

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Blida -1-
Faculté des Sciences de la Nature et de la Vie
Département de Biotechnologies



Mémoire de fin d'études
En vue de l'obtention du Diplôme de Master 2 en Biotechnologie
Option : Biotechnologie Microbienne

Les biofilms chez *Staphylococcus aureus* : Etude de la régulation et analyse bio-informatique

Réalisé par : OUALHA Mohamed

Devant le jury composé de :

Présidente :	Mme AMMAD F.	MCA	USBD1
Examinatrice :	Mme BENOUSSAID N.	MCB	USBD1
Promoteur :	Mr BENCHABANE M.	Professeur	USBD1

Année universitaire 2021/2022



Dédicaces



Je dédie ce travail,

À ma très chère mère, avec tout mon amour pour son soutien et ses encouragements. Je lui dois ma réussite, mon éducation, ma fierté. J'espère rester à la hauteur de ses espoirs. Que Dieu la protège.

À la personne qui s'est toujours sacrifié pour me voir réussir, mon père qui je porte son nom avec fierté et dignité

À mes deux sœurs, que j'aime autant.

À ma grand-mère, mon oncle

À l'âme de ma tante, mon grand-père et ma grand-mère, que Dieu les garde dans son vaste paradis

À tous mes amis(es) sans exceptions pour leur soutien, leur encouragement et pour tous les moments agréables qu'on a passé ensemble.

*À toute personne qui m'a aidé d'un mot, d'une idée ou d'un encouragement
À ceux qui m'ont tout donné sans rien en retour.*

Remerciements

En tout premier lieu, je remercie le bon Dieu, tout puissant, de m'avoir donné la force, la volonté, et la patience durant toutes les années d'études.

*Mes premiers remerciements vont d'abord à mon promoteur le professeur **BENCHABANE MESSAOUD** pour l'aide qu'il a fourni et les connaissances qu'il m'a transmises. Je le remercie infiniment pour sa grande disponibilité tout au long de ce travail. Travailler avec lui restera à jamais un grand honneur pour moi.*

*Je voudrais remercier également les membres du Jury, Docteur **AMMAD F.** pour l'honneur qu'elle m'a fait pour présider ce jury.*

*Docteur **BENOUSSAID N.** d'avoir accepté d'examiner ce modeste travail.*

*Je désire aussi remercier les responsables de l'administration et l'ensemble du corps enseignant de la Faculté des Sciences de la Nature et de Vie de l'Université de Blida 1. J'exprime ma gratitude et mon profond respect à tous les enseignants que je leur serai toujours reconnaissant. Mme **KRIMI**, Mme **BENCHABANE**, Mme **BOUCHENAK**, Mme **BENKORTEBY**, Mme **TAFFIFET**, Mr **AISSAT** et Mme **DJAMILA** technicienne de laboratoire.*

Ma gratitude à toute la promotion de Master 2 BTM de l'année 2021 - 2022

Résumé

Les biofilms chez *Staphylococcus aureus* : Etude de la régulation et analyse bio-informatique

Staphylococcus aureus, agent causal d'infections nosocomiales et de graves pathologies associées à la formation de biofilms, qui provoquent des problèmes majeurs de santé dans le monde. Comme les biofilms sont des systèmes biologiques dynamiques et complexes, *S. aureus* a développé un vaste réseau de mécanismes de régulation, qui se caractérise par divers facteurs de virulence et une résilience accrue aux antibiotiques et à la réponse immunitaire. Une telle situation exige des défis thérapeutiques remarquables, notamment dans les milieux hospitaliers.

En plus de la présence de riches informations scientifiques, en microbiologie fondamentale et vue l'importance de cette bactérie, de nombreuses études et travaux de recherche ont abouti à des publications scientifiques de haut niveau. En explorant les bases de données notamment NCBI (National Center for Biotechnology Information), notre objectif a porté sur la réalisation d'une investigation *in silico*, analyse bio-informatique sur les protéines ClfB, SarA et les gènes *SigB*, *CodY* associés à la formation du biofilm et de la virulence, et la compréhension des principaux mécanismes de sa formation et de sa régulation chez *S. aureus*. Nos résultats sur la prédiction des structures des protéines, la recherche de similarités de séquences en utilisant les logiciels BLASTp et BLASTn, et la phylogénie des protéines en utilisant les méthodes de calcul des distances selon NJ (Neighbor Joining) et EM (évolution minimale), nous ont permis de mettre en évidence les aspects moléculaires des gènes qui présentent des convergences et des protéines qui présentent des divergences, liés au virulence et impliqués dans la formation du biofilm chez *S. aureus*. Ainsi, les aspects moléculaires analysés représentent des cibles possibles pour développer des stratégies préventives et thérapeutiques, d'une meilleure stabilité et de leur exploitation biotechnologique.

Mots clés : *Staphylococcus aureus*, biofilm, virulence, régulation, *in silico*.

Abstract

Biofilms in *Staphylococcus aureus* : Regulatory studies and bio-informatics analysis

Staphylococcus aureus, the causative agent of hospital-acquired infections and serious pathologies associated with biofilm formation, which cause major health problems worldwide. As biofilms are dynamic and complex biological systems, *S. aureus* has developed an extensive network of regulatory mechanisms, characterized by various virulence factors and increased resilience to antibiotics and immune response. This presents remarkable therapeutic challenges, particularly in the hospital setting.

In addition to the presence of rich scientific information, in fundamental microbiology and given the importance of this bacterium, many studies and research works have resulted in high level scientific publications. Exploring the databases, in particular NCBI (National Center for Biotechnology Information), our objective was to carry out an *in silico* investigation, bioinformatics analysis on the proteins ClfB, SarA and the genes *SigB*, *CodY* associated with the formation of the biofilm and virulence, and the understanding of the main mechanisms of its formation and its regulation in *S. aureus*. Our results on protein structure prediction, sequence similarity search using BLASTp and BLASTn software, and protein phylogeny using NJ (Neighbor Joining) and EM (minimal evolution) distance calculation methods, allowed us to highlight the molecular aspects of the converging genes and diverging proteins related to virulence and involved in biofilm formation in *S. aureus*. Thus, the molecular aspects analyzed represent possible targets for the development of preventive and therapeutic strategies, better stability and their biotechnological exploitation.

Key words : *Staphylococcus aureus*, biofilm, virulence, regulation, *in silico*.

ملخص

الأغشية الحيوية في *Staphylococcus aureus*: دراسة التنظيم وتحليل المعلوماتية الحيوية

Staphylococcus aureus، العامل المسبب لعدوى المستشفيات والأمراض الخطيرة المرتبطة بتكوين الأغشية الحيوية، والتي تسبب مشاكل صحية كبيرة في العالم. نظرًا لأن الأغشية الحيوية عبارة عن أنظمة بيولوجية ديناميكية ومعقدة، فقد طورت بكتريا *S. aureus* شبكة واسعة من الآليات التنظيمية، والتي تتميز بعوامل ضراوة مختلفة وزيادة المرونة في مواجهة المضادات الحيوية والاستجابة المناعية. تتطلب مثل هذه الحالة تحديات علاجية ملحوظة، خاصة في المستشفيات.

بالإضافة إلى وجود معلومات علمية ثرية، في علم الأحياء الدقيقة الأساسي، نظرًا لأهمية هذه البكتيريا، فقد أسفرت العديد من الدراسات والأبحاث عن منشورات علمية رفيعة المستوى. من خلال استكشاف قواعد البيانات على وجه الخصوص NCBI (المركز الوطني لمعلومات التكنولوجيا الحيوية)، يتمثل هدفنا في إجراء تحقيق *in silico* وتحليل المعلومات الحيوية على البروتينات CifB و SarA والجينات *SigB* و *Cody* المرتبطة بتكوين الأغشية الحيوية والفوعة وفهم الآليات الرئيسية لتشكيلها وتنظيمها في *S. aureus*. نتائجه حول تنبؤ بنية البروتين، والبحث عن تشابه التسلسل باستخدام برنامجي BLASTn و BLASTp، ونسالة البروتين باستخدام طرق حساب المسافة وفقًا لـ NJ (انضمام الجار) و EM (الحد الأدنى من التطور)، سمحت لنا بتسليط الضوء على الجوانب الجزيئية المتعلقة للجينات التي تقدم تقاربات و البروتينات التي تظهر اختلافات مرتبطة بالفوعة وتشارك في تكوين الأغشية الحيوية في بكتريا *S. aureus*. وبالتالي، فإن الجوانب الجزيئية التي تم تحليلها تمثل أهدافًا محتملة لتطوير استراتيجيات وقائية وعلاجية، من أجل استقرار أفضل واستغلال التكنولوجيا الحيوية.

الكلمات المفتاحية: *Staphylococcus aureus*، الغشاء الحيوي، الفوعة، التنظيم، *in silico*.

Liste des abréviations

% : Pourcentage

3D : Tri Dimensionnelle

ADN : Acide Désoxyribo Nucléique

AIP : Peptide Auto-Inducteur

AMPs : Anti Microbial Peptides

ARN : Acide Ribo Nucléique

BLAST : Basic Local Alignment Search Tool.

CG : Génome Central

EPS : Substance Polymérique Extracellulaire

FASTA : Fast Adaptive Sharinkage Thresholding Algorithme.

HGT : Transfert Horizontal de Gènes

K : Kilo paire de bases

KDa : KiloDalton

Mb : Millions de paires de bases

ME : Evolution Minimale

MGEs : Génomes Mobiles Express

MSCRAMM : (Microbial Surface Components Recognizing Adhesive Matrix Molecules)

NCBI : National Center for Biotechnology Information

NJ : Neighbor Joining

PDB : Protein Data Bank

PIA : Polysaccharide Intercelulaire Adhesin

PSM : Modulines Solubles dans le Phénol

QS : Quorum Sensing

S. aureus : *Staphylococcus aureus*

SARM : *Staphylococcus aureus* Résistant à la Méthicilline

SCC : Staphylococcal Cassette Chromosome

SI : Séquences d'Insertion

VCG : Génome Central Variable

µm : Micromètre

Glossaire

Terme	Définition
Arthrite septique	Une infection articulaire provoquée par un germe ayant pénétré dans l'articulation soit par voie sanguine, depuis un foyer infectieux situé à distance, soit accidentellement, à la faveur d'une blessure ouverte, voire d'une infiltration
Bactériémie	La présence de bactéries dans le sang
Dermatite atopique	Maladie immunitaire héréditaire se manifestant par des lésions cutanées eczémateuses.
Endocardite infectieuse	Une inflammation causée par une infection d'origine bactérienne ou fongique de la paroi interne du cœur (endocarde) ou des valvules
Immunodéprimé	Se dit d'un sujet incapable d'avoir des réactions immunitaires normales.
Mucoviscidose	Maladie héréditaire, sévère et fréquente, de la sécrétion glandulaire exocrine, atteignant surtout le pancréas et les bronches, et entraînant des troubles digestifs et respiratoires chroniques.
Ostéomyélite	Maladie infectieuse grave, chronique ou aiguë, du tissu osseux.

Liste des figures

Figure 01 : Colonies jaunes de <i>S. aureus</i> sur gélose Mannitol-Sel	03
Figure 02 : <i>S. aureus</i> sous microscope optique au Gx1500 (coloration de Gram) (a), sous microscope électronique à balayage au Gx34000 (b).....	04
Figure 03 : Carte du génome de <i>S. aureus</i> FORC_001.....	05
Figure 04 : Biofilm de <i>S. aureus</i> au microscope électronique (a: Gx50400, b:Gx94500).....	07
Figure 05 : Représentation de la formation de biofilm de <i>S.aureus</i> sur une surface abiotique.....	08
Figure 06 : Les différents gènes et leur codage correspondant à la formation du biofilm chez <i>S. aureus</i>	09
Figure 07 : Réseau de régulation de la formation du biofilm chez <i>S. aureus</i>	13
Figure 08 : Carte disciplinaire de l'informatique, la bio-informatique, la biologie computationnelle, la biologie et des systèmes biologiques.....	17
Figure 09 : Applications des outils bio-informatiques dans divers domaines des sciences biologiques.....	21
Figure 10 : Format FASTA de la séquence qui code pour la protéine ClfB chez <i>S. aureus</i>	28
Figure 11 : Format FASTA de la séquence qui code pour la protéine SarA chez <i>S. aureus</i>	28
Figure 12 : Format FASTA du gène <i>SigB</i> chez <i>S. aureus</i>	29
Figure 13 : Format FASTA du gène <i>CodY</i> chez <i>S. aureus</i>	29
Figure 14 : La recherche structural de la protéine ClfB.....	30
Figure 15 : La recherche structural de la protéine SarA.....	30
Figure 16 : Enregistrement de travaux de <i>staphylococcus aureus</i> aux différentes bases de données de NCBI.....	32
Figure 17 : Représentation graphique du gène <i>SigB</i>	33
Figure 18 : Représentation graphique du gène <i>CodY</i>	33
Figure 19 : Représentation graphique de la partie codante pour la protéine ClfB.....	33
Figure 20 : Représentation graphique de la partie codante pour la protéine SarA.....	33
Figure 21 : Structure 3D de la protéine ClfB de <i>S. aureus</i>	34
Figure 22 : Structure 3D de la protéine SarA de <i>S. aureus</i>	34
Figure 23 : Liste des séquences d'alignement significatif de la protéine ClfB.....	35
Figure 24 : Liste des séquences d'alignement significatif de la protéine SarA.....	36
Figure 25 : Liste des séquences d'alignement significatif du gène <i>SigB</i>	37
Figure 26 : Liste des séquences d'alignement significatif du gène <i>CodY</i>	38
Figure 27 : Arbre phylogénétique de la protéine ClfB.....	39
Figure 28 : Arbre phylogénétique de la protéine SarA.....	40
Figure 29 : Arbre phylogénétique du gène <i>SigB</i>	40
Figure 30 : Arbre phylogénétique du gène <i>CodY</i>	40

Liste des tableaux

Tableau 01 : Principaux régulateurs du biofilm chez <i>S. aureus</i>	12
Tableau 02 : Différentes catégories d'agents antibiofilms avec exemples et mécanisme d'action rapporté contre le biofilm de <i>S. aureus</i>	16
Tableau 03 : Outils sélectionnés pour les analyses de séquence primaire.....	21
Tableau 04 : Outils populaires utilisés pour les analyses phylogénétiques.....	22
Tableau 05 : Outils sélectionnés pour effectuer des analyses de la structure-fonction des protéines.....	22
Tableau 06 : Liste de quelques bases de données populaires	23

Table des matières

Introduction	01
<i>Synthèse bibliographique</i>	
I- <i>Staphylococcus aureus</i> : Caractéristiques et biofilms	02
1. Généralité sur <i>Staphylococcus aureus</i>	02
1.1- Etymologie et Taxonomie.....	02
1.2- Ecologie et mode de transmission.....	02
1.3- Caractéristiques morphologiques et culturales.....	03
1.4- Génome de <i>Staphylococcus aureus</i>	04
2. Importance du biofilm chez <i>Staphylococcus aureus</i>	05
2.1- Définition du biofilm.....	05
2.2- Biofilms de <i>Staphylococcus aureus</i>	06
2.3- Formation du biofilm.....	07
2.4- Expression génique de l'état biofilm.....	08
2.5- Régulation et signalisation.....	11
3. Résistance aux antibiotiques	13
4. Virulence et pathogénèse	14
5. Stratégies d'inhibition et de perturbation du biofilm de <i>Staphylococcus aureus</i>	15
II- Analyse bio-informatique (<i>in silico</i>)	17
1. Aperçu et historique	17
2. Objectif et thématiques	19
3. Domaines d'application	19
4. Outils et programmes	20
5. Bases de données biologiques	22
6. Techniques et exploration des données	24
6.1- Association.....	24
6.2- Classification.....	24
6.3- Clustering.....	24
6.4- Arbres de décision.....	25
6.5- Prédiction.....	25
6.6- Modèles séquentiels.....	25

Partie expérimentale

I- Analyse <i>in silico</i>	26
1. Matériels et méthodes	26
1.1- National Center for Biotechnology Information (NCBI).....	26
1.2- Protein Data Bank (PDB).....	26
1.3- Universal Protein Resource (UniProt).....	27
1.4- Séquences génétiques et protéiques	27
1.5- Prédiction des structures	29
1.6- Calcul des taux de similarités.....	30
1.7- Analyse phylogénétique.....	31
1.7.1- Méthode de Neighbor-joining (NJ)	31
1.7.2- Méthode d'évolution minimale (ME)	31
2. Résultats	32
2.1- Travaux de recherche sur <i>Staphylococcus aureus</i>	32
2.2- Localisation et cartographie génétique.....	32
2.3- Structures 3D.....	34
2.4- Bilan BLAST.....	35
2.5- Arbres phylogénétiques.....	39
3. Discussion	41
Conclusion et perspectives	46
Références bibliographiques	

Introduction

Staphylococcus aureus est un pathogène responsable de nombreuses maladies infectieuses humaines (Momtaz et Hafezi, 2014). Cette bactérie, Gram positif, est l'agent causal de la bactériémie, l'endocardite infectieuse, les infections dermiques et des tissus mous, notamment chez les malades chroniques et immunodéprimés. Les infections liées au *S. aureus* sont courantes à la fois dans les milieux communautaires et hospitaliers et leur traitement est relativement difficile à gérer, en raison de l'émergence de souches multirésistantes aux antibiotiques, telles que le SARM (*Staphylococcus aureus* résistant à la méthicilline) (Rasigade et Vandenesch, 2014).

S. aureus est un pathogène qui développe de nombreux facteurs de virulences, notamment les toxines et les adhésines. L'expression de ces facteurs doit être finement régulée pour permettre une réponse appropriée aux conditions environnementales. La bactérie exprime de nombreux facteurs agissant sur des gènes de virulence, avec la particularité de se lier à l'ADN pour activer ou réprimer leur transcription (Mauro, 2017).

Les biofilms représentent des modalités normales en mode de vie chez *S. aureus*. Leur croissance est étroitement régulée par des facteurs génétiques complexes. Les biofilms de cet agent opportuniste, une fois établis, sont récalcitrants aux traitements antimicrobiens et les réponses immunitaires de l'hôte. Souvent contre ce genre d'infections persistantes, accompagnées de formation de biofilms, l'inefficacité thérapeutique conduit à des maladies chroniques (Archer et *al.*, 2011).

Avec les progrès technologiques récents, l'avènement des analyses bio-informatiques a permis d'acquérir de nouvelles connaissances scientifiques déterminantes, constituant une perspective globale à partir de laquelle des principes unificateurs en biologie peuvent être reconnus (Kumar et Chordia, 2017).

L'objectif de notre étude est donc de comprendre les principaux mécanismes de régulation du biofilm de *S. aureus*. Nous ciblons, particulièrement, la connaissance des aspects moléculaires liés aux gènes de virulence, de résistance et de synthèse des protéines impliquées dans la formation du biofilm. Notre investigation *in silico* vise les protéines ClfB et SarA plus les gènes *SigB* et *CodY* associés à la formation du biofilm et de la virulence chez *S. aureus*, afin d'obtenir et de discuter leurs arbres phylogéniques moléculaires.

Synthèse bibliographique

I- *Staphylococcus aureus* : Caractéristiques et biofilms

1. Généralité sur *Staphylococcus aureus*

1.1- Etymologie et Taxonomie

Les staphylocoques ont été découverts entre 1877 et 1878, à partir d'un furoncle purulent et d'une ostéomyélite (Avril *et al.*, 1992) et l'espèce *Staphylococcus aureus*, pathogène majeur chez les humains, a été mis en évidence pour la première fois en 1878 par Robert Koch en Allemagne et en 1880 par Louis Pasteur en France à partir de pus d'abcès (Spicer, 2003).

En 1883, Ogston a créé le nom de « Staphylocoque ». Le genre *Staphylococcus* a été exprimé par une combinaison des mots grecs « staphylos » qui signifie grappe de raisin et « kokkos » qui signifie baie. Et en 1884, l'espèce *Staphylococcus aureus* du latin « aurum » a été différenciée des autres *Staphylococcus* par le médecin allemand Friedrich Julius Rosenbach suite à la couleur dorée de ses colonies qui signifie « or » (Giancarlo, 2013).

La classification des staphylocoques a connu plusieurs remaniements. En 1923, étaient classés au sein de la famille des Streptococcaceae dans la 1^{ère} édition du « Bergey's Manual of Determinative Bacteriology ». En 1926, étaient classés dans la famille des Micrococcaceae selon la 2^{ème} édition. Et lors de la 9^{ème} édition et jusqu'à présent, les staphylocoques se classent comme suit (Prescott *et al.*, 2010) :

- Règne : Bacteria ;
- Division : Firmicutes ;
- Classe : Bacilli ;
- Ordre : Bacillales ;
- Famille : Staphylococcaceae ;
- Genre : *Staphylococcus* ;
- Espèce : *Staphylococcus aureus*.

