture d'écran de Plateforme TYGS

Outil d'annotation et de soumission rapide DDBJ DFAST

DDBJ Fast Annotation and Submission Tool (DFAST) (<https://dfast.ddbj.nig.ac.jp/>) est un pipeline d'annotation du génome bactérien intégré avec des méthodes d'évaluation de la qualité et de la taxonomie. DFAST est développé de manière à ce que toutes les procédures requises pour la soumission puissent être effectuées en ligne de manière transparente, ce qui permet de l'utiliser comme un espace de travail en ligne pour préparer les fichiers de soumission au Système de Soumission de Masse (MSS) de DDBJ (**Tanizawa et al., 2018**).

The screenshot shows the DFAST web interface in a browser. The address bar displays 'dfast.ddbj.nig.ac.jp/dfc/'. The page title is 'DFAST Prokaryotic genome annotation pipeline'. The main content area includes a 'Query File (Fasta format, up to 15Mbyte)' section with a 'Choose File' button and a 'No file chosen' status. To the right of this is a checkbox for 'Demo mode (Sample annotation for E.coli O26)'. Below this is a 'Name/Title for the Job' field with '(optional)' placeholder text. A blue arrow points to this field. Further down is a 'Mail Address' field with the placeholder 'E-mail notification will be sent to this address when the job is completed. (optional)'. Below the mail address are three checkboxes: 'Perform Taxonomy Check', 'Perform Completeness Check', and 'Perform GTDB Taxonomy Assignment'. At the bottom of the form is a '▼ Advanced Options' button and a blue 'Run' button.

Figure 7: présenter une capture d'ecran de DDBJ DFAST

Plateforme BV-BRC

La plateforme BV-BRC (<https://www.bv-brc.org>) fournit un accès à des ressources de bioinformatique bactérienne et virale, offrant un ensemble de données complet sur les microorganismes, en particulier ceux liés aux agents pathogènes et aux maladies infectieuses. Elle propose également des outils de calcul pour l'analyse des données, y compris des analyses prédictives utilisant des techniques d'intelligence artificielle (Olson et al., 2023).

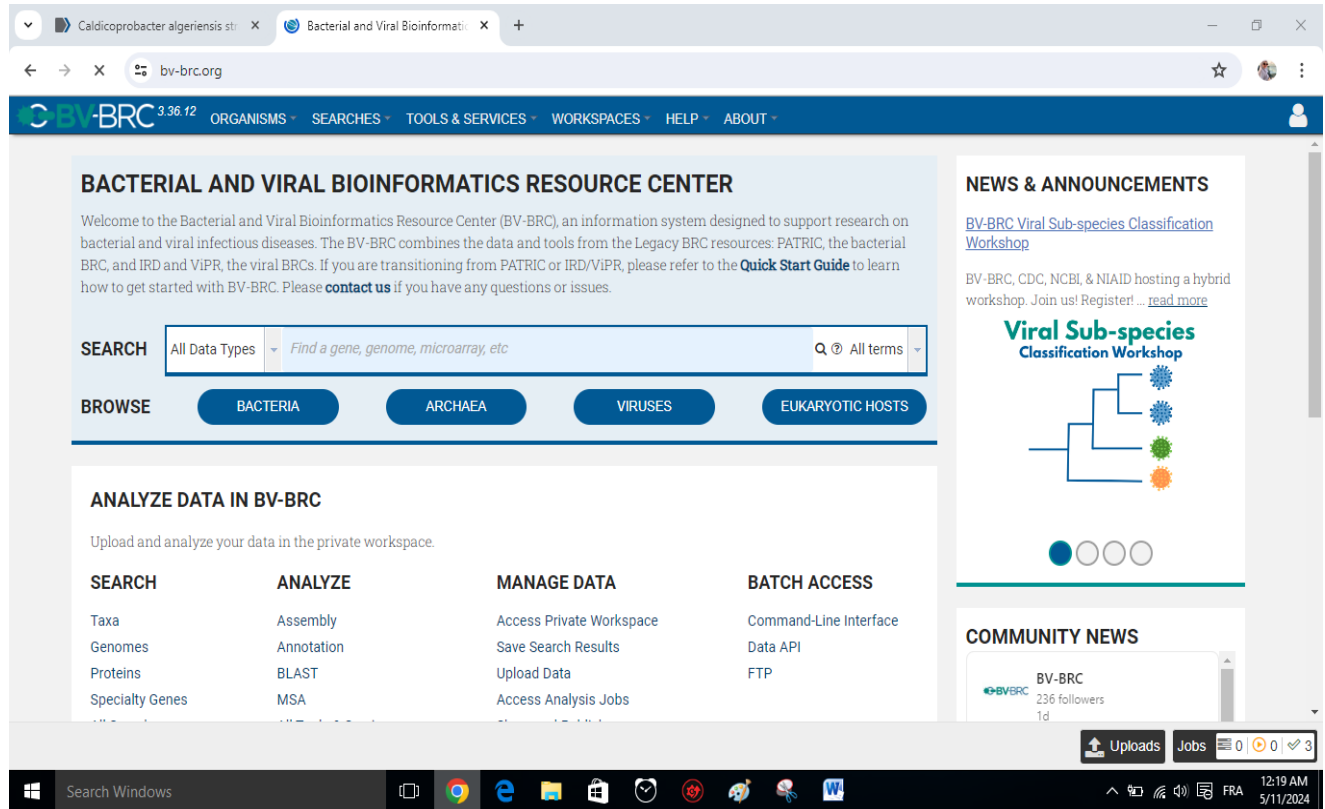


Figure 8: présenter une capture d'ecran de Plateforme BV-BRC

Serveur RAST

Le serveur RAST (<https://rast.nmpdr.org>) offre un service complet d'annotation pour identifier et annoter les gènes dans les génomes bactériens. Il classe les fonctions des gènes et les regroupe en sous-systèmes fonctionnels (COGs), fournissant des informations sur la fonctionnalité génomique (Aziz et al., 2008).

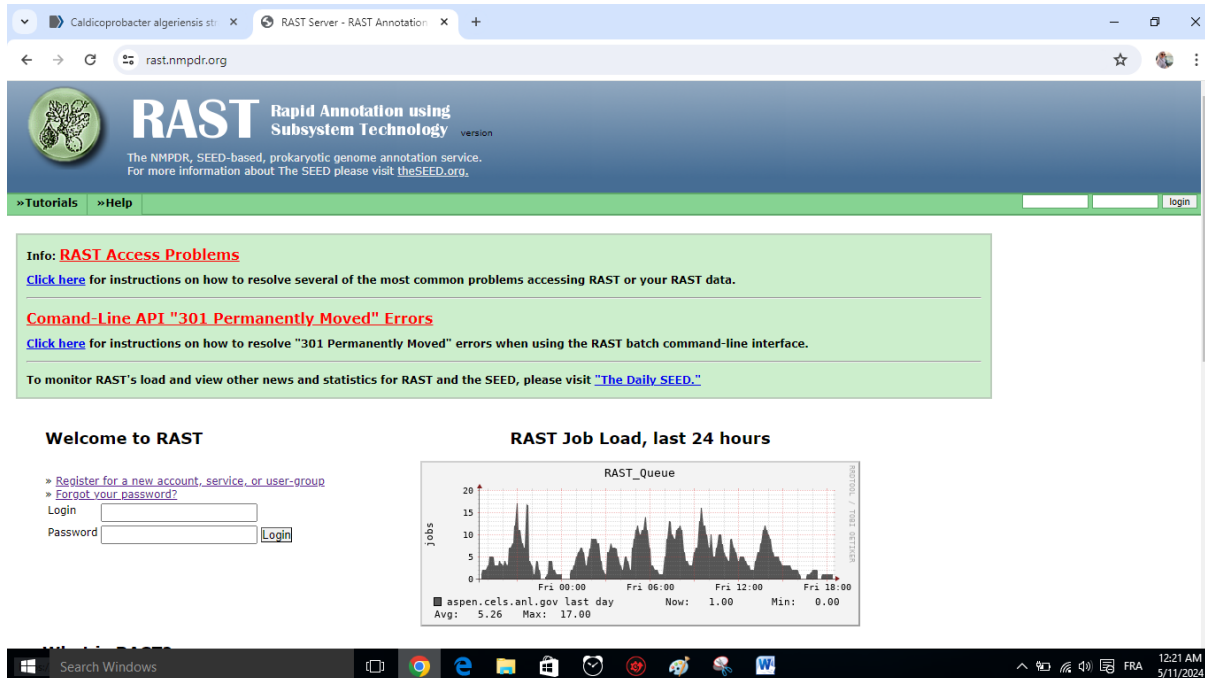


Figure 9: présenter une capture d'ecran de serveur RAST

Programme Prokka Serveur RAST

Le programme Prokka, intégré à la plateforme mondiale de bioinformatique Galaxy (<https://usegalaxy.eu>), analyse les séquences génomiques pour identifier les gènes avec précision. Il est particulièrement utile pour identifier les gènes liés à la virulence, à la pathogénicité et à l'infectiosité (Seemann et al., 2014).

