



HAL
open science

Analyse comparative de génomes complets de souches pathogènes et de portage de *Staphylococcus lugdunensis* et caractérisation du système de sécrétion Ess/type VII

Jérémie Lebeurre

► **To cite this version:**

Jérémie Lebeurre. Analyse comparative de génomes complets de souches pathogènes et de portage de *Staphylococcus lugdunensis* et caractérisation du système de sécrétion Ess/type VII. Génétique. Normandie Université, 2018. Français. NNT : 2018NORMR081 . tel-01985179

HAL Id: tel-01985179

<https://theses.hal.science/tel-01985179>

Submitted on 17 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité Sciences de la vie et de la santé

Préparée au sein de l'Université de Rouen Normandie

Analyse comparative des génomes complets de souches pathogènes et de portage de *Staphylococcus lugdunensis* et caractérisation du système de sécrétion Ess/type VII

Présentée et soutenue par
Jérémie LEBEURRE

Thèse soutenue publiquement le 20 Décembre 2018
devant le jury composé de

Madame Sarah DUBRAC	Chargé de Recherche, Institut Pasteur de Paris	Rapporteur
Monsieur Philippe LANOTTE	Professeur, Université de Tours	Rapporteur
Monsieur Patrice FRANÇOIS	Privat-Docent, Hôpital Universitaire de Genève	Examinateur
Madame Laleh MAJLESSI	Directeur de recherche, Institut Pasteur de Paris	Examinateur
Madame Martine PESTEL-CARON	Professeur, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par le Professeur Martine PESTEL-CARON,
Laboratoire EA2656 Groupe de Recherche sur l'Adaptation Microbienne



REMERCIEMENTS

C'est avec un grand plaisir que je rédige ces remerciements qui sonnent la fin de la rédaction de ce manuscrit, mais également la fin de trois années de travail.

J'adresse tout d'abord mes vifs remerciements aux membres du jury qui ont accepté de juger ce manuscrit.

Au Docteur Sarah Dubrac et au Professeur Philippe Lanotte pour me faire l'honneur d'avoir accepté d'être rapporteurs de ce travail de thèse.

Au Docteur Maleh Majlessi et au Docteur Patrice François pour me faire l'honneur d'avoir accepté de participer à ce jury de soutenance de thèse en tant qu'examineurs. Merci à vous Patrice François et Seydina de m'avoir accueilli dans votre laboratoire pour l'apprentissage du langage bio-informatique. Merci pour votre précieuse contribution à ce travail et à toute votre équipe pour m'avoir si bien reçu durant mon séjour à Genève.

Je remercie également le Pr Éric Mallet et la Fondation Charles Nicolle qui ont financé ce projet de recherche, sans qui ces travaux n'auraient pas été possibles.

Au Professeur François Caron, merci de m'avoir accueilli au sein de votre équipe de recherche pour réaliser mon stage de Master 2 et mon doctorat.

Au Docteur Gilles Prévost et au Docteur Xavier Argemi de l'EA-7290 Virulence bactérienne précoce. Un grand merci pour les discussions scientifiques très enrichissantes et nos collaborations très fructueuses.

Au Docteur Olivier Lesouhaitier et à Thomas Clamens du Laboratoire de Microbiologie – Signaux et Micro-environnement (LMSM EA 4312). Merci pour votre importante contribution à ce travail, votre aide et votre disponibilité.

Au Professeur Thierry Frébourg et au Docteur Isabelle Tournier de l'équipe de génétique humaine Inserm U1245 pour leur contribution dans l'initiation du projet de séquençage des génomes de *Staphylococcus lugdunensis*.

Au Professeur Pascal Cosette, au Docteur Julie Hardouin et à Clément Guillou de la Plateforme Instrumentale en Sciences Séparatives et Analytiques de Rouen (PISSARO). Merci pour votre collaboration et pour votre aide.

Lors de cette thèse, j'ai eu la chance de rencontrer de belles personnes avec lesquelles j'ai eu le plaisir de collaborer.

Au Professeur Martine Pestel-Caron pour avoir dirigé cette thèse. Martine, je ne saurai jamais assez vous remercier pour votre soutien, votre aide, vos encouragements, votre optimisme (qui me fait souvent défaut), votre présence tout au long de la thèse et ce, malgré un emploi du temps chargé. Merci de m'avoir accueilli au sein du laboratoire, il y a quatre ans déjà pour mon stage de Master 2 et de m'avoir fait confiance en me permettant de réaliser cette thèse. Veuillez trouver ici l'expression de toute ma gratitude.

Aux enfants terribles du GRAM que j'ai eu le plaisir de côtoyer au quotidien : le Dr SD, la binôme de choc du "lugdu" ! Que d'aventures et de péripéties pendant ces trois années avec toi Sandrine ! Des sessions "BioNum" aux Maï-Taïs stuttgartois, du château de Sissi aux réunions de crise GATC, jusqu'aux soirées madrilènes. Un grand merci pour ta sympathie, ton aide et ton soutien constant. Au Patron, Julien (ou Jean-Phi pour les intimes !). Merci pour tes goûts musicaux, grâce à toi, je n'entendrais plus Johnny Hallyday et Franky Vincent comme avant. Merci aussi pour tes grands moments de craquage et ta bonne humeur chaque jour au labo. A Stéphanie, la grande sœur du GRAM. Merci pour ta gentillesse, ton aide et ton écoute. Merci (ou bravo) à toi d'avoir supporté mes moments de folie, mes expressions désuètes et mon absence de second degré (pourtant je crois m'être amélioré, non ?). A Marie, la californienne dont la "yogattitude" et la zénitude sont à toutes épreuves. Merci pour ces petites pauses "cup of tea" toujours dans la joie et la bonne humeur. Je compte sur toi pour refaire la déco dans le bureau à ton arrivée (mais ne le dis pas à Sandrine !). A Agnès et à nos incroyables difficultés à communiquer. Merci pour ton soutien et pour avoir écouté mes nombreuses plaintes et doléances ces dernières semaines. A mon tour de te soutenir pendant la rédaction ! Courage !

Ludo, Elodie, Fabienne, Estelle et Cindy, merci pour votre soutien, votre aide et votre bonne humeur. Un grand merci Ludo pour toujours avoir pris "cinq minutes" pour répondre à mes questions !

A tous les étudiants qui ont contribué à ces travaux à l'occasion d'un stage. Merci Charline, Amandine, Camille et Léo pour votre enthousiasme pour la recherche et pour votre bonne humeur !

Une pensée toute particulière à Roger, à mon frère et à toi maman. Merci pour m'avoir toujours soutenu et encouragé dans la poursuite de mes études. Merci également de m'avoir permis de m'expatrier dans la pluvieuse Normandie malgré des moments difficiles.

A Jeannie et à Gaston, mes grands-parents de substitution. Merci d'avoir toujours répondu présents dans les bons comme dans les mauvais moments depuis toutes ces années. Vivement Noël pour rentrer à la maison et manger du chocolat au coin de la cheminée avec vous !

A Lucile et à Antoine, merci pour votre soutien sans faille depuis toutes ces années. Vous savez déjà toute l'affection que je vous porte.

SOMMAIRE

LISTE DES FIGURES.....	1
LISTE DES TABLEAUX.....	5
LISTE DES ABREVIATIONS.....	6
INTRODUCTION GENERALE.....	7
DONNEES BIBLIOGRAPHIQUES	8
1. STAPHYLOCOCCUS LUGDUNENSIS.....	9
1.1. Caractères généraux.....	9
1.1.1. Caractères microbiologiques.....	9
1.1.2. Habitat.....	10
1.1.3. Sensibilité aux antibiotiques et aux métaux lourds.....	12
1.2. Infections à S. lugdunensis	14
1.2.1. Infections de la peau et des tissus mous.....	15
1.2.2. Infections ostéo-articulaires.....	15
1.2.3. Bactériémies et endocardites infectieuses.....	17
1.2.4. Autres infections.....	18
1.3. Les facteurs de virulence.....	19
1.3.1. Facteurs de colonisation	19
1.3.2. La formation de biofilm.....	25
1.3.3. Les hémolysines ou toxines cytolytiques	29
1.3.4. Le locus <i>isd</i> , système d'acquisition du fer.....	30
1.3.5. La <i>lugdunine</i>	32
1.3.6. Autres facteurs de virulence.....	33
1.4. Le locus <i>agr</i>, système de régulation globale de la virulence.....	36
2. SEQUENÇAGE HAUT DEBIT ET VIRULENCE DE S. LUGDUNENSIS	39
2.1. Le séquençage haut débit.....	39
2.1.1. Le séquençage <i>Illumina</i>	39
2.1.2. Le séquençage <i>Pacific Biosciences</i>	41
2.1.3. Analyse des données brutes de séquençage haut débit	41
2.2. Analyses génomiques de S. lugdunensis.....	44
2.2.1. Les données génomiques de <i>S. lugdunensis</i>	44
2.2.2. Identification de facteurs de virulence.....	46
2.2.3. Identification d'éléments génétiques mobiles.....	46
2.2.4. Identification de facteurs de restriction	48
2.2.5. Analyse du pan-genome de <i>S. lugdunensis</i>	52

3.	LE SYSTEME DE SECRETION Ess/TYPE VII.....	54
3.1.	Origine et identification du système de sécrétion Ess/type VII.....	54
3.2.	Le système de sécrétion Ess/type VII de <i>S. aureus</i>	56
3.2.1.	Organisation génétique et transcriptionnelle.....	56
3.2.2.	Structure et fonction des protéines codées.....	60
3.2.3.	Rôle du système de sécrétion Ess/type VII dans la virulence de <i>S. aureus</i>	67
3.3.	Le système de sécrétion Ess/type VII de <i>S. lugdunensis</i>	69
4.	GENOTYPAGE AU SEIN DU GENRE STAPHYLOCOCCUS.....	71
4.1.	L'analyse par MultiLocus Sequence Typing.....	73
4.1.1.	Principe de la méthode.....	73
4.1.2.	MLST & <i>S. aureus</i>	75
4.1.3.	MLST & SCN.....	77
4.2.	L'analyse par Multi-Virulence Locus Sequence Typing (MVLST).....	81
4.3.	Le génotypage basé sur l'étude des Variable Number Tandem Repeats.....	83
4.3.1.	Le génotypage par Multiple-Loci VNTR Analysis.....	84
4.3.2.	Typage de <i>S. aureus</i> par spa-typing.....	89
	TRAVAUX PERSONNELS.....	93
1.	ANALYSE GENOMIQUE COMPARATIVE DE SOUCHES PATHOGENES ET DE PORTAGE DE <i>S. LUGDUNENSIS</i>	94
1.1.	Objectifs et principaux résultats de l'étude.....	94
1.2.	Publication n°1.....	98
2.	CARACTERISATION DU SYSTEME DE SECRETION Ess/TYPE VII.....	142
2.1.	Objectifs et principaux résultats de l'étude.....	142
2.2.	Publication n°2.....	145
3.	DISCUSSION ET PERSPECTIVES.....	172
	TRAVAUX ASSOCIES.....	190
1.	APPORT PHYLOGENETIQUE DE L'ANALYSE DE VNTR.....	191
1.1.	Objectifs et principaux résultats de l'étude.....	191
1.2.	Publication n°3.....	195
1.3.	Publication n°4.....	207
2.	DISCUSSION ET PERSPECTIVES.....	251
	CONCLUSION GENERALE.....	261
	REFERENCES BIBLIOGRAPHIQUES.....	263

LISTE DES FIGURES

Figure 1 : Observation macroscopique de différents morphotypes de colonies de *S. lugdunensis* après une nuit d'incubation à 37°C.

Figure 2 : Classification clinique et épidémiologique des différentes espèces de staphylocoques, basée sur la présence de la coagulase libre comme facteur de virulence.

Figure 3 : Infections de la peau et des tissus mous à *S. lugdunensis*.

Figure 4 : Observation macroscopique des deux morphotypes des colonies d'une souche responsable d'une infection ostéoarticulaire à *S. lugdunensis*.

Figure 5 : Illustration des étapes de production, de maturation et d'ancrage d'une protéine de surface à la paroi bactérienne.

Figure 6 : Représentation schématique de la protéine Fbl de *S. lugdunensis* et de la protéine ClfA de *S. aureus*.

Figure 7 : Représentation schématique de la protéine vWbl de *S. lugdunensis*.

Figure 8 : Adhésion de *S. lugdunensis* à l'endothélium vasculaire médiée *in vitro* et *in vivo* par le vWf.

Figure 9 : Représentation schématique des protéines SlsA à G potentiellement ancrées à la paroi bactérienne.

Figure 10 : Représentation schématique de la structure des autolysines de type Atl au sein du genre *Staphylococcus*.

Figure 11 : (A) Comparaison de la capacité de la souche parentale de *S. lugdunensis* DSM4804 et de son mutant DSM4804 Δ atlL à former du biofilm. (B) Virulence comparée de ces deux souches dans un modèle d'infection du nématode *Caenorhabditis elegans*.

Figure 12 : Pathogénèse des infections sur cathéter et facteurs influençant la formation de biofilm.

Figure 13 : Voie de synthèse du polysaccharide intercellular adhesin (PIA) par le locus *ica*.

Figure 14 : Évaluation de la matrice extracellulaire du biofilm de *S. lugdunensis* et *S. epidermidis* par microscopie confocale à balayage laser.

Figure 15 : Comparaison de l'organisation génétique du locus *ica* de *S. lugdunensis*, *S. aureus*, *S. epidermidis* et *S. caprae*.

Figure 16 : Activité hémolytique synergique de *S. lugdunensis* par CAMP-test.

Figure 17 : Représentation schématique des composants du système Isd impliqués dans le transport de l'hème chez *S. aureus*.

- Figure 18** : Comparaison du *locus isd* de *S. aureus* et de *S. lugdunensis*.
- Figure 19** : Modèle de contribution de l'autolysine IsdP à l'acquisition de l'hème chez *S. lugdunensis*.
- Figure 20** : Évaluation de l'activité anti-microbienne de la lugdunine vis-à-vis de *S. aureus*.
- Figure 21** : Évaluation de la capacité des macrophages à éradiquer les souches de *S. lugdunensis* HKU09-01, de *S. aureus* USA300 et de *Micrococcus luteus*.
- Figure 22** : Survie comparée de la souche de *S. lugdunensis* HKU09-01 sauvage ou délétée du gène *oatA* dans un modèle murin d'infection systémique.
- Figure 23** : Modèle d'activation de la transcription de l'ARNIII du *locus agr* chez *S. aureus*.
- Figure 24** : Modèle de régulation par le système *agr* de la formation de biofilm chez *S. aureus*.
- Figure 25** : Implication des principaux facteurs de virulence de *S. lugdunensis* dans les infections ostéoarticulaires et les endocardites infectieuses.
- Figure 26** : Principe de la *Bridge* PCR pour la formation de *clusters* d'ADN lors du séquençage *Illumina*.
- Figure 27** : Principe de l'incorporation des quatre terminateurs de séquence fluorescents réversibles utilisés lors du séquençage *Illumina*.
- Figure 28** : Principe du séquençage *Pacific Biosciences Single Molecule Real Time*.
- Figure 29** : Principe de l'assemblage *de novo* de *reads* courts.
- Figure 30** : Principe du *Hierarchical Genome-Assembly Process* de longs *reads* obtenus par séquençage *Pacific Biosciences*.
- Figure 31** : Mécanisme général des Systèmes de Restriction-Modification (SRM).
- Figure 32** : Organisation générale du *locus Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR) associé aux protéines Cas (CRISPR/Cas).
- Figure 33** : Description des étapes d'adaptation, d'expression et d'interférence du *Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR) associé aux protéines Cas (CRISPR/Cas).
- Figure 34** : Organisation génomique du système de sécrétion de type VII (SST7) chez différentes espèces bactériennes.
- Figure 35** : Le système de sécrétion de type VII (SST7) et le *locus ess* de *S. aureus*.
- Figure 36** : Représentation de l'organisation génétique du *locus ess* codant le système de sécrétion de type VII (SST7) chez quatre souches de *S. aureus*.
- Figure 37** : Étude de l'organisation transcriptionnelle du *locus ess* de cinq souches de *S. aureus* par *Reverse Transcription* couplée à la PCR (RT-PCR).

Figure 38 : Profils de sécrétion des protéines EsxA et EsxC de cinq souches de *S. aureus* obtenus par *Western blot*.

Figure 39 : Structure de la protéine transmembranaire EssC.

Figure 40 : Modèle hexamérique des domaines ATPases SpoIIIE-Ftsk de la protéine transmembranaire EssC.

Figure 41 : Analyse des rôles des domaines de la protéine transmembranaire EssC.

Figure 42 : Modèle de la synthèse et de la sécrétion de la toxine EsaD de *S. aureus*.

Figure 43 : Localisation du gène *esxX* dans le *locus ess* de la souches RJ-ST398 de *S. aureus* de *Sequence Type* (ST) 398.

Figure 44 : La protéine EsxX induit la lyse des polynucléaires neutrophiles (PNN) et participe à l'échappement immunitaire de la souche de *S. aureus* RJ-ST398.

Figure 45 : Modèle de formation d'abcès hépatiques et rénaux chez des souris infectées par la souche de *S. aureus Newman* sauvage ou les mutants dont les gènes *esxA*, *esxB* et *essC* sont inactivés.

Figure 46 : Modèle de formation d'abcès rénaux chez des souris infectées par injection rétro-orbitale par la souche de *S. aureus Newman* sauvage ou par les mutants dont les gènes *esaB* et *esaC* (*esxC*) sont inactivés.

Figure 47 : Contribution du système de sécrétion de type VII des souches de *S. aureus* COL, SA113 et RN6390 à la colonisation nasale et à la virulence dans des modèles de colonisation (A) et de pneumonie (B-D) chez la souris.

Figure 48 : Comparaison de l'organisation du *locus ess* de *S. aureus* et de *S. lugdunensis*.

Figure 49 : Principe général du typage d'une souche par *MultiLocus Sequence Typing* (MLST).

Figure 50 : Dendogramme construit par la méthode *neighbor-joining* représentant les relations phylogénétiques entre les 87 isolats de *S. lugdunensis* caractérisés par *MultiLocus Sequence Typing*.

Figure 51 : Localisation des gènes inclus dans les schémas de *MultiLocus Sequence Typing* (MLST) et de *Multi-Virulence Locus Sequence Typing* (MVLST) sur le génome de la souche de *S. lugdunensis* N920143.

Figure 52 : Dendogramme construit par la méthode *neighbor-joining* représentant les relations phylogénétiques entre les 94 isolats de *S. lugdunensis* caractérisés par *Multi-Virulence Locus Sequence Typing*.

Figure 53 : Organisation générale d'un *Variable Number Tandem Repeat* (VNTR).

Figure 54 : Représentation des différents types de répétitions en tandem (TR) (A) et des différents types de *Variable Number Tandem Repeat* (VNTR) (B).

Figure 55 : *Minimum Spanning Tree* des 1681 isolats de *S. aureus* typés par *Multiple Locus Variable Number Tandem Repeat (VNTR) Analysis (MLVA)*.

Figure 56 : *Minimum Spanning Tree (MST)* basé sur les données de typage de *Multiple Locus Variable Number Tandem Repeat (VNTR) Analysis* pour 40 isolats de *S. aureus*.

Figure 57 : Représentation des domaines composant le gène codant la protéine A de *S. aureus*.

Figure 58 : Distribution des *spa-types* de *S. aureus* les plus prévalents à travers le monde.

Figure 59 : Expression différentielle des protéines identifiées comme surexprimées dans le surnageant de culture de la souche sauvage DSM WT et de la souche délétée du gène Δ essC.

LISTE DES TABLEAUX

Tableau I : Répartition et proportions des souches de *S. lugdunensis* et de *S. aureus* isolées de différentes niches de portage chez 75 sujets sains.

Tableau II : Comparaison des données démographiques et cliniques des patients présentant une infection ostéoarticulaire à *S. lugdunensis*, *S. aureus* ou *S. epidermidis*.

Tableau III : Liste des 25 génomes de *S. lugdunensis* disponibles dans la base de données *National Center for Biotechnology information* (le 03/10/18).

Tableau IV : Distribution des gènes codant des facteurs de virulence putatifs au sein de plusieurs espèces de staphylocoques.

Tableau V : *Sequence types* des clones épidémiques hospitaliers de *S. aureus* résistants à la méticilline et leur caractéristiques moléculaires.

Tableau VI : Distribution des *spa-types* les plus prévalents à travers le monde en termes de nombre de souches de *S. aureus* sensibles (SASM) et résistants à la méticilline (SARM).

Tableau VII : Valeurs de *fold change* obtenues pour les protéines identifiées comme surexprimées dans le surnageant de culture de la souche sauvage DSM WT ou chez la souche délétée du gène Δ essC.

LISTE DES ABREVIATIONS

aa : acides aminés	MLST : <i>MultiLocus Sequence Typing</i>
Aap : <i>Accumulation-associated protein</i>	MLVA : <i>Multiple-Locus Variable Number Tandem Repeat Analysis</i>
Ac : Anticorps	MSCRAMM : <i>Microbial Surface Component Recognizing Adhesive Molecule</i>
ADN : Acide DésoxyriboNucléotidique	MT : <i>MLVA-Type</i>
ADNe : ADN extracellulaire	MVLST : <i>Multi-Virulence Locus Sequence Typing</i>
agr : <i>accessory gene regulator</i>	NCBI : <i>National Center for Biotechnology information</i>
PAI : Peptide Auto-Inducteur	NEAT : <i>NEAr iron Transporter</i>
ARN : Acide RiboNucléotidique	NGS : <i>Next Generation Sequencing</i>
Bap : <i>Biofilm associated protein</i>	Pacbio : <i>Pacific Biosciences</i>
Bhp : <i>Bap homolog protein</i>	pb : paire de bases
CC : Complexe Clonal	PCR : <i>Polymerase Chain Reaction</i>
CCv : Complexe Clonal associé à la Virulence	PIA : <i>Polysaccharide Intercellular Adhesin</i>
CDS : séquence codante (<i>Coding Sequence</i>)	PNN : PolyNucléaire Neutrophile
CFP-10 : <i>Culture Filtrate Protein of 10 kDa</i>	PSM : <i>Phenol-Soluble Modulín</i>
CifA : <i>Clumping factor A</i>	PFGE : <i>Pulsed Field Gel Electrophoresis</i>
CMI : Concentration Minimale Inhibitrice	RNA-seq : Séquençage des ARN
CRISPR : <i>Clustered Regularly Interspaced Short Palindromic Repeat</i>	RT : <i>Reverse Transcription</i>
CRISPR/Cas : CRISPR associé aux protéines Cas	SaPI : Ilots de Pathogénicité de <i>S. aureus</i>
DI : Index de Diversité de Simpson	SARM : <i>S. aureus</i> Résistant à la Méricilline
DLV : <i>Double Locus Variant</i>	SASM : <i>S. aureus</i> Sensible à la Méricilline
DO : Densité Optique	SCN : Staphylocoque à Coagulase Négative
DUF : <i>Domain of Unknown Function</i>	SCV : <i>Small Colony Variant</i>
EGM : Élément Génétique Mobile	Sls : <i>S. lugdunensis surface protein</i>
EI : Endocardite Infectieuse	SLUSH : <i>Staphylococcus lugdunensis Synergetic Hemolysin</i>
ESAT-6 : <i>Early Secreted Antigen Target-6 kDa</i>	SLV : <i>Single Locus Variant</i>
ess : <i>ESAT-6 secretion system</i>	SMRT : <i>Single Molecule Real Time</i>
Fbl : <i>Fibrinogen binding protein</i>	SNP : <i>Single Nucleotide Polymorphism</i>
FHA : <i>ForkHead-Associated</i>	SRM : Système de Restriction-Modification
HGAP : <i>Hierarchical Genome-Assembly Process</i>	SSC : <i>Staphylococcal Cassette Chromosome</i>
HPG : Hydrolase du PeptidoGlycane	SST7 : Système de Sécrétion de Type VII
IA : Index d'Association	ST : <i>Sequence Type</i>
IAS : Infection Associée aux Soins	T/AT : Toxine/AntiToxine
ica : <i>intercellular adhesion</i>	TR : Répétitions en Tandem (<i>Tandem Repeat</i>)
IMAGE : <i>Iterative Mapping and Assembly for Gap Elimination</i>	TRF : <i>Tandem Repeat Finder</i>
IOA : Infection OstéoArticulaire	TRST : <i>Tandem Repeat Sequence Typing</i>
IPTM : Infection de la Peau et des Tissus Mous	TRT : <i>TRST-Type</i>
isd : <i>iron-regulated surface determinant</i>	vapE : <i>virulence-associated protein E</i>
kb : kilobase	VNTR : <i>Variable Number Tandem Repeat</i>
kDa : kiloDalton	VT : <i>Virulence Type</i>
LPXTG : Leucine-Proline-X-Thréonine-Glycine	vWbl : protéine de liaison au facteur von Willebrand
Mb : Millions de paires de bases	vWf : facteur von Willebrand
MC : <i>MLVA-Complex</i>	
MEC : Matrice Extra-Cellulaire	
MLEE : <i>MultiLocus Enzyme Electrophoresis</i>	

INTRODUCTION GENERALE

Ce travail de thèse s'inscrit dans l'une des thématiques de recherche développées par le "Groupe de Recherche sur l'Adaptation Microbienne" EA2656, qui consiste à étudier d'une part la diversité génétique de *Staphylococcus lugdunensis* et d'autre part les mécanismes qui sous-tendent la virulence particulière de cette espèce de staphylocoque à coagulase négative.