1.2- Ecologie et mode de transmission

Le réservoir naturel des staphylocoques est l'homme et les animaux à sang chaud. Cependant, lorsqu'elles sont éliminées dans le milieu extérieur, ces bactéries sont très résistantes et sont fréquemment retrouvées dans l'environnement. *S. aureus* est à la fois une bactérie commensale et un pathogène des humains. Environ 30 % de la population humaine est colonisée par *S. aureus*. Simultanément, il s'agit d'une cause majeure de bactériémie et d'endocardite infectieuse ainsi que d'infections ostéoarticulaires, cutanées et des tissus mous,

pleuropulmonaires. Son site de colonisation préférentiel chez l'homme est la muqueuse nasale où 20 à 30% des adultes sont porteurs de *S. aureus* au niveau des fosses antérieures du nez, 20% sont également présent au niveau digestif et entre 8 à 15% au niveau vaginal. A partir des sites porteurs, *S. aureus* peut coloniser les territoires cutanés, en particulier les zones humides (aisselles et périnée) et les mains. La transmission intra- ou inter-humaine s'opère généralement par contact direct (manuportage). Plus rarement, elle peut être indirecte à partir d'une source environnementale (vêtements, draps, matériels médicaux) (Wertheim et *al.*, 2005).

1.3- Caractéristiques morphologiques et culturales

Les cellules bactériennes de *S. aureus* apparaissent sous forme de cocci Gram positif de 0,5 à 1,5 μm de diamètre, groupées en amas, immobiles, non sporulées et généralement non capsulées. Ce sont des aéro-anaerobies, oxydase négative et produisent plusieurs enzymes, telles que la coagulase, la DNAase et la catalase, ce qui permet de les distinguer des autres staphylocoques et cocci Gram positif (Denis et *al.*, 2011).

La culture de ses bactéries est obtenue après 24 heures d'incubation sur milieux ordinaires. *S. aureus* peut également être cultivée en présence de fortes concentrations de sel (gélose Chapman). Généralement les colonies observées sont lisses, rondes, bombées, brillantes et fermentent le mannitol. La pigmentation en jaune doré peut être observée chez certaines souches d'où le nom « aureus » (Avril et *al.*, 2000) (Figure 01 et 02).

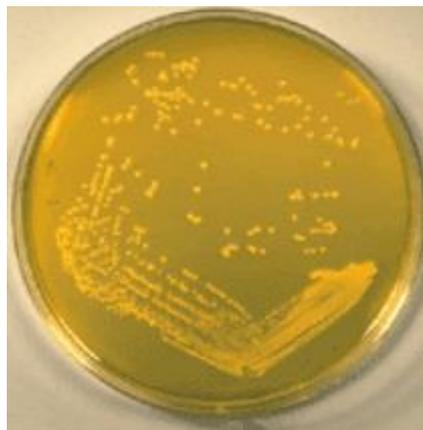


Figure 01 : Colonies jaunes de *S. aureus* sur gélose Mannitol-Sel (Ghias et *al.*, 2016).

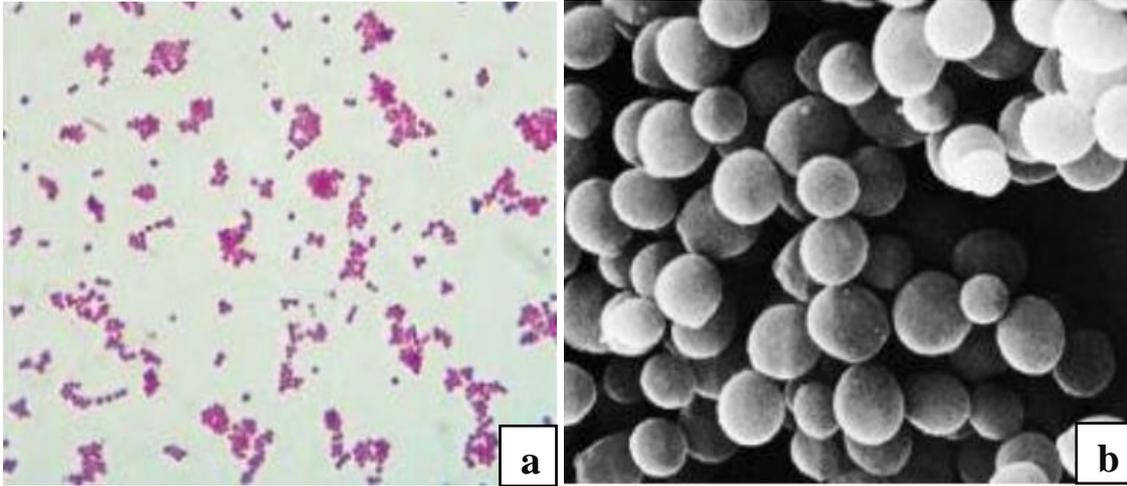


Figure 02 : *S. aureus* sous microscope optique au Gx1500 (coloration de Gram) (a), sous microscope électronique à balayage au Gx34000 (b) (Prescott et *al.* 2002).

1.4- Génome de *Staphylococcus aureus*

La première séquence complète du génome de *S. aureus* était déjà disponible en 2001 (Kuroda et *al.*, 2001). Après cette date, de nombreux autres génomes complets de diverses souches de *S. aureus* ont été décrits. Le génome bactérien est d'environ 2,8 Mb et porte 2800 gènes (McCarthy et *al.*, 2012) (Figure 03). Trois parties différentes du génome de *S. aureus* se distinguent (McCarthy et *al.*, 2012 ; Lindsay, 2014) :

- Le génome central (CG), est hautement conservé chez tous les isolats de cette espèce (> 97%) et représente environ 75% du génome correspondant aux gènes domestiques.
- Le génome central variable (VCG), représente 10% du génome et diffère entre les isolats par la présence ou l'absence de gènes accessoires ou par des polymorphismes.
- Les génomes mobiles express (MGEs), comprennent les bactériophages, les plasmides, les îlots de pathogénicité « pathogenicity islands », le staphylococcal cassette chromosome (SCCs), les transposons et les séquences d'insertion (SI). Ces séquences génomiques pourraient représenter environ 15% du génome. Les MGEs peuvent se déplacer entre les bactéries par des mécanismes de transfert horizontal de gènes (HGT), peuvent varier dans les différentes lignées génétiques et contiennent généralement des gènes de pertinence clinique, comme c'est le cas des gènes de virulence ou des gènes de résistance aux antimicrobiens et aux métaux lourds, ou ceux impliqués dans l'adaptation écologique de l'hôte.

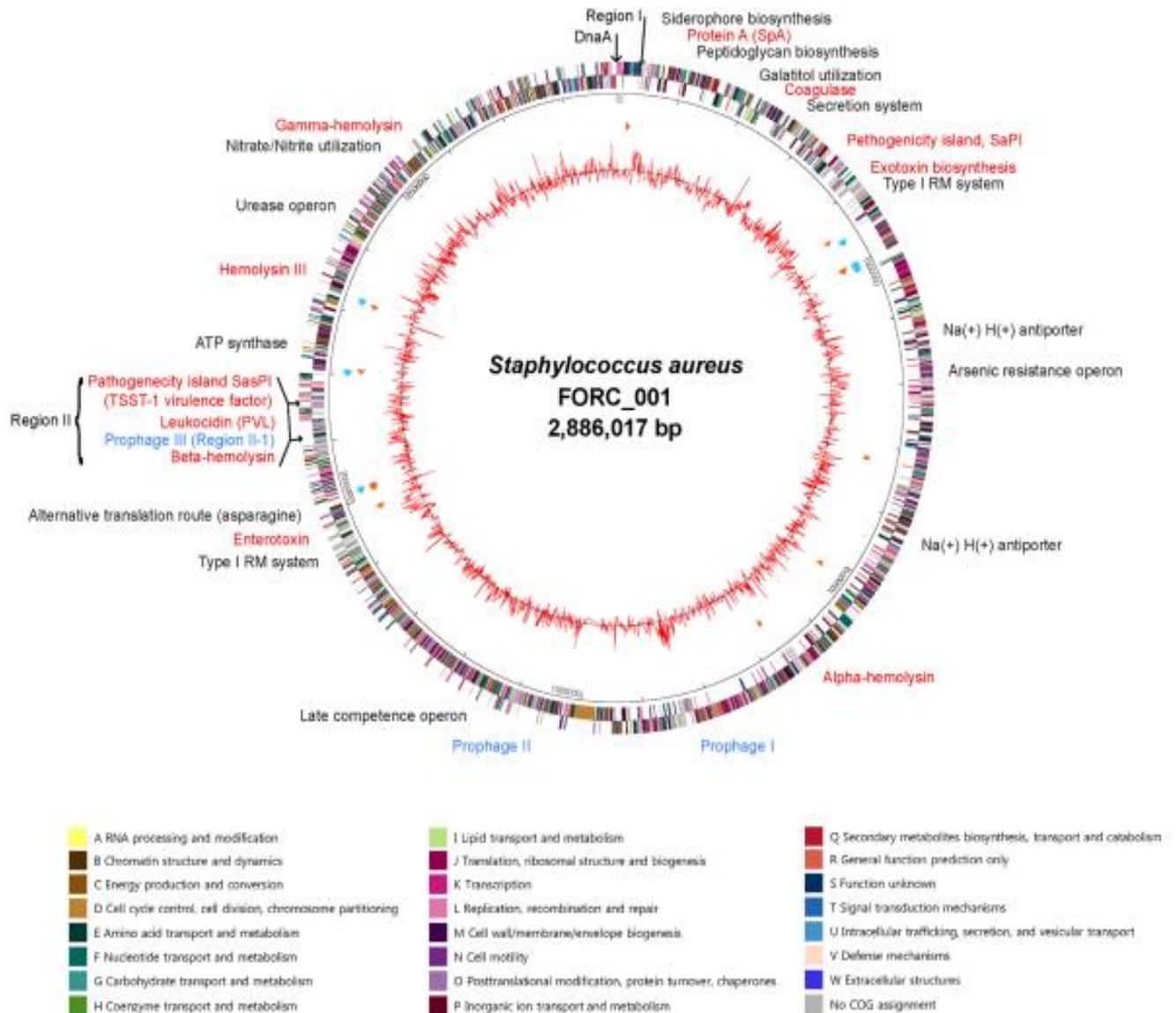


Figure 03 : Carte du génome de *S. aureus* FORC_001 (Sooyeon et al., 2014).

2. Importance du biofilm chez *Staphylococcus aureus*

2.1- Définition du biofilm

Le terme biofilm, proposé en 1978, désigne une communauté de microorganismes enrobés d'une matrice hydratée, riche en polymères extracellulaires, en situation de contact avec une surface (Lebeaux et Ghigo, 2012).

En 1999, la définition de base proposée pour le biofilm était « Une communauté structurée de cellules bactériennes enfermées dans une matrice polymérique, autoproduite, adhérentes à une surface inerte ou vivante ». Les composants de la matrice peuvent être des exopolysaccharides,

des protéines, des acides nucléiques ou d'autres substances extrapolymeriques, censées fournir aux cellules de nombreux avantages par rapport aux cellules planctoniques (Balaban et al., 2008).

La matrice assure la stabilité structurelle et la protection du biofilm. De plus, presque tous les microorganismes, notamment les bactéries (plus de 90%), peuvent former un biofilm (Al-Shuneigat et al., 2014).

Bien que les biofilms bactériens sont bénéfiques dans la plupart des environnements, ils se développent sur des implants ou lors d'infections chroniques constituent des réservoirs de pathogènes à l'origine de nombreuses infections nosocomiales. Malgré la mise en œuvre des mesures préventives, les biofilms sont difficiles à éradiquer en raison de leur tolérance caractéristique à des doses élevées d'antibiotiques (Lebeaux et Ghigo, 2012).

2.2- Biofilms de *Staphylococcus aureus*

Les biofilms représentent des modalités normales en mode de vie chez *S. aureus*. Leur formation offre à cette bactérie pathogène plusieurs avantages de survie par rapport à ses homologues planctoniques, telles que des interactions cellule-cellule bénéfiques, la protection accrue contre les agressions extérieures et la capacité de dispersion accrue. Par conséquent, les biofilms augmentent la persistance et la prolifération de *S. aureus* dans les habitats biotiques et abiotiques ; ce qui est très préoccupant dans le cas des hôpitaux et des installations alimentaires. Le développement de biofilms augmente la persistance de cette bactérie pathogène dans des environnements qui peuvent présenter un risque pour la santé humaine (Alexandra, 2017).

Des biofilms formés par *S. aureus* ont été détectés sur diverses surfaces biotiques et abiotiques, y compris, les tissus humains, les dispositifs médicaux à demeure (les cathéters implantés, les valves cardiaques artificielles, les prothèses osseuses et articulaires), les produits alimentaires et les installations de transformation des aliments (Vázquez-Sánchez et al., 2012) (Figure 04).

Les cellules de *S. aureus* qui sont enfermées et protégées par des biofilms présentent des caractères phénotypiques différents par rapport aux cellules sous leur forme planctonique. Les cellules de *S. aureus* associées aux biofilms sont plus résistantes aux antibiotiques et présentent des différences de taille et de croissance cellulaire, d'expression génique et de production de protéines, par rapport à leurs homologues en état libre (Otto, 2018).

Les cellules de *S. aureus* associées à des biofilms ont quatre états métaboliques différents : elles peuvent soit se développer de manière aérobie, être fermentatives, dormantes, ou même mortes (Archer et *al.*, 2011). En outre, la matrice polymère extracellulaire du biofilm protège les cellules contre les antibiotiques. Et les cellules dormantes où à croissance métabolique lente accentuent la résistance aux antimicrobiens (Lister et *al.*, 2014).

Les cellules de *S. aureus* enfermées dans un même biofilm se développent à des vitesses différentes, certaines se développent à une vitesse plus rapide par rapport à d'autres. Ces cellules sont de plus petites tailles et atteignent leur taille normale une fois libérées lors de la dispersion du biofilm (Moormeier et *al.*, 2017).

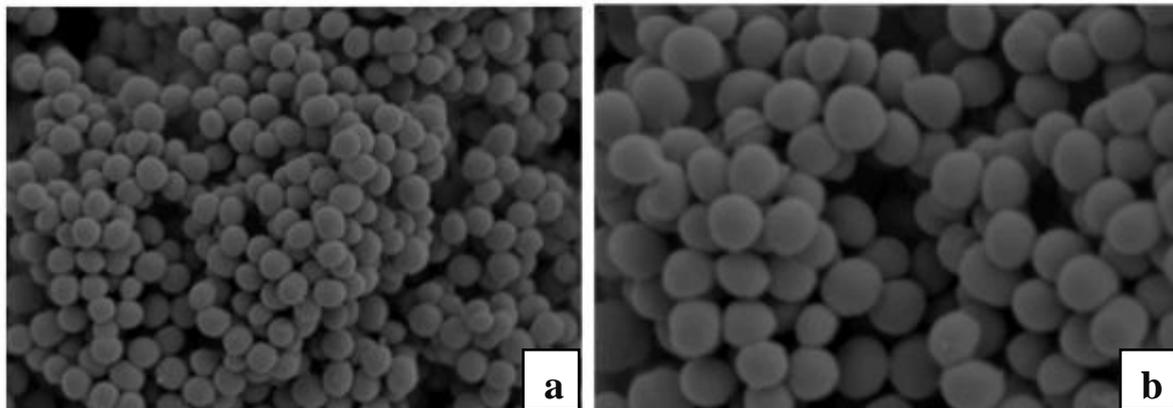


Figure 04 : Biofilm de *S. aureus* au microscope électronique (a : Gx50400, b : Gx94500) (Qi et *al.*, 2021).

2.3- Formation du biofilm

Le développement du biofilm staphylococcique est un processus complexe qui peut être divisé en trois phases principales : l'attachement initial, la production de matrice extracellulaire et prolifération cellulaire ainsi que la structuration du biofilm et le détachement des cellules. *S. aureus* se fixe à une surface biotique ou abiotique en utilisant une gamme de facteurs et de mécanismes différents, y compris les adhésines de surface attachées à la sortase (Foster et *al.*, 2019).

La formation de biofilm est initiée lorsque des cellules planctoniques flottant librement se fixent à la surface disponible en entamant sa colonisation (Petrova et Sauer, 2012). L'adhérence de *S. aureus* à une surface est influencée par les interactions hydrophobes et hydrophiles entre la surface cellulaire de *S. aureus* et toute surface biotique ou abiotique. L'adhésion, aux surfaces

hydrophobes, se réalise à l'aide de nombreuses macromolécules à liaison faible, tandis que son adhérence aux surfaces hydrophiles implique des macromolécules de liaison moins nombreuses mais plus fortes (Maikranz et *al.*, 2020).

La formation de microcolonies est suivie de la formation d'une substance polymérique extracellulaire (EPS) qui se développe en un biofilm complètement mature. Au sein de ce dernier, les cellules bactériennes qui y résident libèrent certains produits chimiques, à savoir des acides aminés D et des enzymes dégradant l'EPS (alginate lyase) pour briser et disperser le biofilm. Ces cellules planctoniques sont prêtes soit à recoloniser le même site, soit à se fixer sur un site différent et à reproduire le processus pour former un nouveau biofilm (Idrees et *al.*, 2021). Les différentes étapes impliquées dans la formation d'un biofilm bactérien sont présentées dans la figure 05.

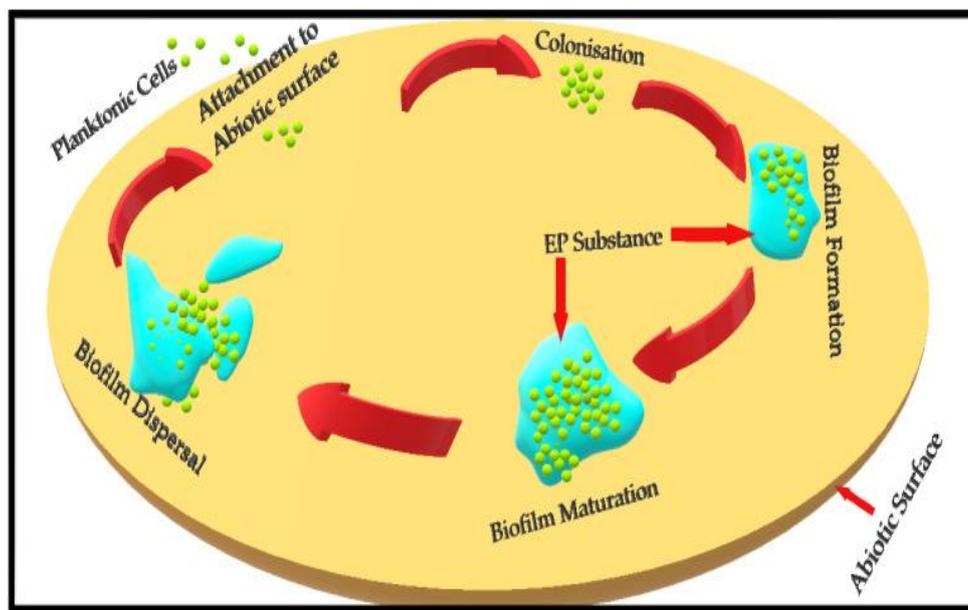


Figure 05 : Représentation de la formation de biofilm de *S. aureus* sur une surface abiotique. (Idrees et *al.*, 2021).

2.4- Expression génique de l'état biofilm

Les cellules de *S. aureus* associées aux biofilms présentent des expressions géniques altérées, signifiant qu'une régulation ascendante et descendante des gènes a été observée dans les cellules résidant à l'intérieur d'un biofilm chez cette bactérie. L'expression différentielle des

gènes explique la variation de la taille des cellules au sein d'un biofilm, leurs taux de croissance et la production de protéines (Idrees et *al.*, 2021).

Comme de nombreuses autres fonctions microbiennes, la formation de biofilms microbiens est également codée par certains gènes associés (Sauer, 2003). Chez *S. aureus*, la formation du biofilm est principalement codée par 12 gènes différents : le gène des protéines de liaison au fibrinogène (*fib*), les gènes des protéines de liaison à la fibronectine (*fnbA* et *fnbB*), les gènes d'adhésion intercellulaire (*icaA*, *B*, *C* et *D*), le facteur d'agglutination (*ClfA* et *B*), la protéine de liaison à l'élastine (*ebps*), la protéine de liaison à la laminine (*eno*) et le gène de la protéine de liaison au collagène (*cna*) (Nourbakhsh et Namvar, 2016). Les différents gènes codant pour les étapes correspondantes lors de la formation du biofilm de *S. aureus* sont énumérés dans la figure 06.

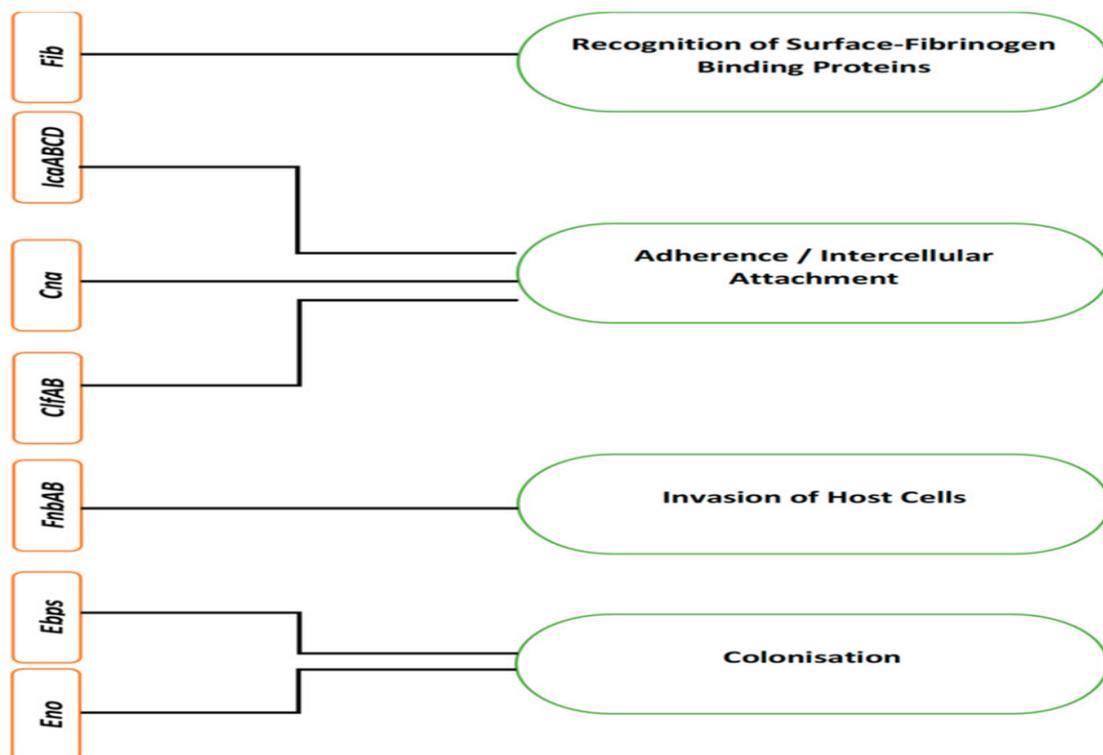


Figure 06 : Les différents gènes et leur codage correspondant à la formation du biofilm chez *S. aureus* (Idrees et *al.*, 2021).