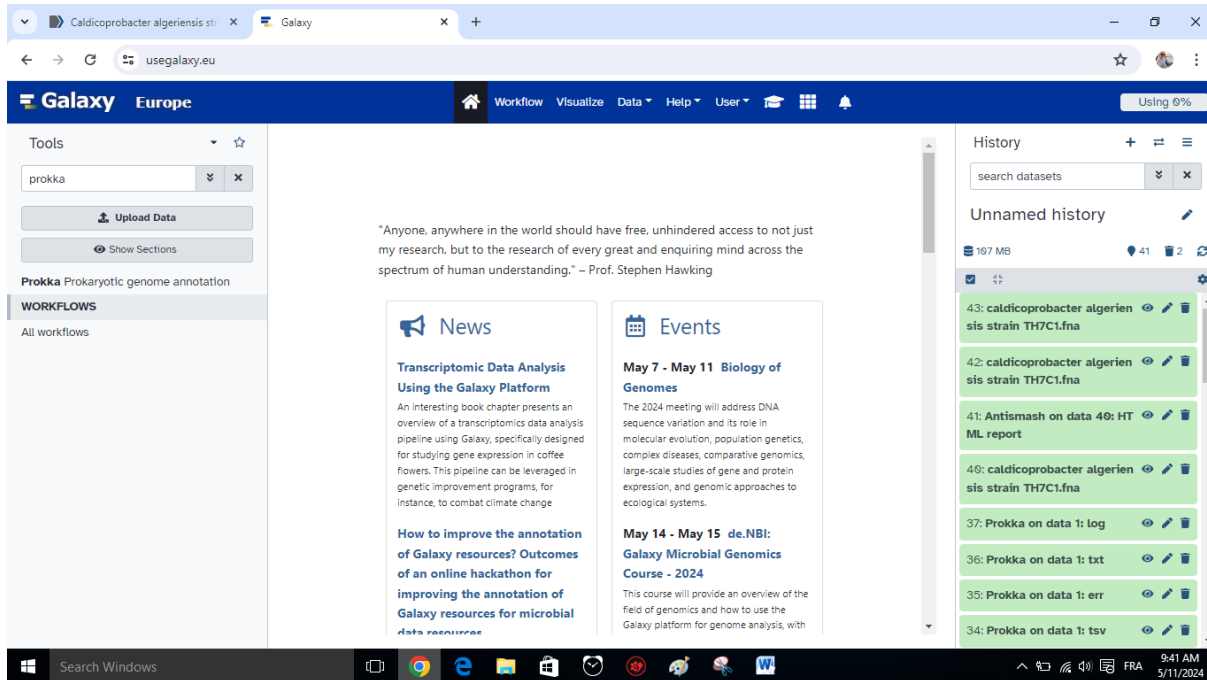


Figure 10 : présenter une capture d'ecran de *Programme Prokka Serveur RAST*

Serveur antiSMASH

Le serveur antiSMASH (<https://antismash.secondarymetabolites.org>) est un outil en ligne utilisé pour identifier des ensembles de gènes biosynthétiques codant pour des métabolites secondaires dans le génome complet de l'organisme étudié, facilitant l'exploration de la production de

métabolites secondaires (Blin et al., 2023).

The figure consists of two screenshots of the antiSMASH web interface. The top screenshot shows the main submission page. On the left, a sidebar displays server status: 'working', 22 running jobs, 0 queued jobs, and 1897081 jobs processed. The main area has tabs for 'Nucleotide input' and 'Results for existing job'. Below these are buttons for 'Load sample input' and 'Open example output'. A 'Notification settings' section includes an email field with 'a23@gmail.com' and a red arrow pointing to it. The 'Data input' section has 'Upload file' and 'Get from NCBI' buttons, followed by a text field containing 'caldicoprobacter algeriensis strain TH7C' and a red arrow pointing to it. Below this is a 'Browse' button and a checkbox for 'Upload extra annotations'. The 'Detection strictness' section shows a slider set to 'relaxed' with a red arrow pointing to it. The bottom screenshot shows the 'Extra features' section. It includes a 'Submit' button and a list of features: 'KnownClusterBlast' (checked), 'MIBiG cluster comparison' (unchecked), 'Cluster Pfam analysis' (unchecked), 'TFBS analysis' (checked), 'ClusterBlast' (unchecked), 'ActiveSiteFinder' (checked), 'Pfam-based GO term annotation' (unchecked), 'SubClusterBlast' (checked), 'RREFinder' (checked), and 'TIGRFam analysis' (unchecked). A note at the bottom states: 'Please be considerate in your use of antiSMASH. Help us keep antiSMASH available for everybody by limiting yourself to 5 concurrent jobs. Need to run more? See the antiSMASH guide for instructions for getting your own antiSMASH installation.'

Figure 11: présenter une capture d'ecran de *Serveur antiSMASH*

Méthode

Le fichier de séquence génomique complète de la souche bactérienne étudiée, à savoir *Caldicoprobacter algeriensis* souche TH7C1, a été récupéré de la base de données NCBI au format FASTA, avec l'adresse numérique du génome étant le numéro d'accèsion ASM2383852v1. Isolé d'une source chaude terrestre à Guelma, en Algérie, cet isolat bactérien a subi un séquençage génomique complet (WGS) utilisant la technologie Illumina NovaSeq, réalisé par le Laboratoire de Biotechnologie Microbienne, Enzymatique et Biomolécules du Centre de Biotechnologie de Sfax.

Après l'acquisition du génome, une série de processus et d'analyses informatiques a été menée en utilisant divers programmes et plateformes approuvés académiquement (Dif et Zitouni, 2023). L'approche méthodologique est décrite comme suit:

- Étude taxonomique : Le génome de la souche GMI12077 a été téléchargé sur la plateforme TYGS pour confirmer son espèce bactérienne, en tirant parti des algorithmes de la plateforme et de sa vaste base de données taxonomique pour une identification précise des espèces. Les valeurs d'hybridation ADN-ADN numérique (dDDH) ont été calculées à l'aide du calculateur de distance génomique.
- Traitement avec la plateforme BV-BRC : En utilisant le service d'annotation de cette plateforme, nous avons obtenu des informations sur les caractéristiques générales de la souche bactérienne et sa représentation circulaire.
- Traitement avec le serveur RAST : Ce serveur a facilité la classification fonctionnelle des gènes dans le génome de la souche bactérienne en sous-systèmes fonctionnels (COGs), fournissant des informations sur la fréquence des gènes dans chaque catégorie fonctionnelle.
- Traitement avec le programme Prokka : L'utilisation de ce programme a permis une exploration plus approfondie des gènes dans le génome, en mettant l'accent sur l'identification des gènes significatifs associés à la virulence, à la pathogénicité et à l'infectiosité.

- Traitement avec le serveur antiSMASH : Grâce aux résultats de ces traitements, des clusters génétiques au sein du génome responsables des métabolites secondaires ont été identifiés, éclairant ainsi les voies biochimiques et les fonctions potentielles.

A red decorative border with a double-line effect and rounded corners, enclosing the title text.

Résultats et Discussion

Résultats

1. Analyses phylogénétiques et phylogénomiques

En utilisant TYGS pour la classification taxonomique, la souche étudiée (TH7C1) a été assignée à l'espèce *Caldicoprobacter algeriensis* (**Figures 12, 13**). Cette vue interactive de l'arbre affiche des informations sur les données génomiques sous-jacentes, la formule de distance GBDP et l'algorithme utilisé. Des annotations supplémentaires sont affichées sur le côté droit du visualiseur, y compris les clusters d'espèces et de sous-espèces, la teneur en G+C du génome (en %), les valeurs δ , la longueur totale de la séquence (en pb) et le nombre de protéines. Les alignements comprennent 15 souches.