Les travaux présentés ici sont consacrés à l'analyse comparative des génomes complets de souches pathogènes et de portage de *Staphylococcus lugdunensis* et à la première caractérisation moléculaire et phénotypique du système de sécrétion Ess/type VII de cette espèce.

La première partie de ce manuscrit est constituée d'une analyse des données bibliographiques de la littérature concernant :

- L'espèce *S. lugdunensis* : ses principales caractéristiques bactériologiques, sa sensibilité aux antibiotiques et l'état actuel des connaissances sur son pouvoir pathogène et sa virulence.
- Le séquençage haut débit et son application chez *S. lugdunensis*.
- Le système de sécrétion Ess/type VII : son origine et sa caractérisation chez *Staphylococcus aureus*, son organisation génétique et transcriptionnelle, son rôle dans la virulence et l'état actuel des connaissances de ce système chez *S. lugdunensis*.
- Les méthodes de typage moléculaires utilisées pour les études micro- et macro-épidémiologiques au sein du genre *Staphylococcus*.

Dans une seconde partie, sont présentés nos travaux expérimentaux portant sur l'analyse génomique comparative de souches pathogènes et de portage de *S. lugdunensis* et l'étude du rôle du système de sécrétion Ess/type VII dans sa virulence. Enfin, dans une troisième partie, nous présentons les travaux que nous avons menés à contribution égale avec le Dr Sandrine Dahyot. Ils ont consisté à développer de nouvelles méthodes de génotypage chez *S. lugdunensis* fondées sur l'analyse du polymorphisme de motifs répétés en nombre variable sur le génome.

DONNEES BIBLIOGRAPHIQUES

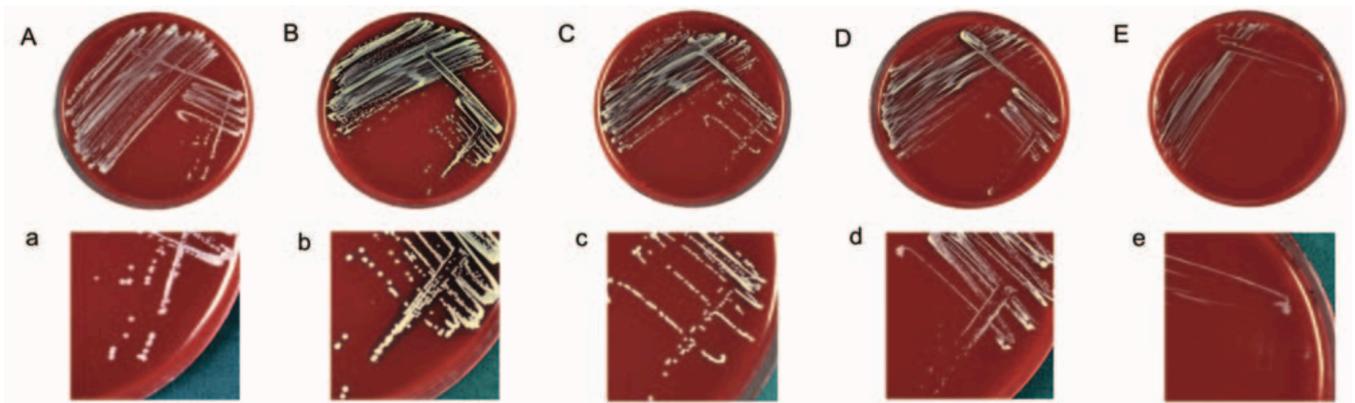


Figure 1 : Observation macroscopique de différents morphotypes de colonies de *S. lugdunensis* après une nuit d'incubation à 37°C⁴.

Les géloses A à D représentent un phénotype dit "classique" caractérisé par des colonies pigmentées : A, de couleur crème ; B-D, de couleur jaune. Les cadrans a à d montrent le caractère hémolytique variable des colonies : a, non-hémolytique ; b, hémolyse totale ; c et d, hémolyse faible. La gélose E/e représente des variants à petites colonies non pigmentées et non hémolytiques.

1. *Staphylococcus lugdunensis*

1.1. Caractères généraux

1.1.1. Caractères microbiologiques

Décrit pour la première fois en 1988¹, *S. lugdunensis* doit son nom à la ville dans laquelle il a été découvert : Lyon (du latin : *Lugdunum*).

S. lugdunensis est une bactérie appartenant à la famille des *Staphylococcaceae* au sein du *phylum* des *Firmicutes*. A l'examen direct, *S. lugdunensis* se présente sous la forme de *cocci* à Gram positif de 0,8 à 1 µm de diamètre, isolés, en paires ou en amas¹.

Cette bactérie est aérobie-anaérobie facultative, non-exigeante, avec un caractère halotolérant et mésophile, la rendant cultivable à des températures entre 30 et 45°C². *S. lugdunensis* apparaît sous la forme de colonies de 1 à 4 mm de diamètre, lisses, brillantes, plates, à bords réguliers. Ces colonies apparaissent parfois muqueuses ou rugueuses et mates avec des contours irréguliers. Ces variations dépendent des souches ou du milieu de culture utilisé^{1,3}. Les colonies d'abord blanches se pigmentent de couleur crème ou jaune voire dorée après plusieurs jours d'incubation^{3,4} (**Figure 1**).

Là aussi, des variations existent, car sur un même milieu, certaines souches produisent des colonies non pigmentées ou avec des morphotypes différents (dissociation de colonies). Parmi ces variants, il existe des variants à petites colonies (*Small Colony Variant* [SCV]) non pigmentées et non hémolytiques, avec un métabolisme ralenti, puisqu'elles peuvent n'apparaître qu'après une incubation de 48 voire 72 heures^{5,6}.

Sur des géloses supplémentées avec du sang de mouton ou de cheval, une hémolyse de type β (hémolyse totale) peut être observée autour des colonies¹. Cette hémolyse totale est un caractère quasi-constant, puisque présent chez 85%⁷ voire 95%⁸ des souches testées. Par ailleurs, il a été mis en évidence une activité hémolytique synergique, semblable à celle observée avec la β-hémolysine de *Staphylococcus aureus*⁹. Sur ces milieux, il se dégage des colonies de *S. lugdunensis*

une forte odeur de salaison (odeur de saucisson) caractéristique de la décomposition des acides gras des érythrocytes³.

Enfin, comme tout staphylocoque, *S. lugdunensis* produit une catalase et comme tout Staphylocoque à Coagulase Négative (SCN), cette espèce ne produit pas de coagulase libre ou staphylocoagulase. Toutefois, même si cette espèce ne produit pas non plus de coagulase liée, elle possède une protéine homologue à la protéine ClfA de *S. aureus* capable d'interagir dans les tests commerciaux dédiés à l'identification des protéines de surface spécifiques de *S. aureus*. En effet, des faux positifs peuvent résulter de la présence d'une protéine de liaison au fibrinogène ou "*fibrinogen binding protein*" (Fbl)^{10,11}. Ces faux positifs associés à des colonies β-hémolytiques de couleur jaunâtre peuvent facilement orienter le diagnostic en faveur d'une souche de *S. aureus* si d'autres tests phénotypiques ou moléculaires ne sont pas réalisés.

Toutefois, ces erreurs d'identification tendent maintenant à disparaître du fait de l'implantation de la spectrométrie de masse de type *Matrix-Assisted Laser Desorption/Ionization – Time Of Flight* qui permet aujourd'hui des identifications bactériennes fiables au rang de l'espèce¹².

1.1.2. Habitat

Chez l'Homme

Comme de nombreux staphylocoques, *S. lugdunensis* est une bactérie du microbiote cutané avec des niches privilégiées telles que les régions inguinale, périnéale, mammaire et plantaire^{13,14}. Une étude française de portage de *S. lugdunensis* a été réalisée par un prélèvement inguinal chez une cohorte de 140 patients admis au service des urgences sur une période de trois mois¹³. Les résultats ont montré un portage inguinal de 22% (23% des femmes et 19% des hommes), et pour 68% de ces patients, un portage inguinal simultané bilatéral. Les femmes de plus de 65 ans présentaient le plus fort taux de portage (51%), alors que chez les hommes, la fréquence la plus élevée a été retrouvée chez les sujets de moins de 65 ans¹³.

Location ^a	<i>S. lugdunensis</i>			<i>S. aureus</i>		
	Males (n = 33)	Females (n = 42)	Total (%)	Males (n = 33)	Females (n = 42)	Total (%)
Nose	6	1	7 (9.3)	8	6	14 (18.7)
Axilla	9	6	15 (20)	1	1	2 (2.7)
Perineum	9	2	11 (14.7)	0	1	1 (1.3)
Groin	13	11	24 (32)	0	0	0 (0)
Breast	3	1	4 (5.3)	1	1	2 (2.7)
Toe	1	3	4 (5.3)	0	0	0 (0)
Nail 1st toe	10	8	18 (24)	0	1	1 (1.3)
Unspecified	6	5	11 (14.7)	1	0	1 (1.3)
Total	57	37	94	11	10	21

^aOne swab per location from each of the 75 subjects.

Tableau I : Répartition et proportions des souches de *S. lugdunensis* et de *S. aureus* isolées de différentes niches de portage chez 75 sujets sains¹³.

Bieber *et al.*¹⁴ ont réalisé une étude comparée des sites de portage de *S. lugdunensis* et de *S. aureus* au sein d'une cohorte de 75 volontaires sains. Alors que *S. aureus* était rarement isolé à partir de prélèvements cutanés, *S. lugdunensis* y était isolé fréquemment (66,7% *versus* 21,3%). Le même constat a pu être fait au niveau de l'aîne (32% *versus* 0%), de l'ongle du premier orteil (24% *versus* 1,3%) et au niveau axillaire (20% *versus* 2,7%). *A contrario*, *S. lugdunensis* était très rarement isolé au niveau des fosses nasales (9%), où *S. aureus* était prédominant avec un portage de 19%¹⁴. Au final, *S. lugdunensis* était rencontré plus fréquemment que *S. aureus* dans tous les sites explorés, hormis dans les fosses nasales¹³ (**Tableau I**).

S. lugdunensis appartient donc à la flore cutanée normale, avec une prédilection pour la partie inférieure du corps et notamment l'aîne. Il a été suggéré que *S. lugdunensis* serait un commensal des aires cutanées possédant des glandes sudoripares apocrines comme les aisselles, la région génito-anale, le conduit auditif, les paupières ou les auréoles mammaires¹⁵.

Chez l'animal

L'Homme n'est pas le seul à être colonisé par *S. lugdunensis*, puisque ce dernier a également été isolé chez plusieurs espèces animales domestiques (chien, chat et oiseau) mais aussi chez des animaux d'élevage¹⁶. *S. lugdunensis* a été isolé à partir des fosses nasales et d'aisselles de chien, mais avec une fréquence très faible (1,7%)¹⁷. Même si peu de données ont été rapportées chez l'animal, des cas d'infections à *S. lugdunensis* ont été décrits. Par exemple, le cas notable d'une endocardite infectieuse chez un chien¹⁶ ou encore des mammites infectieuses chez la chèvre¹⁸.

1.1.3. Sensibilité aux antibiotiques et aux métaux lourds

A la différence des principales autres espèces de staphylocoques (*S. aureus*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, ...), *S. lugdunensis* est particulièrement sensible à la quasi-totalité des antibiotiques anti-staphylococciques^{2,3,19}. En revanche, le taux de résistance à la pénicilline G varie considérablement selon les études publiées allant de 48%²⁰ à 69%²¹, probablement du fait de problèmes méthodologiques de détection de pénicillinase (discordances entre les méthodes de diffusion en milieu gélosé et les méthodes automatisées en milieu liquide)²⁰. En effet, selon les études, la résistance varie de 7 à 24% en France et atteint même des taux de 48% aux États-Unis²⁰ voire 69% en Asie²¹. La résistance à cet antibiotique est liée à la production d'une pénicillinase codée par le gène *blaZ*^{20,22}.

Les souches de *S. lugdunensis* peuvent aussi être résistantes aux β -lactamines du fait de l'acquisition du gène *mecA* porté par un élément génétique mobile appelé *Staphylococcal Cassette Chromosome* (SSC) *mec*, comme chez *S. aureus*. Plusieurs études ont décrit des souches résistantes de *S. lugdunensis* à la méticilline présentant une cassette SCC*mec*, notamment en Asie^{23,24}. Il existe quatre types de la cassette SCC*mec* (II, IV, V et Vt) à ce jour chez *S. lugdunensis*^{23,24}. La résistance à la méticilline reste particulièrement peu élevée : 8% des isolats ont été catégorisés comme résistants à la méticilline dans une étude internationale, taux le plus bas observé parmi l'ensemble des SCN analysés²⁵. Toutefois, des études asiatiques ont décrit l'émergence de souches résistantes à la méticilline : 20% selon Lin et al.²¹ et 32% des souches isolées à Taïwan dans un contexte d'infections associées aux soins selon Yeh et al.²⁶

Cette résistance émergente est préoccupante, et ce d'autant que cette espèce présente une tolérance aux glycopeptides, classe thérapeutique principalement utilisée en cas de résistance à la méticilline. Cette tolérance se caractérise par un défaut de bactéricidie vis-à-vis de souches pourtant sensibles à l'activité bactériostatique (Concentration Minimale Inhibitrice [CMI] normale) ; elle a été observée pour 46% (6/13) des souches testées dans le travail de Bourgeois et al.²⁷.

Par ailleurs, quelques rares études font état de résistance plasmidique aux métaux lourds. En effet, le plasmide pLUG10 a été identifié dans une étude chez 60% (18/30) des souches de *S. lugdunensis*²⁸. Il porte les gènes *cadB* et *cadX* codant des protéines impliquées dans la résistance au cadmium^{28,29}. Ce plasmide conférerait une résistance au cadmium à des niveaux élevés (CMI supérieure ou égale à 125 µM)²⁸.

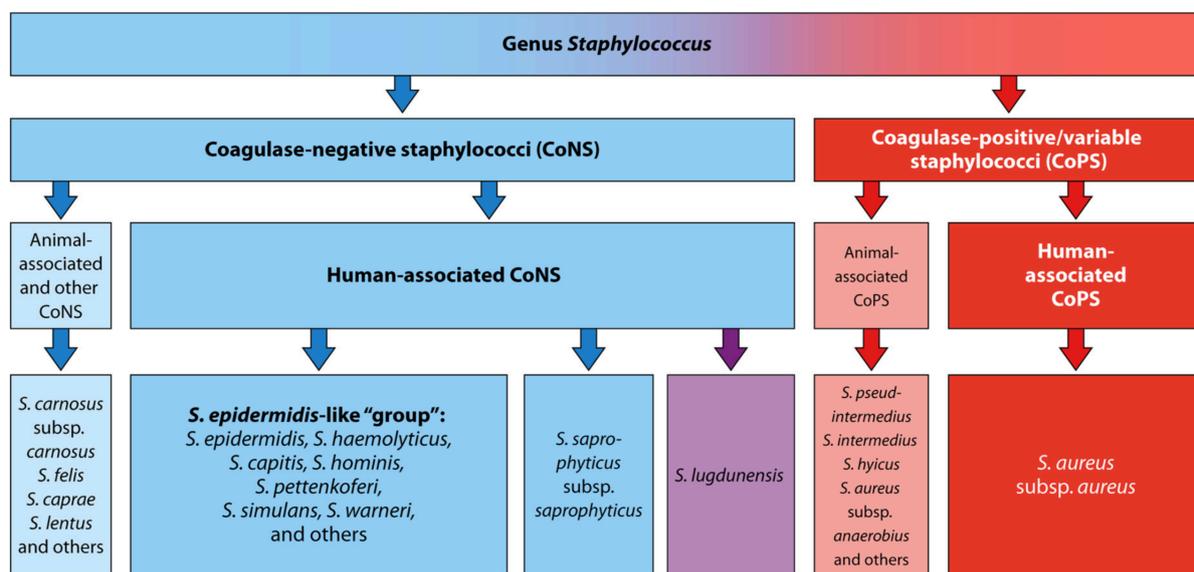


Figure 2 : Classification clinique et épidémiologique des différentes espèces de staphylocoques, basée sur la présence de la coagulase libre comme facteur de virulence³⁰.

1.2. Infections à *S. lugdunensis*

L'implication des SCN en pathologie humaine a longtemps été sous-estimée. En effet, ces derniers contaminant régulièrement les prélèvements en lien avec leur réservoir naturel cutané et du fait des méthodes d'identification complexes, leur identification en routine au rang d'espèce n'était pas systématiquement effectuée ou fiable.

Bactérie pathogène opportuniste, *S. lugdunensis* se distingue pourtant des autres SCN par une virulence particulière. En effet, de nombreuses observations cliniques décrivent les infections liées à cette espèce comme plus proches de celles causées par *S. aureus* que de celles dues aux autres SCN, aussi bien en termes d'agressivité tissulaire que d'évolution clinique. Ainsi, Becker *et al.*³⁰ lui accordent-ils une place intermédiaire entre *S. aureus*, *Staphylococcus saprophyticus* et les autres SCN (**Figure 2**).

En 2006, Tan *et al.*³¹ ont étudié la pathogénicité des isolats de *S. lugdunensis*, et montré que *S. lugdunensis* est le SCN le plus fréquemment isolé d'infections cliniquement significatives. Cette espèce représente en effet 45% des isolats de SCN cliniquement significatifs contre 33% pour *S. epidermidis* et 11% pour *S. haemolyticus*. Il s'agit de l'espèce de SCN la plus virulente puisque jusqu'à 91% des isolats de *S. lugdunensis* sont associées à une infection significative³¹ ; pourcentage confirmé en 2011 par Wu *et al.*³² avec un taux similaire d'infection à *S. lugdunensis*, égal à 83%.

Depuis sa découverte en 1988, des souches de *S. lugdunensis* ont été isolées dans des infections de localisations variées mais principalement des infections de la peau et des tissus mous (IPTM) et des infections invasives particulièrement sévères et inhabituelles pour un SCN comme des infections ostéoarticulaires (IOA) et des endocardites infectieuses (EI).



Figure 3 : Infections de la peau et des tissus mous à *S. lugdunensis*.

A, abcès dorsal avec cellulite suite à une morsure d'insecte³⁴.

B, érythème cutané et écoulement cicatriciel après une dermolipectomie abdominale³⁶.

1.2.1. Infections de la peau et des tissus mous

La majorité des infections à *S. lugdunensis* sont des IPTM. Dans une étude rétrospective monocentrique américaine, il a été montré que sur un total de 143 infections à *S. lugdunensis*, 55,4% étaient des IPTM³³. Une autre étude rétrospective réalisée à Taïwan rapporte que 41 sur 49 des isolats issus d'infections communautaires étaient des IPTM³². En outre, le premier isolat clinique de *S. lugdunensis* décrit provenait d'un abcès de ganglion axillaire¹.

Les IPTM à *S. lugdunensis* se caractérisent par des infections cutanées profondes telles que des abcès ou des surinfections de plaies, mais aussi des cellulites ou des paronychies en particulier chez les patients débilisés (diabétiques)¹⁵. Ces infections peuvent être de nature communautaire³⁴ mais elles surviennent dans 75% des cas après un geste chirurgical ou un traumatisme^{35,36} (**Figure 3**). Toutefois, ces infections peuvent aussi se présenter sous des formes superficielles telles que des folliculites et des pustuloses chez les patients immunocompétents³⁷.

Les localisations des IPTM sont diverses. Ont été rapportées dans plusieurs études, des infections dans la partie inférieure du corps (au niveau périnéal, inguinal et plantaire), mais aussi au niveau mammaire et du cuir chevelu^{2,37,38}. La prévalence de *S. lugdunensis* comme commensal dans ces différentes régions du corps peut expliquer la localisation de ces IPTM³⁷.

S. lugdunensis a aussi été décrit comme l'espèce prédominante isolée dans les nodules et abcès des patients atteints de la maladie de Verneuil suggérant un rôle clé de cette espèce dans l'évolution et l'aggravation de cette pathologie^{37,39}.

1.2.2. Infections ostéo-articulaires

Il a été rapporté dans plusieurs études cliniques que *S. lugdunensis* pouvait être responsable d'IOA. L'étude rétrospective de Douiri *et al.*⁴⁰ a rapporté que sur un total de 215 isolats de *S. lugdunensis*, 39,9% étaient responsables d'une IOA. Par ailleurs, l'étude prospective de Argemi *et al.*¹¹ rapporte que parmi 129 isolats de *S. lugdunensis*, 36,4% étaient impliqués dans une IOA.

	<i>S. lugdunensis</i> (n=28)	<i>S. aureus</i> (n=30)	<i>S. epidermidis</i> (n=30)	p-Value
Age (years)	67.5 (58–78)	60.5 (44–75)	67 (61–77)	0.34
Number of males	13 (46.4%)	20 (66.7%)	20 (66.7%)	0.196
Prosthesis site				
Knee	16 (57%)	18 (60%)	11 (36.7%)	0.145
Hip	10 (36%)	9 (30%)	16 (53%)	0.158
Other	2 (7%)	3 (10%)	3 (10%)	
Clinical signs				
Fever	15 (53.6%)	9 (30%)	4 (13.3%)	0.004
Local signs of inflammation	13 (46.4%)	7 (23.3%)	4 (13.3%)	0.015
Surgery type				
Irrigation and debridement	11 (39.3%)	10 (33.3%)	5 (16.7%)	0.144
One-stage surgery	6 (21.4%)	12 (40%)	16 (53.3%)	0.044
Two-stage revision	11 (39.3%)	8 (26.7%)	9 (30%)	0.568
Delay from surgery to infection (weeks)	12 (3–56)	44 (12–144)	84 (44–192)	0.0449
Samples				
Number of positive samples	3.5 (1.5–5)	4 (3–5)	4 (3–5)	0.449
Total number of samples	5 (2–6)	4 (3–5)	5 (4–6)	0.413
Treatment duration (weeks)	7	7	7	
Outcome positive	89%	83%	97%	0.233

Tableau II : Comparaison des données démographiques et cliniques des patients présentant une infection ostéoarticulaire à *S. lugdunensis*, *S. aureus* ou *S. epidermidis*⁴⁴.