Les gènes susmentionnés codent pour différentes protéines de surface pour provoquer l'adhérence de *S. aureus* à l'hôte, la pénétration de *S. aureus* dans l'hôte et sa colonisation, conduisant finalement à la formation du biofilm et à la virulence. Chez *S. aureus*, le

gène *fib* facilite et code la reconnaissance des protéines de liaison au fibrinogène de surface, tandis que les protéines de liaison au collagène codées par leurs gènes *cna* correspondants favorisent l'adhérence à la surface (Paharik et Horswill, 2016).

Les gènes d'adhérence intercellulaire *icaABCD* codent le processus d'adhérence de cellule à cellule et initient la formation de biofilm (Chen et al., 2020). La co-expression des gènes *fnbAB* facilite également la formation de biofilm chez *S. aureus*. Les gènes *fnbA* et *fnbB* de *S. aureus* ne sont pas impliqués dans le processus d'adhérence mais ils contribuent néanmoins au développement du biofilm, *fnAB* agissent comme des invasives et facilitent la pénétration de *S. aureus* dans les cellules hôtes (O'Neill et al., 2008).

Les gènes des facteurs d'agglutination *ClfA* et *ClfB* codent pour des protéines ancrées dans la paroi cellulaire qui se fixent au fibrinogène de surface de l'hôte (Wolz et al., 2002). Cette fixation des facteurs d'agglutination A et B, codés par les gènes *ClfA*, facilite la colonisation de l'hôte par *S. aureus*, favorise la formation du biofilm et provoque la virulence par évitement immunitaire via la liaison du fibrinogène soluble (Herman-Bausier et al., 2018). Les facteurs de répétition sérine-aspartate C et D (*SdrCD*) facilitent l'attachement aux cellules épithéliales desquamées et la colonisation nasale, tandis que *SdrE* provoque l'évitement immunitaire en se liant au facteur H du complément (Foster et al., 2014). Ces protéines sont codées pour leurs rôles respectifs par les gènes *Sdr* correspondants (Sabat et al., 2006). Le codage des protéines de liaison à l'élastine et des protéines de liaison à la laminine par leurs gènes respectifs, les gènes *ebps* et *eno*, facilitent la colonisation de l'hôte et la formation de biofilm (Kot et al., 2018).

Une fois le biofilm totalement mature, il se disperse et libère les cellules sessiles qui sont prêtes à repeupler soit le même site (primaire) ou un nouveau site (secondaire). Cette dispersion du biofilm est régulée par quatre gènes différents du système de régulation des gènes accessoires (*Agr*) (Moormeier et Bayles, 2017). Les gènes *Agr* codant pour la dispersion du biofilm comprennent *AgrA*, *AgrB*, *AgrC* et *AgrD* (Le et Otto, 2015). La dispersion régulée du biofilm est provoquée par l'induction de différentes protéases et modulines solubles dans le phénol (PSM) (Lister et Horswill, 2014). Ces protéases et PSM dispersent le biofilm en agissant comme surfactants (Saggu et al., 2019).

2.5- Régulation et signalisation

La communication de cellule à cellule chez *S. aureus* est régulée par le gène régulateur accessoire « *Agr* », un système qui provoque des adaptations génétiques pour la communication dans les cellules bactériennes (Le et Otto, 2015). *S. aureus* possède un système *Agr* distinctif qui régule l'expression de différentes toxines et facteurs de virulence et contrôle également l'interaction bactérie-hôte au site d'infection (Tang et al., 2008).

Le système *Agr* est constitué de deux paires de gènes, à savoir *AgrA-AgrC* et *AgrB-AgrD*, et s'active en réponse à la sécrétion d'un peptide auto-inducteur (AIP) (Grundstad et al., 2019). Les gènes *AgrB* et *AgrD* régulent l'expression et le transport de l'AIP. Une fois qu'une quantité suffisante d'AIP est agrégée dans l'environnement extracellulaire environnant, le système à deux composants d'*AgrA* et d'*AgrD* s'active, déclenchant la communication intracellulaire et le quorum sensing (QS) (Yu et al., 2017).

Le système *Agr* QS régule principalement la production de protéases et de modulines solubles dans le phénol (PSM), qui sont les principaux contributeurs à la maturation et au désassemblage du biofilm chez *S. aureus* (Le et al., 2019).

La formation de biofilm est un comportement de groupe social et chaque étape, de l'attachement initial à la diffusion du biofilm mature est étroitement réglementée (Parsek et Greenberg, 2005). L'un des principaux régulateurs du développement du biofilm est le système QS (Quorum Sensing), qui est chez *S. aureus* responsable du passage de la formation du biofilm à son désassemblage (Le et Otto, 2015). Outre le QS, un grand nombre d'indices environnementaux (pH, disponibilité des nutriments et débit des fluides) ainsi que des systèmes de régulation (*Sae*, *SarA*, *Rot*) sont impliqués dans la régulation de la formation du biofilm (Kavanaugh et Horswill, 2016).

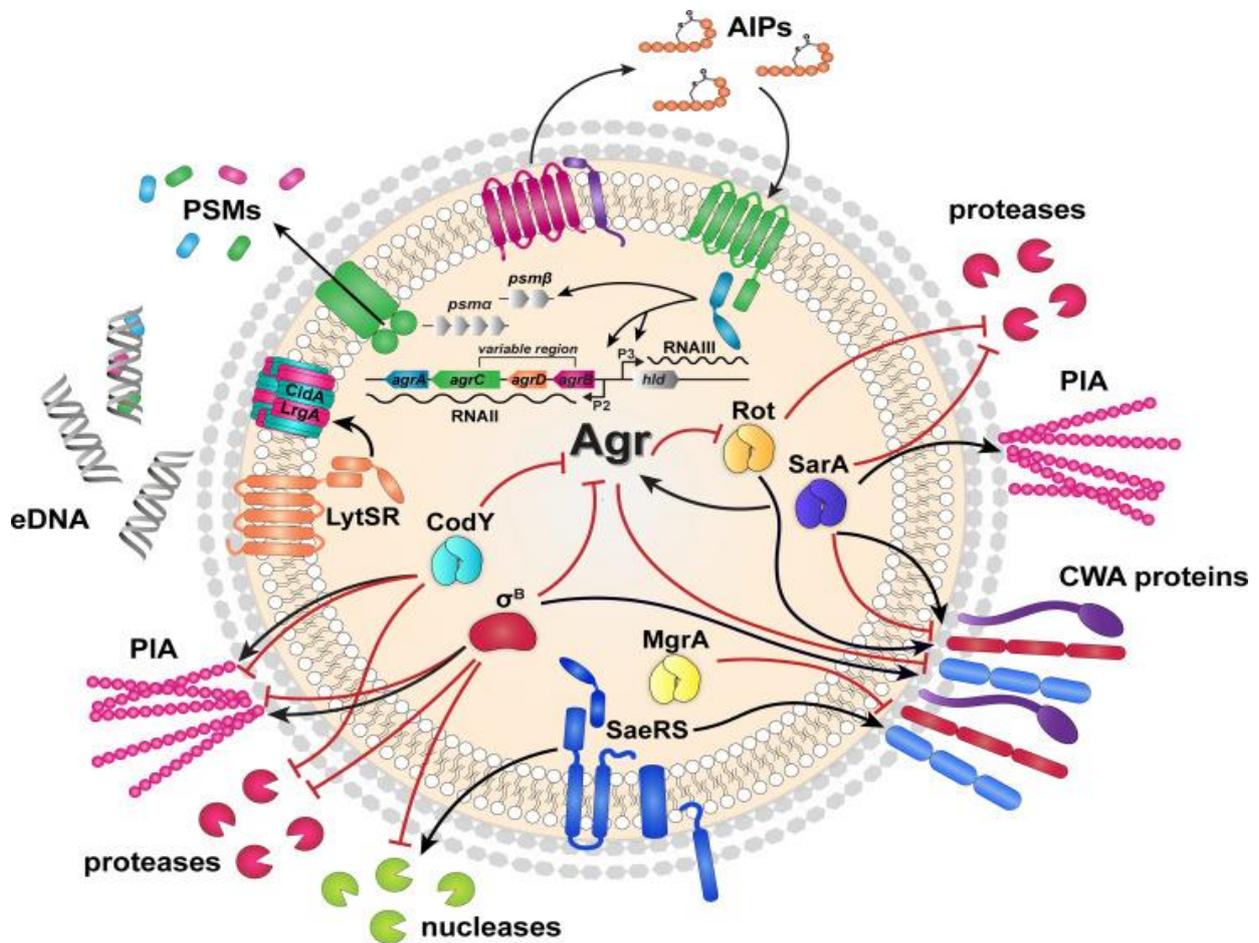
Un aperçu complet des processus de régulation les plus importants, connus pour contribuer à la formation, à la structuration et au démontage des biofilms chez *S. aureus* est présenté dans le tableau 01.

Tableau 01 : Principaux régulateurs du biofilm chez *S. aureus*.

Régulateur	Fonction	Participation au développement du biofilm	Référence
<i>Agr</i>	Système de communication cellule-cellule (quorum sensing), TCS (Two-Component System) avec des AIP comme molécules de signalisation ; stimulation de la production de protéases et des PSM ; répression des adhésines.	Important pour la fixation initiale et le démontage du biofilm	(Ythier et al., 2012)
<i>SaeRS</i>	TCS ; stimulation des adhésines et des nucléases	Important pour la maturation du biofilm; régule l'événement "exode" dans le modèle de biofilm à cinq phases	(Moormeier et al., 2014)
<i>SarA</i>	régulateur cytoplasmique de la famille SarA; répression de la production de protéases et stimulation de PIA (Polysaccharide Intercellulaire Adhesin); régulation des adhésifs	Important pour la formation de biofilm principalement par la répression des protéases ; régulateur positif de la formation de biofilm	(Loughran et al., 2014)
<i>Rot</i>	régulateur cytoplasmique de la famille SarA; principal effecteur de l'Agr QS ; stimulation des adhésines et répression des protéases	Important pour la formation de biofilm principalement par la répression des protéases ; régulateur positif de la formation de biofilm.	(Gimza et al., 2019)
<i>MgrA</i>	régulateur cytoplasmique de la famille SarA; principal effecteur de l'ArlRS TCS ; répression des adhésines	Important pour la formation de biofilm principalement par la répression des adhésines ; régulateur négatif de la formation de biofilm	(Kwiecinski et al., 2019)
<i>SigB</i>	Facteur sigma alternatif pour la réponse au stress ; stimulation des adhésines et répression des protéases et des nucléases ; régulation dépendante de la souche de PIA	Important pour la fixation initiale et le démontage du biofilm	(Kiedrowski et al., 2011)
<i>CodY</i>	Régulateur cytoplasmique de la réponse métabolique ; régulation dépendante de la souche de PIA ; répression des protéases	Important pour la formation de biofilm dépendant du PIA	(Mlynek et al., 2020)
<i>LytSR</i>	TCS régulant la lyse cellulaire ; stimulation de LrgAB	Important pour la formation de biofilm dépendant de l'eDNA (extracellular DNA)	(Sharma-Kuinkel et al., 2009)

Plusieurs autres systèmes de régulation sont importants au cours du processus de développement du biofilm, et certains sont interconnectés avec le système « *Agr QS* ». Ce réseau permet à *S. aureus* d'affiner sa réponse aux conditions environnementales changeantes et de moduler le développement du biofilm. Un schéma simplifié du système *Agr QS* et de son

réseau réglementaire avec les régulateurs les plus importants est présenté dans la figure 07, montrant une vue d'ensemble simplifiée du système de détection du quorum *Agr*, son interaction régulée avec les régulateurs du biofilm les plus importants (*LytSR*, *SigB*, *CodY*, *SaeRS*, *MgrA*, *SarA* et *Rot*) et leur influence sur les déterminants moléculaires de la formation du biofilm (Schilcher et al., 2020).



Les flèches noires indiquent une stimulation, et les flèches rouges émoussées une répression.

Figure 07 : Réseau de régulation de la formation du biofilm chez *S. aureus* (Schilcher et al., 2020).

3. Résistance aux antibiotiques

Comme toute autre bactérie formant un biofilm, chez *S. aureus* son biofilm peut ralentir ou empêcher la diffusion des médicaments antimicrobiens, empêchant ainsi l'accès des médicaments aux cellules nichées à l'intérieur du biofilm (Donlan, 2000). La formation du biofilm est provoquée par différents facteurs, tels que la privation de nutriments, les changements de température et la déshydratation (Rizzato et al., 2019). Étant donné que le

biofilm protège les cellules enrobées, *S. aureus* l'utilise comme l'un de ses outils utiles pour résister aux antibiotiques qui lui sont généralement bactéricides, en situation cellulaire planctonique flottante (Ou et *al.*, 2020). Les cellules bactériennes recouvertes d'un biofilm résistent aux antibiotiques, 10 à 1000 fois plus, que leurs formes planctoniques correspondantes (Reffuveille et *al.*, 2017).

En qualité d'agent pathogène nosocomial et en raison de sa capacité à former des biofilms sur les surfaces biotiques et abiotiques, les infections à *S. aureus* sont compliquées et difficiles à éradiquer, en particulier celles causées par les souches résistantes à la méthicilline (SARM) (Rasmussen et *al.*, 2011). La capacité d'attachement et de production de biofilms des souches de SARM a un rôle majeur dans l'amélioration des facteurs de virulence. (Nourbakhsh et Namvar, 2016).

Les personnes atteintes de certaines affections cutanées, telles que la dermatite atopique, sont très sensibles à la colonisation par *S. aureus* et donc à la formation de son biofilm (Sonesson et *al.*, 2017). Les médicaments antibactériens deviennent moins pénétrables à travers le biofilm, permettant aux cellules bactériennes de survivre en présence de médicaments à des concentrations réduites (Donlan, 2000).

En plus du rôle majeur des biofilms bactériens dans la résistance aux médicaments, d'autres facteurs, tels que l'évasion du système immunitaire inné de l'hôte, la plasticité et l'adaptabilité du génome par l'évolution des gènes et l'échange de matériel génétique, contribuent également au développement et à l'augmentation de la résistance aux antimicrobiens (Otto, 2018).

4. Virulence et pathogénèse

Les biofilms staphylococciques peuvent résister à la réponse immunitaire de l'hôte. D'abord, il a été proposé que la déficience des phagocytes pour tuer les staphylocoques dans un biofilm est principalement due à la barrière de diffusion physique d'un biofilm où les cellules immunitaires sont à peine capables de pénétrer à travers la matrice. Cependant, les leucocytes humains sont capables d'adhérer et de pénétrer dans la matrice du biofilm, même dans des conditions imitant le cisaillement physiologique (Leid et *al.*, 2002). Pour échapper au système immunitaire de l'hôte, *S. aureus* repose plutôt sur l'attaque directe des cellules immunitaires de l'hôte avec des facteurs sécrétés (les nucléases, les PSM et les hémolysines) et en biaisant l'immunité de l'hôte vers des réponses anti-inflammatoires qui favorisent la persistance des staphylocoques (Ricciardi et *al.*, 2018).

La pathogénèse de *S. aureus* est le résultat de l'expression de différents facteurs de virulence tels que des toxines, des immunomodulateurs et des exoenzymes (Oogai et *al.*, 2011). L'expression de ces facteurs est régulée et codée par des gènes particuliers (Gordon et Lowy, 2008). Parmi les facteurs de virulence chez *S. aureus*, les toxines sont les facteurs les plus importants qui protègent la bactérie en évitant toute élimination possible par le système de défense de son hôte (Otto, 2014).

Outre les facteurs de virulence susmentionnés, il existe d'autres facteurs, tels que l'attachement à l'hôte et le facteur de persistance (impliqué dans la formation du biofilm) qui, lorsqu'ils sont régulés par leurs gènes correspondants, provoquent différentes maladies infectieuses. Les facteurs de virulence impliqués dans l'attachement de *S. aureus* à la surface de l'hôte, régulés par les gènes *ClfAB*, *fnbAB*, *cna* et *ica*, et ceux impliqués dans la persistance, codés par les gènes *ica* locus et *hemB*, sont responsables d'endocardite, d'arthrite septique, de mucoviscidose et de rechute infections (Oogai et *al.*, 2011).

Le régulateur *SarA* module directement l'expression de certains facteurs de virulence et augmente l'expression d'*Agr*, en courbant la région promotrice d'*Agr* pour améliorer l'interaction AgrA-dimère (Reyeet *al.*, 2011). *SarA* induit également l'expression de *hla*, *fnbA* et *fnbB* mais inhibe la transcription de la protéine A (*spa*) (Xiong *al.*, 2006).

5. Stratégies d'inhibition et de perturbation du biofilm de *Staphylococcus aureus*

Les biofilms sont une cause majeure de résistance aux antimicrobiens et de complications dans le traitement des infections liées à *S. aureus* (Agostinho et *al.*, 2009). Il est donc important de traiter ces biofilms pour minimiser et réduire les risques de résistance aux antimicrobiens en adoptant des stratégies antibiofilm adéquates. Ces stratégies sont principalement de deux types : l'inhibition ou la prévention de la formation de nouveaux biofilms et la dispersion ou l'éradication des biofilms existants (Bhattacharya et *al.*, 2015).

Différents agents antibiofilms, y compris des médicaments, des excipients fonctionnels et d'autres molécules antibiofilms (à la fois naturelles et synthétiques) sont signalés dans l'inhibition du biofilm et sa perturbation chez *S. aureus* (Tableau 02).

Tableau 02 : Différentes catégories d'agents antibiofilms avec exemples et mécanisme d'action rapporté contre le biofilm de *S. aureus* (Idrees et al., 2021).

Agent antimicrobien	Exemple	Mécanisme d'action contre le biofilm de <i>S. aureus</i>
AMPs (Antimicrobial peptides)	NA-CATH: ATRA1-ATRA1	Prévention de l'expression génique qui code la formation de protéines liées au biofilm. Attachement à l'eDNA. (extracellular DNA)
Excipients fonctionnels	D-amino acids	Perturbation de l'eDNA
Extraits de Plantes et Huiles Essentielles	Extraits d'ail et de ginseng	Perturbation du quorum sensing bactérien
	Huile de feuilles de <i>Melaleuca bracteata</i>	Inhibition de la formation de protéines liées au biofilm et perturbation de la détection du quorum
Enzymes	DNase I, dispersine B	Dégradation de l'eDNA
	a-amylase	Dégradation des exopolysaccharides.
Nanoparticules	Argent, oxyde de zinc.	Inconnu.
Antibiotiques	Azithromycine	Inhibition de la formation de protéines liées à l'EPS et perturbation de la détection du quorum.
	Ciprofloxacine, rifampicine, amoxicilline, clindamycine, vancomycine, etc.	À l'exception de l'azithromycine, le mécanisme d'action de l'activité antibiofilm de la plupart de ces antibiotiques n'est pas entièrement compris.
Ionophores		Inconnu.

II- Analyse bio-informatique (*in silico*)

1. Aperçu et historique

La bio-informatique, un domaine de recherche interdisciplinaire, au centre de la biologie moléculaire moderne, où des méthodes informatiques sont développées et utilisées pour transformer les données biologiques en connaissances et les traduire pour des applications biologiques (Larrañaga et *al.*, 2006).

La bio-informatique est la forme abrégée de "Biological Informatics" et elle est considérée comme une association des sciences biologiques et de l'informatique. De nombreux scientifiques utilisent le terme de biologie computationnelle (Figure 08). Au lancement du grand projet du génome humain en 1991 aux Etats Unis (instituts de santé), cette discipline a gagné plus de popularité et d'intéressement scientifique (Nhgri, 2018). La bio-informatique fusionne la biologie, l'informatique et les technologies de l'information pour former une seule discipline. Elle couvre de nombreux domaines des sciences biologiques, en particulier la biologie moderne, à savoir la génomique, la transcriptomique, la protéomique, la génétique et l'évolution (Kumar et Chordia, 2017).