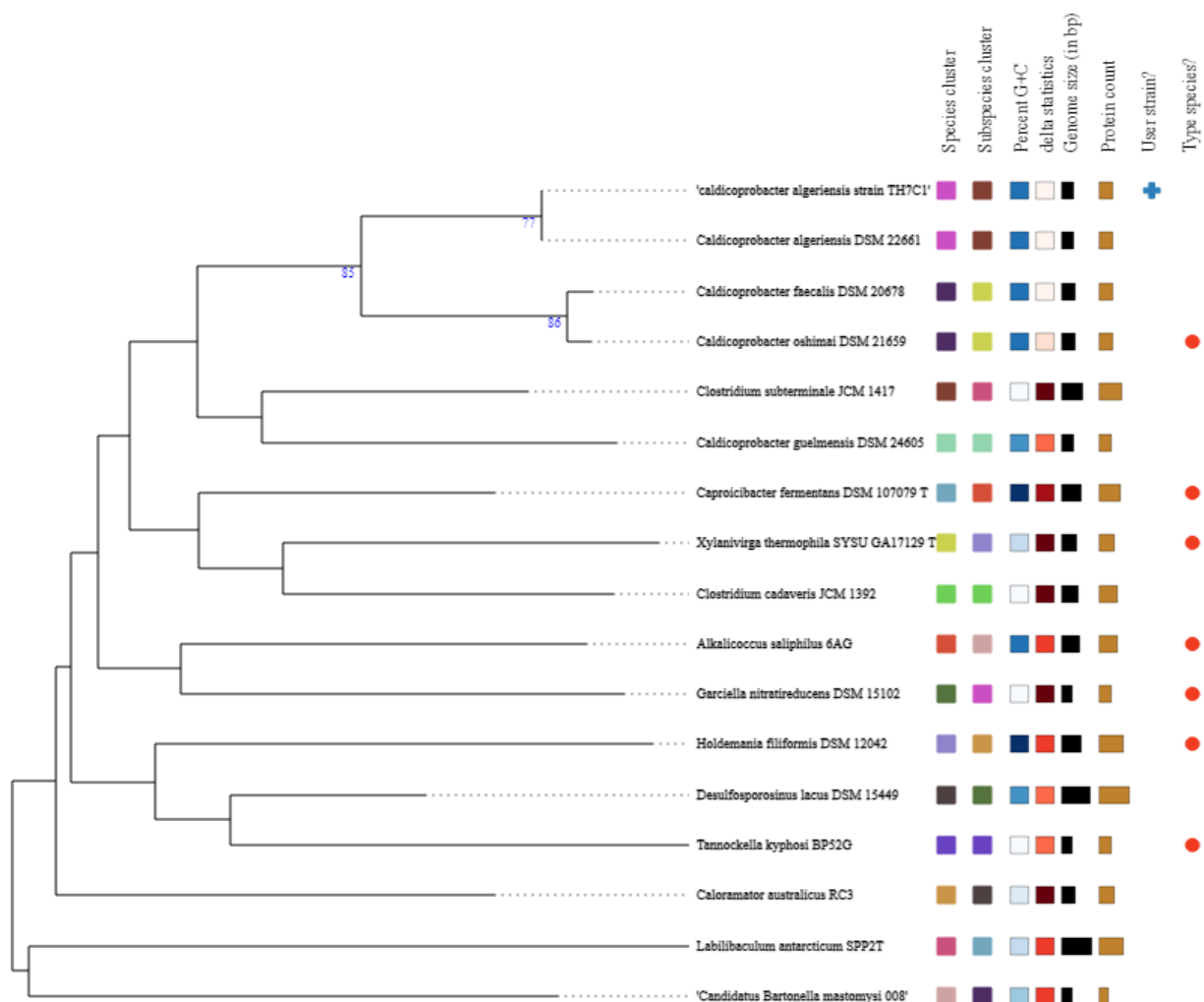


Figure 12. Arbre phylogénétique basé sur l'ADNr 16S généré par TYGS montrant la parenté de *Caldicoprobacter algeriensis* souche TH7C1 avec les souches types les plus proches, y compris celles qui n'ont pas encore été soumises à l'analyse de séquence génomique.

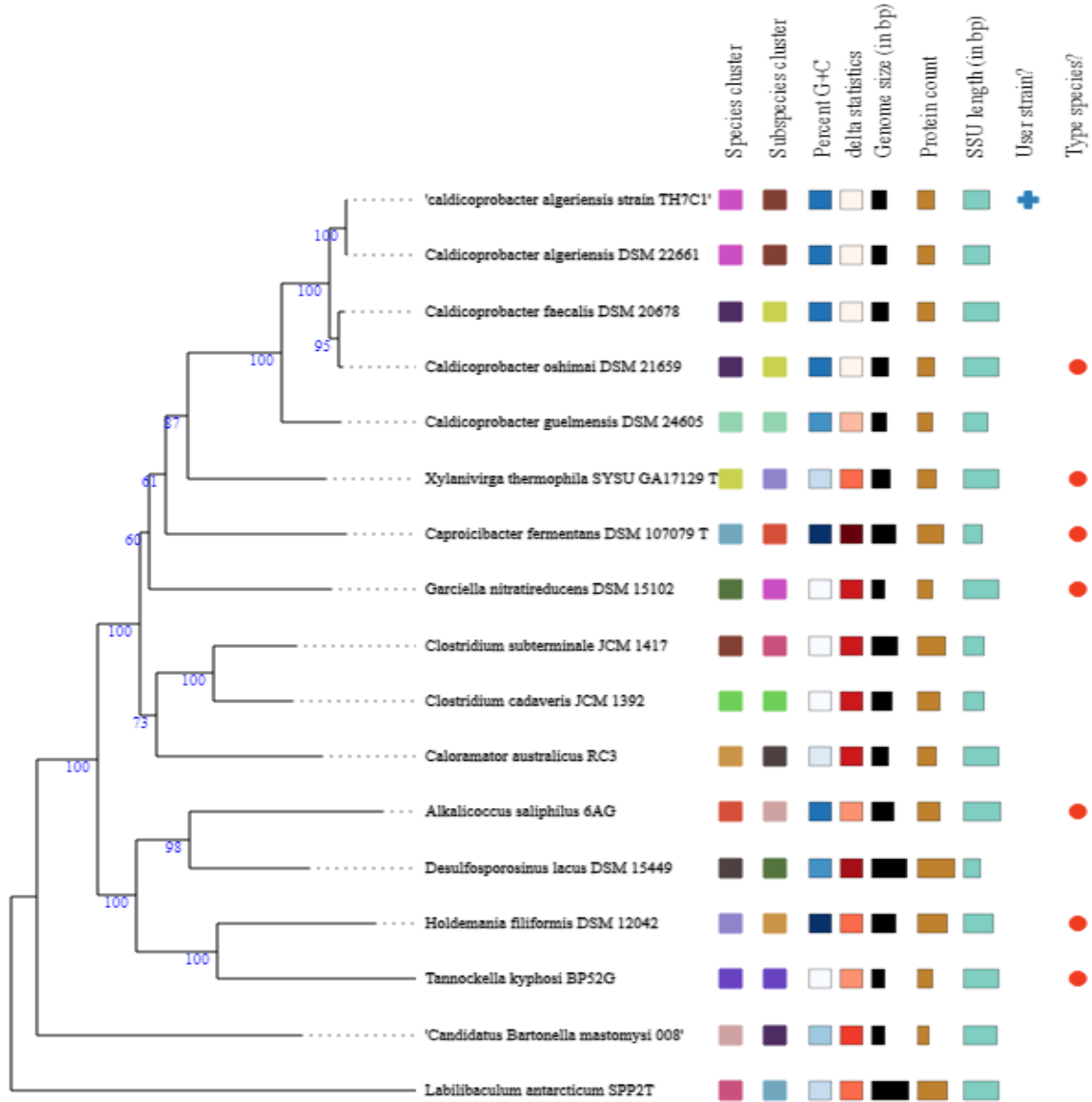


Figure 13 : Arbre phylogénomique basé sur les séquences génomiques dans l'arbre TYGS inféré avec FastME 2.1.6.1 (Meier-Kolthof et Göker, 2019) à partir de l'approche Genome BLAST Distance Phylogeny (GBDP); les distances sont calculées à partir des séquences génomiques. Les longueurs des branches sont échelonnées selon la formule de distance GBDP d5. Les nombres au-dessus des branches sont des valeurs de support pseudo-bootstrap GBDP > 60 % issues de 100 répliques, avec un support de branche moyen de 93,7 %. L'arbre a été enraciné au point médian (Meier-Kolthof et al., 2013).

2. Caractéristiques du génome

Le génome de *Caldicoprobacter algeriensis* souche TH7C1 est constitué de 2 535 023 pb avec une teneur en GC de 44,9 %. Le génome comporte 2 425 gènes codants, dont 50 pour les ARNt, 3 pour les ARNr, 5 pour les ARNnc et 1 région prophagique (**Tableau 03, Figure 14**).

Tableau 4: Caractéristiques du génome de *Caldicoprobacter algeriensis* souche TH7C1. Les caractéristiques génétiques sont le résultat de la sortie combinée de l'annotation DFAST.

Caractéristiques	Chromosome
Longueur totale (pb)	2,535,023
Nombre de séquences (Contigs)	45
Le contenu en GC (%)	44.9%
N50	121,172
Le taux de lacunes (%)	0.0%
Nombre de CDS	2,425
Nombre de ARNr	3
Nombre de ARNt	50
Nombre de CRISPR	1
Ratio de codage (%)	85.3%