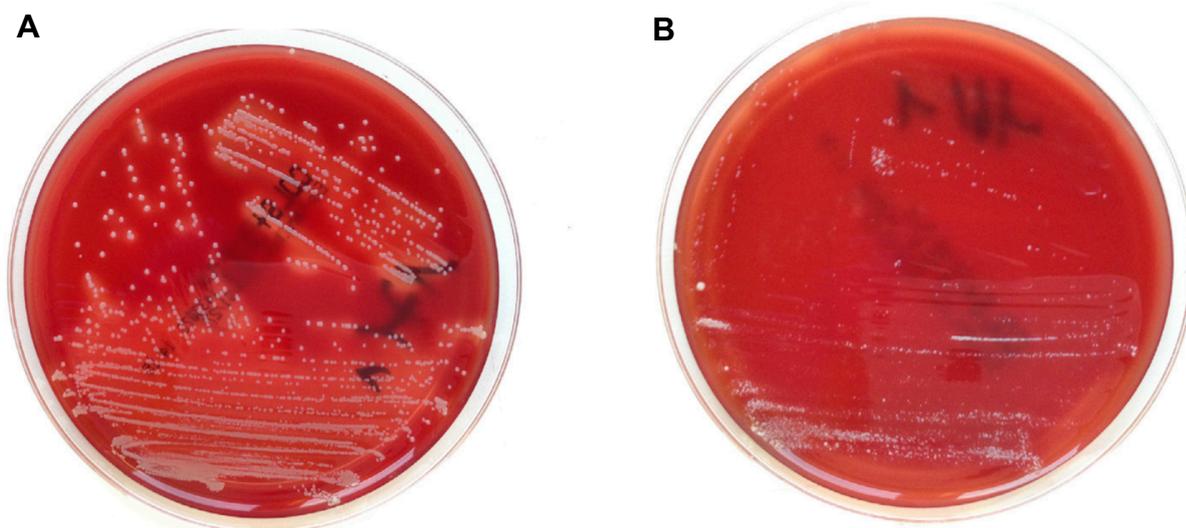


Figure 4 : Observation macroscopique des deux morphotypes des colonies d'une souche responsable d'une infection ostéoarticulaire à *S. lugdunensis*⁶.

A, morphotype classique des colonies de *S. lugdunensis* au premier épisode.
B, variant à petites colonies isolé lors de l'épisode de récurrence.

Ainsi, *S. lugdunensis* peut être responsable d'IOA telles que des ostéomyélites ou des IOA sur matériel, secondaires à un geste chirurgical ou à la pose d'une prothèse articulaire. L'étude rétrospective de Seng *et al.*⁴¹ chez 47 patients atteints d'IOA a montré que 33 (70,2%) d'entre eux présentaient une IOA sur matériel et neuf (19,6%) présentaient une ostéomyélite. D'autre part, Douri *et al.*⁴⁰ ont montré que 25 IOA (64,1%) étaient des infections sur prothèse de genou, de la hanche et de l'épaule.

Plusieurs études rapportent une plus forte prévalence d'IOA à *S. lugdunensis* sur prothèse de genou plutôt que sur prothèse de hanche (89% *versus* 11%)⁴², comparativement aux IOA à *S. aureus* qui surviennent aussi bien sur les deux types de prothèse (56,3% pour le genou *versus* 43,7% pour la hanche)⁴³. Par ailleurs, le délai d'apparition de l'IOA à *S. lugdunensis* est plus court après l'acte chirurgical qu'avec *S. aureus* et *S. epidermidis*. En effet, le délai avant l'apparition de l'infection avec *S. lugdunensis* est de 12 semaines en moyenne, alors qu'il faut compter 44 et 48 semaines pour *S. aureus* et *S. epidermidis* respectivement⁴⁴. En outre, les IOA à *S. lugdunensis* sont significativement plus fébriles et plus inflammatoires que celles à *S. aureus* et *S. epidermidis*⁴⁴ (**Tableau II**). De plus, l'évolution d'une IOA à *S. lugdunensis* est moins favorable (89%) qu'une IOA à *S. epidermidis* (97%)⁴⁴. Dans tous les cas, il est nécessaire, en plus de l'antibiothérapie, de recourir à la chirurgie pour traiter l'infection voire remplacer la prothèse^{11,43}.

Enfin, l'équipe de l'hôpital universitaire de Strasbourg a fait état d'un cas de récurrence d'une infection sur prothèse du genou avec la même souche de *S. lugdunensis* (même antibiogramme et même génotype)⁶. La particularité de cette récurrence était la morphologie des colonies observée lors du second épisode. Ces dernières se présentaient sous la forme de SCV, de petite taille et non hémolytiques avec une croissance lente de plus de 72h⁶ (**Figure 4**). *A contrario*, lors du premier épisode, les colonies présentaient un aspect dit "classique" après 12h de culture sur un milieu identique.

1.2.3. Bactériémies et endocardites infectieuses

S. lugdunensis peut aussi être responsable de bactériémie. Cependant, cette manifestation clinique est très rare puisqu'elle représente à peine 1% des bactériémies à SCN (entre 0,3 et 0,87%)⁴⁵⁻⁴⁷. En effet, il s'agit du septième SCN le plus isolé de bactériémies contrairement à *S. epidermidis* rencontré dans 30,3% des cas⁴⁶. Bien que rares, des bactériémies à *S. lugdunensis* associés à des chocs septiques ont été décrits^{48,49}.

La prévalence des bactériémies à *S. lugdunensis* est faible mais celles-ci se compliquent plus fréquemment d'une EI que pour les autres staphylocoques. Dans une étude rétrospective, Non *et al.*⁵⁰ ont observé que sur une population de 74 patients présentant une bactériémie à *S. lugdunensis*, 15% d'entre eux ont développé une EI. De plus, parmi les patients présentant une ou plusieurs hémocultures positives à *S. lugdunensis*, 6,3% à 27% des patients présentaient une EI⁵⁰.

L'incidence des EI à *S. lugdunensis* est principalement issue des données de revues rétrospectives. En effet, une méta-analyse des cas d'EI à *S. lugdunensis* décrits entre 1988 et 2008 a recensé 67 cas dans la littérature scientifique⁵¹. De leur côté, Anguera *et al.*⁵² ont identifié lors de leur étude rétrospective, dix patients présentant une EI à *S. lugdunensis* parmi 912 cas d'EI entre 1990 et 2003, soit une prévalence de 1,1%. Patel *et al.*⁵³ ont, eux, identifié quatre isolats de *S. lugdunensis* parmi 89 isolats de staphylocoques collectés de patients avec une EI. Dans cette étude, *S. lugdunensis* était responsable de 18% des EI à SCN et de 44% des EI sur valves natives.

La méta-analyse de Liu *et al.*⁵¹ a montré que cette espèce était principalement responsable d'infection au niveau des valves cardiaques natives (78,7%) contrairement aux SCN qui infectent principalement les valves prothétiques. *S. lugdunensis* est aussi responsable d'infections sur valves prothétiques (11,9%) ou dispositifs intracardiaques (*pacemakers*, défibrillateurs) (7,5%). Une fois sur deux, la porte d'entrée du germe n'est pas déterminée⁵¹.

Les EI à *S. lugdunensis* sont des infections sévères qui entraînent une destruction tissulaire importante associée à la formation d'abcès et de végétations

volumineuses^{2,51}. Le taux de mortalité des EI à *S. lugdunensis* est significativement plus élevé (38,8 à 50%) que celles à *S. aureus* (14,5%) et *S. epidermidis* (20%)⁵². La chirurgie est alors essentielle en plus de l'antibiothérapie. Elle est d'ailleurs plus fréquente pour une EI à *S. lugdunensis* (70%) qu'à *S. epidermidis* (60%) ou *S. aureus* (36,9%), pourtant réputé très virulent⁵².

1.2.4. Autres infections

S. lugdunensis a également été isolé dans d'autres types de présentations cliniques rares telles que des fasciites nécrosantes³⁶, des infections du système nerveux central (abcès cérébral³, infection de dérivation ventriculopéritonéale⁵⁴ et méningite⁵⁵) des infections oculaires (endophtalmie⁵⁶ et kératite⁵⁷), des péritonites⁵⁸ ou encore des infections urinaires⁵⁹.

1.3. Les facteurs de virulence

La diversité et la sévérité des infections à *S. lugdunensis* décrites dans la littérature témoigne du fort pouvoir pathogène de cette espèce et a conduit plusieurs équipes à étudier les mécanismes qui sous-tendent cette pathogénicité par l'identification de facteurs de virulence.

1.3.1. Facteurs de colonisation

L'aptitude de *S. lugdunensis* à coloniser des surfaces biotiques ou abiotiques suggère une forte capacité d'adhésion. Cette aptitude à coloniser est une étape indispensable dans le processus de développement des infections car elle permet l'implantation des bactéries sur les tissus de l'hôte avant leur invasion *via* des protéines appelées adhésines ou encore *Microbial Surface Component Recognizing Adhesive Molecule* (MSCRAMM).

Les adhésines sont des protéines de surface liées de façon covalente à la paroi et sont capables de se lier soit directement à un récepteur membranaire de la cellule, soit à des facteurs de la Matrice Extra-Cellulaire (MEC) de l'hôte. La MEC est une structure très organisée, constituée d'un réseau complexe de plusieurs composants et en contact étroit avec les cellules qui les sécrètent. La MEC est constituée de nombreux composants tels que le collagène, la fibronectine, l'élastine, la laminine...⁶⁰ Ceux-ci se lient à des récepteurs cellulaires membranaires comme les intégrines de type α et β , et constituent ainsi un premier point d'ancrage à l'hôte pour les bactéries⁶¹. Il a été montré par Paulsson *et al.*⁶² que *S. lugdunensis* est capable de se lier à de nombreux composants de la MEC tels que le collagène de type I et IV, le fibrinogène, la laminine, mais avec une affinité inférieure à celle de *S. aureus*.

Cette capacité de liaison des MSCRAMM est liée à la présence de plusieurs domaines protéiques essentiels⁶³ : (i) un peptide signal d'adressage YSRIK/GXXE (Tyrosine – Sérine – Arginine – Isoleucine – Lysine/Glycine – X – X – Acide glutamique) au système de sécrétion *sec*, (ii) un ou plusieurs domaines de liaison aux composants de la MEC, (iii) un motif d'ancrage à la paroi LPXTG (Leucine – Proline – X – Thréonine – Glycine) et (iv) un signal de sortie chargé positivement. Lorsqu'une

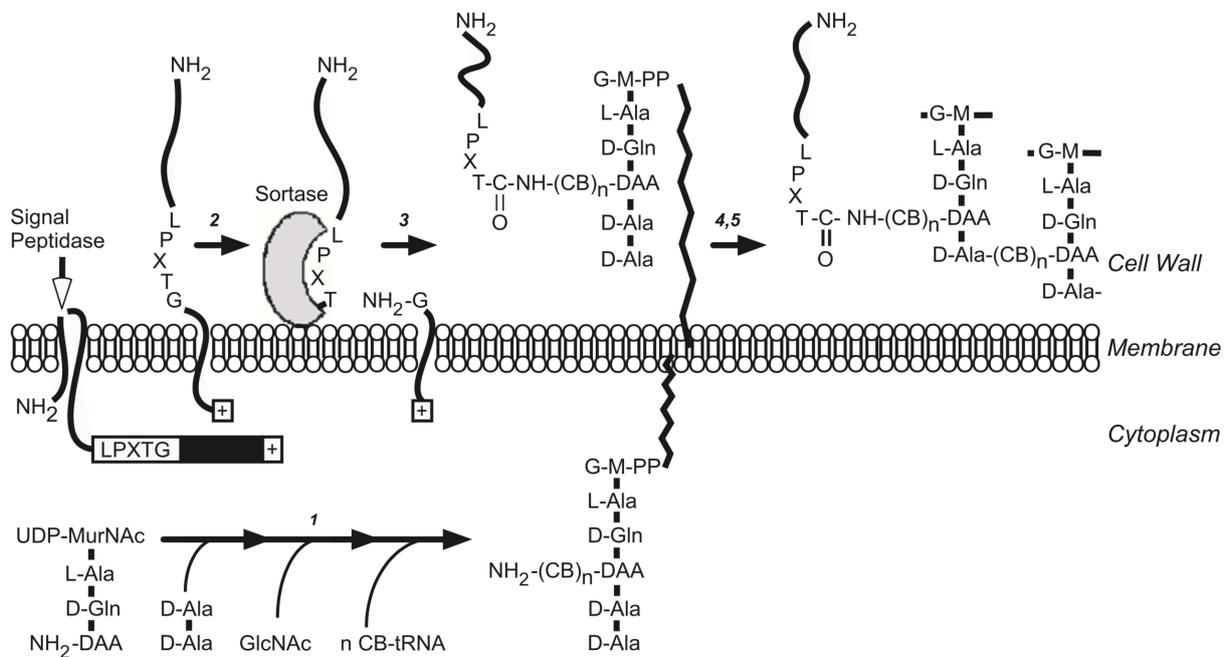


Figure 5 : Illustration des étapes de production, de maturation et d'ancrage d'une protéine de surface à la paroi bactérienne⁶⁴.

(1) La synthèse de la paroi bactérienne commence dans le cytoplasme bactérien par l'assemblage des précurseurs formant le lipide II, transféré à travers la membrane plasmique. (2) Les protéines de surface sont synthétisées en tant que précurseurs dans le cytoplasme. Ces précurseurs portent, en position N-terminale, un peptide signal et, en position C-terminale, un domaine d'ancrage à la paroi. Ce domaine est constitué d'un motif d'ancrage LPXTG, suivi d'un domaine hydrophobe (boîte noire) et d'une queue de résidus chargés positivement (+). Après clivage du peptide signal, permettant la translocation du précurseur, le domaine hydrophobe et la queue chargée retiennent la protéine de surface au niveau de la membrane. (3) Au niveau de la paroi, le précurseur est clivé entre la thréonine et la glycine du motif LPXTG par une sortase A (transpeptidase ancrée à la membrane), (4,5) qui catalyse une réaction de transpeptidation et de transglycosylation permettant l'attachement de la protéine au peptidoglycane par l'intermédiaire du lipide II.

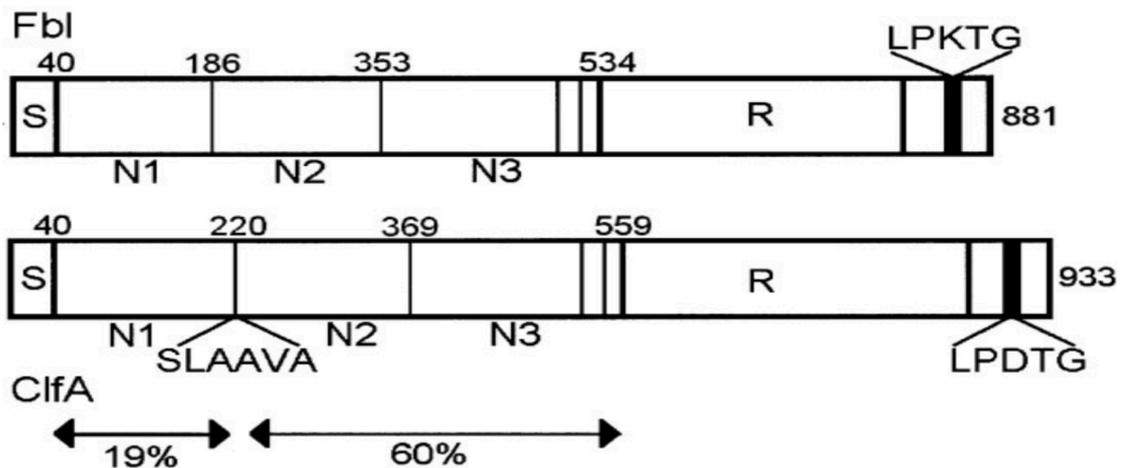


Figure 6 : Représentation schématique de la protéine Fbl de *S. lugdunensis* et de la protéine ClfA de *S. aureus*⁶⁶.

Depuis la partie N-terminale, les protéines Fbl et ClfA sont successivement composées d'un peptide signal (S), un domaine de liaison au fibrinogène (motifs N1, N2 et N3), un domaine de motifs répétés SDSA et SD pour Fbl et ClfA respectivement, et un domaine d'ancrage à la paroi LPXTG. A la différence de Fbl, ClfA possède un site de clivage par une métalloprotéase (SLAAVA) entre les motifs N1 et N2. Les pourcentages d'homologie des motifs N1, N2 et N3 entre les deux protéines sont représentés sous les flèches.

adhésine est traduite, le signal d'adressage en position N-terminale permet sa translocation du cytoplasme vers la paroi *via* le système *sec*. Ce signal est ensuite clivé par une peptidase alors qu'une protéine de type sortase A va permettre l'attachement de la protéine à la paroi (**Figure 5**)⁶⁴. En effet, la sortase A permet de cliver la liaison entre la Thréonine et la Glycine du motif LPXTG et catalyse une réaction de transpeptidation qui permet l'ancrage de la protéine à la paroi bactérienne⁶⁵.

Chez *S. lugdunensis*, cette capacité d'adhésion est imputée à ce jour à une protéine de liaison au fibrinogène (Fbl)^{66,67} et une protéine de liaison au facteur *von Willebrand* (vWbl)^{68,69}.

1.3.1.1. La protéine de liaison au fibrinogène, Fbl

La *Fibrinogen binding protein*, Fbl, est une protéine appartenant à la famille des MSCRAMM. Elle est codée par le gène *fbl*, retrouvé chez toutes les souches de *S. lugdunensis*⁸ et utilisé comme marqueur spécifique d'identification de l'espèce par *Polymerase Chain Reaction* (PCR)¹⁰.

La protéine Fbl est composée de 881 acides aminés (aa) et présente une homologie de 58% avec le *Clumping factor A* (ClfA), un important facteur de virulence de *S. aureus*, responsable de la liaison au fibrinogène et aux plaquettes⁷⁰. Une organisation similaire a été décrite entre les protéines Fbl et ClfA (**Figure 6**)⁶⁶. Les domaines de liaison de la protéine Fbl présentent une identité protéique partielle ; 60% d'identité pour les motifs N2 et N3 mais seulement 19% pour le motif N1.

Les domaines R des gènes *clfA* et *fbl* contiennent de nombreux motifs répétés de 18 paires de bases (pb) codant un hexapeptide de type SD (Sérine – Acide aspartique) et SDSDSA (Sérine – Acide aspartique ... – Alanine) respectivement⁶⁶. Ces motifs de 18 pb sont présents en nombre variable. L'équipe de Hartford *et al.*⁷¹ a montré que la capacité d'une souche de *S. aureus* à adhérer à du fibrinogène immobilisé est proportionnelle à la longueur du domaine R du gène *clfA*. En d'autres termes, plus le nombre de motifs répétés est important, plus l'adhésion au fibrinogène est importante⁷¹.

Le clonage du gène *fbI* chez une espèce incapable de lier le fibrinogène, *Lactococcus lactis*, a montré l'implication de Fbl dans la liaison au fibrinogène soluble et immobilisé⁶⁶. Ces travaux ont aussi montré l'inhibition de la liaison de *L. lactis* au fibrinogène en présence d'anticorps (Ac) anti-Fbl. De même, les Ac anti-Fbl inhibent l'adhésion médiée par ClfA, confirmant la très forte homologie des deux protéines⁶⁶.

Plus récemment, d'autres travaux ont confirmé ces résultats par la construction de mutants de délétion du gène *fbI* chez trois souches de *S. lugdunensis* ; la délétion entraîne une abolition totale de l'adhérence des souches au fibrinogène immobilisé ou soluble et une restauration est observée après complémentation du gène *fbI*⁷². L'ensemble de ces résultats suggère que la liaison au fibrinogène est uniquement médiée par Fbl.

Ainsi, Fbl est la protéine majeure de liaison au fibrinogène et constitue à ce titre un facteur de virulence primordial de cette espèce, seul SCN à ce jour à exprimer une protéine homologue de ClfA. Toutefois, même si le gène *fbI* est présent chez toutes les souches de *S. lugdunensis*, une étude de Marlinghaus *et al.*⁷² a montré que sur une collection de 104 souches, seulement 13% liaient fortement le fibrinogène, 14% modérément et 73% ne le liaient pas. Ces résultats soulèvent des questions quant à l'expression du gène, la production et la putative maturation de la protéine mais aussi la possible production d'une capsule chez certaines souches, qui masquerait ainsi l'accessibilité du fibrinogène à cette protéine de liaison.

1.3.1.2. La protéine de liaison au facteur von Willebrand, vWbl

Le facteur *von Willebrand* (vWf) est une glycoprotéine plasmatique jouant un rôle primordial dans l'hémostase primaire. Cette protéine est synthétisée par les cellules endothéliales et les mégacaryocytes puis stockée par les granules alpha plaquettaires. En cas de lésion vasculaire, le vWf se lie au collagène de la paroi vasculaire lésée par l'intermédiaire des thrombocytes pour faciliter l'adhésion et l'agrégation plaquettaire. Par ailleurs, il permet également le transport du facteur VII et assure son activité coagulante⁶⁸.

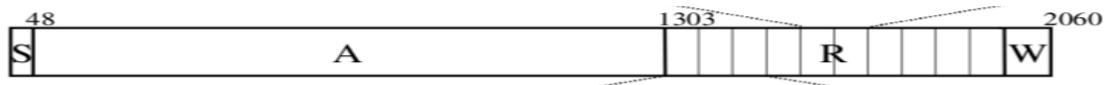


Figure 7 : Représentation schématique de la protéine vWbl de *S. lugdunensis*⁶⁸.

Depuis la région N-terminale, la protéine est successivement composée d'un peptide signal (S), d'une région non-répétée comportant le motif RGD de liaison aux intégrines (A), d'un domaine de 10 séquences répétées (R) et d'un domaine d'ancrage à la paroi LPXTG (W).

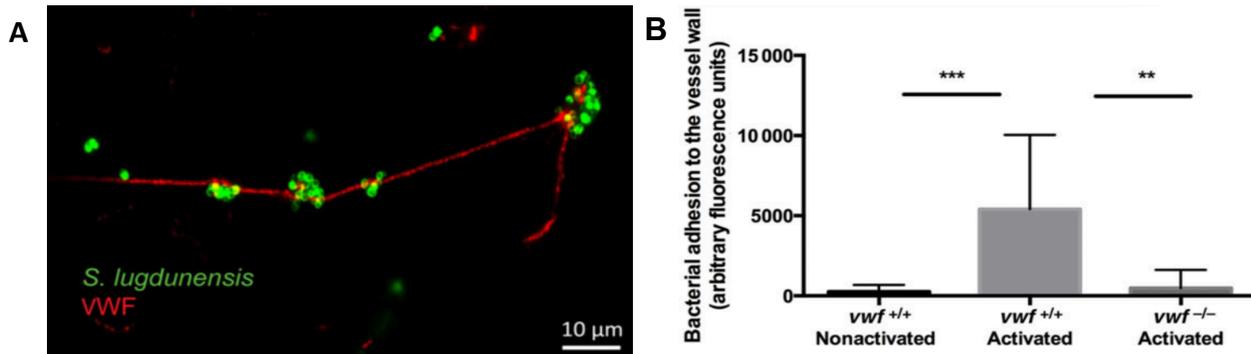


Figure 8 : Adhésion de *S. lugdunensis* à l'endothélium vasculaire médiée *in vitro* et *in vivo* par le vWf⁶⁹.

A, Visualisation par immunofluorescence de la liaison directe de *S. lugdunensis* (vert) au vWf (rouge) dans du plasma humain dans une chambre de micro-circulation.

B, Modèle murin de perfusion veineuse mésentérique. Après activation de l'endothélium vasculaire, un *inoculum* de $3 \cdot 10^7$ UFC de *S. lugdunensis* couplé à un fluorophore a été administré *via* un cathéter veineux jugulaire. L'adhésion de *S. lugdunensis* à l'endothélium vasculaire n'a été visualisée qu'après son activation ; elle est nulle pour les souris dépourvues de vWf (*vwf*^{-/-}). Les valeurs de significativité statistique **P < 0,01 et ***P < 0,001 ont été obtenues avec le test Mann-Whitney.

L'un des facteurs de virulence de *S. lugdunensis* est capable de se lier au vWf⁶⁸. Il s'agit de la protéine vWbl, codée par le gène *vwbl*, appartenant elle aussi à la famille des MSCRAMM. Cette adhésine est composée de 2060 aa et comporte tous les domaines fonctionnels permettant d'assurer sa fonction (peptide signal et motif d'ancrage à la paroi). Elle comporte surtout en position C-terminale un domaine de dix séquences répétées de taille variable permettant l'attachement au vWf *via* une séquence de liaison de 24 aa⁶⁸ (**Figure 7**). Il est à noter, en position N-terminale, un motif RGD (Arginine-Glycine-Asparagine), motif retrouvé chez les protéines capables de se lier aux intégrines^{2,68}.