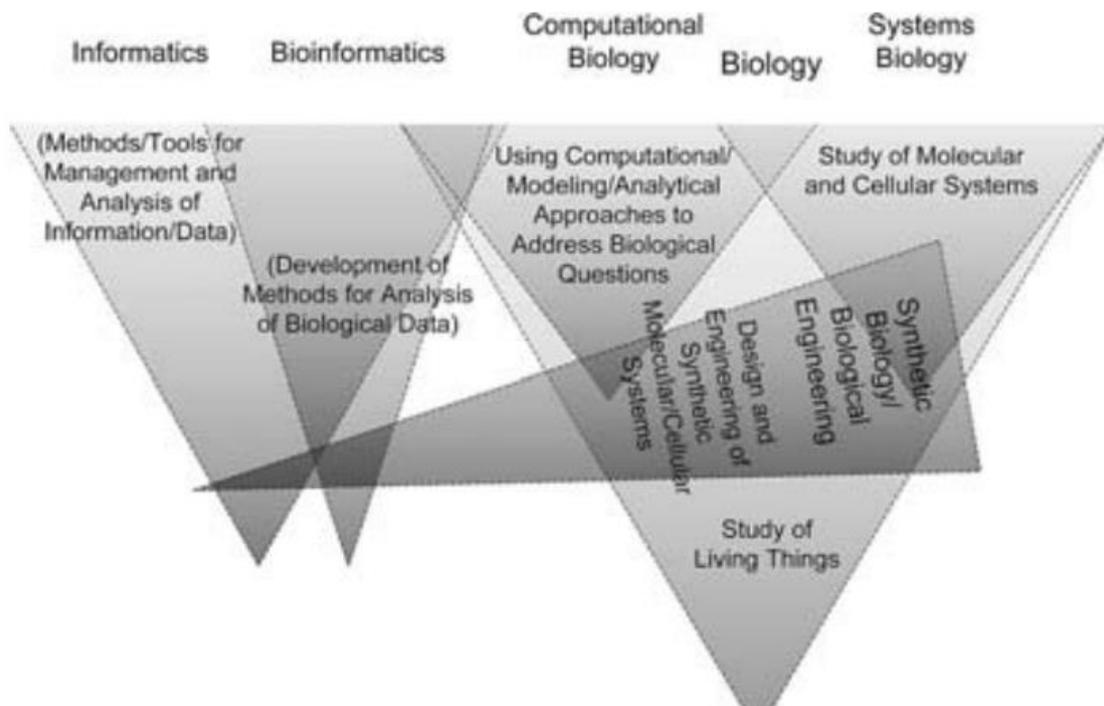


Figure 08 : Carte disciplinaire de l'informatique, la bio-informatique, la biologie computationnelle, la biologie et des systèmes biologiques (Hallam, 2013).

En 1978, le dictionnaire Anglais d'Oxford a attribué la première utilisation du terme "bio-informatique" par la biologiste Paulien Hogeweg, mais cette discipline, en tant que sous-domaine de la biologie, n'a été reconnue que dans les années 1990 (Hallam, 2013).

Vers la fin des années 1980 et le début des années 1990 ont également vu la création de plusieurs institutions clés, dont le NCBI en 1988 et l'Institut Européen de Bio-informatique en 1993 (Hallam, 2013).

Dans les années 1990, l'utilisation des ordinateurs en biologie s'est transformée en un ensemble de compétences reconnues et spécialisées dans la gestion et l'analyse de grands volumes de données (Hallam, 2013).

En 1992, le professeur et directeur du Centre de Pharmacologie Computationnelle et des Biosciences Computationnelles, Lawrence E. Hunter a utilisé des publications et des listes de diffusion de conférences pour créer une base de données de chercheurs intéressés par l'intelligence artificielle et la biologie moléculaire. La conférence que Hunter a organisée autour de sa liste est devenue la première réunion ISMB (Intelligent Systems for Molecular Biology), qui s'est tenue à Washington, DC, en 1993 et a été parrainée conjointement par la Bibliothèque Nationale de Médecine et la National Science Fondation (Hallam, 2013). C'est également au début des années 1990 que l'on a commencé à mettre en place des cours de bio-informatique spécialement conçus pour les étudiants de premier et de deuxième cycle (Hallam, 2013).

En 1995, le premier séquençage complet du génome d'un organisme vivant (*Haemophilus influenzae*) a été séquencé par l'Institut de Recherche Génomique (TIGR) (Fleischmann et al., 1995).

En 1999, selon un rapport, il y avait vingt et un programmes de bio-informatique aux États-Unis (Hallam, 2013).

Durant la période 2000 à 2016, il y avait relance de l'ancien défi informatique, avec la naissance du séquençage de deuxième génération (également appelé séquençage de nouvelle génération ou NGS) qui a permis de séquencer des milliers, voire des millions de molécules d'ADN en une seule fois (Goodwin et al., 2016).

2. Objectif et thématiques

La bio-informatique exploite les capacités de stockage et de calcul des ordinateurs pour développer des outils pour collectionner, trier, récupérer et analyser les données biologiques à grande échelle. De nombreuses ressources informatiques sont accessibles via internet, qui permet leur consultation. L'objectif central de la bio-informatique est de rassembler en un même lieu toute l'information disponible, sous forme de banques ou de bases de données, qui permettent la manipulation et l'analyse des données par les ordinateurs à l'aide d'algorithmes appropriés (Botham et *al.*, 2017).

L'objectif principal de la bio-informatique est de permettre la découverte de nouvelles connaissances biologiques ainsi que de créer une perspective globale à partir de laquelle des principes unificateurs en biologie peuvent être reconnus (Kumar et Chordia, 2017).

Les thèmes traités par la bio-informatique sont (Vert, 2004) :

- Modélisation et représentation de la connaissance en bases de données.
- Méthodes de comparaison de chaînes de caractères (formats mots et des textes).
- Algorithmes et techniques d'alignement de séquences (alignement multiple de séquences).
- Identification de motifs et modèles pour des séquences multiples
- Analyse et interprétation : Techniques de dataminings (la fouille des données).
- Représentation graphique des surfaces et des volumes, et comparaison structurelle 3D
- Simulations moléculaires.
- Analyses statistiques afin de fournir une mesure objective pour la signification des résultats.
- Réalisation des interfaces web pour faciliter l'accès aux banques de données à travers le monde.

3. Domaines d'application

La bio-informatique peut être appliquée à de nombreux domaines des sciences de la vie et des biotechnologies confrontés à l'étude et à l'utilisation du vivant. Elle joue un rôle important et croissant dans de nombreuses industries, allant de la recherche biomédicale jusqu'à l'agroalimentaire, en passant par l'énergie et l'environnement.

En industries, les entreprises pharmaceutiques et biotechnologiques sont les premiers utilisateurs de la bio-informatique. En effet, elles utilisent de plus en plus de technologies à haut

débit, comme la protéomique ou le séquençage, pour étudier les systèmes biologiques qui les intéressent. Elles s'appuient naturellement sur les outils et les méthodes de la bio-informatique pour décoder l'information biologique cachée dans les multiples données qu'elles génèrent et ainsi faciliter la traduction de ces données en avancées médicales (Vert et *al.*, 2013).

4. Outils et programmes

Les principaux outils d'un bio-informaticien sont les logiciels informatiques et l'internet. L'activité fondamentale de la bio-informatique est l'analyse des séquences d'ADN et des protéines à l'aide de divers programmes et de bases de données disponibles sur le World Wide Web. La bio-informatique est une discipline en évolution et les experts en bio-informatique utilisent des logiciels complexes pour récupérer, trier, analyser, prédire et stocker les données biologiques, souvent présentées en séquences nucléiques (ADN, ARN) et en protéines (Ardeshir, 2002).

Les outils informatiques sont couramment utilisés pour la caractérisation des gènes, la détermination des propriétés structurelles et physico-chimiques des protéines, les analyses phylogénétiques et la réalisation de simulations pour étudier comment les biomolécules interagissent dans une cellule vivante. Bien que ces outils ne puissent pas générer d'informations aussi fiables que l'expérimentation, qui est coûteuse, longue et fastidieuse, les analyses *in silico* peuvent néanmoins faciliter la prise de décision éclairée pour mener une expérience coûteuse (Mehmood et *al.*, 2014).

L'application des outils bio-informatiques dans divers domaines des sciences biologiques est assez diversifiée (Figure 09). Parmi les outils utilisés, on cite : les analyses de séquences d'ADN/protéines (Tableau 03), les études phylogénétiques (Tableau 04), la prédiction des structures 3D des molécules de protéines (Tableau 05), les interactions et simulations moléculaires ainsi que la conception de médicaments (Mehmood et *al.*, 2014).

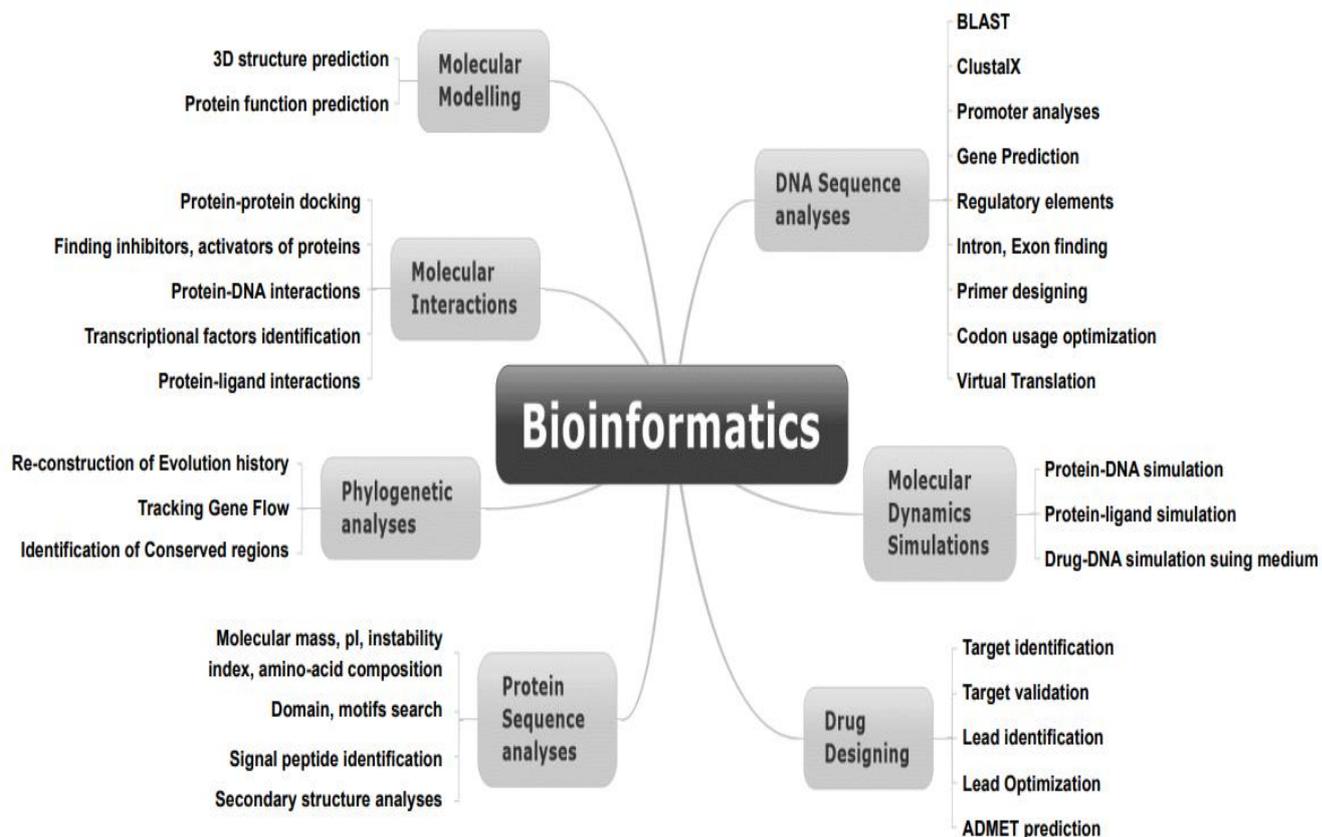


Figure 09 : Applications des outils bio-informatiques dans divers domaines des sciences biologiques (Mehmood et *al.*, 2014).

Tableau 03 : Outils sélectionnés pour les analyses de séquence primaire (Mehmood et *al.*, 2014).

Outil	Description
BLAST	Utilisé pour la recherche de séquences d'ADN ou de protéines basée sur l'identité
HMMER	Des séquences de protéines homologues peuvent être recherchées dans les bases de données respectives à l'aide de cet outil
Clustal Omega	Des alignements de séquences multiples peuvent être effectués à l'aide de ce programme.
Sequerome	Utilisé pour le profilage de séquence
ProtParam	Utilisé pour prédire les propriétés physico-chimiques des protéines
JIGSAW	Trouver des gènes et prédire les sites d'épissage dans les séquences d'ADN sélectionnées
novoSNP	Utilisé pour trouver la variation d'un seul nucléotide dans la séquence d'ADN
ORF Finder	Les gènes putatifs peuvent être soumis à cet outil pour trouver le cadre de lecture ouvert (ORF)
PPP	Outil de prédiction de promoteur procaryote utilisé pour prédire les séquences de promoteur présentes en amont du gène
Virtual Footprint	Le génome procaryote entier (avec un motif régulier) peut être analysé à l'aide de ce programme avec des régions promotrices avec plusieurs motifs régulateurs
WebGeSTer	C'est une base de données contenant des séquences de terminateurs de transcription. Elle est utilisée pour prédire les sites de terminaison des gènes pendant la transcription.
Genscan	Utilisé pour prédire les sites exon-intron dans les séquences génomiques.
Softberry Tools	Plusieurs outils sont spécialisés dans l'annotation des génomes animaux, végétaux et bactériens ainsi que dans la prédiction de la structure et de la fonction de l'ARN et des protéines.

Tableau 04 : Outils populaires utilisés pour les analyses phylogénétiques (Mehmood et *al.*, 2014).

Outil	Description
MEGA	Construit des arbres phylogénétiques pour étudier la proximité évolutive.
MOLPHY	C'est un outil d'analyse phylogénétique moléculaire basé sur la méthode du maximum de vraisemblance
PAML	Un outil d'analyse phylogénétique basé sur le maximum de vraisemblance
PHYLIP	Un package pour les études phylogénétiques.
JStree	Une bibliothèque open source pour visualiser et éditer des arbres phylogénétiques pour améliorer la présentation.
TreeView	Logiciel permettant de visualiser les arbres phylogénétiques, avec la possibilité de changer de vue
Jalview	C'est un éditeur d'alignement et est utilisé pour affiner l'alignement

Tableau 05 : Outils sélectionnés pour effectuer des analyses de la structure-fonction des protéines (Mehmood et *al.*, 2014).

Outil	Description
CATH	Un outil semi-automatique pour l'organisation catégorisée des protéines.
RaptorX	Il permet à l'utilisateur de prédire la structure des protéines sur la base d'un threading à un ou plusieurs modèles.
JPRED	Utilisé pour prédire les structures secondaires des protéines.
PHD	Utilisé pour prédire la structure du réseau neuronal.
HMMSTR	Un modèle de Markov caché pour la prédiction des corrélations séquence-structure dans les protéines.
APSSP2	Prédit la structure secondaire des protéines.
MODELLER	Prédit la structure 3D des protéines sur la base d'une modélisation comparative.
Phyre and Phyre2	Serveurs Web pour la prédiction de la structure des protéines.

5. Bases de données biologiques

La quantité croissante de données provenant des projets sur le génome a nécessité des bases de données informatisées qui présentent une assimilation rapide des formats utilisables et des programmes de logiciels d'algorithmes pour une gestion efficace des données biologiques. En raison de la diversité des données émergentes, il n'existe pas de bases de données uniques et complètes, permettant d'accéder à toutes ces informations. Cependant, un nombre progressif de bases de données contenant des informations utiles pour les chercheurs sont disponibles

(Tableau 06). Les informations fournies par la plupart de ces bases sont d'accès libre et offertes gratuitement aux communautés universitaires (Ardeshir, 2002).

Tableau 06 : Liste de quelques bases de données populaires (Mehmood et *al.*, 2014).

	Base de données	Description
Bases de données de nucléotides	DNA Data Bank of Japan	Elle est membre de l'International Nucleotide Sequence Databases (INSD) et constitue l'une des plus grandes ressources de séquences nucléotidiques.
	European Nucleotide Archive	Il capture et présente des informations relatives aux flux de travail expérimentaux basés sur le séquençage des nucléotides.
	GenBank	Elle est membre de l'International Nucleotide Sequence Databases (INSD) et est une ressource de séquences de nucléotides.
	Rfam	Une collection de familles d'ARN, représentées par de multiples alignements de séquences
Bases de données de protéines	Uniprot	Une des plus grandes collections de séquences protéiques.
	Protein Data Bank	Il s'agit d'une autre ressource majeure de protéines contenant des informations sur les structures déterminées expérimentalement d'acides nucléiques, de protéines et d'autres assemblages complexes.
	Prosite	Fournit des informations sur les familles de protéines, les domaines conservés et les sites actifs des protéines.
	Pfam	Collection de famille de protéines
	SWISS PROT	Une section de la base de connaissances UniProt contenant les séquences de protéines annotées manuellement.
	InterPro	Décrit les familles de protéines, les domaines conservés et les sites actifs.
	Proteomics Identifications Database	Une source publique, contenant des preuves à l'appui de la caractérisation fonctionnelle et de la modification post-traduction des protéines et des peptides.
Bases de données génomique	Ensembl	Il s'agit d'une base de données contenant des génomes annotés d'eucaryotes, y compris l'homme, la souris et d'autres vertébrés
	PIR	Une ressource publique intégrée pour soutenir la recherche en génomique et en protéomique

6. Techniques et exploration des données

De nombreuses techniques d'exploration des données ont été développées et utilisées dans des grands projets biologiques et médicales ; englobant des méthodes d'association, de classification, de regroupement, de conception d'arbres décisionnels, de prédiction et de modélisations séquentielles. Chaque technique a ses règles et ses méthodes, qui déterminent le type de problème qu'elle résout. Les techniques d'exploration des données doivent permettre une organisation des données (métadonnées) pour des analyses efficaces (Abdullahi, 2019).

6.1- Association

L'association est l'une des techniques d'exploration de données les plus connues. Elle permet de découvrir des modèles basés sur la relation entre les variables d'une même transaction. Elle est également connue sous le nom de technique de relation car elle utilise la relation entre les éléments et découvre l'occurrence fréquente de différents éléments qui apparaissent avec les fréquences les plus élevées dans l'ensemble de données. Les règles d'association utilisent les instructions "si-alors" afin de montrer la probabilité de relations entre des éléments de données ou des variables au sein de grands ensembles de données dans différents types de bases de données. Elles ont un certain nombre d'applications et sont largement utilisées pour aider à découvrir les corrélations, les données transactionnelles ou dans les ensembles de données médicales (Abdullahi, 2019).

6.2- Classification

La technique de classification est utilisée pour classer une collection de données en différents groupes, ou classes, afin d'obtenir une prédiction et des analyses précises pour un grand ensemble de données. En outre, la classification peut être utilisée pour alimenter le résultat d'autres techniques, comme les arbres de décision pour déterminer une classification, ou le clustering pour utiliser des attributs communs dans différentes classifications afin d'identifier des clusters (Abdullahi, 2019).

6.3- Clustering

Le clustering est l'une des premières techniques utilisées dans l'extraction de données. Le processus de clustering implique une analyse d'un ou de plusieurs attributs pour identifier les données qui sont similaires les unes aux autres afin de comprendre les différences et les similitudes entre les ensembles de données. Le processus de clustering est parfois appelé

segmentation, car il segmente les données en catégories pour identifier un groupe de résultats corrélés (Abdullahi, 2019).

6.4- Arbres de décision

Les techniques de l'arbre de décision pourraient être appliquées en tant que critères de sélection partielle. En outre, pour aider l'utilisation et la sélection de données spécifiques au sein de la structure globale. Dans l'arbre de décision, à partir d'une question simple qui a deux (ou parfois plus) de réponses. Chaque réponse conduit à une question supplémentaire pour soutenir la classification ou l'identification des données, ainsi il pourrait être catégorisé, ou une prédiction pourrait être faite sur la base de chaque réponse. L'arbre de décision utilisé avec les systèmes de classification souvent pour attacher le type d'information, et avec les systèmes prédictifs, où diverses prédictions pourraient être basées sur l'expérience historique précédente qui aidera à propulser la structure de décision et la sortie (Abdullahi, 2019).

6.5- Prédiction

La prédiction est l'une des techniques d'exploration de données les plus précieuses, car elle est utilisée pour projeter les types de données que vous verrez à l'avenir. Dans de nombreux cas, il suffit de reconnaître et de comprendre les tendances historiques pour établir une prédiction assez précise de ce qui se passera à l'avenir (Abdullahi, 2019).

6.6- Modèles séquentiels

La technique d'analyse des modèles séquentiels vise à explorer ou à identifier des modèles similaires, des événements cohérents ou des tendances dans les données de transaction sur une période donnée (Masood et Khan, 2015).

Partie expérimentale

I- Analyse *in silico*

1. Matériels et méthodes

Notre travail consiste à réaliser une étude *in silico* (bio-informatique) sur les protéines ClfB et SarA plus les gènes *SigB* et *CodY* associés, respectivement, à la formation du biofilm et de la virulence chez *Staphylococcus aureus*. Le travail s'est exécuté en ligne sur la plateforme NCBI (National Center for Biotechnology Information) à l'adresse <https://www.ncbi.nlm.nih.gov>, en utilisant la banque de données PDB (Protein Data Bank,) à l'adresse <https://www.rcsb.org/>.