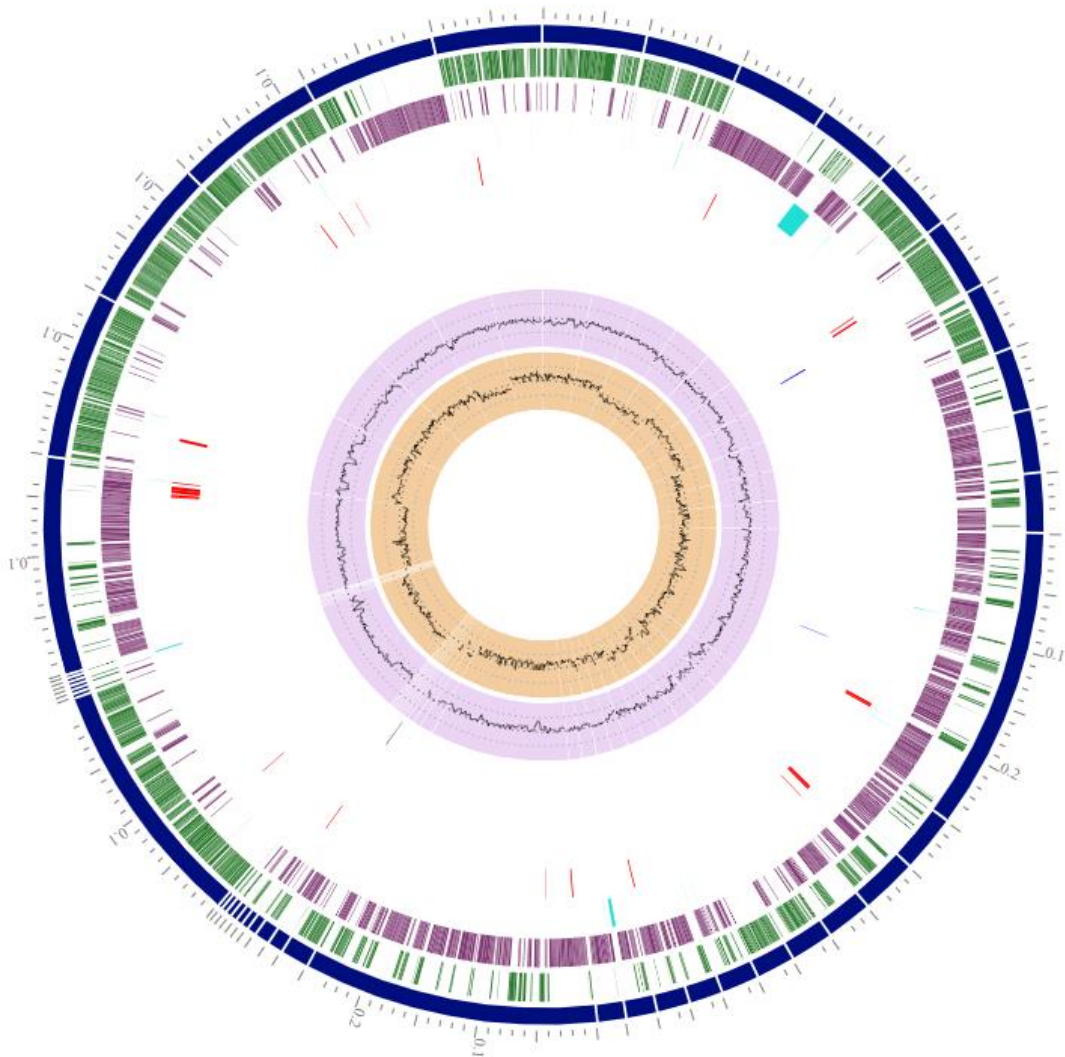


Figure 14. Vue circulaire du génome de la souche TH7C1. Les pistes sont affichées sous forme d'anneaux concentriques, du plus externe au plus interne : 1, position de référence dans le génome ; 2, position et ordre des 94 contigs assemblés ; 3, brins CDS à terme ; 4, CDS-brin inversé ; 5, les caractéristiques non codantes ; 6, gènes de résistance antimicrobienne ; 7, gènes des facteurs de virulence ; 8, gènes pour les transporteurs ; 9, cibles médicamenteuses ; 10, contenu GC ; et 11, asymétrie GC.

3. Groupes de protéines orthologues (COG)

Une annotation complète du génome a été réalisée en utilisant le serveur RAST, et un total de 365 sous-systèmes ont été classifiés avec une couverture de sous-systèmes de 28 % (**Figure 15**). Les caractéristiques des sous-systèmes appartenaient principalement aux acides aminés et dérivés (418 gènes) ; suivis des glucides (233 gènes) ; du métabolisme des protéines (214 gènes) ; des cofacteurs, vitamines, groupes prosthétiques et pigments (186 gènes) ; du transport membranaire (113 gènes) ; du métabolisme de l'ADN (112 gènes) ; de la respiration (111 gènes) ; des nucléosides et nucléotides (103 gènes) et de la réponse au stress (94 gènes). Il est à noter que cette souche possédait 94 gènes correspondant au sous-système de réponse au stress, incluant le stress oxydatif (50 gènes), la réponse au stress - sans sous-catégorie (20 gènes), le stress osmotique (13 gènes), la détoxification (5 gènes) et le stress périplasmique (6 gènes).

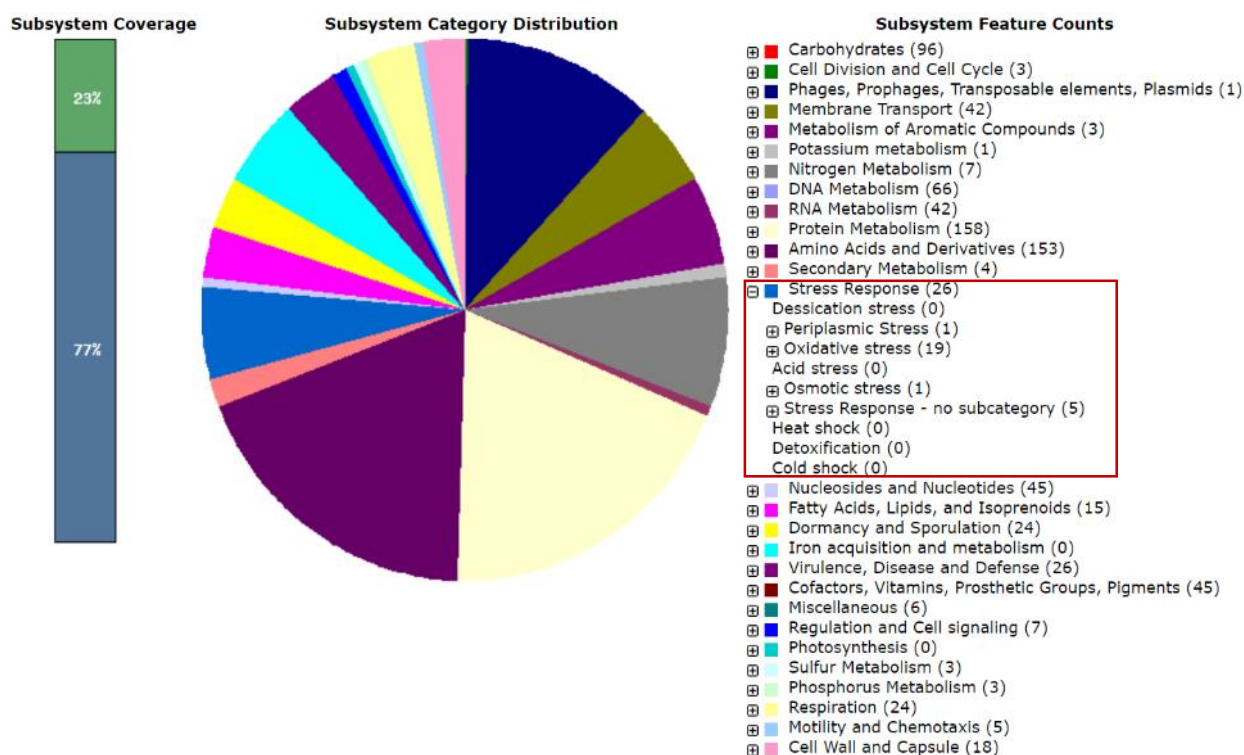


Figure 15: Vue d'ensemble des catégories de sous-systèmes de TH7C1. L'annotation du génome a été réalisée en utilisant le serveur RAST.

4. Éléments génétiques impliqués dans la résistance au stress de TH7C1, ainsi que ses capacités de colonisation du sol et des plantes.

4.1. Tolérance au stress osmotique

L'analyse du génome de la souche TH7C1 de *C. algeriensis* a révélé la présence de plusieurs gènes impliqués dans la tolérance au stress osmotique, notamment : les transporteurs de sodium et de potassium, le système de transport des osmoprotecteurs, la biosynthèse et le transport d'osmolytes (par exemple, la proline, le glutamate, la glutamine) (**Tableau 04**). L'abondance d'éléments génétiques impliqués dans la tolérance au stress osmotique dans le génome de la souche TH7C1 est en accord avec les données fonctionnelles montrant que la souche TH7C1 tolère des concentrations élevées de salinité.

Tableau 5: Gènes impliqués dans la tolérance au stress osmotique.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_01545	Réductase de gamma-glutamyl phosphate	Biosynthèse de la proline
BAODNHJJ_01544	Kinase de glutamate 5	
BAODNHJJ_02278	Réductase de pyrroline-5-carboxylate	
BAODNHJJ_01361	Synthase de glutamate [NADPH] petite chaîne	Biosynthèse du glutamate
BAODNHJJ_01230	Synthase de folylpolyglutamate	
BAODNHJJ_01005	Synthétase de glutamine	Biosynthèse de la glutamine
BAODNHJJ_01027		
BAODNHJJ_00370	Synthétase de NAD(+) dépendante de la glutamine	
BAODNHJJ_00372		
BAODNHJJ_00816	Transporteur de proline régulé par l'osmose OpuE	Système de transport des osmoprotecteurs
BAODNHJJ_01948	Protéine de liaison à l'ATP pour le transport de glutamine GlnQ	Transport de la glutamine
BAODNHJJ_00404	Antiporteur sodium-lithium/proton	

BAODNHJJ_00420	Sous-unité catalytique A de l'ATPase sodique de type V	Transporteurs de sodium
BAODNHJJ_00421	Sous-unité B de l'ATPase sodique de type V	
BAODNHJJ_00418	Sous-unité B de l'ATPase sodique de type C	
BAODNHJJ_00422	Sous-unité B de l'ATPase sodique de type D	
BAODNHJJ_00419	Sous-unité B de l'ATPase sodique de type G	
BAODNHJJ_00416	Sous-unité B de l'ATPase sodique de type K	
BAODNHJJ_00624	Protéine A de capture de potassium du système Ktr	Absorption du potassium
BAODNHJJ_00625	Protéine B de capture de potassium du système Ktr	

4.2. Résistance aux contraintes thermiques et formation de spores

Le génome de la souche TH7C1 de *C. algeriensis* contient plusieurs protéines de choc thermique (csp) et de choc thermique (hsp), ainsi que d'autres protéines chaperonnes impliquées dans la réponse au stress thermique (**Tableau 05**). De plus, le génome contient plusieurs chaperonnes impliquées dans la réponse au stress thermique, y compris dnaJ, dnaK, groES, groEL, ainsi que plusieurs autres gènes de petites protéines hsp.

Tableau 6: Gènes impliqués dans la résistance au stress thermique.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_00612	Protéine de choc thermique 18 kDa	Résistance au stress thermique
BAODNHJJ_00506	Protéine chaperonne ClpB	
BAODNHJJ_00855	Protéine chaperonne DnaJ	
BAODNHJJ_01263		
BAODNHJJ_00856	Protéine chaperonne DnaK	
BAODNHJJ_01265		
BAODNHJJ_00859	Protéine chaperonne de l'hème HemW	
BAODNHJJ_02400	Sous-unité ATPase de la protéase ATP dépendante ClpQ	

Un total de 172 gènes prévus pour être impliqués dans les différentes phases de la formation de spores et des processus de germination des spores ont été détectés dans le génome de TH7C1 (**Tableau 06**).