Le rôle de l'adhésion de *S. lugdunensis* au vWf dans les étapes initiales du développement des EI a récemment été démontré⁶⁹. En effet, les travaux de Liesenborghs *et al.*⁶⁹ ont démontré la co-localisation de *S. lugdunensis* et de vWf par immuno-marquage dans un modèle de chambre de micro-circulation sanguine. Ils ont aussi montré que l'adhésion de *S. lugdunensis* à l'endothélium vasculaire activé est en partie médiée par le vWf dans un modèle murin de perfusion veineuse mésentérique. Le rôle de vWbl a finalement été confirmé dans un modèle murin d'endocardite puisque l'adhésion de *S. lugdunensis* à la valve aortique était significativement réduite en absence de vWf⁶⁹ (**Figure 8**).

De façon intéressante, il a été mis en évidence que seuls *S. aureus* et *S. lugdunensis* possèdent la capacité d'adhésion au vWf, confirmant que *S. lugdunensis* serait le seul SCN à posséder une protéine possédant une telle fonction. Ainsi, la protéine vWbl jouerait un rôle critique dans l'attachement aux lésions endothéliales vasculaires de l'endocarde et pourrait contribuer au développement des végétations et à la pathogenèse des endocardites infectieuses à *S. lugdunensis*.

1.3.1.3. Les protéines de surface putatives SIs

L'analyse du génome complet de la souche N920143 séquencée par Heilbronner *et al.*⁷³ a permis l'identification de sept gènes codant des *S. lugdunensis* surface proteins (SIs) putatives. L'analyse *in silico* a montré que ces protéines possèdent des domaines conservés relatifs à une fonction d'adhésion, à savoir un peptide signal et un motif d'ancrage à la paroi. Les sept protéines, SIsA à G, ont un

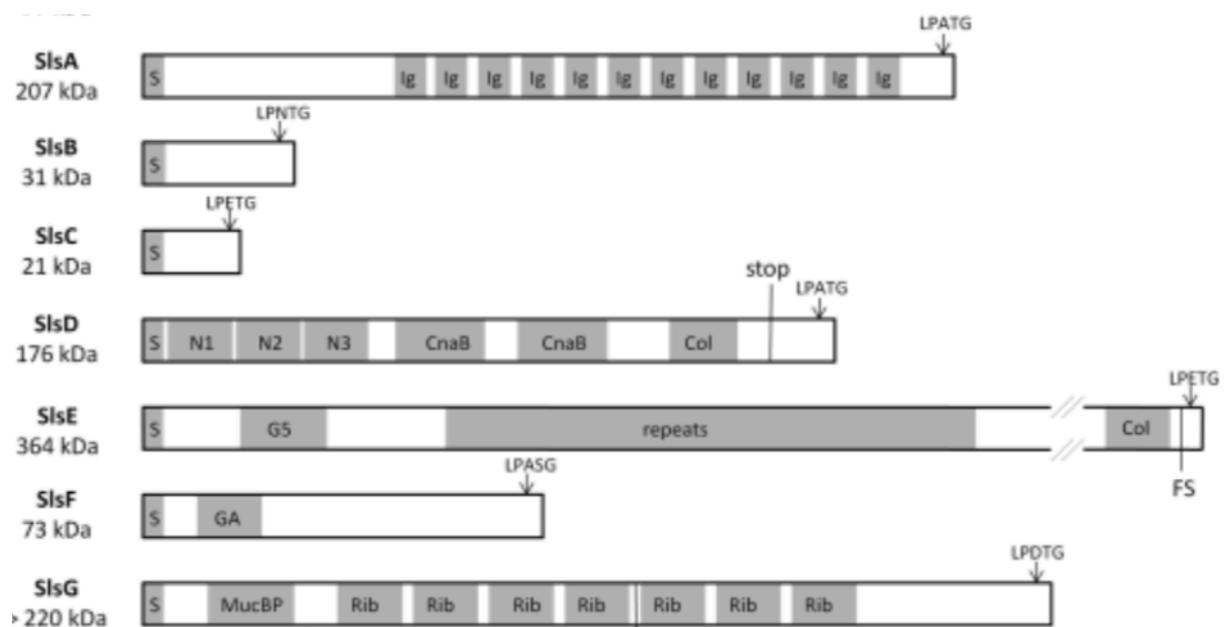


Figure 9 : Représentation schématique des protéines SlsA à G potentiellement ancrées à la paroi bactérienne⁷³.

Les domaines prédits sont indiqués dans des carrés grisés. Sont représentés, le peptide signal (S), les domaines répétés structurellement proches d'une immunoglobuline, le domaine de liaison au fibrinogène (N1, N2 et N3), les domaines B similaires à l'adhésine Cna (CnaB), le domaine de liaison au N-acétylglucosamine (G5), le domaine de liaison à l'albumine (GA), le domaine de liaison aux mucines (MucBP), les domaines de protection contre la réponse immunitaire de l'hôte (Rib) et le domaine d'ancrage à la paroi (LPXTG).

poids moléculaire compris entre 20,8 et 380 kilodaltons (kDa) (**Figure 9**). Parmi elles, cinq présentent une organisation particulière et des motifs putatifs de liaison aux facteurs de l'hôte :

- La protéine SlsA de 1930 aa (207 kDa) comprend 12 domaines répétés non identiques, structurellement proches d'une immunoglobuline G, et localisés entre la région N-terminale et le motif LPXTG.
- La protéine SlsD de 1137 aa (176 kDa) présente en position N-terminale des motifs N1, N2 et N3 de structure similaire au domaine de liaison au fibrinogène de l'adhésine SdrG de *S. epidermidis*⁷⁴. Du côté C-terminal, sont présents deux domaines répétés similaires aux domaines B de l'adhésine Cna permettant la liaison de *S. aureus* au collagène. Toutefois, il a été identifié une mutation non-sens (codon stop) en position C-terminale en amont du motif LPXTG, rendant son ancrage à la paroi bactérienne incertain.
- La protéine SlsE de 3459 aa (364 kDa) comporte en position N-terminale un domaine G5 de liaison au N-acétylglucosamine, retrouvé dans les protéines SasG et Aap qui sont impliquées dans la formation de biofilm de *S. aureus* et *S. epidermidis* respectivement⁶³. Ce domaine est suivi de 46 séquences répétées d'un motif de 31 aa. Cependant, un décalage du cadre de lecture (FS : *Frameshift*) est localisé en position C-terminale, rendant son expression incertaine.
- La protéine SlsF de 659 aa (73kDa) comporte un domaine GA de liaison putatif à l'albumine présentant une organisation similaire à celui de l'espèce *Peptostreptococcus magnus*⁷⁵.
- La protéine SlsG de 2032 aa (220 kDa) présente un motif MucBP de liaison aux mucines (protéines du mucus) et plusieurs motifs répétés Rib, retrouvés dans des protéines de surface du streptocoque du groupe B. Ces motifs pourraient être impliqués dans la protection contre la réponse immunitaire de l'hôte⁷⁶.

Ces MSCRAMM putatives sont particulièrement intéressantes car elles peuvent possiblement jouer un rôle dans l'adhésion aux tissus épithéliaux de l'hôte, dans l'échappement au système immunitaire et dans la formation de biofilm⁷³. A ce jour, aucune équipe n'a caractérisé la fonction de ces protéines, certainement du fait de

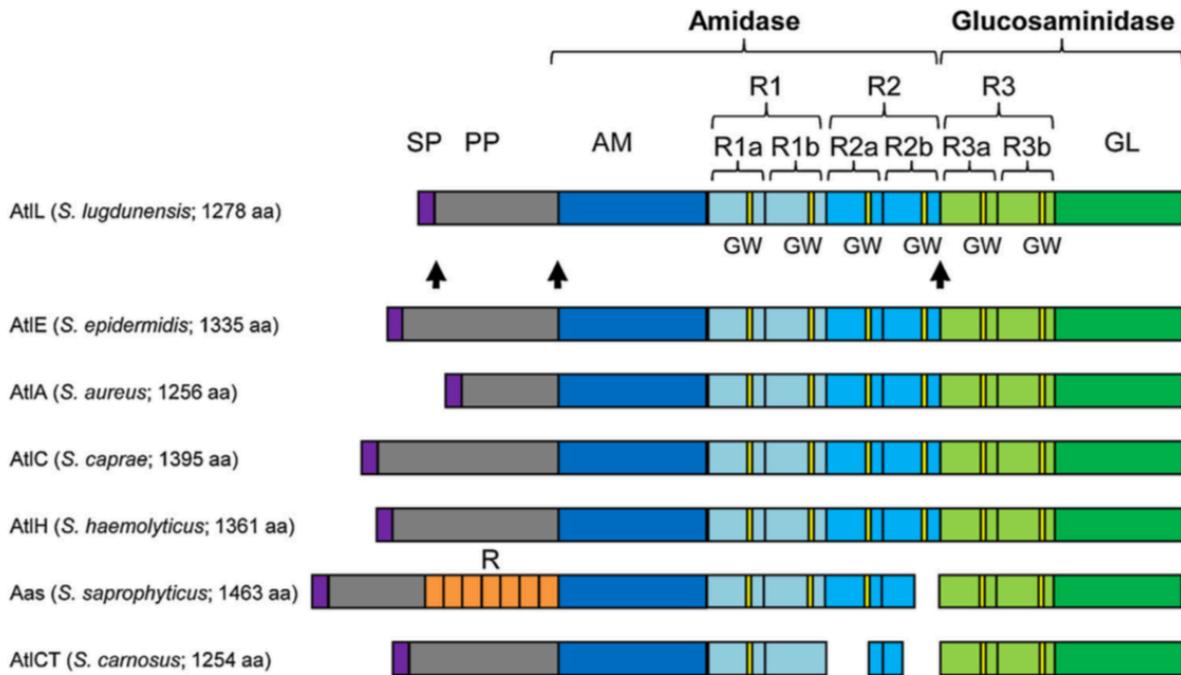


Figure 10 : Représentation schématique de la structure des autolysines de type AtI au sein du genre *Staphylococcus*⁷⁹.

Les autolysines de type AtI comportent deux domaines catalytiques : un domaine de type N-acétylmuramoyl-L-alanine amidase (AM) et un domaine de type glucosamidase N-acétylglucosaminidase (GL). Les domaines à activité enzymatique sont séparés par des séquences répétées R1, R2 et R3 comportant des motifs GW (Glycine – Tryptophane). Les autolysines comportent également en position N-terminale d'un peptide signal (SP) suivi d'un propeptide (PP).

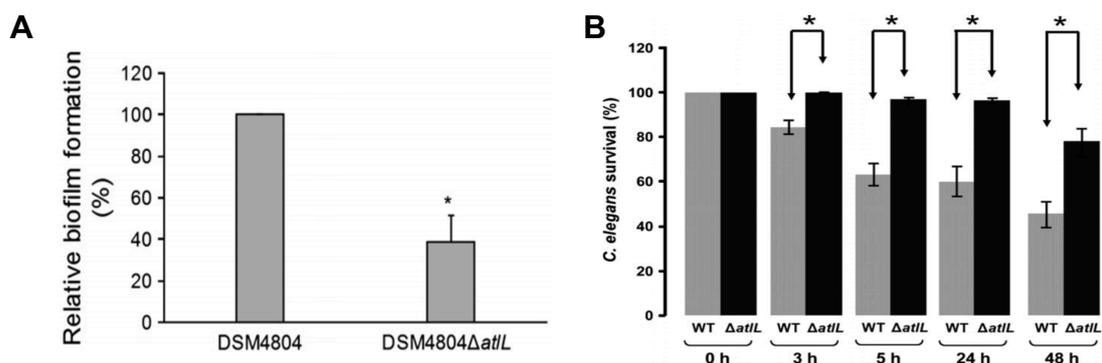


Figure 11 : (A) Comparaison de la capacité de la souche parentale de *S. lugdunensis* DSM4804 et de son mutant DSM4804 Δ atIL à former du biofilm. (B) Virulence comparée de ces deux souches dans un modèle d'infection du nématode *Caenorhabditis elegans*⁸¹.

A, Capacité des souches à former du biofilm après 24h d'incubation en microplaque.
B, Survie du nématode *C. elegans* après 3h, 5h, 24h et 48h d'infection avec la souche sauvage en gris et la souche mutée en noir. La valeur de significativité statistique * $P < 0,001$ a été obtenue avec le test Mann–Whitney.

leur taille importante, de la difficulté de transformation de *S. lugdunensis*, mais aussi des possibles phénomènes de compensation entre elles chez un mutant de délétion.

1.3.1.4. L'autolysine AtIL

Les autolysines ou hydrolases du peptidoglycane (HPG) assurent des fonctions physiologiques essentielles au sein de la cellule bactérienne. Elles sont notamment impliquées dans la croissance cellulaire, le *turnover* du peptidoglycane et le processus de septation cellulaire⁷⁷. La plupart des HPG sont composées de deux régions, l'une possédant l'activité enzymatique et l'autre un domaine de liaison à la paroi de façon non covalente. Toutefois, parmi les HPG, sont décrites les autolysines bifonctionnelles de type "AtI", comportant deux domaines catalytiques. Au sein du genre *Staphylococcus*, les autolysines "AtI" présentent une structure similaire, avec de la région N-terminale à la région C-terminale : un peptide signal (SP), un propeptide (PP), un domaine catalytique de type amidase (N-acétylmuramoyl-L-alanine amidase [AM]), des séquences répétées possédant des motifs GW (Glycine – Tryptophane) (R), et un domaine catalytique de type glucosamidase (N-acétylglucosaminidase [GL])^{78,79} (**Figure 10**).

Chez *S. lugdunensis*, une autolysine bifonctionnelle, nommée AtIL, a été décrite en 2009 par Bourgeois *et al.*⁸⁰ Par la suite, l'implication d'AtIL dans la virulence a été montré grâce à la délétion du gène codant cette autolysine^{79,81}. La délétion affecte la capacité de la souche mutée à former du biofilm^{79,81} ainsi que sa virulence dans un modèle d'infection du nématode *Caenorhabditis elegans* (**Figure 11**)⁸¹. Par ailleurs, AtIL est également impliquée dans la liaison à la fibronectine, la vitronectine et au fibrinogène, mais aussi dans les processus d'adhésion et d'internalisation de *S. lugdunensis* par les endothéliales EA.hy926⁷⁹.

Ainsi, en plus de ses rôles physiologiques essentiels à la bactérie, AtIL joue un rôle significatif dans les processus d'adhérence, d'internalisation cellulaire et de formation de biofilm.

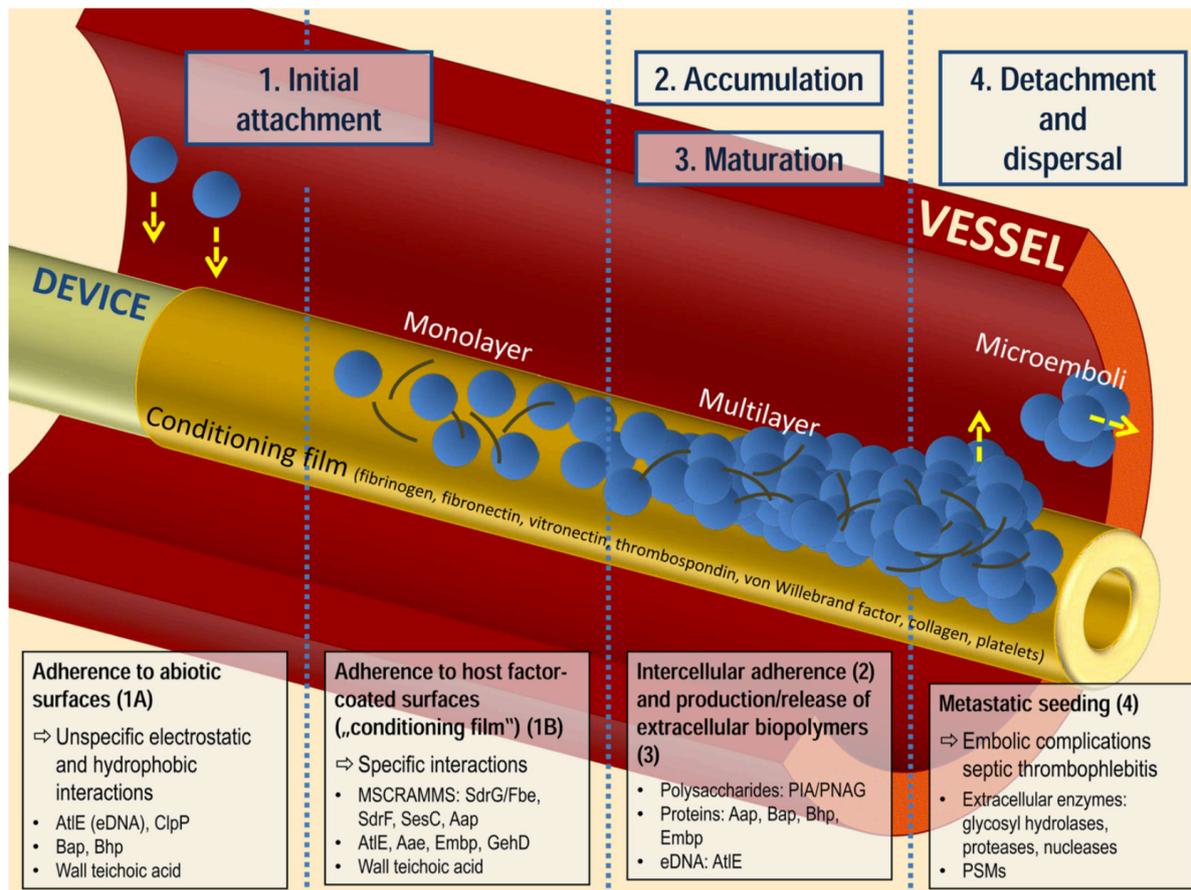


Figure 12 : Pathogénèse des infections sur cathéter et facteurs influençant la formation de biofilm³⁰.

Le processus de formation de biofilm à la surface d'un cathéter intravasculaire est représenté en quatre étapes. Après l'attachement initial, les bactéries s'accumulent à la surface du cathéter et le biofilm se développe par des processus d'adhésion intercellulaire médiée par des exopolymères polysaccharidiques, des protéines et de l'ADN extracellulaire. Une fois le biofilm mature, des bactéries peuvent se détacher et gagner la circulation sanguine pour coloniser de nouvelles surfaces et entraîner des complications emboliques.

1.3.2. La formation de biofilm

Le biofilm est une communauté structurée de cellules bactériennes englobées dans une matrice hétérogène et adhérente. Cette architecture bactérienne confère généralement une moindre sensibilité aux antibiotiques et peut être à l'origine d'échecs thérapeutiques, car ces derniers diffusent peu ou mal dans cette matrice⁸².

Le développement d'un biofilm est un processus complexe et multifactoriel qui peut être divisé en trois étapes³⁰ (**Figure 12**) :

- L'adhésion initiale
- L'accumulation et la maturation du biofilm
- Le détachement et la dispersion du biofilm

1.3.2.1. L'adhésion initiale

L'adhésion initiale est la première étape de la formation du biofilm. Elle commence par un attachement "réversible" puis "irréversible" des cellules bactériennes à une surface biotique ou abiotique telle que les dispositifs médicaux. Cette étape d'adhésion est le produit de liaisons physico-chimiques faibles remplacées par des liaisons médiées par les composants de la surface bactérienne telles que les MSCRAMM ou les autolysines. Ces protéines jouent un rôle prépondérant dans la colonisation qui constitue la première étape du processus infectieux^{30,63}.

1.3.2.2. La maturation du biofilm

Une fois solidement attaché à une surface, le biofilm va se développer lors d'une phase de maturation. Cette phase est caractérisée par une adhésion intercellulaire *via* la production d'une matrice composée de plusieurs éléments. Chez *S. aureus*, le composé majeur de cette matrice est un exopolymère polysaccharadique nommé *polysaccharide intercellular adhesin* (PIA), mais des acides teichoïques, de l'Acide DésoxyriboNucléotidique (ADN) extracellulaire (ADNe) et des protéines de surface peuvent également la composer⁸³. Tous ces composants permettent une structuration du biofilm et une architecture en trois dimensions. Toutefois, l'absence du *locus ica*

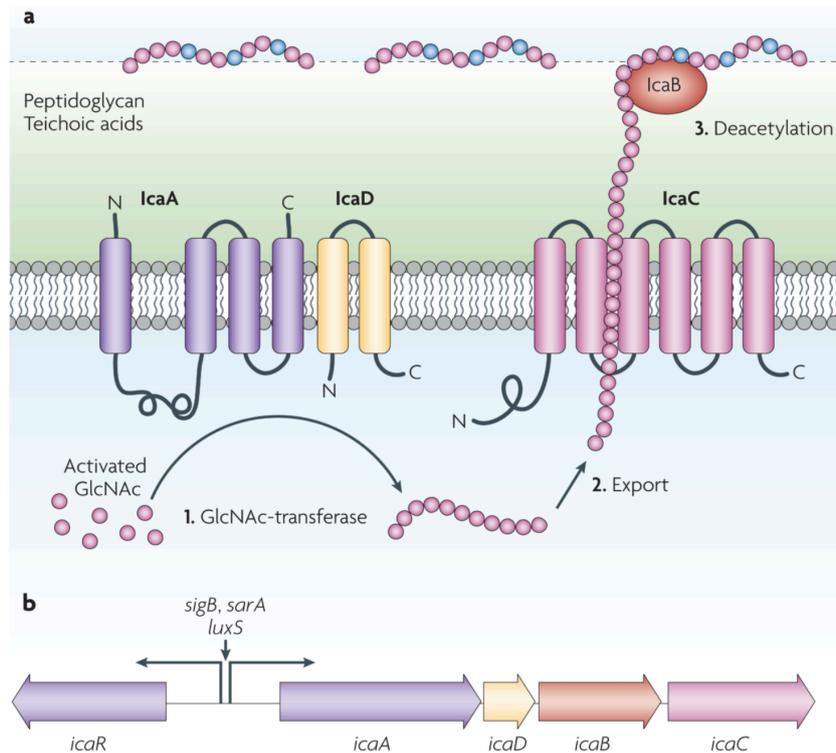


Figure 13 : Voie de synthèse du *polysaccharide intercellular adhesin* (PIA) par le locus *ica*⁸³.

Le locus *ica* est impliqué dans la synthèse du PIA. Les protéines IcaA et IcaD sont des N-acétylglucosamine transférases qui catalysent la formation de longues chaînes de monomères de N-acétylglucosamine. L'élongation et l'export du PIA sont sous la dépendance de la protéine membranaire IcaC jusqu'à sa N-désacétylation par la protéine IcaB qui introduit les charges positives primordiales pour l'adhésion du PIA à la surface bactérienne.

chez certaines espèces de staphylocoques ou chez certaines souches de *S. aureus* et de *S. epidermidis* ne signifient pas qu'elles ne peuvent pas former de biofilm^{84,85}.

Ainsi, il existe deux types de maturation du biofilm : l'adhésion PIA-dépendante et l'adhésion PIA-indépendante.

L'adhésion PIA-dépendante

Chez *S. aureus* et *S. epidermidis*, le PIA est également connu sous le nom de poly-N-acétylglucosamine. Le PIA est le principal composant de la matrice des biofilms des staphylocoques qui donne l'aspect du "slime"⁸⁶. Il porte une charge positive nette qui favorise les interactions avec les acides téichoïques chargés négativement à la surface des bactéries et permet ainsi les interactions électrostatiques avec les cellules bactériennes et donc la maturation du biofilm⁸⁷.

Chez les staphylocoques, le *locus ica* (pour *intercellular adhesion*) est composé de quatre gènes, *icaA*, *icaB*, *icaC*, *icaD*, et d'un régulateur transcriptionnel, *icaR*. Ce *locus* est impliqué dans la synthèse du PIA. Les protéines IcaA et IcaD sont des N-acétylglucosamine transférases qui catalysent la formation de longues chaînes de monomères de N-acétylglucosamine. L'élongation et l'export du PIA sont sous la dépendance de la protéine membranaire IcaC jusqu'à sa N-désacétylation par la protéine IcaB qui introduit les charges positives primordiales pour l'adhésion du PIA à la surface bactérienne⁸³ (**Figure 13**).