1.1- National Center for Biotechnology Information (NCBI)

NCBI est un centre affilié à la Bibliothèque Nationale de Médecine des Instituts Nationaux de la Santé (National Library of Medicine des National Institutes of Health), qui a été créé en 1988 pour développer des systèmes d'information en biologie moléculaire (Sayers et *al.*, 2020).

NCBI fournit une vaste gamme de ressources en ligne, en informations et des données biologiques : la base de données GenBank relative aux séquences d'acides nucléiques et la base de données PubMed dédiée aux citations et aux résumés publiés dans les revues scientifiques des sciences de la vie. Le système « **Entrez** » permet d'effectuer des opérations de recherche et d'extraction de la plupart de ces données à partir de 34 bases de données distinctes (Sayers et *al.*, 2020).

Les ressources supplémentaires qui ont été mises à jour au cours de la dernière année comprennent PMC, Bookshelf, Genome Data Viewer, SRA, ClinVar, dbSNP, dbVar, Pathogen Detection, BLAST, Primer-BLAST, IgBLAST, iCn3D et PubChem. Toutes ces ressources sont accessibles via la page d'accueil du NCBI à l'adresse <https://www.ncbi.nlm.nih.gov> (Sayers et *al.*, 2020).

1.2- Protein Data Bank (PDB)

La banque de données sur les protéines (PDB) est l'une des deux ressources d'archives pour les données expérimentales essentielles à la recherche et à l'éducation biomédicale dans le monde. Elle a été créée en 1971 en tant que première ressource de données numériques en libre accès en biologie (Burley et *al.*, 2017).

L'archive PDB est gérée conjointement par le partenariat Worldwide Protein Data Bank (wwPDB ; wwpdb.org), composé de la Research Collaboratory for Structure Bioinformatics (RCSB) Protein Data Bank, Protein Data Bank Japan (PDBj), Protein Data Bank in Europe

(PDBe) et BioMagResBank (BMRB). L'organisation wwPDB fonctionne dans le cadre d'un accord formel (Burley et *al.*, 2017).

La vision du (RCSB) PDB est de permettre de percée dans la recherche scientifique, la médecine, la découverte de médicaments, la technologie et l'éducation structurelles des systèmes biologiques (Burley et *al.*, 2017).

PDB abrite actuellement plus de 134 000 structures biomoléculaires au niveau atomique déterminées par cristallographie macromoléculaire aux rayons X, spectroscopie RMN et microscopie électronique 3D (Burley et *al.*, 2017).

1.3- Universal Protein Resource (UniProt)

UniProt est une ressource complète pour les séquences de protéines et les données de leurs annotations. Les bases de données UniProt existent pour soutenir la recherche biologique et biomédicale en fournissant un recueil complet de toutes les données de séquence protéiques connues, liées à un résumé des informations fonctionnelles vérifiées expérimentalement ou prédites par calcul sur cette protéine. L'objectif de la base UniProt est de fournir aux utilisateurs un ensemble complet, de haute qualité, librement accessible, de séquences de protéines annotées avec des informations fonctionnelles (The UniProt Consortium. 2020).

UniProt fournit trois outils principaux :

- « **BLAST** » pour la recherche de similarité des séquences,
- « **Align** » pour l'alignement de séquences multiples,
- « **Retrieve/ID Mapping** » pour utiliser une liste d'identifiants pour récupérer les protéines.

Ces outils sont disponibles sur leurs propres pages dédiées sur le site Web d'UniProt à l'adresse <https://www.uniprot.org/> (Pundir et *al.*, 2016).

1.4- Séquences génétiques et protéiques

Sur la plate-forme NCBI, nous avons téléchargé sous format FASTA, les séquences protéiques des gènes *ClfB* et *SarA* ainsi que les séquences génétiques du *SigB* et *CodY* associés à la formation du biofilm et de la virulence chez *S. aureus*, en spécifiant leurs numérotations de concession (NC), (travail réalisé en date du 3 juin 2022 sur le site : <https://www.ncbi.nlm.nih.gov/gene/?term>).

Le format **FASTA** (Fast Adaptive Sharinkage Thresholding Algorithm) est un format de fichier utilisé pour stocker des séquences polymères de nucléotides et d'acides aminés et est compatible avec une grande variété de logiciels de bio-informatique (Hoogstrate et *al.*, 2021).

Les numérotations de concession (NC) des protéines et des gènes concernés sont :

- ClfB (WP_164784326.1) (figure 10),
- SarA (ABD73686.1) (figure 11),
- *SigB* (NC_007795.1 : c2132699-2131929) (figure 12),
- *CodY* (NC_007795.1 : 1179705-1180478) (figure 13).

Les formats FASTA des gènes obtenus sont :

```
>WP_164784326.1 MSCRAMM family adhesin clumping factor ClfB, partial [Staphylococcus aureus]
MKKRIDYLSNKQNKYSIRRFVGTTSVIVGATILFGIGNHQAQASEQSNDTTQSSKNNASADSEKNNMIE
TPQLNTTANDTSDISANTNSANVDSTTKPMSTQTSNTTTTTEPASTNETPQPTAIKNQATAAKMQDQTVPQ
EANSQVDNKTTNDANSIATNSELKNSQTLDLQPQSSPQTISNAQGTSKPSVRTRAVRSLAVAEPVVNAADA
KGTNVNDKVTASNFKLEKTTFDPNQSGNTFMAANFTVTDKVKSGDYFTAKLPDSLGTNGDVDYSNSNNTM
PIADIKSTNGDVVAKATYDILTKTYTFVFTDYVNNKENINGQFSLPLFTDRAKAPKSGTYDANINIADDEM
FNNKITYNYSSPIAGIDKPNGANISSQIIGVDTASGQNTYKQTVFVNPQQRVLGNTWVYIKGYQDKIEES
SGKVSATDTKLRIFEVNDISKLSDSYADPNDSNLKEVTDQFKNRIYYEHPNVASIKFGDITKTYVVLVE
GHYDNTGKNLKTQVIQENVPVTNRDYSIFGWNNEVVRYGGGSADGDSAVNPKDPTPGPPVDPEPSPDP
EPEPSPDPDPDSDSDSGSDSDSGSDSDSESDSDSDSD
```

Figure 10 : Format FASTA de la séquence qui code pour la protéine ClfB chez *S. aureus*.

```
>ABD73686.1 SarA, partial [Staphylococcus aureus]
VTYADKLLSLIKKEFSISFEEFAVLTYISENKEKEYLKDIIINHLNYQPQVVKAVKILSQEDYFDKCRN
EHDERTVLILVNAQQRKKIESLLSRVVK
```

Figure 11 : Format FASTA de la séquence qui code pour la protéine SarA chez *S. aureus*.

```
>NC_007795.1:c2132699-2131929 Staphylococcus aureus subsp. aureus NCTC 8325 chromosome,
complete genome
ATGGCGAAAGAGTCGAAATCAGCTAATGAAATTTACCTGAGCAAATTAACCAATGGATTAAGAACACC
AAGAAAATAAGAATACAGATGCACAGGATAAGTTAGTTAAACATTACCAAAAATAATTGAGTCATTGGC
ATATAAATATTCTAAAGGACAATCACATCACGAAGATTTAGTTCAAGTTGGTATGGTTGGTTAATAGGT
GCCATAAATAGATTTCGATATGTCCTTTGAACGGAAGTTTGAAGCCTTTTTAGTACCTACTGTAATCGGTG
AAATCAAAGATATCTACGAGATAAACTTGGAGTGTACATGTTCCGAGACGTATTAAGAAATTGGGCC
AAGAATCAAAAAGTGAGCGATGAACTAACCGCTGAATTAGAGCGTTCACCTTCTATCAGTGAATAGCT
GATCGATTAGAAGTCTCAGAAGAAGAAGTGTAGAAAGCAATGGAAATGGGACAAAAGTTATAATGCGTTAA
GTGTTGATCATTCCATTGAAGCTGATAAAGATGGTTCAACTGTTACGCTATTAGATATTATGGGGCAACA
AGATGACCATTATGACTTAACTGAAAAACGTATGATTTTAGAAAAATATTACCTATATTATCTGATCGC
GAACGAGAAATCATAAATGTACGTTTATTGAAGGATTGAGTCAAAAAGAGACAGGTGAGCGTATCGGTT
TAAGTCAAATGCATGTATCACGACTTCAGAGAACGGCAATTAAGAAATTACAAGAAGCAGCACATCAATA
G
```

Figure 12 : Format FASTA du gène *SigB* chez *S. aureus*.

```
>NC_007795.1:1179705-1180478 Staphylococcus aureus subsp. aureus NCTC 8325 chromosome,
complete genome
ATGAGCTTATTATCTAAAACGAGAGAGTTAAACACGTTACTTCAAAAACACAAAGGTATTGCGGTTGATT
TTAAAGATGTAGCACAAACGATTAGTAGCGTAACTGTAACAAATGTATTTATTGTATCGCGTCGAGGTAA
AATTTTAGGATCGAGTCTAAATGAATTATTAAGAAAGTCAAAGAATTATTCAAATGTTGGAAGAAAGACAT
ATTCCAAGTGAATATACAGAACGATTAATGGAAGTTAAACAAACAGAATCAAATATTGATATCGACAATG
TATTAACAGTATTCCCACCTGAAAACAGAGAATTATTCATAGATAGTCGTACAACATCTTCCCAATTTT
AGGTGGAGGGGAAAGATTAGGTACATTAGTACTTGGTCGAGTACATGATGATTTAATGAAAATGATTTG
GTAAGTGGTGAATATGCTGCTACAGTTATTGGTATGGAAATCTTACGTGAGAAGCATAGTGAAGTAGAAA
AAGAAGCGCGGATAAAGCTGCTATTACAATGGCAATTAATCATTATCTTATTCTGAAAAAGAAGCGAT
TGAACATATCTTTGAAGAACTTGGCGGTACGGAAGGCCTATTAATCGCATCAAAGTTGCAGATAGAGTT
GGTATTACTAGATCTGTAATTGTAATGCACTACGTAATTAGAAAAGTGTGGTGAATTGAATCACGTT
CTTTAGGAATGAAAGGTACTTTTATTAAGTTAAAAAAGAAAAATTCTTAGATGAATTAGAAAAAGTAA
ATAA
```

Figure 13 : Format FASTA du gène *CodY* chez *S. aureus*.

1.5- Prédiction des structures

Dans notre travail, nous avons consulté la plateforme Protein Data Bank (PDB) pour la prédiction des structures 3D des protéines, par insertion du nom de la protéine et le nom de la bactérie dans le champ de recherche en date du 04 juin 2022 à l'adresse <https://www.rcsb.org/> (figure 14 et 15).

The screenshot shows the RCSB PDB website interface. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. The main header features the RCSB PDB logo and the text "190841 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education". A search bar contains the text "ClfB staphylococcus aureus". Below the search bar, there are several logos for partner organizations like PDB-101, EMDatResource, and Worldwide Protein Data Bank. A sidebar on the left contains navigation options: Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area is titled "A Structural View of Biology" and includes introductory text about the PDB archive. To the right, there is a "May Molecule of the Month" section featuring a 3D molecular model of a protein structure, with the caption "Nicotine, Cancer, and Addiction".

Figure 14 : Recherche structurale de la protéine ClfB.

This screenshot is similar to Figure 14, showing the RCSB PDB website interface. The search bar now contains "SarA staphylococcus aureus". The rest of the page layout, including the navigation bar, sidebar, and main content area with the "A Structural View of Biology" section and the "May Molecule of the Month" section, remains identical to the previous figure.

Figure 15 : Recherche structurale de la protéine SarA.

1.6- Calcul des taux de similarités

Pour réaliser notre étude *in silico* et pour le calcul des taux de similarité entre nos séquences, nous avons utilisé l'outil de recherche de similitude de séquences, Basic Local Alignment Search Tool, ou **BLAST**.

Le programme BLAST développé en 1990, sert à la recherche dans de grandes bases de données de séquences moléculaires, des séquences qui présentent des régions de similitude avec la

séquence entrée, fournie par l'expérimentateur (Gibson et *al.*, 2004). Ce programme est un algorithme de recherche de similitudes locales (Deléage et al. 2015),

Il existe de nombreux types de recherches BLAST, et dans notre travail nous avons utilisé :

- **BLASTp** pour les protéines ClfB et SarA qui compare une séquence d'acides aminés d'intérêt aux séquences d'une base de données de protéines,
- **BLASTn** pour les gènes *SigB* et *CodY* qui compare une séquence d'intérêt aux séquences d'une base de données de nucléotides.

1.7- Analyse phylogénétique

Un arbre phylogénétique (phylogénie) est une forme de classification des espèces, qui traduit les relations de descendance des espèces avec modification de leurs caractères. C'est un graphe non cyclique constitué de plusieurs nœuds qui sont les unités taxonomiques évolutives (OTUs : Operational Taxonomic Unit). Ces nœuds sont reliés entre eux par des branches dont la signification est relative aux types de données utilisées pour remonter la phylogénie (Stuessy, 1983).

Les nouvelles méthodes de construction de l'arbre phylogénique sont basées sur le calcul de distances, obtenus à partir des calculs d'alignement par la méthode de Neighbor-joining et la méthode d'évolution minimale

1.7.1- Méthode de Neighbor-joining (NJ)

La méthode Neighbor-joining, est proposée pour reconstruire des arbres phylogénétiques à partir de données de distance évolutive. Le principe de cette méthode est de trouver des paires d'unités taxonomiques opérationnelles (OTU) qui minimisent la longueur totale des branches à chaque étape de regroupement des OTU, en commençant par un arbre en forme d'étoile. Les longueurs de branches ainsi que la topologie d'un arbre parcimonieux peuvent être obtenues rapidement en utilisant cette méthode (Saitou et *al.*, 1987), qui a été inspirée à partir de la méthode ancienne UPGMA (Unweighted Paired Group Mean Arithmetic) (Sneath et Sokal, 1973).

1.7.2- Méthode d'évolution minimale (ME)

La méthode d'évolution minimale est basée sur l'hypothèse que l'arbre avec la plus petite somme d'estimations de longueur de branche est le plus susceptible d'être le vrai (Rzhetsky et Nei, 1993).

2. Résultats

2.1- Travaux de recherche sur *Staphylococcus aureus*

Nos résultats obtenus lors de la recherche sur la bactérie *Staphylococcus aureus*, via la base de données NCBI en date du 5 juin 2022 (figure 16), révèlent :

- plus de dix-huit-mille publications sur l'option Gènes,
- presque trente-six-millions publications sur Protéines,
- plus de trois-mille publication dans les structures protéiques,
- plus de cent-trente-neuf-mille publication dans PubMed,
- plus deux-cent-sept-mille publications dans PubMed Central,
- plus de deux-mille-huit-cents ouvrages.

Literature	Genes	Proteins
Bookshelf 2,823	Gene 18,816	Conserved Domains 259
MeSH 175	GEO DataSets 10,110	Identical Protein Groups 1,519,552
NLM Catalog 239	GEO Profiles 430,745	Protein 35,958,438
PubMed 139,122	HomoloGene 2	Protein Family Models 1,526
PubMed Central 207,705	PopSet 771	Structure 3,048
Genomes	Clinical	PubChem
Assembly 28,944	ClinicalTrials.gov 797	BioAssays 50,088
BioCollections 0	ClinVar 8	Compounds 3
BioProject 7,213	dbGaP 13	Pathways 220
BioSample 112,631	dbSNP 0	Substances 64
Genome 1	dbVar 0	
Nucleotide 2,378,238	GTR 1	
SRA 106,732	MedGen 69	
Taxonomy 1	OMIM 94	

Figure 16 : Enregistrement de travaux de *staphylococcus aureus* aux différentes bases de données de NCBI.

2.2- Localisation et cartographie génétique

D'après les résultats obtenus à partir de la base de données NCBI, notre analyse *in silico* sur les deux gènes et les deux protéines associés à la formation du biofilm et de la virulence chez *S. aureus* ont exprimé une cartographie génétique et une position par rapport aux autres gènes

Résultats

de *SigB* qui est localisé de 2131929k jusqu'à 2132699k (figure 17) et de *CodY* qui est localisé de 1179705 k jusqu'à 1180478 k (figure 18), et aussi la cartographie génétique des parties codantes pour la protéines ClfB (figure 19) et la protéines SarA (figure 20).

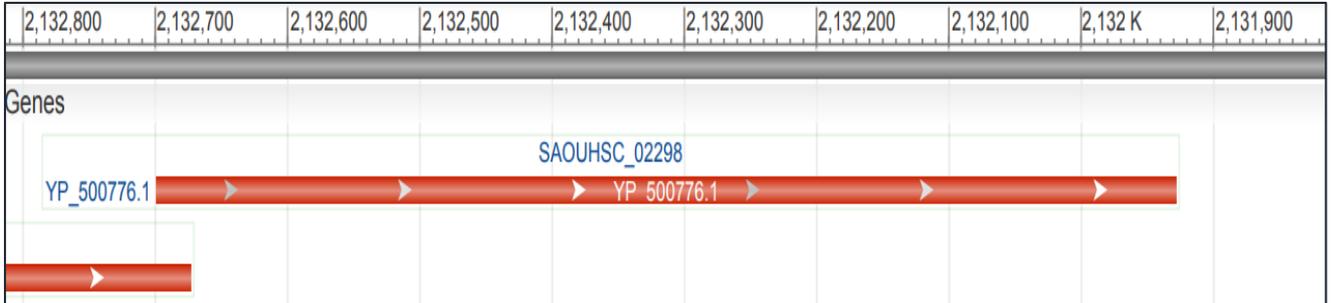


Figure 17 : Représentation graphique du gène *SigB*.

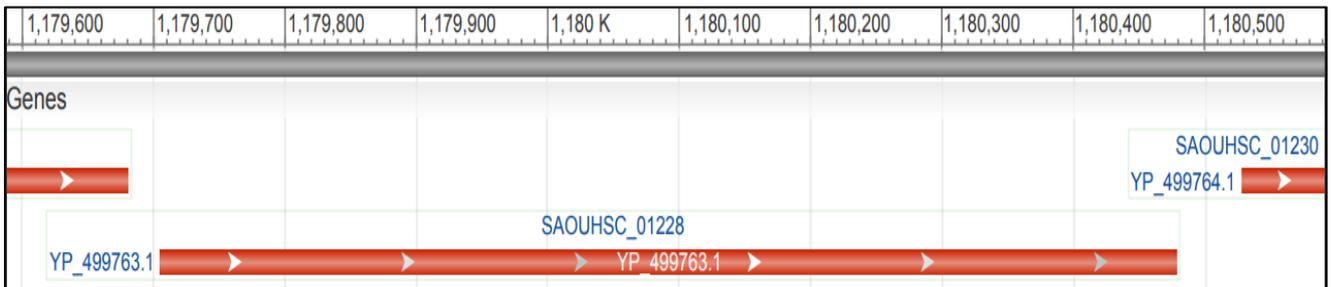


Figure 18 : Représentation graphique du gène *CodY*.

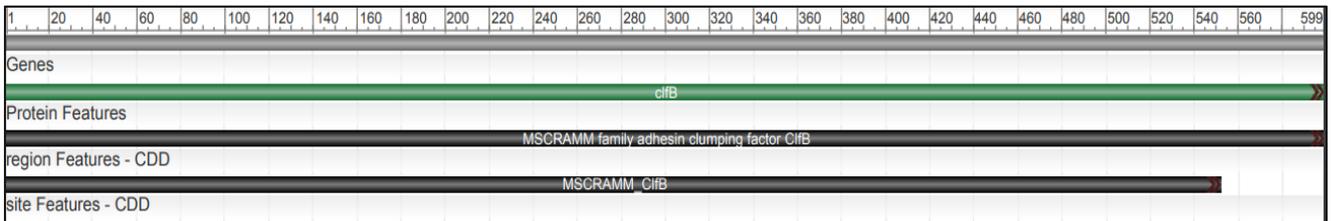


Figure 19 : Représentation graphique de la partie codante pour la protéine ClfB.

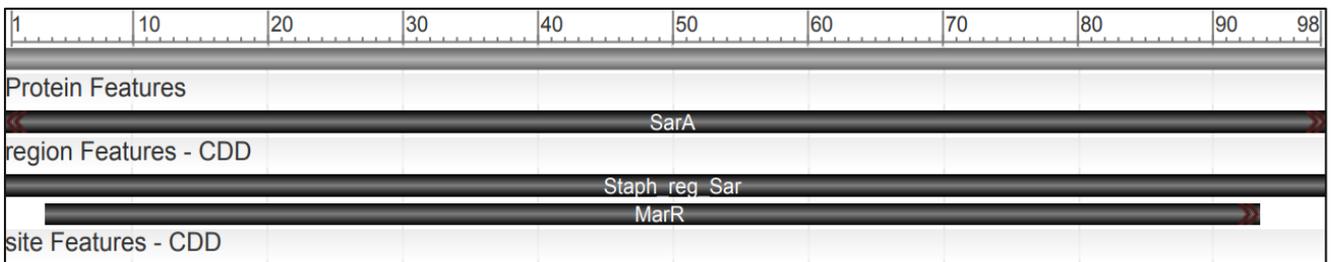


Figure 20 : Représentation graphique de la partie codante pour la protéine SarA

2.3- Structures 3D

La plate-forme PDB (<https://www.rcsb.org>) nous a permis de prédire la structure 3D des protéines ClfB et SarA (figures, 21 et 22).