Tableau 7: Gènes impliqués dans la formation et la germination des spores.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_00121	Protéine de sporulation	Formation et germination des spores
BAODNHJJ_00386		
BAODNHJJ_00397		
BAODNHJJ_00433		
BAODNHJJ_00648		
BAODNHJJ_00950		
BAODNHJJ_01178		
BAODNHJJ_01191		
BAODNHJJ_01192		
BAODNHJJ_01203		
BAODNHJJ_01218		
BAODNHJJ_02271		
BAODNHJJ_02287		
BAODNHJJ_02290		
BAODNHJJ_02337		
BAODNHJJ_02343		
BAODNHJJ_02363		
BAODNHJJ_02364		
BAODNHJJ_02367		
BAODNHJJ_02343	Stage 0 sporulation protein	
BAODNHJJ_00950		
BAODNHJJ_02343		
BAODNHJJ_01178	Protéine de sporulation de stade II E	
BAODNHJJ_02337	Protéine de sporulation de stade II M	
BAODNHJJ_02363	Protéine de sporulation de stade III AF	

BAODNHJJ_02364	Protéine de sporulation de stade III AE	
BAODNHJJ_00121	Protéine de sporulation de stade IV A	
BAODNHJJ_02290	Protéine de sporulation de stade V AD	
BAODNHJJ_00648	Protéine de sporulation de stade V B	
BAODNHJJ_01191		
BAODNHJJ_01218	Protéine de sporulation de stade V D	
BAODNHJJ_01192	Protéine de sporulation de stade V T	
BAODNHJJ_00670	Protéine de germination des spores	
BAODNHJJ_01712		
BAODNHJJ_02196		
BAODNHJJ_02197		
BAODNHJJ_02253		
BAODNHJJ_02254		
BAODNHJJ_02255		
BAODNHJJ_02185	Protéine de la couche de spores	
BAODNHJJ_01754	Protéine de petite taille, soluble dans l'acide, A	
BAODNHJJ_01157	Protéine de maturation des spores B	
BAODNHJJ_01158	Protéine de maturation des spores A	
BAODNHJJ_01896	Lyase des photoproduits des spores	
BAODNHJJ_00396	Enzyme lytique du cortex des spores	
BAODNHJJ_01907		
BAODNHJJ_01206	Protéase de traitement du facteur sigma-E de sporulation	
BAODNHJJ_02344	Protéase SpoIVB	

4-2- Résistance aux métaux lourds

Les gènes de transport/résistance aux métaux lourds sont abondants dans le génome de la souche TH7C1 de *C. algeriensis* (**Tableau 07**). Cela inclut plusieurs gènes de transport (pour le zinc, le cobalt, le cuivre, le cadmium, le manganèse, l'arséniate, le chromate et le fluorure), ainsi que des

gènes codant pour des réductases d'arséniate et des transports de chromate. Le grand nombre de gènes de résistance aux métaux lourds dans le génome de la souche TH7C1 de *C. algeriensis* suggère que cette souche est capable de faire face à des niveaux élevés de métaux lourds.

Tableau 8: Gènes impliqués dans le transport et la résistance aux métaux lourds.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_01706	Transporteur de zinc ZupT	Transport de zinc
BAODNHJJ_00409	Transporteur de zinc ZupT	
BAODNHJJ_01767	Protéine de régulation métallospécifique du zinc	
BAODNHJJ_02431	ATPase de transport de cadmium, zinc et cobalt	Transport de cobalt-zinc-cadmium
BAODNHJJ_01289	Répresseur transcriptionnel sensible au cuivre CsoR	Résistance au cuivre
BAODNHJJ_01292	Chaperon de cuivre CopZ	
BAODNHJJ_01293	ATPase de type P exportatrice de cuivre	
BAODNHJJ_02102	Régulateur transcriptionnel de la résistance à l'arsenic ArsR1	
BAODNHJJ_02103	Protéine de résistance à l'arséniate Acr3	Réduction de l'arséniate
BAODNHJJ_01723	Réductase d'arséniate de glutarédoxine	
BAODNHJJ_00253	Protéine de transport de chromate putative	Transport de chromate
BAODNHJJ_00254	Protéine de transport de chromate	
BAODNHJJ_00963	Protéine du système d'efflux du manganèse MneP	Transport de manganèse
BAODNHJJ_01899	Transporteur de magnésium MgtE	Transporteur de magnésium
BAODNHJJ_00111	Transporteur de l'ion fluoride putatif CrcB	Transport de fluoride
BAODNHJJ_00112		

4.4. Résistance au stress oxydatif

Le génome de la souche TH7C1 de *C. algeriensis* contient de nombreux gènes impliqués dans la réponse au stress oxydatif, notamment un gène de superoxyde dismutase Cu-Zn et un régulateur de l'opéron du peroxyde (**Tableau 08**). Cette collection de gènes indique une forte adaptation et résistance aux espèces réactives de l'oxygène (ERO), qui sont couramment produites et accumulées par les bactéries dans des conditions stressantes et sont également connues pour être impliquées dans la réponse de défense des plantes.

Tableau 9: Gènes impliqués dans la protection contre le stress oxydatif.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_01964	Superoxyde dismutase [Cu-Zn]	Dégradation des radicaux d'anions superoxyde
BAODNHJJ_01057	Catalase au manganèse putative	Dégradation du peroxyde d'hydrogène
BAODNHJJ_01161	Thiol-disulfide oxydoréductase ResA	
BAODNHJJ_02404	Régulateur de l'opéron du peroxyde	
BAODNHJJ_02352	Synthase de diphosphate de farnésyle	
BAODNHJJ_00119	Acyltransférase de glycérol-3-phosphate	
BAODNHJJ_00351		

4.5. Acquisition d'azote, de soufre et de phosphore

La souche TH7C1 de *C. algeriensis* possède des gènes de réductase du nitrate et du nitrite, ainsi que plusieurs gènes impliqués dans le transport du nitrate/nitrite (**Tableau 09**).

Les gènes du système de transport du phosphate, ainsi que les gènes de la phosphatase alcaline, sont présents dans le génome de la souche TH7C1 (**Tableau 09**), et contribuent à ses principales capacités d'acquisition de phosphore organique.

Tableau 10: Gènes impliqués dans le métabolisme de l'azote, du soufre et du phosphore.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_01031	Sous-unité de transfert d'électrons de la réductase du nitrate assimilatoire	Réduction du nitrate dissimilatoire
BAODNHJJ_00439	Transporteur d'ammonium	Transporteur d'ammonium
BAODNHJJ_02050	Réductase de phosphoadénosine phosphosulfate	Réduction du sulfate
BAODNHJJ_02052		
BAODNHJJ_02004		
BAODNHJJ_01736	Protéine régulatrice transcriptionnelle de la synthèse de phosphatase alcaline SphR	Assimilation du phosphate
BAODNHJJ_02464	Phosphatase du phosphoglycolate	
BAODNHJJ_00197	Phosphatase CheY-P CheC	
BAODNHJJ_00411	Phosphatase protéique-arginine	
BAODNHJJ_00791	Phosphatase protéique PrpC	
BAODNHJJ_00982	Phosphatase tyrosine-protéine YwqE	
BAODNHJJ_01063	Phosphatase d'acide sucré	
BAODNHJJ_01081	Transporteur de glycérol-3-phosphate	Transport de phosphate
BAODNHJJ_01400	Protéine de perméase du système de transport du phosphate PstC	
BAODNHJJ_01401	Protéine de perméase du système de transport du phosphate PstA	

4.6. Acquisition de fer

Le génome de la souche TH7C1 de *C. algeriensis* contient des gènes impliqués dans le transport du fer, notamment la protéine de régulation de l'absorption du fer (BAODNHJJ_02397) et la protéine de liaison à l'ATP du transport d'enterobactine ferrique FepC (BAODNHJJ_02412).

4.7. Motilité, chimiotaxie et fixation aux surfaces végétales

Les opérons de biosynthèse et de régulation des flagelles ont été détectés dans le génome (Tableau 08). Plusieurs gènes de chimiotaxie, tels que cheABYWR et motAB, cinq protéines de chimiotaxie acceptant le méthyle et un homologue du transducteur aéro-tactique HemAT à base d'hème ont été détectés (Tableau 08). De plus, le génome de la souche TH7C1 de *C. algeriensis* contient de nombreux gènes impliqués dans la production d'exopolysaccharides, ce qui pourrait faciliter la fixation bactérienne aux racines des plantes. En fait, deux clusters contenant des gènes de biosynthèse des polysaccharides ont été trouvés (Tableau 08). De plus, plusieurs gènes impliqués dans la production d'acides téichoïques, qui jouent un rôle dans la formation de biofilms et la colonisation des racines, ont également été identifiés (Tableau 10).