L'adhésion PIA-indépendant

La formation du biofilm peut également ne pas dépendre du PIA. Dans ce cas, la formation de biofilm et les interactions entre les bactéries peuvent être médiées par des protéines de surface. Chez *S. aureus* et *S. epidermidis*, des protéines ont été identifiées dans la matrice qui compose le biofilm en plus des MSCRAMM déjà impliquées dans le processus d'adhésion initiale⁸⁸.

Chez *S. aureus*, à partir de souches très adhérentes et fortement productrices de biofilm, a été identifiée la *Biofilm Associated Protein* (Bap)⁸⁶. Cette protéine est impliquée dans la formation de biofilm et dans l'adhésion intercellulaire. En outre, la protéine Bap est capable d'interagir avec le PIA lorsqu'il compose le biofilm. Cependant, en absence de ce composé, la protéine Bap à elle seule peut induire la formation de biofilm⁸⁶. Par ailleurs, il a été démontré chez *S. aureus* que la protéine A ou encore les protéines FnbpA et FnbpB de liaison à la fibronectine influeraient sur la formation de biofilm. En effet, O'Neil *et al.*⁸⁹ ont montré qu'un mutant du gène *bap* ou un double mutant des gènes *fnbpA* et *fnbpB* étaient incapables de former du biofilm.

Chez *S. epidermidis*, une protéine homologue de la protéine Bap a été identifiée, la *Bap homolog protein* (Bhp)⁸⁸, jouant le même rôle que chez *S. aureus*. Une protéine a également été identifiée dans le biofilm de *S. epidermidis*, nommée *Accumulation-associated protein* (Aap) jouant un rôle important dans les biofilms PIA-indépendant⁸⁸. Dans ces biofilms, la protéine Aap contribuerait à la fois à l'étape d'adhésion initiale mais aussi à la phase de maturation en créant des liens entre les cellules bactériennes par la formation de "fibres" à leur surface^{63,88}.

Enfin l'ADNe généré par la lyse bactérienne agit comme une "glue" dans le biofilm grâce à sa charge négative et contribue ainsi à sa structuration⁹⁰.

1.3.2.3. Le détachement du biofilm et la dispersion du biofilm

Le détachement du biofilm va permettre la dissémination des bactéries dans l'organisme de l'hôte et la colonisation de nouvelles surfaces. Cette dissémination peut être la résultante de forces mécaniques comme le flux sanguin ou l'action de protéases, de glycosyl hydrolases, de *phenol-soluble modulins* (PSM) et de nucléases qui dégradent la matrice du biofilm. Cette dispersion peut conduire à des complications emboliques de l'infection³⁰ (**Figure 12**).

Tout ce processus d'adhésion et de formation de biofilm est sous le contrôle de mécanismes de communications entre les bactéries appelés "*quorum sensing*" sous le contrôle du régulateur global de la virulence *agr* (*accessory gene regulator*)⁸⁸ (**cf paragraphe 1.4**).

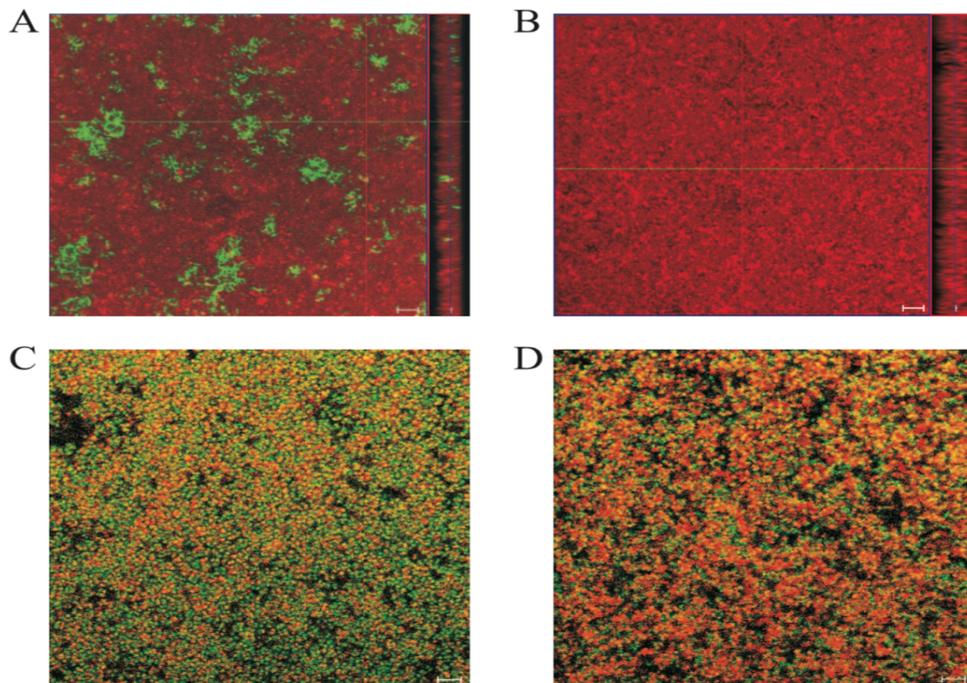


Figure 14 : Évaluation de la matrice extracellulaire du biofilm de *S. lugdunensis* et *S. epidermidis* par microscopie confocale à balayage laser⁹¹.

A/B, Visualisation de la présence ou de l'absence de *polysaccharide intercellular adhesin* (PIA) dans la matrice du biofilm de *S. epidermidis* RP62A (**A**) et de *S. lugdunensis* IDRL-2640 (**B**). Le PIA est coloré en vert et les cellules en rouge par des marqueurs fluorescents. De larges structures de PIA sont visibles dans le biofilm de *S. epidermidis* mais pas dans celui de *S. lugdunensis*.

C/D, Détection de protéines extracellulaires dans le biofilm de *S. epidermidis* RP62A (**C**) et de *S. lugdunensis* IDRL-2640 (**D**). Le marquage SYPRO Ruby permet la coloration des protéines en rouge et le marquage Syto-9, celle des bactéries en vert. La quantité de protéines est significativement plus importante pour le biofilm de *S. lugdunensis* que celui de *S. epidermidis*.

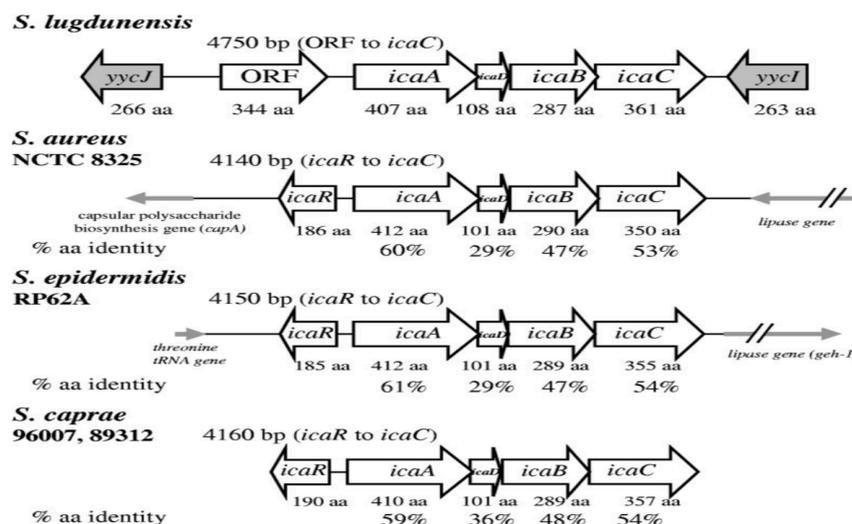


Figure 15 : Comparaison de l'organisation génétique du locus *ica* de *S. lugdunensis*, *S. aureus*, *S. epidermidis* et *S. caprae*⁹¹.

Une région chromosomique de 7,6 kb englobant les gènes *icaADBC* de *S. lugdunensis* a été séquencée à partir des isolats IDRL-2414 et IDRL-2664. Un gène codant une glycosyl hydrolase putative de 344 acides aminés (ORF) a été identifié à la position où le gène *icaR* est présent chez les autres espèces de staphylocoques. Les pourcentages d'identité protéique entre les protéines Ica de *S. lugdunensis* et celles des autres staphylocoques sont indiquées sous les flèches.

1.3.2.4. Le biofilm de *S. lugdunensis*

Frank *et al.*⁹¹ ont étudié la matrice du biofilm de *S. lugdunensis* et montré sa nature essentiellement protéique. En effet, *via* des traitements par la protéinase K, la matrice du biofilm a été déstabilisée. La composition du biofilm de *S. lugdunensis* a également été confirmée par une étude de microscopie confocale où le marquage SYPRO Ruby a révélé une importante quantité de protéines (**Figure 14**)⁹¹.

De manière surprenante, il existe chez *S. lugdunensis* un homologue du *locus ica* partageant une identité protéique de 29 à 61% avec celui de *S. aureus* et de *S. epidermidis* (**Figure 15**)⁹¹. Cependant, il ne semble pas être exprimé par *S. lugdunensis* puisque son biofilm est PIA-indépendant. A ce jour, les protéines qui le composent n'ont toujours pas été identifiées.

En 2015, Rajendran *et al.*⁹² se sont intéressés aux gènes impliqués dans la formation de biofilm. Pour ce faire, une banque de mutants de *S. lugdunensis* a été créée par insertion aléatoire d'un transposon (Tn917) dans le génome⁹². Parmi les 5000 mutants d'insertion obtenus, un seul présentait une capacité réduite à former du biofilm. Le séquençage du site d'insertion du transposon a alors permis d'observer que le transposon était inséré dans le gène *comEB*, impliqué dans la compétence bactérienne. Par ailleurs, la matrice du biofilm de la souche mutante comportait une quantité inférieure d'ADNe par rapport à la souche sauvage, non expliquée par une diminution de l'expression des gènes d'autolysines, ni par une diminution de la viabilité cellulaire ou par une diminution de l'activité autolytique. Ainsi, le gène *comEB* serait impliqué dans la formation de biofilm *via* un mécanisme de libération d'ADN indépendant de la lyse bactérienne⁹².

La capacité à produire du biofilm ne semble pas être une caractéristique "obligatoire" de cette espèce. Une étude évaluant la capacité de 38 souches (22 souches isolées d'IPTM, neuf d'infections profondes et sept d'infections sur matériel) à produire du biofilm en microplaque a montré que seulement 36,8% des souches produisaient du biofilm⁹³. De plus, cette capacité variait en fonction du type d'infection puisque 57,1% des souches responsables d'infections sur matériel produisaient du biofilm contre 36,4 et 22,2% des souches responsables d'IPTM et d'infections profondes, respectivement⁹³. Une étude récente sur la capacité de former du biofilm

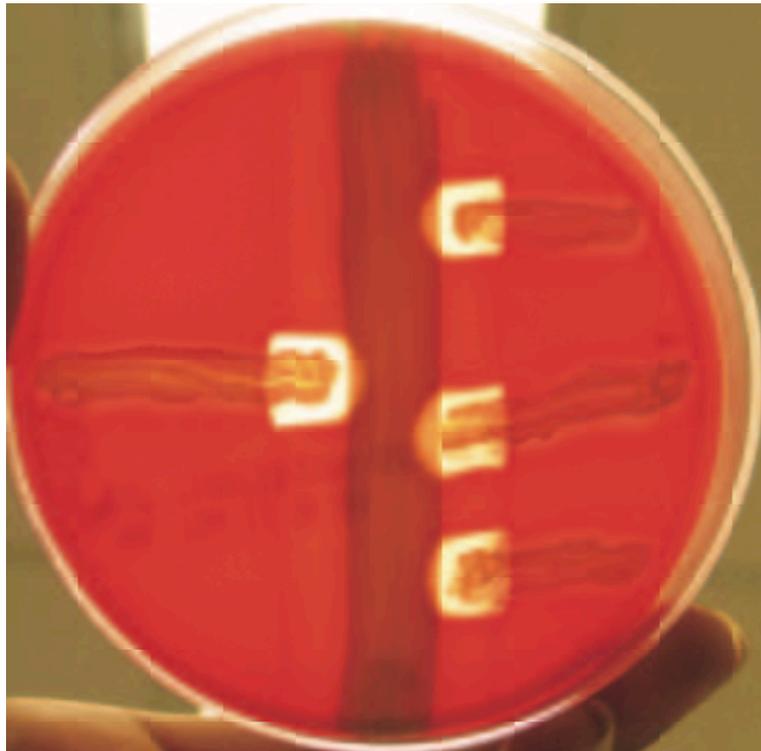


Figure 16 : Activité hémolytique synergique de *S. lugdunensis* par CAMP-test².

Strie verticale de la souche RN4220 de *S. aureus* et stries perpendiculaires des souches de *S. lugdunensis* IDRL-5258 (à gauche) et IDRL24-14 (à droite).

par *BioFilm Ring Test*[®] a montré que 71,4% des 28 souches responsables d'IOA étudiées étaient capables de former du biofilm en moins de deux heures et que 100% l'étaient après quatre heures d'incubation⁹⁴.

La capacité de *S. lugdunensis* à former du biofilm peut donc être considérée comme un réel facteur de virulence, en particulier dans les infections sur matériel^{2,94}. Les facteurs impliqués à ce jour sont AtIL, ComEB et IsdC (cf **paragraphe 1.3.4**).

1.3.3. Les hémolysines ou toxines cytolytiques

Au sein du genre *Staphylococcus*, *S. aureus* est le seul qui est capable de produire de nombreuses toxines endommageant les membranes des cellules de l'hôte *via* la formation de pores conduisant à la lyse des cellules. Parmi ces toxines, figurent les hémolysines qui induisent la lyse des érythrocytes. Il a été décrit chez *S. aureus*, plusieurs protéines, comme les α -, β - et δ -hémolysines ou encore les PSM, conférant à cette espèce une activité hémolytique⁹⁵.

Aucun gène codant une α -, β - et δ -hémolysine n'a été identifié au sein du génome de *S. lugdunensis*⁹⁶. Toutefois, une activité hémolytique agissant en synergie avec une β -hémolysine exogène a été mise en évidence chez *S. lugdunensis* et peut être révélée par un CAMP-test². Ce dernier consiste à réaliser, sur une gélose au sang de mouton, une strie centrale avec une bactérie productrice d'une β -hémolysine exogène (*S. aureus* par exemple) et à inoculer les souches à tester sous forme de stries perpendiculaires² (**Figure 16**).

Cette activité hémolytique synergique a été observée pour une majorité de souches de *S. lugdunensis*, 95% dans une première étude⁹⁷ et 77% dans une seconde⁹⁸. Cette activité est due à la présence d'une δ -hémolysine codée par l'opéron *Staphylococcus lugdunensis Synergetic Hemolysin (slush)*. Ce locus *slush* est composé de trois gènes, *slush-A*, *slush-B* et *slush-C*, responsables de la synthèse de trois peptides à l'activité δ -toxin-like⁹⁹.

Il existe une forte homologie de séquence entre les peptides SLUSH et les peptides de type PSM de plusieurs espèces de staphylocoques. En effet, Rautenberg *et al.*¹⁰⁰ ont montré une importante similarité en terme de taille, de

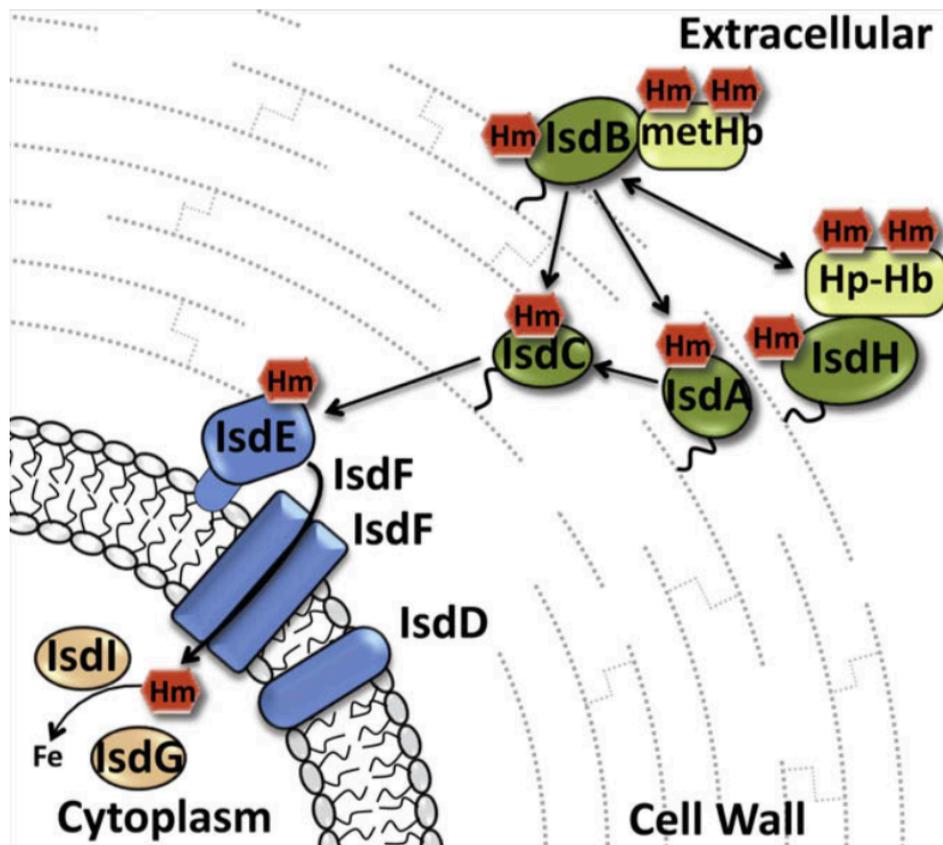


Figure 17 : Représentation schématique des composants du système Isd impliqués dans le transport de l'hème chez *S. aureus*¹⁰³.

séquence et de charge de ces peptides avec les PSM de type β de *S. aureus* et *S. epidermidis*. Ceci suggère que les peptides SLUSH appartiendraient à la famille des PSM de type β et seraient sous le contrôle du *locus agr*¹⁰⁰.

Il a aussi été décrit chez *S. lugdunensis* un gène partageant une identité de 85% avec le gène *hly* de *S. aureus*, codant une β -hémolysine⁷³. Ce gène pourrait coder chez *S. lugdunensis* une protéine à activité putative de sphingomyélinase, responsable de la lyse de la sphingomyéline, composant majeur de la membrane des érythrocytes¹⁰¹. Ce gène est fortement prévalent puisqu'il a été retrouvé chez 97% des 58 souches analysées par Szabados *et al.*⁸.

1.3.4. Le locus *isd*, système d'acquisition du fer

La survie bactérienne est fonction de la capacité des bactéries à résister au système immunitaire de l'hôte, mais elle repose aussi sur leur capacité à s'adapter et à ajuster leur métabolisme afin de se développer et de répondre à leur environnement immédiat. Cette aptitude à survivre et s'adapter est rendue possible par la présence de facteurs de *fitness*, dont plusieurs concourent à l'acquisition du fer. En effet, le fer est un élément essentiel à la croissance des bactéries et pas seulement chez les staphylocoques¹⁰².

Il existe chez les staphylocoques plusieurs systèmes de capture du fer. Parmi ceux-ci figure celui codé par le *locus Iron-regulated Surface Determinant (isd)*. C'est le seul système de capture du fer à ce jour identifié chez *S. lugdunensis*, qui est d'ailleurs le seul SCN à le posséder⁷³.

Initialement décrit chez *S. aureus*, ce *locus*, composé de 10 gènes, code un ensemble de protéines permettant la libération du fer de l'hémoglobine et son transport jusque dans le cytoplasme de la bactérie¹⁰³ (**Figure 17**). Parmi elles, les protéines IsdA, IsdB, IsdC et IsdH ancrées à la paroi *via* un motif LPXTG ou NPQXN/P. Les protéines de surface IsdB et IsdH permettent l'extraction et la liaison de l'hémoglobine grâce à des domaines protéiques particuliers appelés NEAT (pour *NEAr iron Transporter*). L'hème va ensuite être transféré aux protéines IsdA et IsdC pour son internalisation *via* les transporteurs membranaires IsdD, IsdE et IsdF. Enfin, les

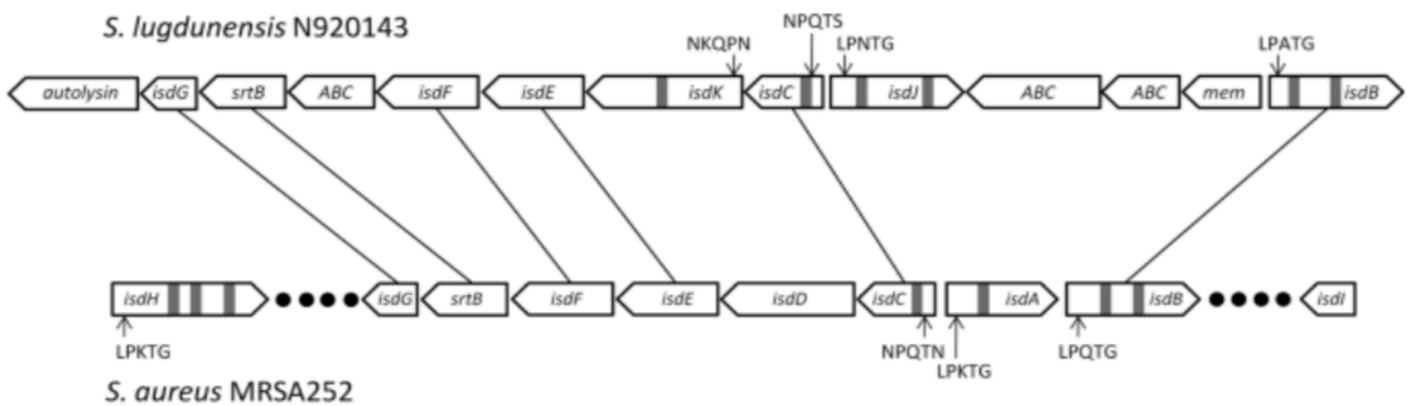


Figure 18 : Comparaison du locus *isd* de *S. aureus* et de *S. lugdunensis*⁷³.

Les motifs NEAT sont représentés par des rectangles grisés. Les gènes orthologues sont reliés par un trait noir et le pourcentage d'identité des protéines codées par ces gènes est de : IsdB 36,8% ; IsdC 57,6% ; IsdE, 74,7% ; IsdF, 57,7% ; SrtB, 58,2% ; IsdG, 68,2%. Les motifs d'ancrage à la paroi LPXTG et NPQXN sont indiqués par des flèches. Les gènes *isdH* et *isdI* de *S. aureus* sont situés hors du locus *isd*, comme suggéré par des ronds noirs.

mono-oxygénases cytoplasmiques IsdG et IsdI vont cliver le noyau porphyrine de l'hème pour libérer le fer ferreux (Fe²⁺) à des fins métaboliques¹⁰³.

Même si le *locus isd* de *S. aureus* présente une organisation similaire et un fort pourcentage d'identité (de 36,8% à 74,4% selon les protéines) avec celui de *S. lugdunensis*, il existe quelques différences^{73,104} (**Figure 18**) :

- Un nombre de gènes différents selon l'espèce. Le *locus isd* de *S. lugdunensis* comprend trois gènes supplémentaires codant des transporteurs *ATP-Binding Cassette*, totalement absents du *locus isd* de *S. aureus*
- La présence de seulement deux protéines ancrées à la paroi *via* un domaine LPXTG chez *S. lugdunensis* (IsdB et IsdJ) contre trois chez *S. aureus* (IsdA, IsdB et IsdH). Ces protéines, toutes composées de domaines NEAT, sont ancrées au peptidoglycane par l'intermédiaire de la sortase A^{105,106}.
- Le gène *isdD*, absent chez *S. lugdunensis*, est remplacé par le gène *isdK*, codant une protéine comprenant un domaine NEAT.
- La présence du gène *isdP* codant l'autolysine IsdP chez *S. lugdunensis*¹⁰⁵, à la place de celui codant la mono-oxygénase IsdI.
- Enfin, la localisation de ce *locus* au sein des génomes de *S. aureus* et *S. lugdunensis* est différente, suggérant une acquisition indépendante du *locus* par ces deux espèces. En effet, celui de la souche de *S. aureus* MRSA252 est localisé au centre du chromosome, alors que celui de *S. lugdunensis* est proche de l'origine de réplication⁷³.