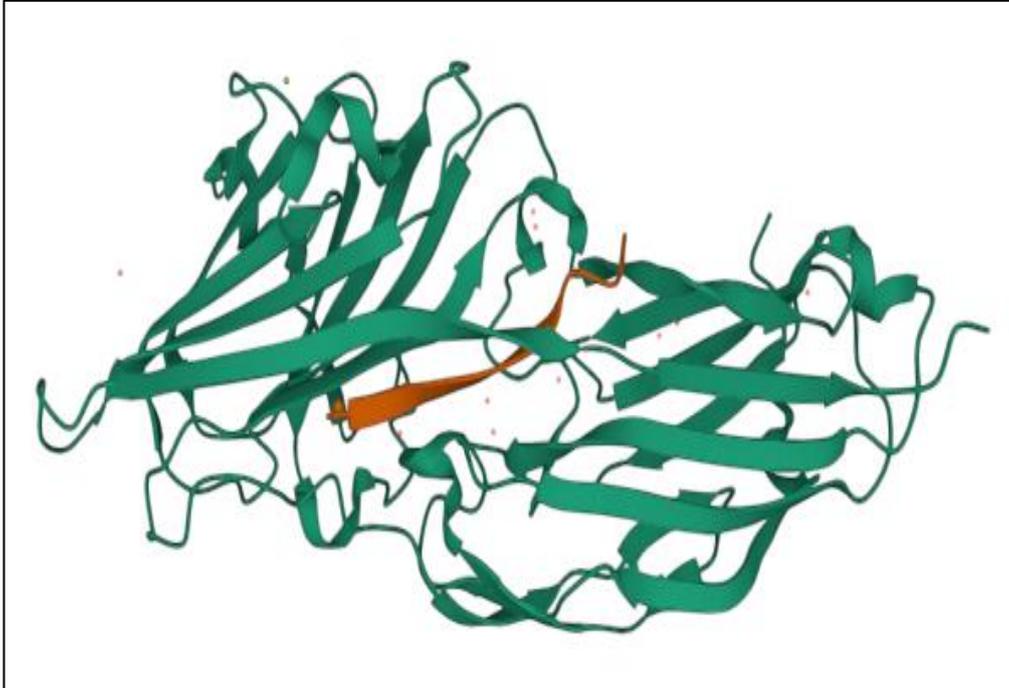


Figure 21 : Structure 3D de la protéine ClfB de *S. aureus*.



Figure 22 : Structure 3D de la protéine SarA de *S. aureus*.

2.4- Bilan BLAST

L'exécution du programme BLAST : BLASTp pour les protéines ClfB et SarA et BLASTn pour les deux gènes *SigB* et *CodY* (travail réalisé en date du 06 juin 2022) a fourni des listes d'un ensemble de séquences similaires à notre requête, avec leurs pourcentages d'identité et numéros d'accession.

Toutes les séquences obtenues pour la protéine ClfB présentent un pourcentage d'identité supérieur ou égal à 99,82% (figure 23) et pour la protéine SarA, le pourcentage d'identité le plus faible est égal à 89,80% mais la majorité de séquences présentent un taux d'identité complète de 100% (figure 24).

	Description	Scientific Name	Per. Ident	Accession
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031924595.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MBG3259365.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_218725149.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_172844374.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_044124426.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_033871896.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MVJ99888.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_237267518.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_088591743.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	99.82%	WP_172844196.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	NGQ07016.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MBG1508216.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MCC0867570.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_033871589.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_058330387.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_191669847.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_086047739.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031909878.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MCC0945014.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031881014.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MBD6775337.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MBD6720912.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_048524646.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031868479.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_053867532.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031875831.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_103483276.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MBD6749613.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_050974304.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031881131.1
<input checked="" type="checkbox"/>	MSCRAMM family adhesin clumping factor ClfB [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_223296624.1

Figure 23 : Liste des séquences d'alignement significatif de la protéine ClfB.

Résultats

	Description	Scientific Name	Per. Ident	Accession
<input checked="" type="checkbox"/>	transcriptional regulator [Staphylococcus aureus]	Staphylococcus aureus	100.00%	NFW64552.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_248499371.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_049882507.1
<input checked="" type="checkbox"/>	TPA: transcriptional regulator [Staphylococcus aureus]	Staphylococcus aureus	100.00%	HAR5648058.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031889640.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_216734356.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_117221146.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Bacteria]	Bacteria	100.00%	WP_001018677.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	98.98%	WP_240053289.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	98.98%	WP_182055592.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	97.96%	WP_130114785.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	98.98%	WP_200645250.1
<input checked="" type="checkbox"/>	transcriptional regulator SarA [Staphylococcus aureus subsp. aureus CO-23]	Staphylococcus aureus subsp. aureus CO-23	100.00%	EID88230.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	98.98%	WP_117216562.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	98.98%	WP_196499591.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_108926651.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Vibrio cholerae O1]	Vibrio cholerae O1	100.00%	MBU5914740.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	MBU6801002.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_129950618.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_031908752.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_191962331.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_216813341.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus saccharolyticus]	Staphylococcus saccharolyticus	90.82%	WP_115313900.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus aureus]	Staphylococcus aureus	100.00%	WP_151134452.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus]	Staphylococcus	90.82%	WP_002451055.1
<input checked="" type="checkbox"/>	global transcriptional regulator SarA [Staphylococcus]	Staphylococcus	89.80%	WP_002441263.1
<input checked="" type="checkbox"/>	transcriptional regulator sarA [Staphylococcus aureus subsp. aureus 68-397]	Staphylococcus aureus subsp. aureus 68-397	100.00%	EEV09834.1
<input checked="" type="checkbox"/>	Staphylococcal accessory regulator A (SarA) [Staphylococcus aureus]	Staphylococcus aureus	100.00%	SUL68049.1

Figure 24 : Liste des séquences d'alignement significatif de la protéine SarA.

En utilisant BLASTn, pour le gène *SigB* de même que *CodY* toutes les séquences présentent un pourcentage d'identité de 100% (figures 25 et 26).

Résultats

	Description	Scientific Name	Per. Ident	Accession
<input checked="" type="checkbox"/>	Staphylococcus aureus TPS3156 DNA_complete genome	Staphylococcus aureus	100.00%	AP023034.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain Dresden-275757 chromosome_complete genome	Staphylococcus aureus	100.00%	CP054876.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain 16405 chromosome_complete genome	Staphylococcus aureus	100.00%	CP053353.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain 14732 chromosome_complete genome	Staphylococcus aureus	100.00%	CP053639.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_403 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047849.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_996 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047830.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_1150 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047786.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_1278 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047822.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_1591 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047809.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_1395 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047820.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain RJ1267 chromosome_complete genome	Staphylococcus aureus	100.00%	CP047321.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain NCCP_16830 chromosome_complete genome	Staphylococcus aureus	100.00%	CP043843.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER02826.3 chromosome_complete genome	Staphylococcus aureus	100.00%	CP030661.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER01533.3 chromosome_complete genome	Staphylococcus aureus	100.00%	CP030675.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER03710.3 chromosome_complete genome	Staphylococcus aureus	100.00%	CP030646.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER01803.3 chromosome_complete genome	Staphylococcus aureus	100.00%	CP030693.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER15890.6A chromosome_complete genome	Staphylococcus aureus	100.00%	CP075572.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER15890.5A chromosome_complete genome	Staphylococcus aureus	100.00%	CP075574.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER15890.4A chromosome_complete genome	Staphylococcus aureus	100.00%	CP075576.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER15890.3A chromosome_complete genome	Staphylococcus aureus	100.00%	CP075578.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain RGB-095930 chromosome_complete genome	Staphylococcus aureus	100.00%	CP077098.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain HL18380 chromosome_complete genome	Staphylococcus aureus	100.00%	CP080553.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain HL17078 chromosome_complete genome	Staphylococcus aureus	100.00%	CP080556.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BSN14S3 chromosome	Staphylococcus aureus	100.00%	CP080055.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BSN14RB chromosome	Staphylococcus aureus	100.00%	CP080054.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BSN14R1 chromosome	Staphylococcus aureus	100.00%	CP080053.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain FDAARGOS_1245 chromosome_complete genome	Staphylococcus aureus	100.00%	CP069565.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain FDAARGOS_1246 chromosome_complete genome	Staphylococcus aureus	100.00%	CP069470.1
<input checked="" type="checkbox"/>	Staphylococcus sp. T93 chromosome_complete genome	Staphylococcus sp. T93	100.00%	CP068247.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain WH39 chromosome_complete genome	Staphylococcus aureus	100.00%	CP060491.1
<input checked="" type="checkbox"/>	Staphylococcus aureus LAC chromosome_complete genome	Staphylococcus aureus LAC	100.00%	CP055225.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain WH52 chromosome_complete genome	Staphylococcus aureus	100.00%	CP060584.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BLR-DV chromosome_complete genome	Staphylococcus aureus	100.00%	CP058312.1

Figure 25 : Liste des séquences d'alignement significatif du gène *SigB*.

Résultats

	Description	Scientific Name	Per. Ident	Accession
<input checked="" type="checkbox"/>	Staphylococcus aureus TPS3156 DNA, complete genome	Staphylococcus aureus	100.00%	AP023034.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain 16405 chromosome, complete genome	Staphylococcus aureus	100.00%	CP053353.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain 14732 chromosome, complete genome	Staphylococcus aureus	100.00%	CP053639.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain SA1428 chromosome, complete genome	Staphylococcus aureus	100.00%	CP048431.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_403 chromosome, complete genome	Staphylococcus aureus	100.00%	CP047849.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_996 chromosome, complete genome	Staphylococcus aureus	100.00%	CP047830.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_1591 chromosome, complete genome	Staphylococcus aureus	100.00%	CP047809.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_967 chromosome, complete genome	Staphylococcus aureus	100.00%	CP047789.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain UP_1395 chromosome, complete genome	Staphylococcus aureus	100.00%	CP047820.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain RJ1267 chromosome, complete genome	Staphylococcus aureus	100.00%	CP047321.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER02826.3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP030661.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER01533.3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP030675.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER03710.3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP030646.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER01803.3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP030693.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER00573.3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP030681.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain ER01570.3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP030653.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain RGB-095930 chromosome, complete genome	Staphylococcus aureus	100.00%	CP077098.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain HL18380 chromosome, complete genome	Staphylococcus aureus	100.00%	CP080553.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain HL17078 chromosome, complete genome	Staphylococcus aureus	100.00%	CP080556.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BSN14S3 chromosome	Staphylococcus aureus	100.00%	CP080055.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BSN14RB chromosome	Staphylococcus aureus	100.00%	CP080054.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BSN14R1 chromosome	Staphylococcus aureus	100.00%	CP080053.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain V8 chromosome, complete genome	Staphylococcus aureus	100.00%	CP079715.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain FDAARGOS_1245 chromosome, complete genome	Staphylococcus aureus	100.00%	CP069565.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain FDAARGOS_1246 chromosome, complete genome	Staphylococcus aureus	100.00%	CP069470.1
<input checked="" type="checkbox"/>	Staphylococcus sp. T93 chromosome, complete genome	Staphylococcus sp. T93	100.00%	CP068247.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain WH39 chromosome, complete genome	Staphylococcus aureus	100.00%	CP060491.1
<input checked="" type="checkbox"/>	Staphylococcus aureus LAC chromosome, complete genome	Staphylococcus aureus LAC	100.00%	CP055225.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain WH52 chromosome, complete genome	Staphylococcus aureus	100.00%	CP060584.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain MRSA - AMRF 3 chromosome, complete genome	Staphylococcus aureus	100.00%	CP062471.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain MRSA - AMRF 5 chromosome, complete genome	Staphylococcus aureus	100.00%	CP062467.1
<input checked="" type="checkbox"/>	Staphylococcus aureus strain BLR-DV chromosome, complete genome	Staphylococcus aureus	100.00%	CP058312.1

Figure 26 : Liste des séquences d'alignement significatif du gène *CodY*.

2.5- Arbres phylogénétiques

Suite aux résultats obtenus, après l'analyse *in silico*, des protéines ClfB, SarA et les gènes *SigB*, *CodY* en sollicitant les programmes BLASTp et BLASTn, nous avons obtenus des arbres phylogénétiques sous forme des dendrogrammes à l'aide de la plate-forme NCBI (travail exécuté en date du 07 juin 2022).

Pour les protéines ClfB et SarA nous avons utilisé la méthode basée sur le calcul de l'évolution minimale (EM) (figure 27 et 28). Pour les gènes *SigB* et *codY* la méthode s'est basée sur l'algorithme (NJ, neighbor joining) reliant les paires les plus proches (figures 29 et 30).

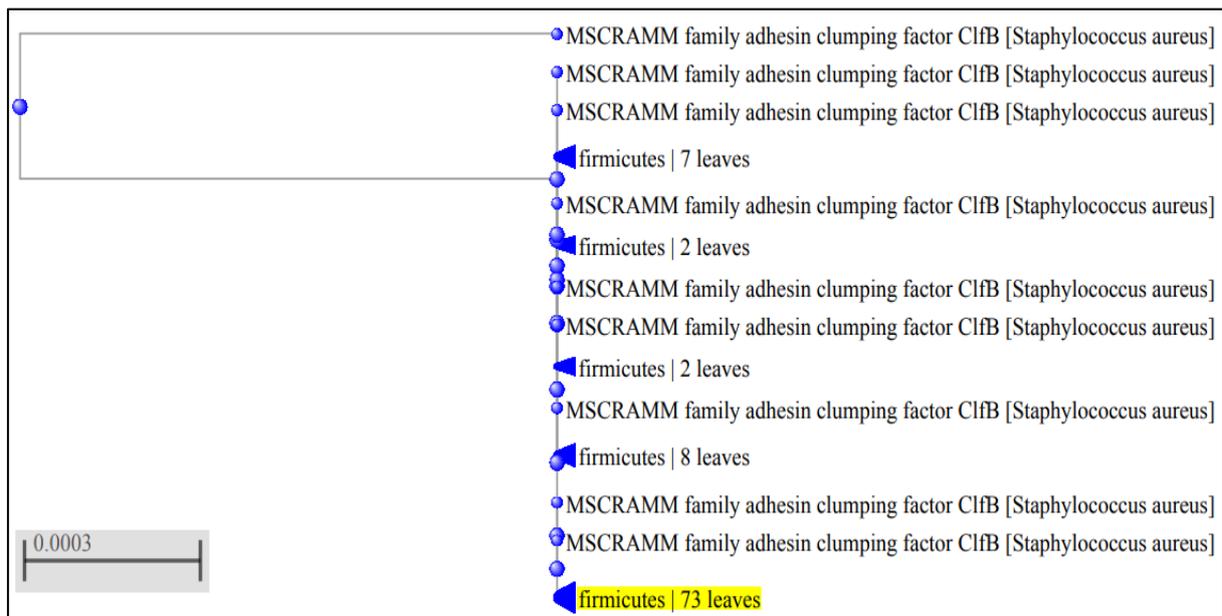


Figure 27 : Arbre phylogénétique de la protéine ClfB

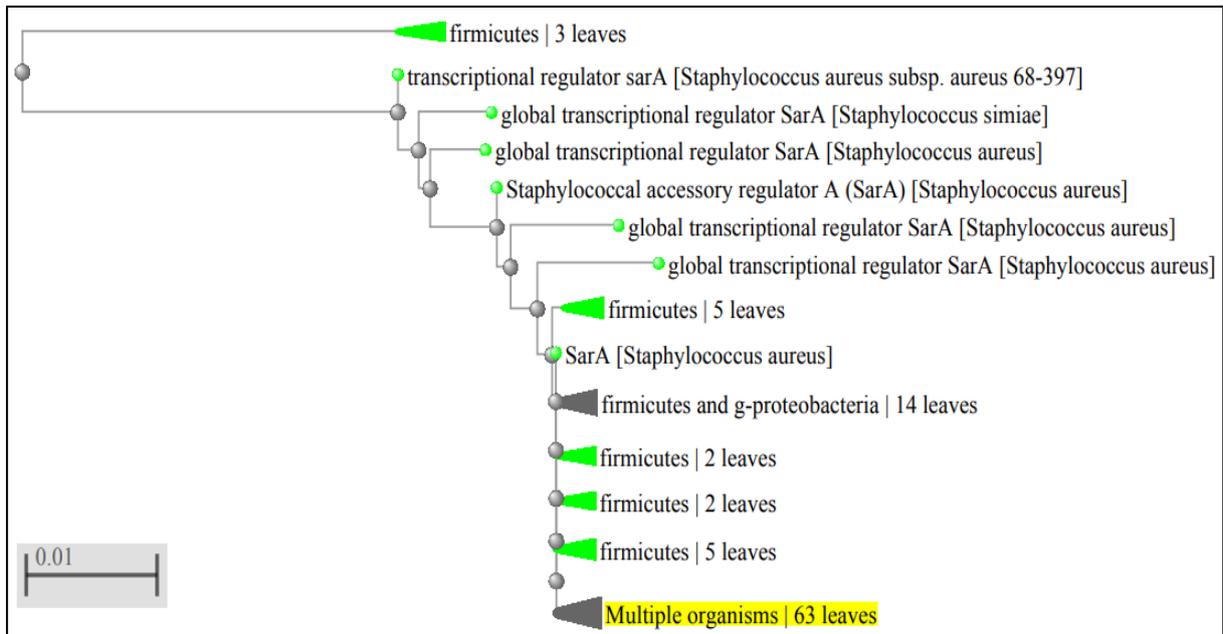


Figure 28 : Arbre phylogénétique de la protéine SarA.

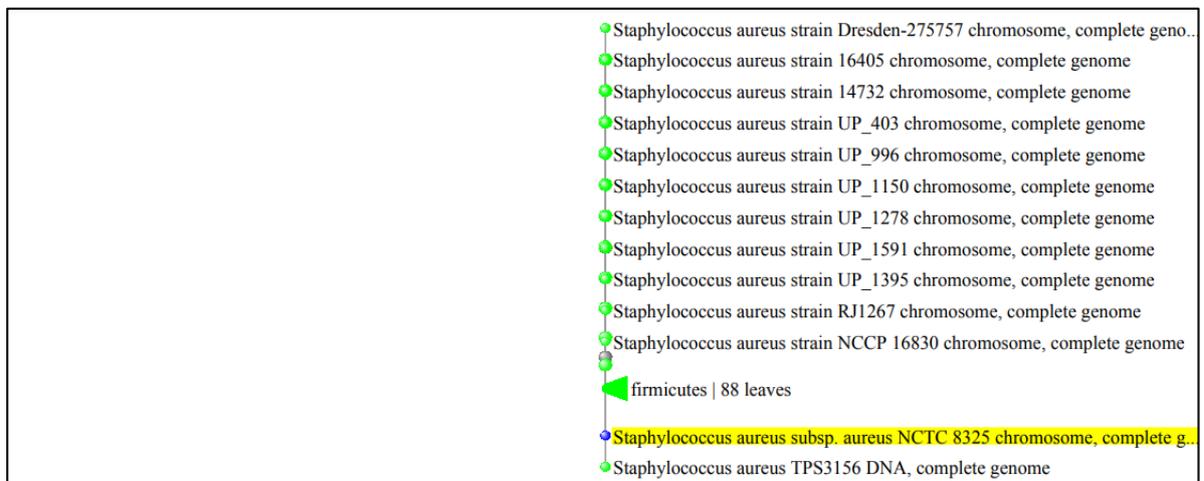


Figure 29 : Arbre phylogénétique du gène *SigB*.

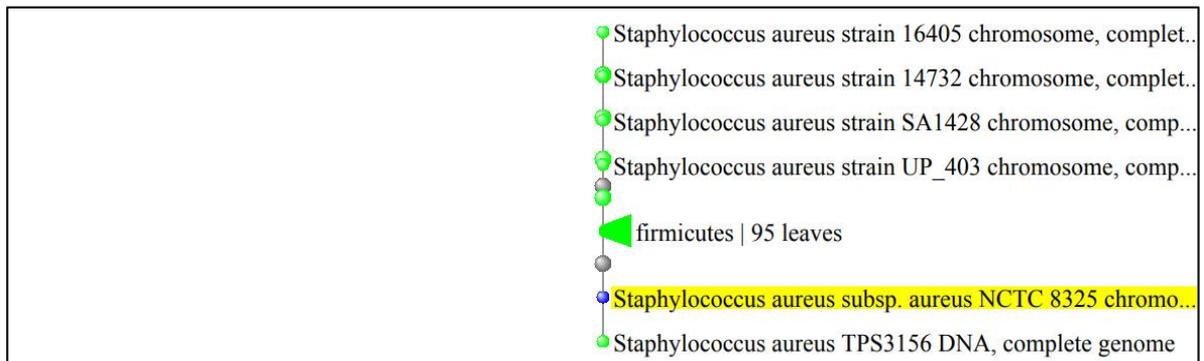


Figure 30 : Arbre phylogénétique du gène *CodY*.