Tableau 11: Gènes impliqués dans la chimiotaxie et la motilité.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_01954	Protéine de chimiotaxie CheW	Chimiotaxie
BAODNHJJ_00199	Protéine de chimiotaxie Che A	
BAODNHJJ_00209	Protéine de chimiotaxie Che Y	
BAODNHJJ_00214	Protéine de chimiotaxie PomA	
BAODNHJJ_00450	Protéine de chimiotaxie acceptant le méthyle 4	
BAODNHJJ_00542	Méthyltransférase de protéine de chimiotaxie	
BAODNHJJ_00544	Protéine de chimiotaxie CheA	
BAODNHJJ_00545	Protéine de chimiotaxie CheW	
BAODNHJJ_01954	Protéine de chimiotaxie CheW	
BAODNHJJ_02310	Méthyltransférase de protéine de chimiotaxie	
BAODNHJJ_00200	Protéine de freinage de flagelle YcgR	Biosynthèse des flagelles
BAODNHJJ_00201	Protéine déterminant le site de la flagelle YlxH	

RESULTATS

BAODNHJJ_00202	Protéine de biosynthèse de flagelle FlhF	
BAODNHJJ_00203	Protéine de biosynthèse de flagelle FlhA	
BAODNHJJ_00204	Protéine de biosynthèse de flagelle FlhB	
BAODNHJJ_00205	Protéine de biosynthèse de flagelle FlhR	
BAODNHJJ_00211	Protéine de commutation du moteur flagellaire FliM	
BAODNHJJ_00224	Protéine de commutation du moteur flagellaire FliG	
BAODNHJJ_00225	Protéine de l'anneau M flagellaire	
BAODNHJJ_00226	Protéine complexe du crochet-base du corps flagellaire FliE	
BAODNHJJ_00227	Protéine de la tige de base du flagelle FlgC	
BAODNHJJ_00228	Protéine de la tige de base du flagelle FlgB	
BAODNHJJ_00435	Protéine de la tige de base du flagelle FlgG	
BAODNHJJ_00436		
BAODNHJJ_01926	Protéine associée au crochet flagellaire 1	
BAODNHJJ_01929	Facteur d'assemblage flagellaire FliW	
BAODNHJJ_01931	Flagelline	
BAODNHJJ_01942	Protéine associée au crochet flagellaire 2	
BAODNHJJ_01943	Chaperonne de sécrétion flagellaire FliS	
BAODNHJJ_02305	Protéine B d'assemblage du lipopolysaccharide	Biosynthèse des polysaccharides
BAODNHJJ_00984	Protéine de biosynthèse présumée de polysaccharide capsulaire YwqC	
BAODNHJJ_02305	Protéine B d'assemblage du lipopolysaccharide	
BAODNHJJ_00975	Protéine de biosynthèse de l'alginate AlgA	Adhérence /
BAODNHJJ_01093	Protéine de biosynthèse des phénazines PhzD	Formation de biofilm
BAODNHJJ_00465	Polyisoprenyl-teichoic acid--peptidoglycan teichoic	Biosynthèse de l'acide téichoïque
BAODNHJJ_00465	acid transferase TagU	

4.8. Éléments génétiques impliqués dans les activités de promotion de la croissance des plantes de la souche *C. algeriensis* TH7C1

Plusieurs souches de *Bacillus*, y compris la souche TH7C1 de *C. algeriensis*, sont connues pour produire des auxines telles que l'AIA, qui jouent un rôle crucial dans la stimulation de la croissance des plantes et les interactions plantes-microbes (Tableau 09). Les polyamines telles que la putrescine, la spermine et la spermidine sont également importantes pour la promotion de la croissance des plantes. De plus, il a été démontré que la production de spermidine par les bactéries réduit l'activité du gène ACC oxydase du tabac (ACO1), responsable de la production d'éthylène, ce qui diminue les niveaux d'éthylène dans les plantes et affecte diverses interactions plantes-microbes. Plusieurs gènes impliqués dans le métabolisme et le transport des polyamines ont été identifiés dans le génome de la souche TH7C1 de *C. algeriensis*, notamment le gène *speB* codant pour l'agmatinase (impliquée dans la production de putrescine), quatre gènes *speE* codant pour la spermidine synthase et plusieurs gènes de transport des polyamines (*potABCD*, *pupP*) (Tableau 11), ce qui souligne l'importance du métabolisme des polyamines dans les capacités de stimulation de la croissance des plantes de la souche TH7C1 de *C. algeriensis*.

Tableau 12: Gènes impliqués dans la production/modulation des phytohormones.

Tag de localisation	Gène et Produit	Voie métabolique
BAODNHJJ_00658	Chaîne alpha de la tryptophane synthase	Production de tryptophane
BAODNHJJ_00659	Chaîne bêta de la tryptophane synthase	
BAODNHJJ_01588	Protéine de transport de la tryptophane putative	
BAODNHJJ_02274	Aspartate aminotransférase	Conversion du tryptophane en intermédiaires d'IAA
BAODNHJJ_00931	Aminotransférase d'acides aminés à chaîne ramifiée 2	
BAODNHJJ_00990	Histidinol-phosphate aminotransférase	
BAODNHJJ_01384	Aminotransférase d'acétylornithine	
BAODNHJJ_01747	Glutamine--fructose-6-phosphate aminotransférase [isomérisation]	
BAODNHJJ_02239	Histidinol-phosphate aminotransférase	
BAODNHJJ_00104	LL-diaminopimélate aminotransférase	

BAODNHJJ_00661	Indole-3-glycérol phosphate synthase	Transformation de l'indole-3-pyruvate en indole-3-acétaldéhyde
BAODNHJJ_01447	Carboxykinase du phosphoénolpyruvate [GTP]	
BAODNHJJ_00238	Aminotransférase de la sérine-pyruvate	
BAODNHJJ_01251	Alcool déshydrogénase aldéhyde	Conversion de l'indole-3-acétaldéhyde en IAA
BAODNHJJ_01336	Déshydrogénase de l'aspartate-semialdéhyde	
BAODNHJJ_00807	N-acétylmuramoyl-L-alanine amidase sle1	Conversion d'IAM en IAA
BAODNHJJ_00395	Décarboxylase de l'arginine	Biosynthèse de la putrescine
BAODNHJJ_01171		
BAODNHJJ_02326	Agmatinase	
BAODNHJJ_00106	Synthase de la N(4)-bis(aminopropyl)spermidine	
BAODNHJJ_02303	Proenzyme de la décarboxylase de la S-adénosylméthionine	Biosynthèse de la spermidine
BAODNHJJ_01673	Protéine de liaison à la spermidine/putrescine du périplasme	Transport de polyamines
BAODNHJJ_01675	Protéine de perméase du système de transport de la spermidine/putrescine PotB	
BAODNHJJ_01676	Protéine de liaison à l'ATP pour l'importation de spermidine/putrescine PotA	
BAODNHJJ_02187	Protéine de liaison à l'ATP pour l'importation de spermidine/putrescine PotA	

5. Les clusters de gènes de biosynthèse et les métabolites secondaires

L'analyse du génome de la souche TH7C1 à l'aide du logiciel antiSMASH a révélé la présence d'un cluster de gènes codant pour des ranthipeptides.



Figure 16: Les clusters de gènes de biosynthèse et les métabolites secondaires

: Les clusters de gènes de biosynthèse et les métabolites secondaires

Les ranthipeptides, définis comme des peptides contenant des radicaux non- α thioéther, sont une nouvelle classe de produits naturels appartenant à la super-famille des peptides synthétisés ribosomalement et modifiés post-traductionnellement (RiPP). Les peptides synthétisés ribosomalement et modifiés post-traductionnellement (RiPPs) sont une classe de produits naturels présentant des caractéristiques structurales et des activités biologiques diverses, comprenant des propriétés antimicrobiennes, antitumorales, immunosuppressives et antinociceptives.

La souche TH7C1 de *C. algeriensis* est une PGPR polyvalente et résistante au stress, avec un grand potentiel pour l'agriculture et la biotechnologie. Son séquençage génomique a révélé la présence de multiples gènes impliqués dans des fonctions importantes concernant la colonisation du sol et des plantes, la promotion de la croissance des plantes et d'autres caractéristiques pertinentes pour l'intérêt biotechnologique (par exemple, la résistance au stress osmotique, thermique, oxydatif, nitrosatif et aux métaux lourds ; le métabolisme des glucides, des acides organiques, des acides aminés, des acides phénoliques et des xénobiotiques, y compris les stéroïdes, les terpénoïdes, les herbicides, les insecticides, les nitronates ; les activités lytiques et antagonistes médiées par des enzymes lytiques, des bactériocines, des phosphonates ; la modulation des COV et de multiples capacités de biosynthèse/modulation des hormones végétales). En fin de compte, les données obtenues dans cette étude apportent de nouveaux éclairages sur les propriétés génomiques qui régissent les interactions bénéfiques réussies entre des micro-organismes du sol importants comme *P. putida* et les plantes, ainsi que leur utilisation potentielle dans le développement de nouvelles applications agricoles et biotechnologiques durables.

Discussion

Discussion :

L'augmentation de la population entraîne une demande croissante de nourriture conforme aux normes de sécurité alimentaire.

L'agriculture est considérée comme le principal moyen de répondre à ces besoins, mais les interventions humaines et les déséquilibres naturels provoquent des phénomènes tels que la sécheresse, la salinité, la pollution et la propagation de maladies.

Utilisation de bactéries favorisant la croissance des plantes (PGPR) :

Les bactéries favorisant la croissance des plantes constituent une alternative écologique à l'amélioration de la santé des plantes.

Ces bactéries améliorent la croissance des plantes en facilitant l'absorption des nutriments et la production d'hormones bénéfiques, ainsi qu'en tolérant la pollution et la salinité.

Séquençage et analyse génétique de nouvelle génération :

La technologie de séquençage de nouvelle génération (NGS) est utilisée pour identifier les traits génétiques et interpréter les données génomiques.

L'analyse des génomes permet d'identifier des gènes ayant une activité bénéfique pour le développement de solutions agricoles durables.

Analyse de la souche *Caldicoprobacter algeriensis* TH7C1 :

La souche a été isolée des eaux thermales d'Algérie et analysée à l'aide de diverses techniques, telles que RAST, Prokka et antiSMASH.

Le génome comprend 2,5 millions de pb, avec une teneur en GC de 44,9 % et contient 2 425 gènes codants.

Résistance au stress environnemental : La souche contient des gènes de résistance à la salinité, à la chaleur et à l'oxydation, ce qui la rend adaptée aux applications environnementales et industrielles.

Gérer les métaux lourds : La souche montre une grande capacité à gérer les sols contaminés, mettant en évidence son potentiel en phytoremédiation.