S. lugdunensis possède ainsi des facteurs facilitant sa survie et son développement dans les tissus de l'hôte. Même si le rôle de ce système de capture du fer dans la pathogénicité des infections à *S. lugdunensis* reste à démontrer, sa capacité à dégrader les molécules d'hème pourrait être un facteur permettant la colonisation et l'invasion des tissus cardiaques. En effet, la souche de *S. aureus Newman* dont les gènes codant les mono-oxygénases IsdG et IsdI ont été inactivés, présente une virulence moindre par rapport à la souche sauvage dans un modèle murin de formation d'abcès systémiques. En effet, une réduction significative de la formation d'abcès cardiaques et rénaux a été observée chez les souris infectées par

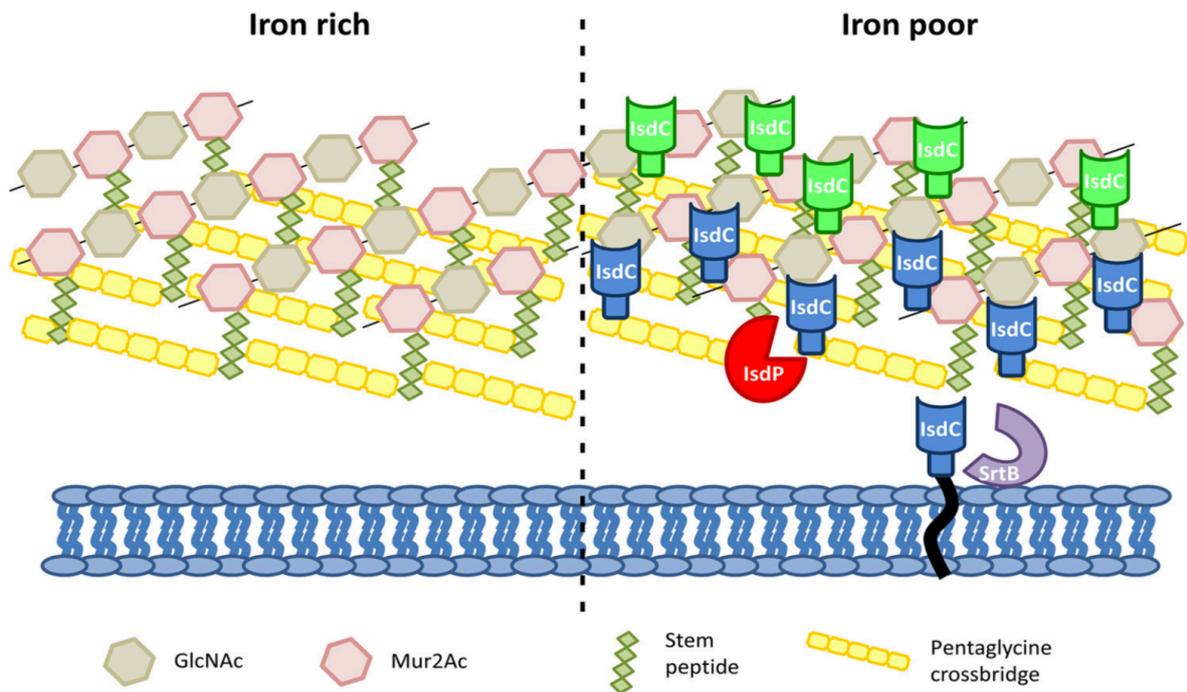


Figure 19 : Modèle de contribution de l'autolysine IsdP à l'acquisition de l'hème chez *S. lugdunensis*¹⁰⁵.

Dans des conditions restreintes en fer, IsdP est exprimée et remodèle la paroi en clivant les ponts pentaglycines permettant l'attachement de la protéine IsdC au peptidoglycane (**en bleu**) via la sortase B. Une sous-population de protéines IsdC (**en vert**) est également présentée à la surface de la bactérie. L'action coordonnée de l'autolysine et de la sortase permet de saturer la paroi avec de nombreuses protéines IsdC, de sorte que la protéine IsdC exposée en surface puisse lier l'hème extracellulaire et l'acheminer de la paroi à la membrane plasmique. Dans des conditions riches en fer, les gènes *isdC* et *isdP* ne sont pas transcrits.

la souche mutante, avec une diminution de la charge bactérienne de deux unités logarithmiques au niveau cardiaque et d'une unité logarithmique au niveau rénal¹⁰⁷.

Il a également été montré l'implication de ce *locus*, et notamment de la protéine IsdC dans la formation de biofilm dans des conditions limitantes en fer chez *S. lugdunensis*¹⁰⁸. La protéine IsdC est liée à la paroi, non-exposée à la surface, mais confinée au sein du peptidoglycane afin d'assurer la fonction de transfert de l'hème jusqu'à la membrane bactérienne¹⁰⁵. Ce confinement dans le peptidoglycane est rendu possible par une sortase B, codée par le gène *srtB*, présent au sein du *locus isd*. Dans le cas de la délétion complète du *locus isd*, de la délétion seule du gène *isdC* ou du gène *srtB*, il n'est plus possible pour une souche mutante de *S. lugdunensis* de former du biofilm¹⁰⁸.

Farrand *et al.*¹⁰⁵ ont proposé un modèle décrivant la contribution des protéines IsdC et IsdP dans l'acquisition du fer. Ainsi, l'autolysine IsdP, de par son rôle physiologique dans l'hydrolyse du peptidoglycane, faciliterait le processus de liaison et de positionnement de la protéine IsdC à la paroi *via* la sortase B dans des conditions limitantes en fer. En accord avec cette hypothèse, il a été montré que les gènes *isdC* et *isdP* étaient surexprimés dans ces conditions (**Figure 19**)¹⁰⁵. Ainsi, l'autolysine IsdP joue un rôle dans la réorganisation de la paroi bactérienne pour l'acquisition de l'hémoglobine et l'absorption de fer durant l'infection¹⁰⁵.

1.3.5. La lugdunine

En 2016, les travaux de Zipperer *et al.*¹⁰⁹ ont consisté à étudier l'activité anti-*S. aureus* d'une collection de 87 souches de staphylocoques, isolées à partir des fosses nasales. Parmi les 87 isolats, seule la souche de *S. lugdunensis* IVK28 inhibait la croissance de *S. aureus* USA300, et ce uniquement en milieu gélosé dans des conditions limitantes en fer.

L'analyse du surnageant de culture de la souche IVK28 a permis d'identifier une molécule antibiotique codée par le *locus lug*, composé de quatre gènes : *lugA*, *lugB*, *lugC* et *lugD*, organisés en une structure opéronique. Cet opéron pourrait être sous le contrôle d'un régulateur de la famille TetR appelé *lugR*. Le *locus* a été retrouvé chez toutes les souches de *S. lugdunensis* testées (n = 17) et exclusivement chez

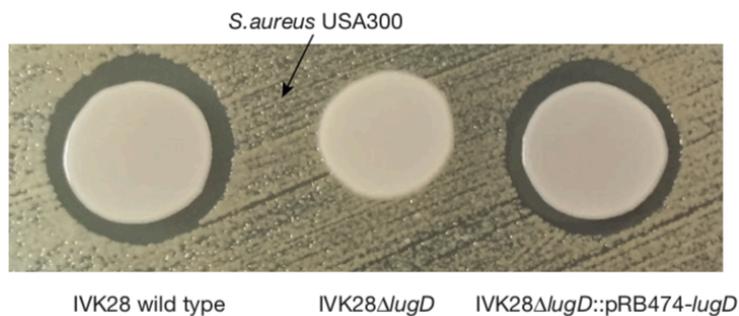


Figure 20 : Évaluation de l'activité anti-microbienne de la lugdunine vis-à-vis de *S. aureus*¹⁰⁹.

Des géloses de milieu minimum ont été inoculées avec la souche de *S. aureus* USA300 et avec des cultures de nuit des souches de *S. lugdunensis* IVK28 sauvage (*wild type*), mutante (IVK28 Δ lugD) et complétée (IVK28 Δ lugD::pRB474-lugD). Les géloses ont ensuite été incubées 24 heures à 37°C avant l'observation d'un diamètre d'inhibition de la croissance de *S. aureus*

S. lugdunensis, suggérant un marqueur spécifique de l'espèce¹⁰⁹. Le faible pourcentage de GC (GC%) de l'opéron *lug* (26,9%) comparativement au 33,8 GC% du génome de *S. lugdunensis*, suggère une potentielle acquisition par transfert horizontal¹⁰⁹.

Des travaux de mutagenèse dirigée ont été effectués chez la souche de *S. lugdunensis* IVK28 afin de déléter le gène *lugD*. La souche mutée ne présentait plus aucune activité anti-microbienne vis-à-vis de la souche de *S. aureus* USA300 comparativement à la souche sauvage, mais le phénotype a été restauré par complémentation avec un gène porteur d'une copie du gène *lugD*¹⁰⁹ (**Figure 20**).

La lugdunine est composée d'un peptide cyclique de cinq aa et d'un hétérocycle thiazolidine ayant une activité anti-microbienne envers les bactéries à Gram positif. Cette molécule a une activité bactéricide envers les bactéries difficiles à traiter comme les souches de *S. aureus* résistantes à la méticilline (SARM) ou les souches de *S. aureus* de sensibilité intermédiaire aux glycopeptides¹⁰⁹. Ainsi, les différences en termes de niches privilégiées de portage de *S. aureus* et de *S. lugdunensis* pourraient être expliquées par la production de la lugdunine¹⁰⁹.

1.3.6. Autres facteurs de virulence

1.3.6.1. Résistance au lysozyme

Le lysozyme est un composant essentiel du système immunitaire inné de l'hôte, lui permettant de se défendre face aux infections bactériennes. Il est le produit de la lyse du lysosome des cellules granulocytaires et monocytaires. Le lysozyme a une activité muramidase qui hydrolyse les chaînes glycanes du peptidoglycane.

Chez *S. lugdunensis* et comme tous les staphylocoques, il a été décrit le gène *oat* qui code une O-acétyl transférase responsable d'une O-acétylation du peptidoglycane¹¹⁰. Cette modification de l'acide N-acétyl-muramique entraîne une résistance au clivage par le lysozyme. *S. lugdunensis* présente la particularité de résister à de fortes concentrations de lysozyme, de l'ordre de 400 mg/ml, et peut ainsi contourner certaines étapes du système immunitaire inné¹¹⁰.

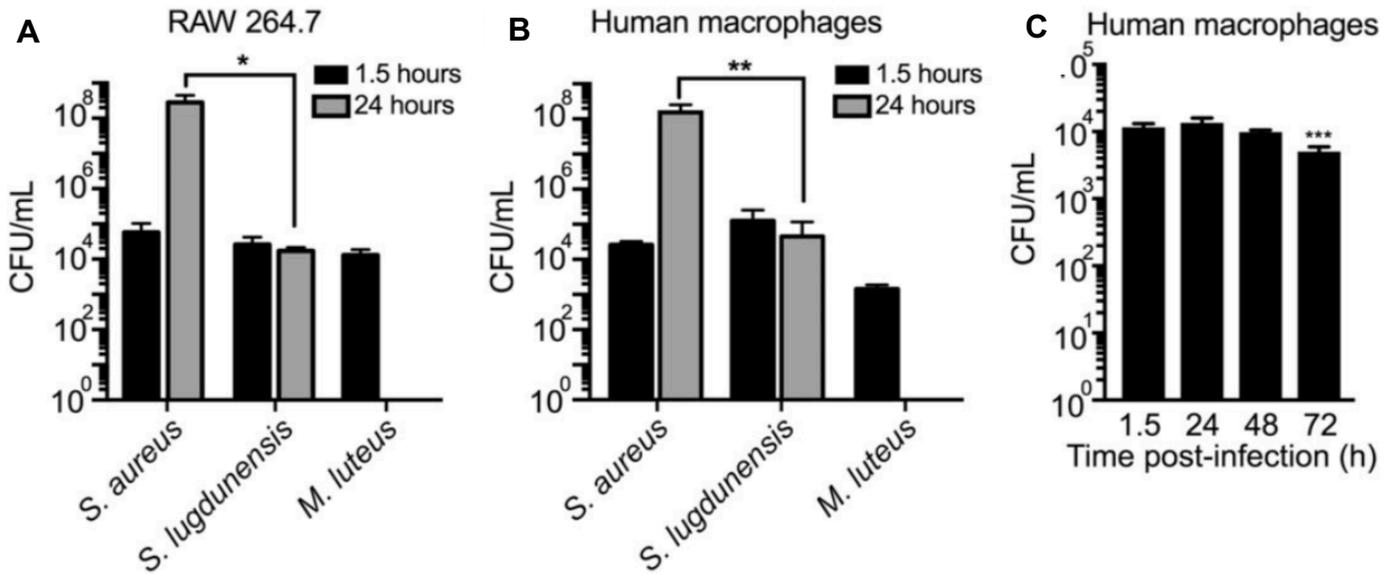


Figure 21 : Évaluation de la capacité des macrophages à éradiquer les souches de *S. lugdunensis* HKU09-01, de *S. aureus* USA300 et de *Micrococcus luteus*¹¹³.

Les cultures de macrophages murins RAW 2664.7 (A) et humains M-CSF (B/C) ont été infectées avec chacune des trois souches. Le nombre de bactéries intracellulaires viables a été évalué après de 1,5 heures et 24 heures (A-C). Ce nombre a aussi été évalué après 48 et 72 heures d'infection pour la souche de *S. lugdunensis* HKU09-01 (C). Les données représentent le nombre d'UFC/mL moyen de trois expériences indépendantes. Les valeurs de significativité statistique ont été obtenues avec le test de Student (A/B) et le test de Dunnett (C) avec *P < 0,05, **P < 0,01 et ***P < 0,001.

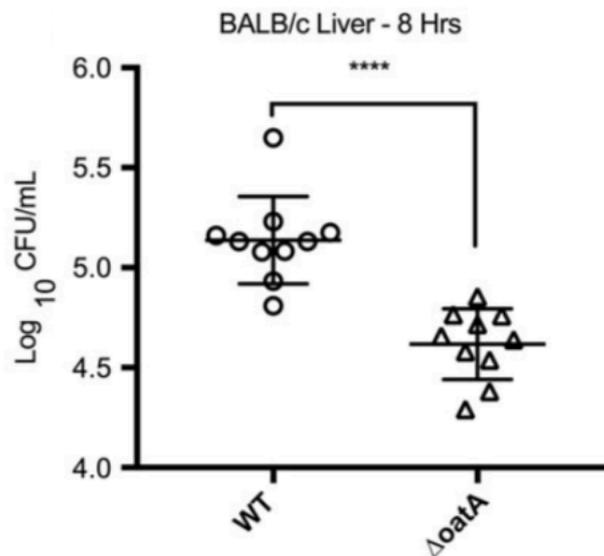


Figure 22 : Survie comparée de la souche de *S. lugdunensis* HKU09-01 sauvage ou délétée du gène *oatA* dans un modèle murin d'infection systémique¹¹³.

Le nombre de bactéries survivantes dans le foie des souris infectées a été évalué huit heures après l'inoculation de la souche HKU09-01 par voie intraveineuse. La valeur de significativité statistique ***P < 0,001 a été obtenue par un test de Student.

Les macrophages participent à la lyse des bactéries par leur capacité de phagocytose mais aussi par la production de composés oxygénés réactifs, l'acidification du phagosome, l'activation de protéases lysosomales et la lyse du peptidoglycane grâce au lysozyme¹¹¹. Il a été montré *in vitro* et *in vivo* que *S. aureus* est capable de persister et de se répliquer dans les phagolysosomes des macrophages¹¹².

Pour explorer les interactions entre les macrophages et *S. lugdunensis*, Flannagan *et al.*¹¹³ ont infecté des cultures de macrophages murins (RAW 264.7) et humains (M-CSF) avec la souche de *S. lugdunensis* HKU09-01. Après 1,5 heures et 24 heures d'infection, un traitement par la gentamicine a été réalisé pour éliminer les bactéries extracellulaires et ensuite évaluer le nombre de bactéries intracellulaires viables après phagocytose. Ainsi, il a été observé que la souche HKU09-01 était capable de persister jusqu'à 72 heures mais pas de se répliquer dans les macrophages (**Figure 21**). Pour conforter ces observations, des macrophages ont parallèlement été infectés par la souche de *S. aureus* USA300 et une souche de *Micrococcus luteus*, connues respectivement pour leur capacité à résister ou non à la lyse par les macrophages¹¹². A la différence de *S. lugdunensis*, la souche de *S. aureus* USA300 est capable de survivre dans les macrophages mais aussi de s'y répliquer puisque le nombre de bactéries observé était nettement supérieur après 24 heures d'incubation. *M. luteus* a, lui, été complètement éliminé par les macrophages après 24 heures.

Pour compléter ce travail, un mutant délété du gène *oat* a été construit. La capacité de la souche mutante, ne présentant plus un peptidoglycane O-acétylé, à survivre dans les macrophages murins a été évaluée *in vitro*. Cette analyse a montré qu'après 24 heures, il y avait une diminution significative de bactéries intracellulaires viables de la souche mutante par rapport à la souche parentale. Ce défaut de persistance intracellulaire de la souche délétée du gène *oat* signifie que l'O-acétylation du peptidoglycane jouerait un rôle clé dans la résistance à la lyse par les macrophages.

Le rôle du gène *oat* a été ensuite étudié dans un modèle *in vivo* d'infection systémique chez la souris (**Figure 22**). Des souris ont été infectées par voie intraveineuse avec la souche sauvage et la souche délétée du gène *oat* puis un

dénombrement des bactéries intra-hépatiques huit heures après l'inoculation a montré que la souche sauvage était significativement plus internalisée que la bactérie mutante, soulignant l'importance de la résistance au lysozyme dans la survie bactérienne¹¹³.

Ainsi, la capacité des macrophages à limiter la croissance bactérienne intracellulaire de *S. lugdunensis* contribue à la lutte du système immunitaire contre l'infection. Néanmoins, l'incapacité de ces cellules de l'immunité à éradiquer complètement les bactéries suggère que *S. lugdunensis* pourrait persister chez l'Homme, les macrophages jouant le rôle de réservoir, à partir duquel les bactéries pourraient essaimer. Ce réservoir pourrait également constituer une protection contre des facteurs extracellulaires du système immunitaire de l'hôte¹¹³.

1.3.6.2. La lugdulysine

Plus récemment, une activité protéolytique a été identifiée chez 50 des 81 souches isolées dans un contexte d'infection avéré à l'hôpital universitaire de Strasbourg¹¹. La protéine responsable de la protéolyse a été purifiée puis analysée par nano-chromatographie liquide couplée à un spectromètre de masse. L'analyse de la séquence peptidique a permis d'identifier une pré-protéine de 37 kDa présentant une homologie de séquence de 62,4% avec la métalloprotéase Shpl (aussi appelée hycolysine) de *Staphylococcus hyicus*. Ainsi, la protéine à activité protéolytique identifiée chez *S. lugdunensis* est une métalloprotéase dépendante du zinc, nommée lugdulysine¹¹.

L'hycolysine de *S. hyicus* est une métalloprotéase isolée d'une espèce bien décrite en médecine vétérinaire, responsable d'infections chez le porc. Le rôle des métalloprotéases dans les pathologies humaines a aussi été décrit. Par exemple, chez *S. aureus*, la métalloprotéase nommée auréolysine est impliquée dans le développement des IOA en participant au remodelage de la MEC. En effet, il a été montré dans un modèle d'ostéomyélite chez la souris que la souche délétée du gène codant la métalloprotéase induisait une destruction osseuse moindre par rapport à celle liée à la souche sauvage¹¹⁴. Il s'agira donc maintenant de démontrer dans un modèle animal son implication dans la virulence de *S. lugdunensis*.

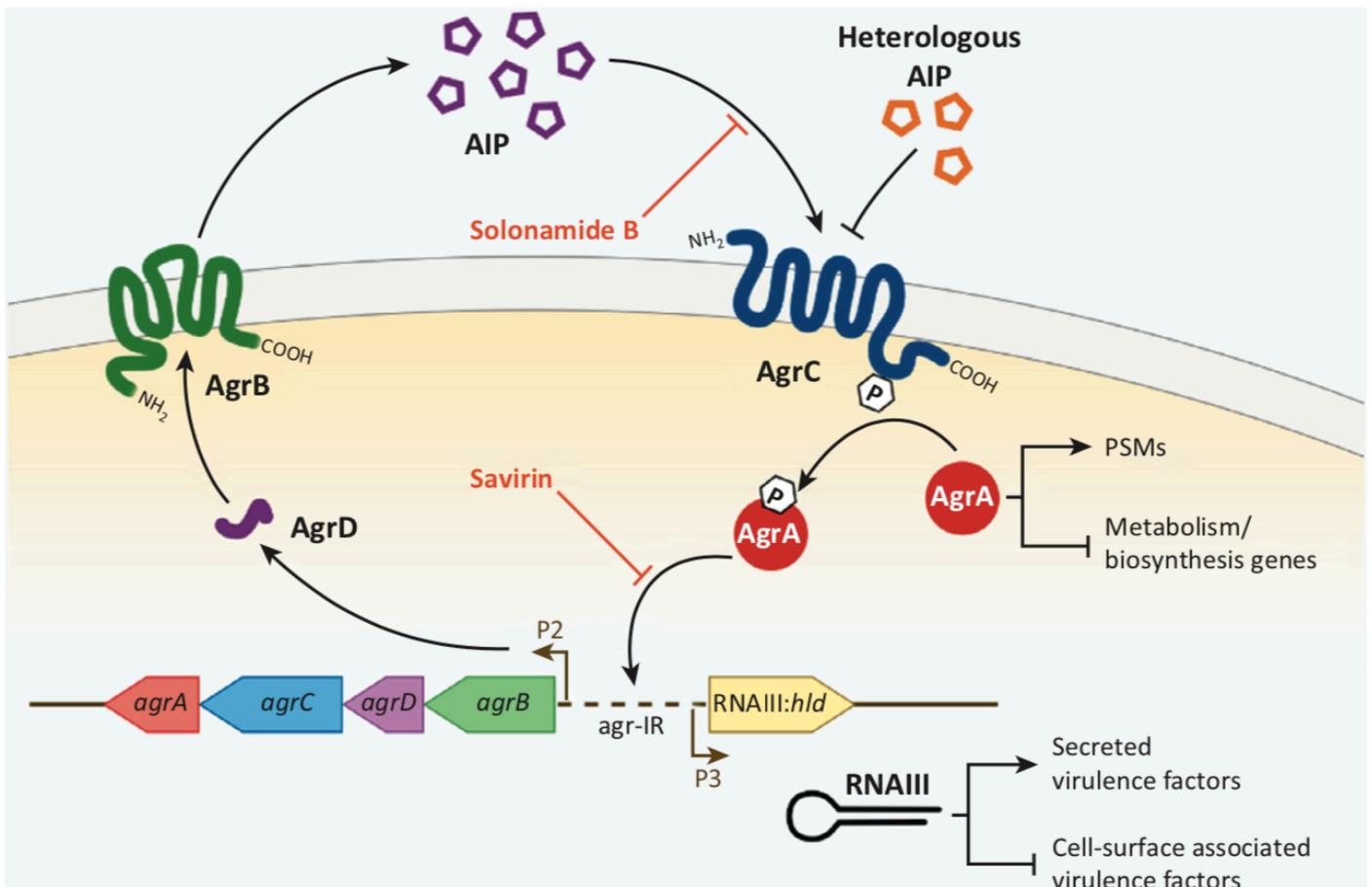


Figure 23 : Modèle d'activation de la transcription de l'ARNIII du locus agr chez *S. aureus*¹¹⁹.