3. Discussion

D'après nos résultats, les arbres phylogénétiques basés sur les méthodes de calcul des distances, des deux protéines analysées *in silico* « ClfB et SarA », présentent une homologie d'une similarité qui pourrait provenir d'un ancêtre commun (Figures 27 et 28). Avec un tel constat, la prédiction phylogénétique obtenue peut suggérer qu'il s'agisse d'une évolution divergente selon les distances génétiques calculées. Les espèces d'une origine ancestrale commune peuvent développer des structures homologues mais avec des fonctions différentes, qui a généré un évènement de spéciation due probablement à une pression évolutive (mutation ou transfert de gènes). Il est utile de rappeler que le processus de spéciation est considéré comme étant un évènement évolutif permettant l'apparition de deux nouvelles espèces (ou plus) à partir d'une espèce ancestrale (Duranton, 2019). Cet évènement est souvent expliqué par l'aboutissement des actions de pressions (écologique et nutritionnelle), conduisant à des évolutions adaptatives à la suite de remaniements génomiques.

Ainsi, Lindsay, (2014) a rapporté dans ses travaux, que l'acquisition ou la perte de MGE (Mobile Genome Express) est un événement important dans l'évolution du génome de *S. aureus*. Un tel évènement évolutif peut se faire par transduction (par des bactériophages, présents dans de nombreux isolats), moins fréquemment par conjugaison (par des plasmides conjugatifs, présents chez certains isolats), et rarement par transformation.

En outre, McCarthy et *al.* (2012) ont cité que le transfert horizontal (HGT : pour Horizontal Gene Transfer) de ces MGE est plus fréquent dans le cas de *S. aureus*, chez les isolats de la même lignée que dans les situations de lignées différentes. La notion de lignée est fondamentale dans l'orientation des évènements mutationnels. Ils ont mis en évidence que des mutations ponctuelles sont également impliquées dans l'évolution du génome de cette bactérie, qui peuvent générer des modifications de la fonction ou de l'expression des gènes, mais en général l'effet dans la cellule est faible. De même, Viana et *al.* (2015) ont indiqué que des mutations ponctuelles dans le génome pourraient modifier le tropisme de l'hôte de ce microorganisme.

La protéine ClfB, une protéine de surface de la famille MSCRAMM (Microbial Surface Components Recognizing Adhesive Matrix Molecules), décrite comme un facteur (B) d'agglutination se liant au fibrinogène, favorisant l'agglutination bactérienne dans le fibrinogène soluble (Xiang et *al.*, 2012). Les travaux de Kang et *al.* (2022) sur la prédiction structurelle de cette protéine ClfB chez *S. aureus* (figure21), ont indiqué qu'elle contenait les structures les plus secondaires de repliement (45,03%), suivies d'un enroulement irrégulier (36,65%), d'un

virage (11,49%) et d'une hélice (5,59%), et la liaison cavité de ClfB comprenait 20 acides aminés, comme L82, D84, D91, T100, D120 et G201. S'ajoute les travaux de Xiang et al. (2012) sur la comparaison structurelle de la protéine ClfB qui ont révélé un motif de liaison riche en glycine-sérine (GSR) conservé dans les ligands, qui a également été trouvé dans d'autres protéines humaines telles que la protéine Engrailed, les protéines TCF20 et Dermokine.

Une étude récente sur l'analyse différentielle et les analyses d'enrichissement ont révélé un enrichissement significatif du gène *ClfB* dans les voies de la pathogenèse et de l'infection par *S. aureus*, ce qui a montré que ce gène possède un rôle essentiel dans le réseau PPI (l'interaction protéine-protéine) et, qui est un gène clé pour la formation de biofilm (Kang et al., 2022).

De plus, Kang et al. (2022) ont identifié 5 sites d'acides aminés de ClfB (D272, R331, I379, K391, E490), potentiels en liaison à la quercétine (QEN). Ces liaisons influencent indirectement les modifications des sites fonctionnels N234, D270, Y273, F328, inhibant la formation du biofilm. Aussi, ils ont mentionné dans leur résultat que le gène *ClfB* était de plus en plus exprimé lors de la formation du biofilm et était impliqué dans l'adhésion cellulaire, la pathogénicité et l'infection.

Nos résultats sur la protéine ClfB exprimés par le bilan BLASTp (figure 23), ont montré que toutes les séquences présentent un pourcentage d'identité égal ou supérieur à 99,82%. Une telle compatibilité identitaire est mentionnée dans les travaux de Lacey et al. (2019), qui ont noté que la séquence de la protéine ClfB varie entre différents complexes clonaux de *S. aureus*. Ils ont souligné que les protéines variantes partagent au moins 94% de similitudes en acides aminés, de plus, le gène *ClfB* est porté par la plupart des souches de *S. aureus*.

Par Ailleurs, Lacey et al. (2019) ont indiqué que la protéine ClfB est une clé déterminante de la colonisation nasale et de l'infection cutanée par *S. aureus* grâce à des interactions de haute affinité avec l'enveloppe cornée dans les narines antérieures. La caractéristique de liaison multiligand de ClfB serait précieuse pour développer de nouvelles stratégies thérapeutiques, telle que l'utilisation du ClfB comme antigène vaccinal ciblé contre les SSTI (Infections de la peau et des tissus mous).

La protéine SarA a été mise en évidence en 1992 par Cheung. Le gène *SarA* est un régulateur majeur chez *S. aureus*, il est sous le contrôle de trois promoteurs (P1, P3 et P2) mais qui produit une seule protéine : SarA. La région promotrice de cette protéine recouvre une grande région d'environ 800 pb. Les promoteurs de SarA sont régulés par différents facteurs : le promoteur

P1, P2 et P3. La production de SarA est constitutive quelles que soient les conditions de cultures de *S. aureus* (Ballal et Manna, 2009).

Les membres de la famille SarA sont des régulateurs transcriptionnels de l'expression de gènes, capables de se réguler entre eux et également de participer à la régulation transcriptionnelle de leur propre gène. Ils partagent également une homologie de séquence avec la protéine MarR ou des homologues chez les bactéries Gram-négative (Mauro, 2017).

SarA est une protéine d'une masse moléculaire d'environ 14,5kDa (124 acides aminés) qui se caractérise par une structure 3D (figure 22) en « winged-helix », composée de 3 hélices α et de 3 feuillets β , qui permettent la liaison de la protéine sur des acides nucléiques. Les protéines «winged-helix» sont les facteurs de transcription les plus retrouvés chez les procaryotes (Aravind et *al.*, 2005).

Nos résultats sur la protéine SarA, après exécution du bilan BLASTp (figure 24) ont montré une variation où la majorité de séquences présentent un pourcentage d'identité de similitude de 100% et le plus faible de 89,80%. Le processus de variation était expliqué par Mauro, (2017) en citant que SarA, comme d'autres membres de sa famille (MgrA, SarZ), est capable de subir différentes modifications post-traductionnelles. De même, Il a montré que le gène *SarA* (facteur de transcription) active l'expression de 76 gènes et réprime 44 gènes inclus dans la régulation des gènes (*agr*, *fnbB*, *hla*, *spa*, etc.) ; donc une régulation globale de SarA peut impliquer d'autres régulateurs transcriptionnels.

SarA favorise l'antibiorésistance de *S. aureus*, qui présente une plus forte sensibilité à une certaine classe d'antibiotiques (les β -lactamines, ciprofloxacine et l'oxacilline) et le développement de cette résistance nécessite l'activation par *SarA* et *mecA* (Mauro, 2017).

Le facteur de transcription SarA est indispensable à la formation de biofilm par l'activation transcriptionnel de l'opéron *icaADBC*. Ce facteur de transcription active également l'expression du gène *bap* qui produit une seconde protéine impliquée dans la formation de biofilm. Ainsi, SarA est un acteur majeur pour la production du biofilm bactérien, en activant les gènes codant pour les protéines nécessaires à sa formation. SarA est le premier membre de la famille des protéines SarA qui joue un rôle vital dans la régulation des gènes de virulence par l'activation de la transcription de *hla* qui produit l'hémolysine α , une toxine qui induit la formation de pores sur les cellules environnantes, plus deux autres toxines Seb et Tsst-1 (Mauro, 2017).

D'après nos résultats, les arbres phylogénétiques des deux gènes, « *SigB* et *CodY* » présentent une analogie qui exprime une similarité qui ne provient pas d'un ancêtre commun donc c'est une évolution convergente (figures 29 et 30).

Pour le gène *SigB* ou facteur alternatif sigma B (σB), il module la réponse au stress de plusieurs bactéries Gram-positives, y compris *S. aureus*. *SigB* est responsable de la transcription de gènes qui peuvent conférer une résistance au stress thermique, oxydatif et antibiotique (Tuchscher et al., 2015). Le système *SigB* est lié au réseau de régulation complexe de *S. aureus*. Il régule directement ou indirectement l'expression de plus de 100 gènes, la maintenance de la paroi cellulaire, le métabolisme intermédiaire, le transport membranaire et la virulence sous des contraintes environnementales, tels que le pH alcalin et acide, le choc thermique, le peroxyde d'hydrogène et les antibiotiques de la paroi cellulaire (Valle et al., 2019).

Nos résultats sur le gène *SigB* par le bilan BLASTn (figure 25) ont montré que toutes les séquences présentent un pourcentage d'identité de 100%. Les travaux de Niles et al. (2009) mentionnent bien que toutes les souches analysées chez *S. aureus* semblent partager le même régulateur *SigB*.

En outre, Valle et al. (2019) ont indiqué que la présence de *SigB* est nécessaire pour construire la matrice du biofilm. Ce facteur contrôle les niveaux de protéines « Ica » par la régulation positive de l'expression d'*icaADBC* au niveau transcriptionnel, il est aussi impliqué dans la régulation de la virulence et de la pigmentation.

De plus, Tuchscher et al. (2015) ont cité que l'opéron *SigB* est un facteur crucial d'adaptation dans les infections chroniques par la favorisation de la persistance intracellulaire bactérienne, qui représente donc une cible possible pour développer des stratégies préventives et thérapeutiques contre les infections chroniques et réfractaires à la thérapie.

Le gène *CodY*, est un régulateur transcriptionnel global qui contrôle directement ou indirectement l'expression de dizaines de gènes dans une variété de bactéries Gram-positives à faible teneur en G+C. Chez les agents pathogènes, y compris *S. aureus*. De plus il relie la disponibilité des nutriments, le métabolisme et ajuste l'expression des gènes de virulence en détectant les acides aminés à chaîne ramifiée (isoleucine, leucine, valine [ILV]) et le GTP. Lors de la liaison de ces métabolites clés, *CodY* est activé en tant que protéine de liaison à l'ADN. *CodY* réprime le locus *Agr*, en bloquant efficacement le signal régulateur de détection de quorum sensing même à faible densité de population tant que les nutriments sont abondants (Mlynek et al., 2018).

L'étude *in vitro* de Montgomery et al. (2012) sur l'effets de *CodY* sur l'expression génique chez *S. aureus* à montrer que la régulation de *CodY* comprend 150 à 200 gènes. La majorité de ces gènes codent pour des protéines qui contrôlent les voies métaboliques, mais *CodY* réprime également de nombreux gènes de virulence au cours de la croissance *in vitro*, et régule l'expression de certains gènes de virulence par liaison directe à l'ADN et régule d'autres indirectement.

Nos résultats sur le gène *CodY* prédits par le bilan BLASTn (figure 26) ont montré que toutes les séquences présentent un pourcentage d'identité de 100%, sont conforme à l'étude réalisée *in silico* de Montgomery et al. (2012) qui ont constaté que la séquence de *CodY* est hautement conservée dans plus de 120 isolats de *S. aureus*.

Egalement, une étude récente de Mlynek et al. (2020) a prouvé que chez *S. aureus* le *CodY* contrôle la production d'importants facteurs de virulence, y compris les enzymes sécrétées et les toxines qui permettent à la bactérie de libérer les nutriments du tissu hôte. De plus, *CodY* régule l'expression des gènes qui module la matrice du biofilm. Il régule positivement les protéines MSCRAMM FnbAB et SasG, qui facilitent la fixation initiale de la cellule bactérienne au tissu hôte, nécessaire à la formation de biofilm et il contrôle les gènes (*ica*, *spsA*, *nuc*, et *hly*). En revanche, *CodY* réprime l'expression des gènes codant pour les protéases sécrétées qui ont un impact négatif sur la formation de biofilm et réprime l'expression de la nucléase (Nuc), qui est nécessaire au stade d'exode du développement du biofilm.

Conclusion et perspectives

Staphylococcus aureus est considéré comme un agent pathogène opportuniste redoutable et une menace infectieuse majeure pour les humains et les animaux, il est responsable d'une diversité d'infections entraînant une morbidité allant des infections superficielles bénignes aux infections sévères à une mortalité élevée. En qualité d'agent pathogène nosocomial *S. aureus* peut causer un large spectre de pathologies en raison de sa grande capacité à former des biofilms.

Comme toutes les bactéries formant des biofilms, son développement chez *S. aureus*, en situation communautaire polymicrobienne, est hautement structuré et intégré dans une matrice adhérente aux surfaces, il s'agit d'un processus complexe qui protège cette bactérie en lui permettant de survivre dans des conditions environnementales hostiles. La formation de biofilm est un comportement de groupe « social » et chaque étape, de l'attachement initial à la diffusion du biofilm mature, est étroitement régulée par des mécanismes moléculaires (Parsek et Greenberg, 2005).

La pathogénèse de *S. aureus* est le résultat de l'expression de différents facteurs de virulence, telles que les toxines, les immunomodulateurs et les exoenzymes. L'expression de ces facteurs, associés à la formation du biofilm, est régulée par des protéines, telles que ClfB et SarA et codée par des gènes particuliers comme *SigB* et *CodY* (Oogai et al., 2011).

Nos résultats, issus de notre investigation *in silico*, indiquent des structures tridimensionnelles bien définies chez les protéines ClfB et SarA. Ces structures ont un rôle majeur dans la régulation et le contrôle de l'expression de gènes de virulence codant pour la formation de biofilms.

Les résultats du bilan BLASTn des gènes *SigB* et *CodY*, chez *S. aureus*, présentent un pourcentage d'identité de 100% au niveau chromosomique bactérien. Ces gènes régulent directement ou indirectement l'expression de plusieurs gènes : 100 gènes pour le réseau de régulation *SigB* et de 150 à 200 gènes pour le régulateur transcriptionnel global *CodY*. La majorité de ces gènes codent pour des protéines qui contrôlent les voies métaboliques et la production d'importants facteurs de virulence, y compris les enzymes sécrétées et les toxines. Egalement ils régulent l'expression des gènes modulant la matrice du biofilm qui assure sa formation. Les biofilms sont une cause majeure de résistance aux antimicrobiens et de complications dans le traitement des infections liées à *S. aureus*, ce qui nécessite des nouvelles stratégies thérapeutiques.

Le développement de la bio-informatique pour manipuler et analyser les données a facilité les analyses génomiques à haut débit, transcriptomiques et protéomiques de la croissance microbienne et de la pathogénèse.

Notre étude *in silico*, en utilisant des outils d'analyse et de prédiction des structures tridimensionnelles a permis de réaliser notre investigation phylogénique moléculaire sur les protéines Clfb et SarA, ainsi que les gènes *SigB* et *CodY*. L'approfondissement dans les connaissances relatives aux aspects moléculaires pourrait initier de nouvelles réflexions sur les stratégies de diagnostic pour des alternatives thérapeutiques prometteuses. A titre d'exemple, le cas de la prédiction de la structure 3D de la protéine ClfB où la caractéristique de liaison multiligand serait très précieuse dans l'utilisation du ClfB comme antigène vaccinal (Lacey et *al.*, 2019). Egalement le cas de SarA qui favorise l'antibiorésistance à plusieurs classe d'antibiotiques, le développement de cette résistance nécessite l'activation par *SarA* alors son désactivation peut être une nouvelle stratégie contre l'antibiorésistance (Mauro, 2017).

La compréhension des aspects moléculaires des gènes *SigB* et *CodY* représente une cible possible pour développer des stratégies préventives et thérapeutiques contre les infections chroniques et réfractaires à la thérapie, comme la sélection d'antigènes appropriés pour un vaccin efficace contre les infections causées par *S. aureus*.

Références bibliographiques

- Abdullahi SO.** 2019. Data Mining Techniques: Review. In : *International Journal of Data Science*. Volume 2, Issue 1.
- Agostinho A, James G, Wazni O, Citron M, Wilkoff BD.** 2009. Inhibition of *Staphylococcus aureus* Biofilms by a Novel Antibacterial Envelope for Use with Implantable Cardiac Devices. *Clin. Transl. Sci.* **2**:193–198. doi: 10.1111/j.1752-8062.2009.00123.x.
- Alexandra F.** 2017. *Staphylococcus aureus*. Ed Alexandra Department of Biological Safety, German Federal Institute for Risk Assessment (BfR), Berlin, Germany. Academic Press. p316. ISBN 978-0-12-809671-0.
- Al-Shuneigat J, Al-Sarayreh S, Al-Saraireh Y, Al-Tarawneh I, Al-Qudah M, Albataineh E.** 2014. Effects of wild Artemisia herba-alba essential oil on biofilm forming bacteria. *Br J Pharm Res.*, 4(19), 2273-2280
- Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM.** 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**, 231-262.
- Archer NK, Mazaitis MJ, Costerton WJ, Leid JG, Powers ME, Shirtliff ME.** 2011. *Staphylococcus aureus* biofilms Properties, regulation and roles in human disease. In : *Landes Bioscience*. Vol 2:5. p445-459.
- Ardeshir B.** 2002. Bioinformatics. In : *Science, medicine, and the future*. 324(7344): 1018–1022. doi: [10.1136/bmj.324.7344.1018](https://doi.org/10.1136/bmj.324.7344.1018).
- Avril JL, Dabernat H, Denis F, Monteil H.** 1992. Bactériologie clinique. 2 nd Edition, Ellipses, Paris.
- Avril JL, Dabernat H, Denis F, Monteil H.** 2000. Bactériologie clinique. Caractères généraux des *Staphylococcus aureus*. Ellipses, édition Paris. P. 7- 28.
- Balaban N, Ren D, Givskov M, Rasmussen TB.** 2008. introduction In "Balaban N., Control of biofilm infections by signal manipulation, Ed. Springer, 06".
- Ballal A et Manna AC.** 2009. Expression of the sarA family of genes in different strains of *Staphylococcus aureus*. *Microbiology* **155**, 2342-2352.
- Bhattacharya M, Wozniak DJ, Stoodley P, Hall-Stoodley L.** 2015. Prevention and treatment of *Staphylococcus aureus* biofilms. *Expert Rev. Anti. Infect. Ther.* 2015;**13**:1499–1516. doi: 10.1586/14787210.1100533.
- Botham KM, Weil A, Rodwell VW, Kennelly PJ, Bender DA.** 2017. Biochimie de harper. De Boeck Superieur.
- Burley S K, Berman HM, Christie C, Duarte JM, Feng Z, Westbrook J, Young J, Zardecki C.** 2017. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. In *Protein Sci.* **27**(1): 316–330. doi: [10.1002/pro.3331](https://doi.org/10.1002/pro.3331).
- Chen Q, Xie S, Lou X, Cheng S, Liu X, Zheng W, Zheng Z, Wang H.** 2020. Biofilm formation and prevalence of adhesion genes among *Staphylococcus aureus* isolates from different food sources. *Microbiologyopen*. **9**:e00946. doi: 10.1002/mbo3.946.
- Deléage G et Gouy M.** 2015. Bioinformatique-2e édition: Cours et applications. Dunod.
- Denis F, Ploy MC, Martin C, Bingen E, Quentin R.** 2011. 2nd Edition. P. 287-320.
- Donlan RM.** 2000. Role of biofilms in antimicrobial resistance. *ASAIO J.* **46**:S47–S52. doi: 10.1097/00002480-200011000-00037.
- Duranton M.** 2019. Intégrer des approches expérimentales et d'évolution moléculaire en génomique de la spéciation afin d'identifier les mécanismes impliqués dans la divergence entre

bar atlantique et loup méditerranéen. Sciences agricoles. Université Montpellier. Français. NNT : 2019MONTG031. tel-02481233.

Fleischmann RD, Adams MD, White O. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512.

Foster TJ, Geoghegan JA, Ganesh VK, Höök M. 2014. Adhesion, invasion and evasion: The many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* 12:49–62. doi: 10.1038/nrmicro3161.

Foster TJ. 2019. The MSCRAMM family of cell-wall-anchored surface proteins of Gram-positive cocci. *Trends Microbiol* 27:927–941. doi: 10.1016/j.tim.2019.06.007.

Ghias W, Sharif M, Yazdani M, Ansari F, Rabbani M. 2016. Isolation and identification of Methicillin and Vancomycin Resistance *Staphylococcus aureus* From Pus Samples of Injured Skin Patients in Lahore, Pakistan. *Biomed Lett* 2(2):103-112.

Giancarlo L. 2013. Étymologie : *Staphylococcus*. *Emerg Infect Dis* [En ligne]. <http://dx.doi.org/10.3201/eid1909.ET1909>.

Gibson G et Muse SV. 2004. Précis de génomique. 1re Édition. De Boeck Supérieur. p347.

Gimza BD, Larias MI, Budny BG, Shaw LN. 2019. Mapping the global network of extracellular protease regulation in *Staphylococcus aureus*. *mSphere* 4:e00676-19. doi:10.1128/mSphere.00676-19.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333–51.

Gordon RJ, Lowy FD. 2008. Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection. *Clin. Infect. Dis.* 46:S350–S359. doi: 10.1086/533591.

Grundstad ML, Parlet CP, Kwiecinski JM, Kavanaugh JS, Crosby HA, Cho YS, Heilmann K, Diekema DJ, Horswill AR. 2019. Quorum Sensing, Virulence, and Antibiotic Resistance of USA100 Methicillin-Resistant *Staphylococcus aureus* Isolates. *mSphere*. doi: 10.1128/mSphere.00553-19.