Production de prébiotiques : La souche produit des substances qui stimulent la croissance des plantes, telles que les auxines et les polyamines, qui améliorent les interactions microbiennes et la croissance des plantes.

Potentiel bactériostatique : Les résultats ont révélé 172 gènes liés à la formation de spores, améliorant l'aptitude de la souche à diverses applications.

Perspectives futures de l'étude :

Les résultats de l'étude du génome bactérien de la souche TH7C1 de *C. algeriensis* représentent un point de départ important pour explorer les possibilités d'application de cette bactérie dans les domaines environnementaux, agricoles et industriels. Compte tenu de ses caractéristiques génétiques avancées, cette race est prometteuse pour réaliser de nombreuses avancées. Parmi les perspectives d'avenir les plus importantes que l'on puisse anticiper figurent les suivantes :

1. Améliorer la production agricole durable

La souche *C. algeriensis* TH7C1 peut être utilisée dans le cadre de techniques d'agriculture biodynamique grâce à sa capacité à favoriser la croissance des plantes par la production de phytohormones (telles que les auxines et les polyamines) et à améliorer la tolérance aux stress environnementaux, tels que la salinité et les températures élevées.

Les bactéries aident les plantes à tolérer les polluants métalliques grâce à des mécanismes d'absorption et d'atténuation des métaux lourds, renforçant ainsi leur potentiel d'assainissement des sols contaminés.

2. Environnement et traitement de la pollution

La souche TH7C1 offre des solutions efficaces pour répondre aux problèmes environnementaux, tels que la pollution par les métaux lourds, grâce à des gènes de résistance à des métaux tels que le zinc, le cuivre et le cadmium.

Leur utilisation dans des projets d'assainissement de l'environnement pour les sols contaminés constitue une approche innovante et efficace pour améliorer la qualité des sols, permettant des utilisations plus larges des sols récupérés.

3. Bioindustries et biotechnologies

La grande adaptabilité de la souche aux températures élevées grâce aux protéines résistantes à la chaleur, telles que DNAj et GroES, ouvre la voie à son utilisation dans des applications industrielles nécessitant une résistance à la chaleur telles que la production d'enzymes.

La souche est une source riche pour la production de molécules bioactives telles que des peptides naturels (Ranthipeptides) qui possèdent des propriétés antimicrobiennes et possèdent de multiples potentiels industriels.

4. Agriculture intégrée et durable

Des bioproduits basés sur cette souche peuvent être développés pour être utilisés comme stimulants naturels de la croissance des plantes ou dans le cadre de stratégies de lutte contre le stress environnemental des plantes.

Se concentrer sur l'intégration de la bactérie *C. algeriensis* et sur les nouvelles technologies peut réduire la dépendance aux engrais chimiques, améliorant ainsi la durabilité environnementale.

5. Développer les études et la recherche en génomique

Une compréhension approfondie des gènes responsables de la résistance au stress environnemental et de la tolérance aux conditions difficiles peut contribuer au génie génétique ciblé pour améliorer les micro-organismes ou d'autres plantes.

Le recours aux technologies de séquençage génomique (NGS) de nouvelle génération constitue une base pour développer de meilleures approches pour exploiter le potentiel biologique de cette souche.

6. Possibilités d'étudier la formation des spores

Les recherches futures sur les mécanismes de sporulation de cette souche ouvrent des perspectives d'application dans les domaines agricoles et industriels, où elle améliore la résistance aux conditions environnementales difficiles et la longue durée de survie.

Conclusion

Conclusion

En conclusion, cette étude approfondie souligne l'importance cruciale des microorganismes, en particulier les Rhizobactéries Promotrices de la Croissance des Plantes (PGPR), dans le développement d'une agriculture durable et résiliente face aux défis environnementaux et agricoles contemporains. Avec l'accroissement continu de la population mondiale et la pression grandissante sur les ressources agricoles, des solutions innovantes sont nécessaires pour garantir la sécurité alimentaire tout en minimisant l'impact sur les écosystèmes. Les effets dévastateurs de la salinisation des sols, de la pollution chimique excessive et de la dégradation des terres fertiles mettent en péril la productivité agricole mondiale. Dans ce contexte, l'utilisation des PGPR offre une alternative écologique prometteuse pour améliorer la croissance des plantes, en renforçant leur résilience face aux stress abiotiques tels que la salinité et la sécheresse.

L'analyse génomique de la souche *Caldicoprobacter algeriensis* TH7C1 a révélé des traits génétiques spécifiques qui facilitent la tolérance aux conditions environnementales difficiles et stimulent la croissance des plantes. Ces caractéristiques en font un candidat idéal pour une application dans les zones arides et semi-arides où l'agriculture est particulièrement affectée par les conditions climatiques extrêmes. De plus, la capacité des PGPR à interagir de manière bénéfique avec les plantes sans provoquer de pollution supplémentaire, grâce à la réduction de l'utilisation d'engrais et de pesticides, représente un pas en avant vers une agriculture plus respectueuse de l'environnement.

Enfin, les découvertes réalisées dans cette étude ouvrent de nouvelles perspectives pour l'amélioration des rendements agricoles, tout en contribuant à la protection de l'écosystème. En intégrant les avancées en matière de biotechnologie et d'analyses génomiques, les chercheurs et agriculteurs peuvent désormais envisager des méthodes plus efficaces et plus durables pour répondre aux besoins croissants en nourriture à l'échelle mondiale. À long terme, l'utilisation des PGPR pourrait transformer les pratiques agricoles traditionnelles, en offrant des solutions écologiques qui contribuent à la santé des sols et à la productivité des cultures, tout en réduisant la dépendance aux intrants chimiques. Cela souligne l'importance continue de la recherche et de l'innovation pour faire face aux défis alimentaires du futur.

Références

Références:

- Alexandratos, N. (2012). *World Agriculture towards 2030/2050: The 2012 revision*.
- Amador, L. A. A. (2019). *Étude bioinformatique des génomes de Porphyromonas*. 274.
- Benson DA, Karsch-Mizrachi I, Lipman DJ . (2008). Nucleic Acids Res. *Sayers EW. GenBank*, p. 53.
- Bouanane-Darenfed et al. (2011, OCT 28). *Caldicoprobacter algeriensis* sp. nov. a new thermophilic anaerobic, xylanolytic bacterium isolated from an Algerian hot spring. *Curr. Microbiol*, p. 14.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+ : architecture et applications. *Bioinformatique BMC*, 10, 1-9.
- de Sousa, D. G., Harvey, L. A., Dorsch, S., & Glinsky, J. V. (2018). Interventions involving repetitive practice improve strength after stroke : A systematic review. *Journal of Physiotherapy*, 64(4), 210-221. <https://doi.org/10.1016/j.jphys.2018.08.004>
- Deng, Y., Chang, L., Yang, M., Huo, M., & Zhou, R. (2016). Gender differences in emotional response : Inconsistency between experience and expressivity. *PLoS ONE*, 11(6). <https://doi.org/10.1371/journal.pone.0158666>
- Dif, G., Belaouni, H. A., Goudjal, Y., Yekkour, A., Djemouai, N., & Zitouni, A. (2021). Potential for plant growth promotion of *Kocuria arsenatis* Strain ST19 on tomato under salt stress conditions. *South African Journal of Botany*, 138, 94-104.
- Dif, G., Belaouni, H. A., Yekkour, A., Goudjal, Y., Djemouai, N., Peñázová, E., ... & Zitouni, A. (2022). Performance of halotolerant bacteria associated with Sahara-inhabiting halophytes *Atriplex halimus* L. and *Lygeum spartum* L. ameliorate tomato plant growth and tolerance to saline stress: from selective isolation to genomic analysis of potential determinants. *World Journal of Microbiology and Biotechnology*, 38(1), 16.
- Drouin, G., Rioux, V., & Legrand, P. (2019). The n-3 docosapentaenoic acid (DPA) : A new player in the n-3 long chain polyunsaturated fatty acid family. *Biochimie*, 159, 36-48. <https://doi.org/10.1016/j.biochi.2019.01.022>
- Edgar, R. C. (2004). MUSCLE : alignement de séquences multiples avec une grande précision et un débit élevé. *Recherche sur les acides nucléiques*, 32(5), 1792-1797.
- El Rhaffouli, H., El Fahime, E. M., Laraoui, A., Bajjou, T., Melloul, M., Obeid, S., Fathallah, L., & Lahlou-Amine, I. (2014). Evolution of the Hemagglutinin Gene of the Influenza A(H1N1)pdm09 Virus in Morocco During Two Influenza Seasons 2009–2011. *Current Microbiology*, 68(3), 372-380. <https://doi.org/10.1007/s00284-013-0463-x>
- Figueira Aburjaile, F., Viana, M. V., Cerqueira, J. C., de Jesus, L., Da Silva, T., Carvalho, R., & Azevedo, V. (2021). Research Article Probiotic potential of novel Brazilian *Lactobacillus crispatus* strains. *Genetics and Molecular Research*, 20. <https://doi.org/10.4238/gmr18900>
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (New York, N.Y.)*, 269(5223), 496-512. <https://doi.org/10.1126/science.7542800>
- Goloboff, P. A., Farris, J. S. et Nixon, K. C. (2008). TNT, un programme gratuit d'analyse phylogénétique. *Cladistique*, 24(5), 774-786.
- Goodfellow et al. (2011). Literacy, literacies, and the digital in higher education.