Le locus *agr* code un système de *quorum sensing* auto-activable du promoteur P2 et de la molécule effectrice, l'ARNIII. Ce dernier régule l'expression de nombreux facteurs de virulence sécrétés comme des toxines et inhibe l'expression des protéines de surface. En outre, la protéine AgrA régule l'expression des *phenol soluble modulins* et de plusieurs gènes du métabolisme.

1.4. Le locus agr, système de régulation globale de la virulence

Chez *S. aureus*, l'expression des différents facteurs de virulence est contrôlée par des systèmes de régulation qui permettent de coordonner leur expression. Il s'agit le plus souvent de systèmes à deux composants qui contrôlent au niveau transcriptionnel l'expression des facteurs de virulence, en réponse directe à un *stimulus* de l'environnement. Il existe chez *S. aureus* plusieurs systèmes de régulation dont le plus étudié est le *locus agr*. Ce dernier est décrit comme un système de *quorum sensing* qui joue un rôle de régulateur global de la virulence^{115,116}.

Le *locus agr* a tout d'abord été caractérisé chez *S. aureus*, puis par analogie de séquence, il a été identifié chez *S. lugdunensis*¹¹⁷. Chez *S. aureus*, le *locus agr* comprend deux opérons sous le contrôle de deux promoteurs distincts nommés P2 et P3, donnant lieu à deux transcrits, l'Acide Ribonucléotidique (ARN) II et l'ARNIII. L'ARNII est composé de quatre cadres de lecture (*agrA*, *agrC*, *agrD*, *agrB*) alors que l'ARNIII contient le cadre de lecture de la β -hémolysine (*hld*)¹¹⁵ (**Figure 23**).

Lors de la phase exponentielle de croissance bactérienne, le gène *agrD* code un peptide auto-inducteur (PAI) qui est maturé et sécrété dans le milieu extérieur *via* la protéine transmembranaire AgrB. Cet PAI s'accumule alors dans le milieu extérieur jusqu'à atteindre une concentration seuil, fonction de la densité bactérienne. Le PAI va alors se fixer à une histidine kinase transmembranaire codée par le gène *agrC*. L'activation d'AgrC induit la phosphorylation de la protéine AgrA entraînant une modification de sa conformation. La protéine AgrA peut alors se fixer au promoteur P2, ce qui active la transcription du *locus agr* et entraîne une boucle d'auto-activation. D'autre part, la protéine AgrA se fixe au promoteur P3, ce qui conduit à la transcription du gène *hld* et à la synthèse de l'ARNIII^{115,116} (**Figure 23**).

L'ARNIII constitue l'effecteur final du *locus agr* qui régule positivement et négativement la transcription de nombreux facteurs de virulence de *S. aureus*. En effet, il régule positivement la transcription de nombreux gènes codant des facteurs de virulence sécrétés (exoprotéases, lipases, entérotoxines, leucocidines, hémolysines...) ^{118,119}. Mais en même temps, il réprime la transcription des gènes codant des protéines d'adhésion (protéine A, FnbpA, FnbpB...) ^{118,119}. Par ailleurs, la protéine AgrA réprime la transcription de plusieurs gènes impliqués dans le

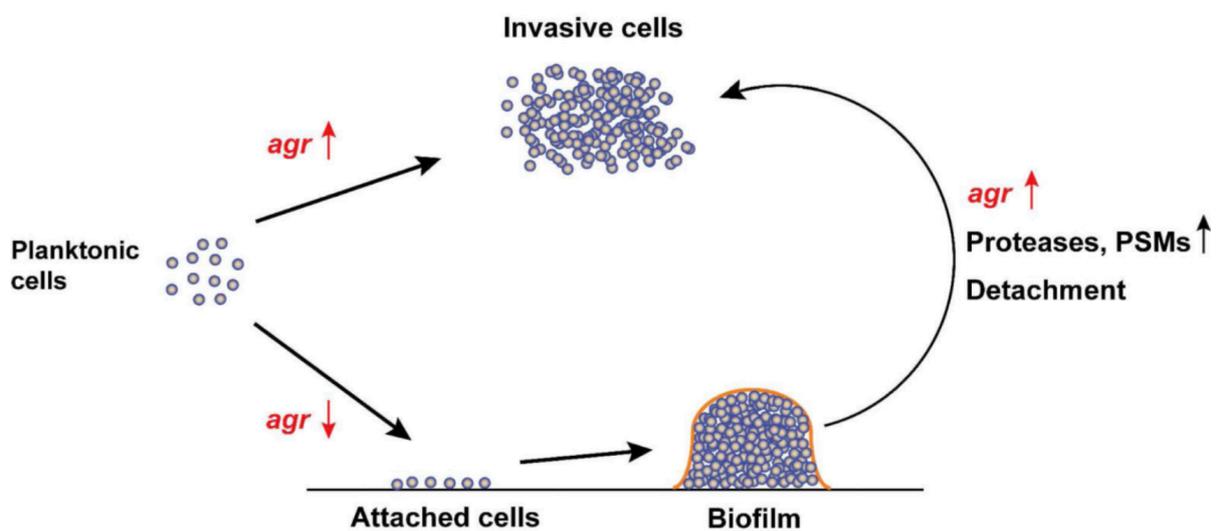


Figure 24 : Modèle de régulation par le système *agr* de la formation de biofilm chez *S. aureus*¹¹⁸.

Lorsque le système *agr* est inhibé ou désactivé, *S. aureus* adhère aux surfaces et forme un biofilm. Quand la densité bactérienne augmente, le locus *agr* s'active pour induire la transcription des gènes codant des protéases et les *phenol soluble modulins* pour revenir à l'état planctonique et coloniser de nouvelles surfaces.

métabolisme de la bactérie comme celui impliqué dans la synthèse des acides aminés ou celle des carbohydrates. *A contrario*, AgrA stimule la transcription des gènes codant des PSM α et β en se fixant sur leur promoteur¹¹⁸.

Chez *S. aureus*, le *locus agr* intervient également dans le développement du biofilm en régulant l'expression des gènes favorisant la colonisation et de ceux favorisant le détachement et la dissémination des bactéries¹¹⁸ (**Figure 24**).

L'expression du *locus agr* suit la courbe de croissance bactérienne. Au début de la phase de croissance, il est inhibé chez les cellules planctoniques ; la production de protéines de surface est alors induite, facilitant l'adhésion bactérienne et la colonisation. Les bactéries peuvent alors adhérer aux surfaces et s'agréger pour former un biofilm mature¹¹⁵. L'augmentation de l'*inoculum* bactérien se traduisant par une augmentation de la concentration de PAI, la transcription du *locus agr* va être activée au cours de la phase de croissance exponentielle. La transcription du système de régulation *agr* va d'une part réprimer la transcription des gènes codant les adhésines et d'autre part promouvoir la transcription des gènes codant des exoenzymes telles que les protéases et les PSM. Les bactéries peuvent alors se détacher du biofilm et essayer pour coloniser de nouvelles surfaces^{43,115}.

Le *locus agr* de *S. lugdunensis* partage une identité nucléotidique de 63% avec celui de *S. aureus* et une organisation similaire avec la présence des promoteurs P2 et P3 et des gènes *agrA*, *agrB*, *agrC* et *agrD*, mais il ne comporte pas le gène *hld*. Bien que l'homologie de séquence de l'ARNIII entre ces deux espèces soit faible, la fonction régulatrice de ce transcrit pourrait être la même. Contrairement à *S. aureus*, la β -hémolysine de *S. lugdunensis* n'est pas codée par le *locus agr*, mais son expression resterait sous son contrôle^{99,117}.

Le *locus agr* de *S. aureus* comporte une région hypervariable qui s'étend de la région 3' du gène *agrB* à la région 5' du gène *agrD*. Ce polymorphisme affecterait en particulier la séquence des gènes codant le PAI et son récepteur. L'étude du polymorphisme de ces gènes a permis d'identifier quatre allèles (*agr-type* I à IV) chez *S. aureus*¹²⁰ et deux allèles (*agr-type* I et II) chez *S. lugdunensis*¹¹⁸. Tseng *et al.*¹²¹ ont confirmé ces résultats par le séquençage du *locus agr* chez 11 souches responsables de bactériémie isolées à l'hôpital universitaire de Taiwan. Ils ont identifié quatre

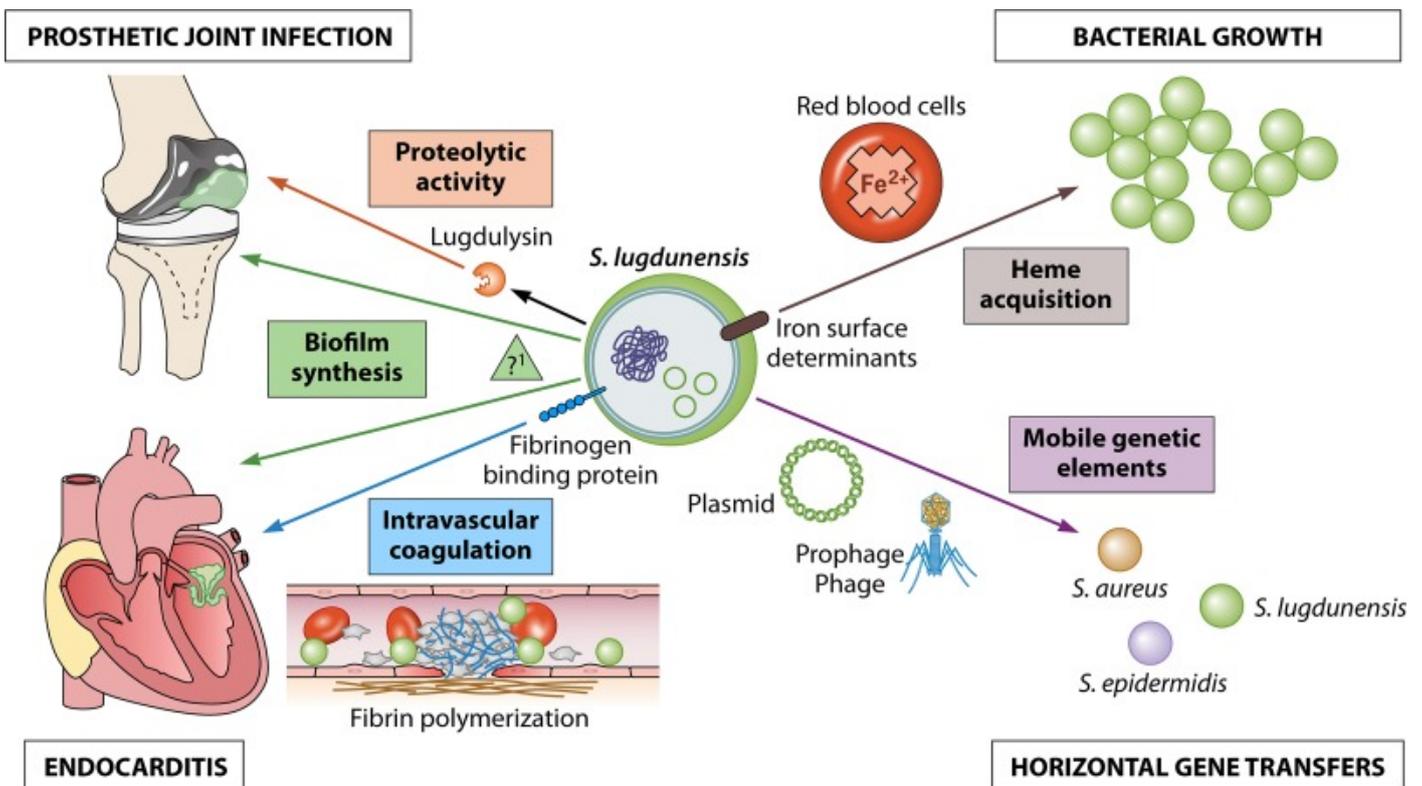


Figure 25 : Implication des principaux facteurs de virulence de *S. lugdunensis* dans les infections ostéoarticulaires et les endocardites infectieuses¹²².

souches de *agr-type* I et sept de *agr-type* II. Aucune différence n'a été observée concernant la production de biofilm PIA-indépendant entre les souches, mais une activité hémolytique et protéolytique plus importante a été observée pour les souches d'*Agr-type* II.

Ainsi, le *locus agr* pourrait être un important régulateur de la virulence de *S. lugdunensis*, mais aucune caractérisation fonctionnelle n'a été réalisée à ce jour chez cette espèce.

Depuis la revue de Frank *et al.*² soulignant combien *S. lugdunensis* était différent des autres SCN, les preuves cliniques, microbiologiques et génétiques s'amoncellent. Cette espèce est responsable de différents tableaux cliniques, dont certains sont parfois sévères, et très comparables à ceux de *S. aureus*. C'est l'espèce de SCN la plus virulente². Il est de ce fait considéré par certains comme "*a wolf in sheep's clothing*". Cependant, les mécanismes reliant les facteurs de virulence identifiés aux formes cliniques décrites restent encore à élucider¹²² (**Figure 25**).

2. Séquençage haut débit et virulence de *S. lugdunensis*

2.1. Le séquençage haut débit

Depuis plus d'une dizaine d'années, le séquençage à haut débit est devenu un outil très utilisé, ayant de nombreuses applications en microbiologie. L'automatisation du séquençage Sanger, qui constituait la première génération de séquençage a laissé place aujourd'hui à plusieurs méthodes de séquençage communément appelées "Next Generation Sequencing" (NGS) basées sur le séquençage parallèle et massif de milliers à millions de petits fragments d'ADN¹²³. Dans la littérature, l'arrivée du NGS s'apparente au début des méthodes d'amplification d'ADN par PCR où la principale limite de son utilisation est l'imagination¹²³. En effet, cette technologie a été utilisée pour le séquençage *de novo* pour notamment la caractérisation de nouveaux génomes bactériens, le génotypage par l'identification de *Single Nucleotide Polymorphisms* (SNP) pour la recherche de mutations conférant des résistances aux antibiotiques, des études de méta-génomique afin de caractériser les communautés multi-espèces d'un site ou d'un environnement donné^{123,124} ...

L'avancée majeure offerte par le NGS est la capacité à produire des volumes de données toujours plus grands à des coûts toujours plus bas au fil des années. Cela a permis de réaliser des études comparatives et évolutives à grande échelle. Pour ce faire, plusieurs méthodes ont été utilisées, proposées par de nombreux constructeurs. Selon les technologies utilisées, les méthodes de séquençage utilisent diverses stratégies pour la préparation des échantillons, le séquençage, la détection des nucléotides mais aussi pour l'assemblage et l'alignement des génomes. Nous décrivons ci-après deux technologies de NGS aux propriétés différentes mais complémentaires utilisées dans notre travail pour le séquençage de génomes complets : le séquençage de seconde génération *Illumina* et le séquençage de troisième génération *Pacific Biosciences*¹²⁴.

2.1.1. Le séquençage *Illumina*

Le séquençage *Illumina* est une technologie très utilisée aujourd'hui que ce soit pour la recherche scientifique ou pour la recherche clinique. Avant de procéder au séquençage, il est nécessaire de constituer une banque ADN appelée librairie d'ADN. Cette collection consiste à fragmenter l'ADN génomique bactérien par sonication (traitement par ultrason). Les millions de fragments d'ADN générés sont ensuite fixés à une surface de séquençage

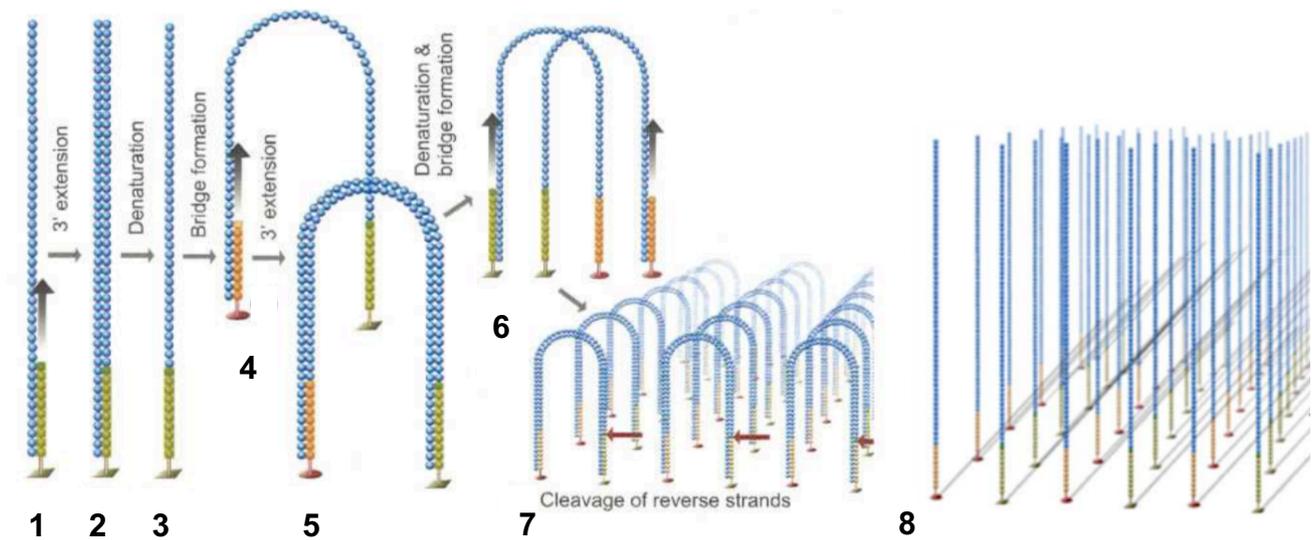


Figure 26 : Principe de la *Bridge PCR* pour la formation de *clusters* d'ADN lors du séquençage *Illumina*.

(1) Les fragments d'ADN fixés sur la *flowcell* servent (2) de matrice à l'ADN polymérase pour l'extension et la synthèse du brin complémentaire. (3) Le brin matrice est éliminé par dénaturation. (4) Les extrémités 3' des brins complémentaires s'hybrident avec des amorces fixées sur la *flowcell* (5) pour synthétiser de nouveaux brins d'ADN complémentaires afin de former des ponts d'ADN double brin. (6) Le pont d'ADN double brin est séparé par dénaturation. (7) Les brins anti-sens sont éliminés par clivage de l'adaptateur libérant l'extrémité 3' des brins sens. (8) Un nouveau cycle de PCR se répète afin de former des *clusters* d'ADN.

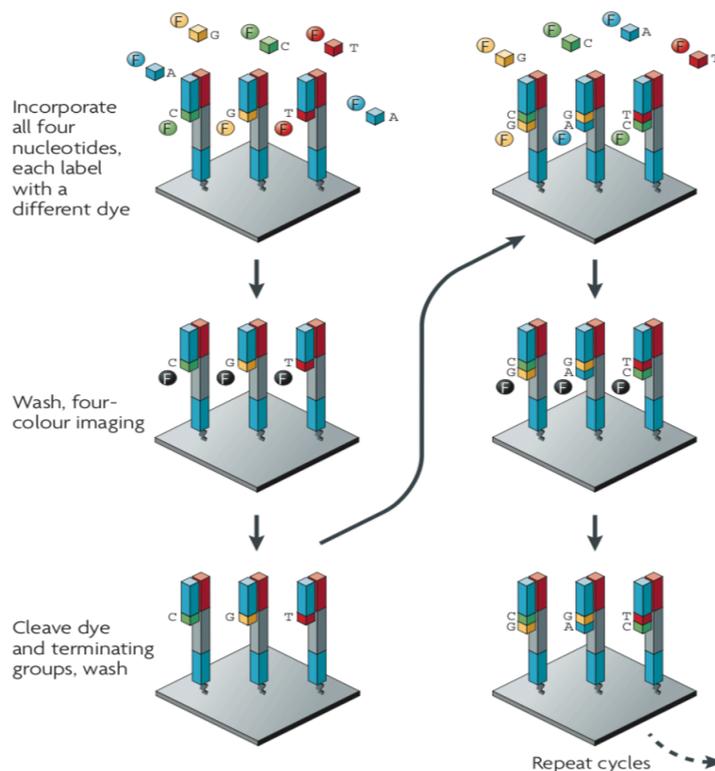


Figure 27 : Principe de l'incorporation des quatre terminateurs de séquence fluorescents réversibles utilisés lors du séquençage *Illumina*¹²³.

Les quatre types de nucléotides terminateurs marqués sont ajoutés dans la *flowcell*. Lors du séquençage, l'ADN polymérase incorpore un nucléotide complémentaire marqué émettant une fluorescence spécifique. Une fois la fluorescence émise et détectée, le terminateur fluorescent est clivé permettant l'ajout d'un nouveau nucléotide marqué.

ou *flowcell*, apparentée à une lame de verre de microscope divisée en huit microcanaux où se déroulent les réactions d'amplification et de séquençage. Pour fixer les fragments d'ADN sur cette lame, des sondes nucléotidiques appelées adaptateurs sont ajoutées à leurs extrémités. Ces adaptateurs vont d'une part leur permettre de se fixer de façon aléatoire sur la *flowcell* par hybridation avec les amorces complémentaires qui en recouvrent la surface et d'autre part permettre la réaction d'amplification en phase solide ou "*Bridge PCR*" (**Figure 26**). Les fragments d'ADN fixés servent de matrice à une ADN polymérase pour la synthèse des brins complémentaires qui seront alors fixés de façon covalente à la *flowcell*. Les brins matrice sont ensuite éliminés par dénaturation et l'extrémité 3' des brins complémentaires s'hybride avec des amorces fixées sur la *flowcell*. L'ADN polymérase synthétise à nouveau le brin complémentaire pour former un pont d'ADN double brin. Les deux brins sont ensuite séparés par dénaturation. A la fin de l'amplification, les brins anti-sens sont éliminés par clivage entre l'adaptateur et le brin d'ADN, libérant l'extrémité 3' du brin sens. Le cycle de "*Bridge PCR*" recommence pour former des *spots* d'ADN clonal appelés *clusters* permettant d'obtenir de nombreux brins d'ADN identiques¹²⁵.

Le séquençage *Illumina* repose sur une variante de la technique de séquençage Sanger appelée "*Cyclic Reversible Termination*", c'est-à-dire un séquençage par synthèse utilisant quatre terminateurs de séquences fluorescents réversibles¹²³ (**Figure 27**). Le séquençage s'effectue sur des millions de *clusters* de séquences ADN simultanément par l'incorporation de quatre terminateurs fluorescents qui vont empêcher que plus d'un nucléotide ne soit ajouté à une séquence lors du cycle de séquençage. Lorsque le nucléotide complémentaire est incorporé, une fluorescence spécifique de la base incorporée est émise. Le fluorophore est ensuite clivé permettant l'ajout d'un nouveau terminateur. La séquence des fragments va ainsi être déterminée au fur et à mesure de l'incorporation des nucléotides¹²³. Il s'agit d'un séquençage "*paired-end*" au cours duquel les fragments sont séquencés depuis leurs deux extrémités, ce qui facilite l'alignement sur un génome de référence.

Selon l'automate de séquençage *Illumina* utilisé, des millions de fragments d'ADN appelés *reads* de séquence d'une taille allant de 75 pb à 300 pb sont obtenus.