Hallam S. 2013. Life Out of Sequence: A Data-driven History of Bioinformatics. Ed The University of Chicago Press, 294 pages.

Herman-Bausier P, Labate C, Towell AM, Derclaye S, Geoghegan JA, Dufrêne YF. 2018. *Staphylococcus aureus* clumping factor A is a force-sensitive molecular switch that activates bacterial adhesion. *Proc. Natl. Acad. Sci. USA.* 115:5564. doi: 10.1073/pnas.1718104115.

Hoogstrate Y, Jenster WJ, Werken HJG. 2021. FASTAFS : fle system virtualisation of random access compressed FASTA fles. In : *BMC Bioinformatics.* 22:535. <https://doi.org/10.1186/s12859-021-04455-3>.

Idrees M, Sawant S, Karodia N, Rahman A. 2021. *Staphylococcus aureus* Biofilm: Morphology, Genetics, Pathogenesis and Treatment Strategies. *Int. J. Environ. Res. Public Health*, 18, 7602. <https://doi.org/10.3390/ijerph18147602>.

Kang X, Ma Q, Wang G, Li N, Mao Y, Wang X, Wang Y, Wang G. 2022. Potential Mechanisms of Quercetin Influence the ClfB Protein During Biofilm Formation of *Staphylococcus aureus*. *Front. Pharmacol.* 13:825489. doi: 10.3389/fphar.2022.825489.

Kavanaugh JS, Horswill AR. 2016. Impact of environmental cues on staphylococcal quorum sensing and biofilm development. *J Biol Chem* 291:12556–12564. doi:10.1074/jbc.R116.722710.

Kiedrowski MR, Kavanaugh JS, Malone CL, Mootz JM, Voyich JM, Smeltzer MS, Bayles

- KW, Horswill AR.** 2011. Nuclease modulates biofilm formation in community-associated methicillin-resistant *Staphylococcus aureus*. PLoS One **6**:e26714. doi:10.1371/journal.pone.0026714.
- Kot B, Sytykiewicz H, Sprawka I.** 2018. Expression of the Biofilm-Associated Genes in Methicillin-Resistant *Staphylococcus aureus* in Biofilm and Planktonic Conditions. *Int. J. Mol. Sci.* **19**:3487. doi: 10.3390/ijms19113487.
- Kumar A, Chordia N.** 2017. Role of Bioinformatics in Biotechnology. *Res Rev Biosci.* **12**(1):116. 2017 Trade Science Inc.
- Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya, K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K.** 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357** (9264), 1225–1240Springer, USA, 350.
- Kwiecinski JM, Crosby HA, Valotteau C, Hippensteel JA, Nayak MK, Chauhan AK, Schmidt EP, Dufrêne YF, Horswill AR.** 2019. *Staphylococcus aureus* adhesion in endovascular infections is controlled by the ArlRS-MgrA signaling cascade. PLoS Pathog **15**:e1007800. doi:10.1371/journal.ppat.1007800.
- Lacey KA, Mulcahy ME, Towell AM, Geoghegan JA, McLoughlin RM.** 2019. Clumping factor B is an important virulence factor during *Staphylococcus aureus* skin infection and a promising vaccine target. PLoS Pathog **15**(4): e1007713. <https://doi.org/10.1371/journal.ppat.1007713>
- Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano J, Armañanzas R, Santafé G, Pérez A, Robles V.** 2006. Machine learning in bioinformatics, *Briefings in Bioinformatics*, Volume 7, Issue 1, Pages 86–112, <https://doi.org/10.1093/bib/bbk007>.
- Le KY, Otto M.** 2015. Quorum-sensing regulation in staphylococci-an overview. *Front Microbiol.* **6**:1174. doi:10.3389/fmicb.2015.01174.
- Le KY, Villaruz AE, Zheng Y, He L, Fisher EL, Nguyen TH, Ho TV, Yeh AJ, Joo HS, Cheung GYC, Otto M.** 2019. Role of phenol-soluble modulins in *Staphylococcus epidermidis* biofilm formation and infection of indwelling medical devices. *J Mol Biol* **431**:3015–3027. doi:10.1016/j.jmb.2019.03.030.
- Lebeaux D et Ghigo JM.** 2012. Infections associées aux biofilms, Quelles perspectives thérapeutiques issues de la recherche fondamentale *Med Sci.*, **28**, 727-39.
- Leid JG, Shirtliff ME, Costerton JW, Stoodley P.** 2002. Human leukocytes adhere to, penetrate, and respond to *Staphylococcus aureus* biofilms. *Infect Immun* **70**:6339–6345. doi:10.1128/iai.70.11.6339-6345.2002.
- Lindsay JA.** 2014. *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *Int. J. Med. Microbiol.* **304**, 103–109.
- Lister JL, Horswill AR.** 2014. *Staphylococcus aureus* biofilms: Recent developments in biofilm dispersal. *Front. Cell. Infect. Microbiol.* **4**:178. doi: 10.3389/fcimb.2014.00178.
- Loughran AJ, Atwood DN, Anthony AC, Harik NS, Spencer HJ, Beenken KE, Smeltzer MS.** 2014. Impact of individual extracellular proteases on *Staphylococcus aureus* biofilm formation in diverse clinical isolates and their isogenic sarA mutants. *Microbiologyopen* **3**:897–909. doi:10.1002/mbo3.214.
- Maikranz E, Spengler C, Thewes N, Thewes A, Nolle F, Jung P, Bischoff M, Santen L,**

- Jacobs K.** 2020. Different binding mechanisms of: *Staphylococcus aureus* to hydrophobic and hydrophilic surfaces. *Nanoscale*. **12**:19267–19275. doi: 10.1039/D0NR03134H.
- Masood MA et Khan MNA.** 2015. Clustering Techniques in Bioinformatics. *International journal Modern Education and Computer Science*.7 (1): 38-46. DOI: 10.5815/ijmecs.2015.01.06.
- Mauro T.** 2017. Étude de la régulation transcriptionnelle de deux ARN régulateurs de *Staphylococcus aureus* : Implication d'un facteur de transcription de la famille SarA. Thèse de doctorat. Faculté de Pharmacie. Université Européenne de Bretagne. p188.
- McCarthy AJ, Lindsay JA., Loeffler A,** 2012. Are all meticillin-resistant *Staphylococcus aureus* (MRSA) equal in all hosts Epidemiological and genetic comparison between animal and human MRSA. *Vet. Dermatol.* 23 (4), 267–275.
- Mehmood MA, Sehar U, Ahmad N.** 2014. Use of Bioinformatics Tools in Different Spheres of Life Sciences. *J Data Mining Genomics Proteomics* 5: 158. doi:10.4172/2153-0602.1000158.
- Mlynek KD, Bullock LL, Stone CJ, Curran LJ, Sadykov MR, Bayles KW, Brinsmade SR.** 2020. Genetic and biochemical analysis of CodY-Mediated Cell aggregation in *Staphylococcus aureus* reveals an interaction between extracellular DNA and polysaccharide in the extracellular matrix. *J Bacteriol* **202**:e00593-19. doi:10.1128/JB.00593-19.
- Mlynek KD, Sause WE, Moormeier DE, Sadykov MR, Hill KR, Torres VJ, Bayles KW, Brinsmade SR.** 2018. Nutritional regulation of the Sae two-component system by CodY in *Staphylococcus aureus*. *J Bacteriol* 200:e00012- 18. <https://doi.org/10.1128/JB.00012-18>.
- Momtaz H et Hafezi L.** 2014. Meticillin-resistant *Staphylococcus aureus* isolated from Iranian hospitals: virulence factors and antibiotic resistance properties. *Bosn. J. Basic Med. Sci.*, 14(4) : 219-226.
- Montgomery CP, Boyle-Vavra S, Roux A, Ebine K, Sonenshein AL, Daumb RS.** 2012. CodY Deletion Enhances *In Vivo* Virulence of Community-Associated Methicillin-Resistant *Staphylococcus aureus* Clone USA300. In : *American Society for Microbiology*. p2382–2389. Volume 80 Number 7. doi:10.1128/IAI.06172-11.
- Moormeier DE, Bayles KW.** 2017. *Staphylococcus aureus* biofilm: A complex developmental organism. *Mol. Microbiol.* **104**:365–376. doi: 10.1111/mmi.13634.
- Moormeier DE, Bose JL, Horswill AR, Bayles KW.** 2014. Temporal and stochastic control of *Staphylococcus aureus* biofilm development. *mBio* **5**:e01341-14. doi:10.1128/mBio.01341-14.
- NHGRI.** 2018. Human Genome Project Completion: Frequently Asked Questions. National Human Genome Research Institute (NHGRI). <https://www.genome.gov/11006943/Human-Genome-Project-Completion-Frequently-AskedQuestions>.
- Niles P, Donegan, Ambrose, Cheung L.** 2009. Regulation of the mazEF Toxin-Antitoxin Module in *Staphylococcus aureus* and Its Impact on sigB Expression. In : *journal of bacteriology*. p. 2795–2805. Vol. 191, No. 8. doi:10.1128/JB.01713-0.
- Nourbakhsh F, Namvar AE.** 2016. Detection of genes involved in biofilm formation in *Staphylococcus aureus* isolates. *GMS Hyg. Infect. Control.* **11**:Doc07. doi: 10.3205/dgkh000267.
- O'Neill E, Pozzi C, Houston P, Humphreys H, Robinson DA, Loughman A, Foster TJ, O'Gara JP.** 2008. A Novel *Staphylococcus aureus* Biofilm Phenotype Mediated by the Fibronectin-Binding Proteins, FnBPA and FnBPB. *J. Bacteriol.* **190**:3835. doi: 10.1128/JB.00167-08.

- Oogai Y, Matsuo M, Hashimoto M, Kato F, Sugai M., Komatsuzawa H.** 2011. Expression of virulence factors by *Staphylococcus aureus* grown in serum. *Appl. Environ. Microbiol.* **77**:8097–8105. doi: 10.1128/AEM.05316-11.
- Otto M.** 2014. *Staphylococcus aureus* toxins. *Curr. Opin. Microbiol.* **17**:32–37. doi: 10.1016/j.mib.2013.11.004.
- Otto M.** 2018. Staphylococcal Biofilms. *Microbiol. Spectr.* **6** doi: 10.1128/microbiolspec.GPP3-0023-2018.
- Ou C, Shang D, Yang J, Chen B, Chang J, Jin F, Shi C.** 2020. Prevalence of multidrug-resistant *Staphylococcus aureus* isolates with strong biofilm formation ability among animal-based food in Shanghai. *Food Control.* **112**:107106. doi: 10.1016/j.foodcont.2020.107106.
- Paharik AE, Horswill AR.** 2016. The Staphylococcal Biofilm: Adhesins, Regulation, and Host Response. *Microbiol. Spectr.* ;**4** doi: 10.1128/microbiolspec.VMBF-0022-2015.
- Parsek MR, Greenberg EP.** 2005. Sociomicrobiology: the connections between quorum sensing and biofilms. *Trends Microbiol* **13**:27–33. doi:10.1016/j.tim.2004.11.007.
- Prescott LM, Harley JP, Klein D, Bacq-Calberg CM, Dusart J.** 2002. « Microbiologie », 5^e édition, De Boeck.
- Prescott LM, Harley JP, Klein D.** 2010. Microbiologie. 2^{ème} Edition Française. De Boeck Université.
- Pundir S, Martin MJ, O'Donovan C, The UniProt Consortium.** 2016. UniProt Tools. In *Curr Protoc Bioinformatics.* ;**53**: 1.29.1–1.29.15. doi:10.1002/0471250953.bi0129s53.
- Qi M, Liu Q, Liu Y, Yan H, Zhang Y, Yuan Y.** 2021. *Staphylococcus aureus* biofilm inhibition by high voltage prick electrostatic field (HVPEF) and the mechanism investigation. In *International Journal of Food Microbiology.* Volume 362, 2 February 2022, 109499.
- Rasigade JP, Vandenesch F.** 2014. *Staphylococcus aureus* : a pathogen with still unresolved issues. *Infect Genet Evol.* **21**:510-4.
- Rasmussen RV, Fowler VG, Skov R, Bruun NE.** 2011. Future challenges and treatment of *Staphylococcus aureus* bacteremia with emphasis on MRSA. *Future Microbiol.* **6**:43–56. doi: 10.2217/fmb.10.155.
- Reffuveille F, Josse J, Vallé Q, Gangloff CM, Gangloff SC.** 2017. *The Rise of Virulence and Antibiotic Resistance in Staphylococcus aureus.* InTech; Rijeka, Croatia: *Staphylococcus aureus* Biofilms and their Impact on the Medical Field; p187. Chapter 11.
- Reyes D, Andrey DO, Monod A, Kelley WL, Zhang G, Cheung AL.** 2011. Coordinated regulation by AgrA, SarA, and SarR to control *agr* expression in *Staphylococcus aureus*. *J Bacteriol* **193**:6020–6031. doi:10.1128/JB.05436-11.
- Ricciardi BF, Muthukrishnan G, Masters E, Ninomiya M, Lee CC, Schwarz EM.** 2018. *Staphylococcus aureus* evasion of host immunity in the setting of prosthetic joint infection: biofilm and beyond. *Curr Rev Musculoskelet Med* **11**:389–400. doi:10.1007/s12178-018-9501-4.
- Rizzato C, Torres J, Kasamatsu E, Camorlinga-Ponce M, Bravo MM, Canzian F, Kato I.** 2019. Potential role of biofilm formation in the development of digestive tract cancer with special reference to helicobacter pylori infection. *Front. Microbiol.* **10**:846. doi: 10.3389/fmicb.2019.00846.
- Rzhetsky A et Nei M.** 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution*, **10**(5), 1073–1095. <https://doi.org/10.1093/oxfordjournals.molbev.a040056>.

- Sabat A, Melles DC, Martirosian G, Grundmann H, Van Belkum A, Hryniewicz W.** 2006. Distribution of the serine-aspartate repeat protein-encoding sdr genes among nasal-carriage and invasive *Staphylococcus aureus* strains. *J. Clin. Microbiol.* **44**:1135–1138. doi: 10.1128/JCM.44.3.1135-1138.2006.
- Saggu SK, Jha G, Mishra PC.** 2019. Enzymatic Degradation of Biofilm by Metalloprotease From Microbacterium sp. SKS10. *Front. Bioeng. Biotechnol.* **7**:192. doi: 10.3389/fbioe.2019.00192.
- Saitou N et Nei M.** 1987. The neighbor-joining method : a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* **4**(4).406-425pp. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
- Sauer K.** 2003. The genomics and proteomics of biofilm formation. *Genome Biol.* **4**:219. doi: 10.1186/gb-2003-4-6-219.
- Sayers EW, Beck J, Bolton E, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O’Leary N, Phan L, Rangwala SH, Schneider VA, Skripchenko Y, Wang J, Ye J, Trawick BW, Pruitt KD, Sherry ST.** 2020. Database resources of the National Center for Biotechnology Information. In : *Nucleic Acids Research*. Vol. 49, p10–17. Doi : 10.1093/nar/gkaa892.
- Schilcher K, Horswill AR.** 2020. Staphylococcal biofilm development: structure, regulation, and treatment strategies. *Microbiol Mol Biol Rev* **84**:e00026-19. <https://doi.org/10.1128/MMBR.00026-19>.
- Sharma-Kuinkel BK, Mann EE, Ahn JS, Kuechenmeister LJ, Dunman PM, Bayles KW.** 2009. The *Staphylococcus aureus* LytSR two-component regulatory system affects biofilm formation. *J Bacteriol* **191**:4767–4775. doi:10.1128/JB.00348-09.
- Sneath PHA et Sokal RR.** 1973. Numerical Taxonomy : The Principles and Practice of Numerical Classification, W.H. Freeman and Company, San Francisco, 576 pp.
- Sonesson A, Przybyszewska K, Eriksson S, Mörgelin M, Kjellström S, Davies J, Potempa J, Schmidtchen A.** 2017. Identification of bacterial biofilm and the *Staphylococcus aureus* derived protease, staphopain, on the skin surface of patients with atopic dermatitis. *Sci. Rep.* **7**:8682–8689. doi: 10.1038/s41598-017-08046-2.
- Sooyeon L, Dong-Hoon L, Woori K, Hakdong S, Hye-Jin K, Jong-eun L, Gun EL, Heebal K, Sang-Ho C, Sangryeol R, Ju-Hoon L.** 2014. Comparative Genomic Analysis of *Staphylococcus aureus* FORC_001 and *S. aureus* MRSA252 Reveal the Characteristics of Antibiotic Resistance and Virulence Factors for Human Infection. In : *Journal of Microbiology and Biotechnology*. **25**(1), 98–108. doi: 10.4014/jmb.1410.10005.
- Sooyeon L, Dong-Hoon L, Woori K, Hakdong S, Hye-Jin K, Jong-eun L, Gun Eui L, Heebal K, Sang-Ho C, Sangryeol R, Ju-Hoon L.** 2014. Comparative Genomic Analysis of *Staphylococcus aureus* FORC_001 and *S. aureus* MRSA252 Reveals the Characteristics of Antibiotic Resistance and Virulence Factors for Human Infection. In *Journal of Microbiology and Biotechnology*. **25**(1), 98–108 <http://dx.doi.org/10.4014/jmb.1410.10005>.
- Spicer W.** 2003. Pratique clinique en bactériologie, mycologie et parasitologie. Flammarion Médecine-Sciences, Paris. 28-29.
- Stuessy TF.** 1983. PHYLOGENETIC TREES IN PLANT SYSTEMATICS. *SIDA, Contributions to Botany*, **10**(1), 1–13.
- Tang J, Zhou R, Wang H, Zeng Z.** 2008. Accessory gene regulator in *Staphylococcus* biofilm formation and infection. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* **33**:1066–1070.

- The UniProt Consortium.** 2020. UniProt: the universal protein knowledgebase in 2021. In *Nucleic Acids Res.* ; 49(D1): D480–D489. doi: 10.1093/nar/gkaa1100.
- Tuscherr L, Bischoff M, Lattar SM, Noto Llana M, Pfortner H, Niemann S.** 2015. Sigma Factor SigB Is Crucial to Mediate *Staphylococcus aureus* Adaptation during Chronic Infections. *PLoS Pathog* 11(4): e1004870. doi:10.1371/journal.ppat.1004870.
- Valle J, Echeverz M, Lasa I.** 2019. σ B inhibits poly-N-acetylglucosamine exopolysaccharide synthesis and biofilm formation in *Staphylococcus aureus*. *J Bacteriol* 201:e00098-19. <https://doi.org/10.1128/JB.00098-19>.
- Vázquez-Sánchez, D, Cabo ML, SAÁ-Ibusquiza P, Rodríguez-Herrera JJ.** 2012. Incidence and characterization of *Staphylococcus aureus* in fishery products marketed in Galicia (Northwest Spain). *Int. J. Food Microbiol.* 157, 286–296.
- Vert JP.** 2004/2005. « Introduction à la biologie moléculaire et à la bioinformatique » cours de MasterRecherche M2.
- Vert JP.** 2013. Les applications industrielles de la bioinformatique. In *Annales des Mines-Realites industrielles* (No. 1, pp. 17-23). ESKA.
- Viana D, Comos M, McAdam PR, Ward MJ, Selva L, Guinane CM, González-Muñoz, BM, Tristan A, Foster SJ, Fitzgerald JR, Penadés JR.** 2015. A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nat. Genet.* 47 (4), 361–366.
- Wertheim HF, Melles DC, Vos MC, van Leeuwen W, van Belkum A, Verbrugh HA, Nouwen JL.** 2005. The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect Dis* 5:751–762. doi:10.1016/S1473-3099(05)70295-4.
- Wolz C, Goerke C, Landmann R, Zimmerli W, Fluckiger U.** 2002. Transcription of Clumping Factor A in Attached and Unattached *Staphylococcus aureus* In Vitro and during Device-Related Infection. *Infect. Immun.* 70:2758. doi: 10.1128/IAI.70.6.2758-2762.2002.
- Xiang H, Feng Y, Wang J, Liu B, Chen Y.** 2012. Crystal Structures Reveal the Multi-Ligand Binding Mechanism of *Staphylococcus aureus* ClfB. *PLoS Pathog* 8(6): e1002751. doi:10.1371/journal.ppat.1002751.
- Xiong YQ, Willard J, Yeaman MR, Cheung AL, Bayer AS.** 2006. Regulation of *Staphylococcus aureus* alpha-toxin gene (hla) expression by agr, sarA, and sae in vitro and in experimental infective endocarditis. *J Infect Dis* 194:1267–1275. doi:10.1086/508210.
- Ythier M, Resch G, Waridel P, Panchaud A, Gfeller A, Majcherczyk P, Quadroni M, Moreillon P.** 2012. Proteomic and transcriptomic profiling of *Staphylococcus aureus* surface LPXTG-proteins: correlation with agr genotypes and adherence phenotypes. *Mol Cell Proteomics* 11:1123–1139. doi:10.1074/mcp.M111.014191.
- Yu Q, Lepp D, Gohari IM, Wu T, Zhou H, Yin X, Yu H, Prescott JF, Nie SP, Xie MY.** 2017. The Agr-like quorum sensing system is required for pathogenesis of necrotic enteritis caused by *Clostridium perfringens* in poultry. *Infect. Immun.* 85 doi: 10.1128/IAI.00975-16.