- Gupta, S., Davoodi, H., & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Economics of Governance*, 3(1), 23-45.
- Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K. et Schuster, S. C. (2005). Phylogénie procaryote du génome entier. *Bioinformatique*, 21(10), 2329-2335.
- Hodkinson, B. P., & Grice, E. A. (2015). Next-Generation Sequencing : A Review of Technologies and Tools for Wound Microbiome Research. *Advances in Wound Care*, 4(1), 50-58. <https://doi.org/10.1089/wound.2014.0542>
- Ilie, M., Marinescu, F., Ghita, G., Deak, G., Tanase, S., & Raischi, M. (2014). Assessment of heavy metal in water and sediments of the Danube River. *Journal of environmental protection and ecology*, 15, 825.
- Javier-Menes R, Muxi L. (2002). *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as electron acceptor, and emended description of the genus *Anaerobaculum*. *Int J Syst Evol Microbiol*. pp. 157-164.
- Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., & He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, 7(12), 885-887. <https://doi.org/10.1038/nchembio.687>
- Joseph Felsenstein. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, pp. 783-791.
- Kreft, Ł., Botzki, A., Coppens, F., Vandepoele, K., & Van Bel, M. (2017). PhyD3 : une visionneuse d'arbres phylogénétiques avec prise en charge étendue du phyloXML pour la visualisation de données de génomique fonctionnelle. *Bioinformatique*, 33(18), 2946-2947.
- Kumar, V., Singh, S., Singh, J., & Upadhyay, N. (2015). Potential of Plant Growth Promoting Traits by Bacteria Isolated from Heavy Metal Contaminated Soils. *Bulletin of Environmental Contamination and Toxicology*, 94(6), 807-814. <https://doi.org/10.1007/s00128-015-1523-7>
- Lacoste, C., Fabre, A., Pécheux, C., Lévy, N., Krahn, M., Malzac, P., Bonello-Palot, N., Badens, C., & Bourgeois, P. (2017). Le séquençage d'ADN à haut débit en pratique clinique. *Archives de Pédiatrie*, 24(4), 373-383. <https://doi.org/10.1016/j.arcped.2017.01.008>
- Le Mignon, G., Blum, Y., Demeure, O., Diot, C., Bihan-Duval, E., Roy, P., & Lagarrigue, S. (2010). Apports de la génomique fonctionnelle à la cartographie fine de QTL. *INRAE Productions Animales*, 23. <https://doi.org/10.20870/productions-animales.2010.23.4.3313>
- Lefort, V., Desper, R. et Gascuel, O. (2015). FastME 2.0 : un programme d'inférence phylogénétique complet, précis et rapide basé sur la distance. *Biologie moléculaire et évolution*, 32(10), 2798-2800.
- Maune MW, Tanner RS . (2012). Description of *Anaerobaculum hydrogeniformans* sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus *Anaerobaculum*. *Int J Syst Evol Microbiol* , pp. 832-838.
- Meier-Kolthoff, J. P. et Göker, M. (2019). TYGS est une plateforme automatisée à haut débit pour une taxonomie génomique de pointe. *Nat Commun* 10 : 2182.
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. et Göker, M. (2013). Délimitation des espèces basée sur la séquence du génome avec des intervalles de confiance et des fonctions de distance améliorées. *Bioinformatique BMC*, 14, 1-14.

- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14, 60. <https://doi.org/10.1186/1471-2105-14-60>
- Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., ... et Klenk, H. P. (2014). Séquence complète du génome du DSM 30083 T, de la souche type (U5/41 T) d'*Escherichia coli*, et d'une proposition de délimitation des sous-espèces dans la taxonomie microbienne. *Normes en sciences génomiques*, 9, 1-19.
- Meinel, S., Datta, A., Kamali, S., & Rashed, A. (2017). Phenomenology of $\Lambda_{\text{b}} \rightarrow \Lambda_{\text{c}} \tau \overline{\nu}_{\tau}$ using lattice QCD calculations. *Journal of High Energy Physics*, 2017(8), 131. [https://doi.org/10.1007/JHEP08\(2017\)131](https://doi.org/10.1007/JHEP08(2017)131)
- Meldrum, R. C., Petkovsek, M. A., Boutwell, B. B., & Young, J. T. N. (2017). Reassessing the relationship between general intelligence and self-control in childhood. *Intelligence*, 60, 1-9. <https://doi.org/10.1016/j.intell.2016.10.005>
- Mercier, J. (2017). *Logique paracohérente pour l'annotation fonctionnelle des génomes au travers de réseaux biologiques*. 186.
- Metzger, B., Souzy, M., Lhuissier, H., Méheust, Y., & Le Borgne, T. (2020). Velocity distributions, dispersion and stretching in three-dimensional porous media. *Journal of Fluid Mechanics*, 891, A16. <https://doi.org/10.1017/jfm.2020.113>
- Ojuederie, O. B., & Babalola, O. O. (2017). Microbial and Plant-Assisted Bioremediation of Heavy Metal Polluted Environments : A Review. *International Journal of Environmental Research and Public Health*, 14(12), 1504. <https://doi.org/10.3390/ijerph14121504>
- Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S. et Phillippy, A. M. (2016). Mash : estimation rapide de la distance du génome et du métagénome à l'aide de MinHash. *Biologie du génome*, 17, 1-14.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R. L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., & Buell, C. R. (2007). The TIGR Rice Genome Annotation Resource : Improvements and new features. *Nucleic Acids Research*, 35(suppl_1), D883-D887. <https://doi.org/10.1093/nar/gkl976>
- Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M. et Stamatakis, A. (2010). Combien de répétitions bootstrap sont nécessaires ?. *Journal de biologie computationnelle*, 17(3), 337-354.
- Pericard, P. (2017). *Algorithmes pour la reconstruction de séquences de marqueurs conservés dans des données de métagénomique*.
- Ponsart, C., Le Bourhis, D., Knijn, H., Fritz, S., Guyader-Joly, C., Otter, T., Lacaze, S., Charreaux, F., Schibler, L., Dupassieux, D., & Mullaart, E. (2014). Reproductive technologies and genomic selection in dairy cattle. *Reproduction, Fertility and Development*, 26(1), 12. <https://doi.org/10.1071/RD13328>
- Poux S & Gaudet P. (2017). Best Practices in Manual Annotation with the Gene Ontology. *Methods in Molecular Biology (Clifton, N.J.)*. https://doi.org/10.1007/978-1-4939-3743-1_4
- Rees GN, Patel BKC, Grassia GS, Sheehy AJ. (1997). *Anaerobaculum thermoterrenum* gen. nov., sp. nov., a novel thermophilic bacterium which ferments citrate. *Int J Syst Bacteriol*, pp. 150-154.

- Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marçais, G., Pop, M., & Yorke, J. A. (2012). GAGE : A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22(3), 557-567. <https://doi.org/10.1101/gr.131383.111>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467.
- Slobodkina GB, Reysenbach A-L, Panteleeva AN. et al. (2012). *Deferrisoma camini* gen. nov., sp. nov., a moderately thermophilic, dissimilatory iron(III)-reducing bacterium from a deep-sea hydrothermal vent that forms a distinct phylogenetic branch in the Deltaproteobacteria. *Int J Syst Evol Microbiol*, p. 63.
- Soustschek E, Winter J, Schindler F, Kandler O. (1984, OCT). *Acetomicrobium flavidum*, gen. nov., sp. nov, a thermophilic anaerobic bacterium from sewage sludge, forming acetate, CO₂ and H₂ from glucose. *Syst Appl Microbiol . Systematic and Applied Microbiology*, pp. 377-390.
- Stamatakis, A. (2014). RAxML version 8 : un outil d'analyse phylogénétique et de post-analyse de grandes phylogénies. *Bioinformatique*, 30(9), 1312-1313.
- Swofford, D. L. (2003). PAUP^{^*} Analyse phylogénétique par parcimonie (^{^*} et autres méthodes). Variante 4. <http://paup.csit.fsu.edu/>.
- Tanizawa, Y., Fujisawa, T. et Nakamura, Y. (2018). DFAST : un pipeline flexible d'annotation du génome procaryote pour une publication plus rapide du génome. *Bioinformatique*, 34(6), 1037-1039.
- Thakur, L., Singh, K. K., Shanker, V., Negi, A., Jain, A., Matlashewski, G., & Jain, M. (2018). Atypical leishmaniasis : A global perspective with emphasis on the Indian subcontinent. *PLOS Neglected Tropical Diseases*, 12(9), e0006659. <https://doi.org/10.1371/journal.pntd.0006659>
- Therrien-Laperrière, S. (2018). *Développement d'un outil bio-informatique pour l'annotation des associations entre gènes et métabolites basée sur les voies métaboliques*. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/20735>
- Vannier, T. (2017). *Dynamique de la structure des génomes et de leur biogéographie dans l'océan: analyses comparatives des données métagénomiques du projet Tara Oceans pour l'étude de la microalgue Bathycoccus et des communautés planctoniques globales* (Doctoral dissertation, Université Paris Saclay (COMUE)).
- Vincent, A. T., Derome, N., Boyle, B., Culley, A. I., & Charette, S. J. (2017). Next-generation sequencing (NGS) in the microbiological world : How to make the most of your money. *Journal of Microbiological Methods*, 138, 60-71. <https://doi.org/10.1016/j.mimet.2016.02.016>
- Washington, DC, USA, A. S. for M. (s. d.). Retraction for Bai et al., “DNA Methyltransferase 3b Regulates Nerve Growth Factor-Induced Differentiation of PC12 Cells by Recruiting Histone Deacetylase 2”. *Molecular and Cellular Biology*, 42(5), e00545-21. <https://doi.org/10.1128/mcb.00545-21>
- Winter et al. (1987). Gait patterns in spastic hemiplegia in children and young adults. p. 34.
- Yokoyama et al. (2010). Caldicoprobacteraceae. *The Interim Register of Marine and Nonmarine Genera*, p.