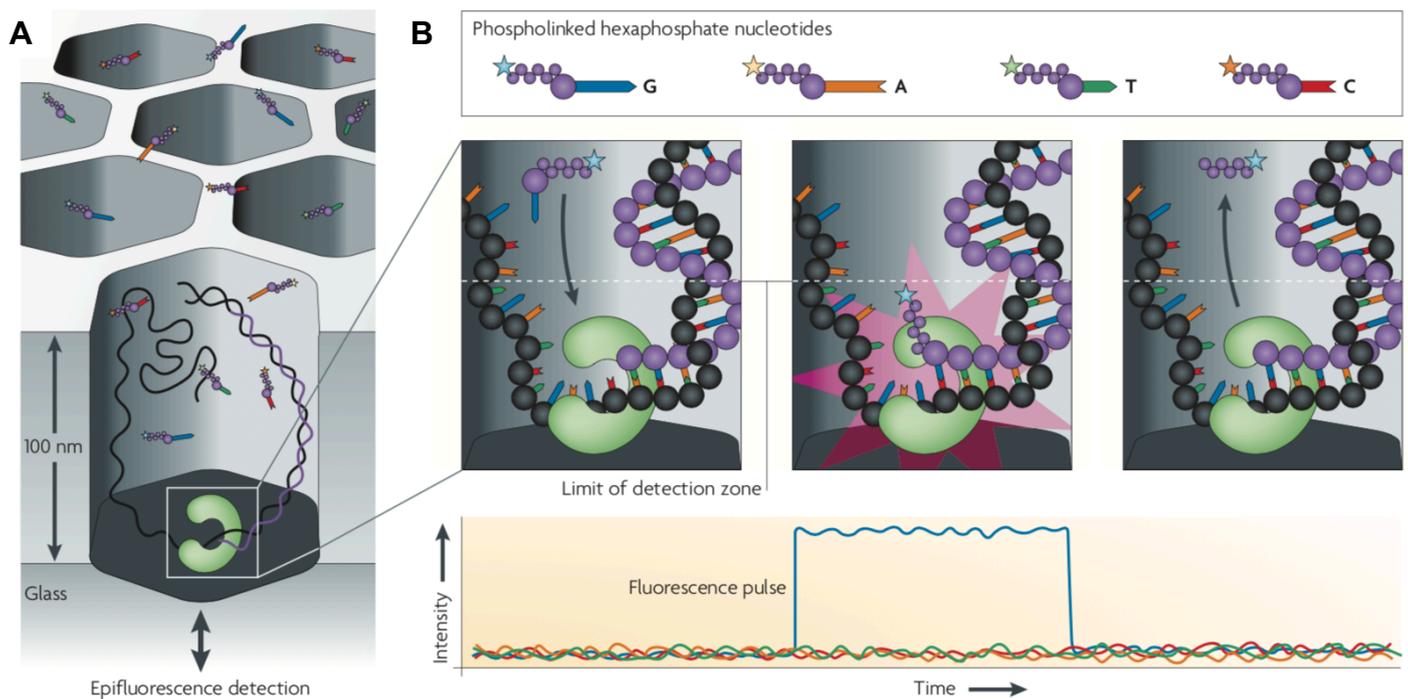


Figure 28 : Principe du séquençage *Pacific Biosciences Single Molecule Real Time*¹²³.

A, Détecteur *Zero Mode Waveguides* (ZMW) dans lequel est fixée une ADN polymérase couplée à une molécule d'ADN.

B, A chaque incorporation d'un nucléotide lié à un fluorophore, un pic de fluorescence est mesuré dont la durée et l'intensité est spécifique d'un nucléotide.

2.1.2. Le séquençage Pacific Biosciences

La différence majeure entre la deuxième et la troisième génération de séquenceurs repose sur leur capacité à séquencer directement les molécules d'ADN sans amplification préalable. Cette capacité offre l'avantage d'accéder aux modifications des bases comme la détection des méthylations. C'est le cas de la technologie de séquençage *Pacific Biosciences (Pacbio) Single Molecule Real Time (SMRT)*. Cette technologie séquence une molécule unique d'ADN en temps réel. La réaction de séquençage se déroule dans des cellules nanophotométriques contenant 7500 détecteurs *Zero Mode Waveguides (ZMW)* de 100 nm de diamètre¹²³ (**Figure 28**). Chacun de ces ZMW contient une ADN polymérase immobilisée qui incorpore des nucléotides liés à un fluorophore. L'intervalle de temps entre chaque signal fluorescent et la durée de ce signal sont propres à chaque nucléotide et permettent leur identification¹²⁶. L'avantage de cette technologie de séquençage est qu'elle produit des *reads* de séquence d'une taille moyenne de 10000 pb. Cependant, cette technologie a l'inconvénient de présenter un taux d'erreurs lors du séquençage de l'ordre de 10 à 15%. Ce taux d'erreur tend à être limité par les nouveaux automates de séquençage *Pacbio*¹²⁴.

2.1.3. Analyse des données brutes de séquençage haut débit

Une fois le séquençage effectué, un énorme volume de données a été produit, de l'ordre de plusieurs millions de *reads* qu'il est nécessaire de traiter par analyse bio-informatique. Par analogie, les *reads* correspondent à des pièces de *puzzle* qu'il faut assembler pour reconstituer le génome bactérien. Il existe deux types d'approches pour reconstituer un génome bactérien : l'alignement des *reads* de séquence sur un génome de référence ou l'assemblage *de novo*¹²⁷⁻¹²⁹.

L'alignement sur un génome de référence

L'alignement des *reads* sur un génome de référence s'apparente à assembler des pièces d'un *puzzle* en s'aidant d'un modèle. En effet, il s'agit d'aligner les *reads* sur un génome de référence par complémentarité. Cette approche est rapide, peu coûteuse en termes de moyens et permet une analyse d'un grand nombre de génomes. Toutefois,

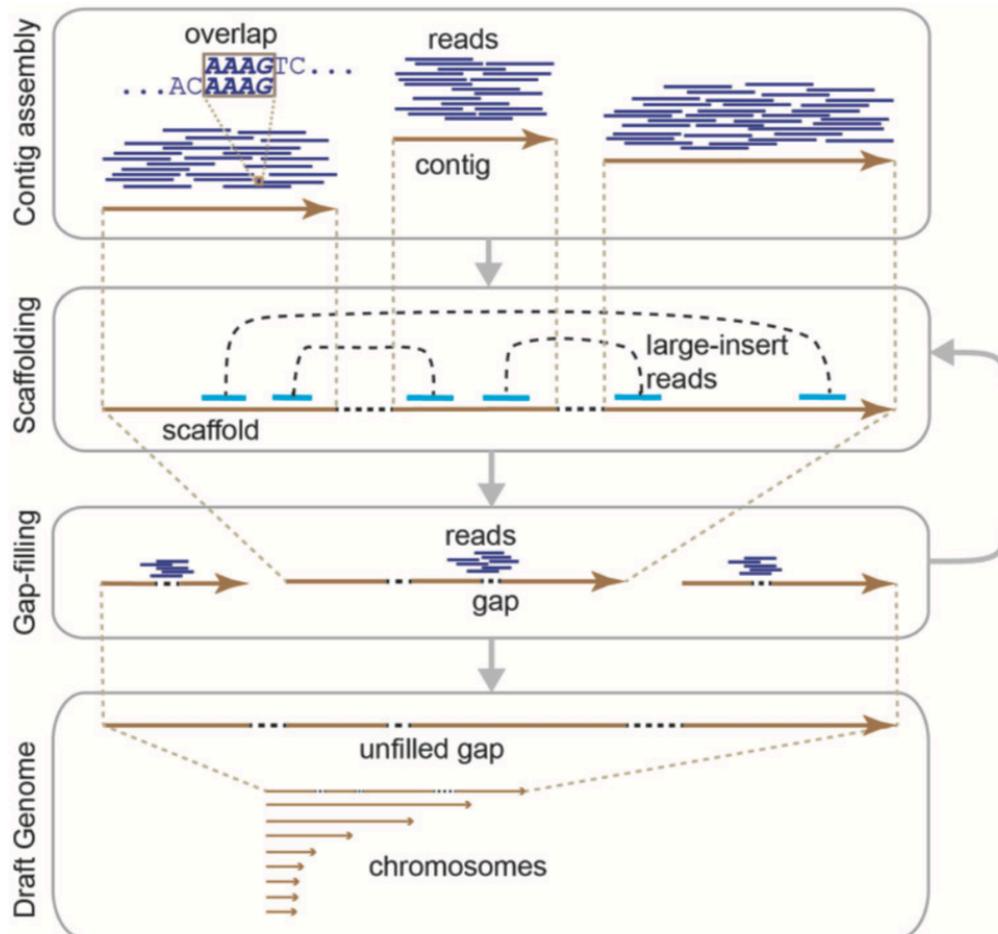


Figure 29 : Principe de l'assemblage *de novo* de reads courts¹²⁹.

Les grandes étapes de l'assemblage sont : l'assemblage des *reads* en *contigs*, la formation de *scaffolds* et le remplacement des *gaps*. Par assemblage des *reads*, des *contigs* sont obtenus puis orientés et ordonnés en *scaffolds*. Au sein de ces *scaffolds*, il peut y avoir des *gaps* qui peuvent être remplacés par un assemblage localisé des *reads* aux extrémités d'un *gap* pour le remplacer par la séquence manquante. Si des *gaps* persistent, le génome obtenu est qualifié de *draft genome*.

s'agissant d'aligner des séquences sur un génome de référence, cette méthode ne permet pas l'identification de nouvelles structures génomiques. En effet, si des séquences du génome étudié sont absentes du génome de référence, elles ne seront pas alignées et seront éliminées¹²⁷.

L'assemblage de novo

L'assemblage de génome *de novo* quant à lui ne s'appuie pas sur un génome de référence. Cette approche est plus longue et mobilise plus de moyens car plusieurs étapes sont nécessaires pour reconstituer le génome bactérien. Ces étapes consistent en : (i) l'assemblage des *reads* en *contigs*, (ii) la formation de *scaffolds* et (iii) le remplacement des *gaps* ("trous") (**Figure 29**)¹²⁹. Ainsi, cette approche consiste dans un premier temps en un appariement des *reads* par chevauchement et complémentarité qui permet d'obtenir de longues séquences génomiques consensus correspondant à des parties du génome bactérien, appelées *contigs*. En assemblant les *contigs* "bout à bout", il est alors possible d'obtenir des *scaffolds*, un ensemble de *contigs* orientés et ordonnés. Une fois les *scaffolds* formés, il peut rester des *gaps* entre les *contigs*. Néanmoins, ces *gaps* peuvent être remplacés en effectuant de nouveaux assemblages localisés, au niveau des extrémités d'un *gap*. Les génomes pour lesquels il reste des *gaps*, c'est-à-dire auxquels il manque toujours des séquences, sont appelés *draft genomes*¹²⁹.

Pour le traitement des données de séquençage haut débit deuxième génération, des *pipelines* ont été proposés listant pour chaque étape des logiciels afin de reconstituer le génome séquencé^{130,131}. Ainsi, une fois les données de séquençage obtenues, la première étape consiste à vérifier la qualité des *reads*. Le logiciel FASTQC est couramment utilisé pour évaluer la qualité globale du séquençage et les caractéristiques générales des *reads*. Cette étape consiste à éliminer les séquences de mauvaise qualité et les *reads* partiellement dérivés des adaptateurs utilisés pour la préparation de la librairie d'ADN. L'élimination des séquences de mauvaise qualité limite les problèmes d'assemblage et la formation de *gaps*. Ensuite, pour l'étape d'assemblage *de novo* de nombreux logiciels existent à ce jour comme *Spades*, *Velvet*, *Edena*... Chacun des logiciels d'assemblage a été développé pour traiter des caractéristiques particulières du génome et, par conséquent, ils diffèrent en termes de fonctionnement, de performance et de rendement. Ainsi, l'utilisateur peut comparer les

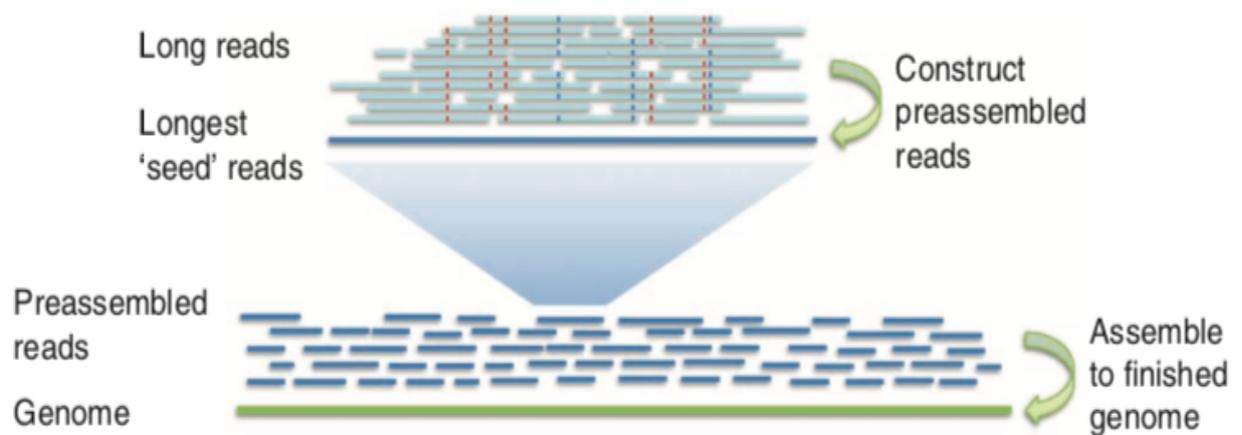


Figure 30 : Principe du *Hierarchical Genome-Assembly Process* de longs *reads* obtenus par séquençage *Pacific Biosciences*¹³².

Les *reads* les plus longs servent de référence pour l'alignement des *reads* plus courts. Cette étape permet de former des *reads* pré-assemblés qui seront assemblés entre eux pour former une séquence consensus correspondant au génome bactérien complet et circularisé.

assembleurs pour choisir le plus approprié ; celui qui permettra un assemblage de qualité avec un nombre de *contigs* limité.

Une fois l'étape d'assemblage effectuée, les séquences obtenues peuvent présenter des *gaps*. Une façon de combler ces *gaps* consiste à utiliser le logiciel "*Iterative Mapping and Assembly for Gap Elimination*" (IMAGE) ou à effectuer un nouvel assemblage localisé. Enfin, il est possible d'annoter le génome assemblé, c'est-à-dire d'utiliser un logiciel qui va identifier et déterminer le nombre de séquences codantes (CDS) et d'éléments fonctionnels au sein des séquences génomiques. De tels logiciels, comme "*Rapid Annotation Subsystem Technology*" (RAST) ou Prokka, vont prédire la fonction des CDS par recherche de similarité avec des CDS d'espèces bactériennes proches ou par l'identification de domaines protéiques spécifiques.

La démarche décrite ci-dessus est utilisée pour l'assemblage de *reads* courts, obtenus avec les séquençages de deuxième génération comme le séquençage *Illumina*. Pour l'assemblage des données de *reads* longs obtenus par séquençage de troisième génération comme le séquençage PacBio, il a été mis au point une méthode d'assemblage appelée *Hierarchical genome-assembly process* (HGAP)¹³² (**Figure 30**). Le principe du HGAP repose sur un pré-assemblage à partir des *reads* les plus longs utilisés en tant que séquences de référence pour l'alignement des *reads* plus courts. Les *reads* pré-assemblés sont alors assemblés entre eux et permettent de générer une séquence consensus. Ainsi, cette approche permet de reconstituer un génome complètement et de le circulariser avec un faible taux d'erreur¹³².

2.2. Analyses génomiques de *S. lugdunensis*

A côté de données génomiques générales, le séquençage de génomes complets de *S. lugdunensis* a permis l'identification de facteurs de virulence, d'éléments génétiques mobiles (EGM) et de facteurs de restriction.

2.2.1. Les données génomiques de *S. lugdunensis*

Le premier séquençage complet de génome de *S. lugdunensis* a été publié en 2010 par Tse *et al.* à partir de la souche HKU09-01 isolée d'un pus prélevé par écouvillonnage lors d'une ITPM¹³³. Le séquençage du génome de cette souche a été effectué avec la technologie de pyroséquençage 454 de Roche® (*GS FLX System*) et a permis d'obtenir 30 *contigs* d'une taille supérieure de 500 pb. Le génome a pu être circularisé après des *walk* PCR et séquençage Sanger des *gaps* entre ces *contigs*. Ce génome d'une taille de 2,6 millions de pb (Mb) présente un GC% de 33,9% et 2489 CDS¹³³.

Le second génome séquencé est celui de la souche N920143 isolée à partir d'un abcès de sein⁷³. Ce génome a été obtenu par pyroséquençage 454 de Roche® et par séquençage *Illumina* (GAIIx). Une combinaison des assemblages des données de séquençage a permis d'obtenir 69 *contigs* d'une taille supérieure à 500 pb. Le génome de la souche N920143 n'a pas été circularisé à la différence de celui de la souche HKU09-01. En effet, l'utilisation du logiciel IMAGE a permis de réduire le nombre de *contigs* à 9, et les auteurs ont renoncé à identifier les *gaps* restant devant l'absence de séquences d'intérêt dans ces régions par comparaison avec le génome de la souche HKU09-01. Le génome de la souche N920143 présente les mêmes caractéristiques que celui de la souche HKU09-01, à savoir une taille de 2,6 Mb, un GC% de 33,9% et 2447 CDS⁷³.

Heilbronner *et al.*⁷³ ont réalisé une comparaison de ces deux génomes et montré que ceux-ci sont très conservés puisqu'il existe une homologie de 95,4%. Toutefois, plusieurs différences entre les deux génomes ont été observées. Il a effectivement été identifié dans le génome de la souche HKU09-01 un transposon portant le gène *bla* associé à une résistance aux β -lactamines par la production d'une pénicillinase et un îlot génomique portant des gènes codant des protéines de résistance au cadmium. Par ailleurs, une portion de 32 kilobases (kb) du génome de la souche HKU09-01 comprenant le *locus isd* est dupliquée. Enfin, un des gènes codant les peptides hémolytiques Slush est absent du

Souche	Assemblage	Scaffolds	Taille (Mb)	GC%	Gènes	Protéines
HKU09-01	Complete Genome	1	2.65837	33.9	2567	2425
N920143	Complete Genome	1	2.59589	33.8	2498	2383
FDAARGOS_141	Complete Genome	1	2.57183	33.8	2465	2350
FDAARGOS_143	Complete Genome	1	2.61388	33.9	2527	2384
FDAARGOS_222	Complete Genome	1	2.54032	33.8	2430	2301
Klug93G-4	Complete Genome	1	2.58415	33.8	2501	2358
FDAARGOS_377	Complete Genome	1	2.57163	33.8	2479	2351
FDAARGOS_381	Complete Genome	2	2.5707	33.7781	2482	2344
NCTC 7990	Complete Genome	1	2.54770	33.9	2456	2316
NCTC 12217	Complete Genome	1	2.54033	33.8	2433	2297
VISLISI_21	Chromosome	1	2.54616	33.7	2437	2344
VISLISI_22	Chromosome	2	2.57089	33.7953	2459	2353
VISLISI_25	Chromosome	1	2.49164	33.8	2397	2293
VISLISI_27	Chromosome	2	2.60094	33.6953	2508	2391
VISLISI_33	Chromosome	2	2.667	33.6997	2584	2465
VISLISI_37	Chromosome	1	2.57952	33.7	2482	2391
C_33	Chromosome	3	2.50069	33.868	2405	2292
NCTC 12217	Scaffold	20	2.51951	33.7	2393	2287
M23590	Scaffold	8	2.52367	33.7	2414	2297
ACS-027-V-Sch2	Scaffold	4	2.61101	33.9	2499	2385
UCIM6116	Scaffold	47	2.55438	33.6	2440	2347
MJR7738	Scaffold	41	2.50183	33.7	2368	2293
VCU139	Contig	69	2.52414	33.7	2459	2244
VCU150	Contig	28	2.53605	33.7	2411	2322
VCU148	Contig	56	2.51451	33.7	2426	2254

Tableau III : Liste des 25 génomes de *S. lugdunensis* disponibles dans la base de données *National Center for Biotechnology information* (le 03/10/18).

génomique de la souche HKU09-01. Aucun plasmide n'a été identifié, mais une région phagique, appelée Φ SL1, a été décrite dans le génome de la souche N920143.

Non seulement les génomes de *S. lugdunensis* sont très conservés, mais une comparaison génomique avec d'autres staphylocoques a montré qu'il existe aussi une forte homologie génomique au sein du genre *Staphylococcus*. Le génome de la souche N920143 partage aussi 78,3% d'identité avec celui de la souche de *S. haemolyticus* JCSC1435, 77,8% avec celui de la souche de *S. aureus* SARM252, 74,7% d'identité avec celui de la souche de *S. epidermidis* RP62a⁷³.

Récemment, une équipe française a réalisé un séquençage complet du génome de sept souches de *S. lugdunensis* collectées lors de l'étude VISLISI : 6 souches responsables d'infections diverses (endocardite, bactériémie, infection d'une prothèse de genou, abcès de sein et abcès de foie) et une souche isolée d'un contexte de colonisation⁹⁶. Le génome de ces souches a été séquençé par technique *Illumina* (HiSeq 2500) avec des *reads* de 125 pb. L'assemblage des données de séquençage a permis d'obtenir selon les souches entre 19 et 69 contigs d'une taille supérieure à 500 pb. Les génomes ont été circularisés à l'aide du logiciel IMAGE qui effectue un assemblage des *reads* à partir des extrémités des *contigs* afin d'identifier les séquences manquantes. Ces génomes ont une taille comprise entre 2,5 et 2,6 Mb selon les souches avec un GC% de 33,6 à 33,8%. Ils sont eux aussi très conservés puisque des pourcentages d'homologie compris entre 94 et 97% ont été obtenus⁹⁶.

A ce jour, sur le site *National Center for Biotechnology information* (NCBI), il existe seulement 25 séquences génomiques de *S. lugdunensis*, dont 17 correspondent à des génomes complets et circularisés (**Tableau III**). Les huit autres séquences ne sont pas circularisées et leur assemblage n'a permis de former que des *scaffolds* ou de simples *contigs*. Seuls neuf génomes de *S. lugdunensis* ont fait l'objet d'une publication à ce jour^{73,96,133}. Comparativement à *S. aureus* et à *S. epidermidis*, ce nombre de données génomiques est faible puisque pour ces espèces sont disponibles sur le site de NCBI 10011 et 482 séquences génomiques respectivement.

	<i>S. lugdunensis</i> N920143	<i>S. lugdunensis</i> HKU09-01	<i>S. aureus</i> MRSA252	<i>S. epidermidis</i> RP62a	<i>S. haemolyticus</i>	<i>S. saprophyticus</i>	<i>S. carnosus</i>
NRPS 1	+	+	+	+	–	–	–
NRPS 2	+	+	–	–	–	–	–
NRPS 3	+	+	–	–	–	–	–
<i>isd</i> locus	+	Duplicated	+	–	–	–	–
<i>sst</i> locus	Duplicated	Duplicated	+	+	+	+	+
<i>cap</i> locus (PS capsule)	+	+	+	–	+	+	+
<i>cap</i> locus (PGA capsule)	+	+	–	+	+	+	–
<i>esx</i> locus	+ one gene FS	+ one gene FS	+	–	–	–	–
Streptolysin S-like toxin	+ one gene FS	+ one gene FS	– (RF122)	–	–	–	–
Lantibiotic resistance locus	+	+	–	–	–	–	–
<i>agr</i> locus	+	+	+	+	+	+	+
<i>ica</i> locus	+	+	+	+	–	+	–
<i>mprF/dlt</i>	+	+	+	+	+	+	+
CRISPR region	+	+	–	+	–	–	–
β -Haemolysin	+	+	+	+	–	–	–
Putative haemolysin III	+	+	+	+	+	+	+

Tableau IV : Distribution des gènes codant des facteurs de virulence putatifs au sein de plusieurs espèces de staphylocoques⁷³.

Un signe “+” correspond à la présence du gène. L’abréviation “FS” signifie un décalage du cadre de lecture dans la séquence du gène.

2.2.2. Identification de facteurs de virulence

L'analyse du génome de la souche N920143 a mis en évidence des gènes codant des facteurs de virulence ou des protéines putativement associées à la virulence de *S. lugdunensis*⁷³. C'est par l'analyse de ce génome qu'ont notamment été identifiés le *locus isd* impliqué dans l'acquisition du fer et les gènes codant les protéines de surfaces SIs.

L'analyse génomique de la souche de *S. lugdunensis* N920143 a aussi permis l'identification de plusieurs *loci* ayant un rôle putatif dans la colonisation et la virulence de cette espèce dont certains présents chez *S. aureus*, mais absents chez d'autres SCN⁷³ (**Tableau IV**).

2.2.3. Identification d'éléments génétiques mobiles

Les éléments génétiques mobiles (EGM) jouent un rôle central dans l'adaptation bactérienne et constituent un moyen de transférer du matériel génétique entre les espèces bactériennes. Les EGM tels que les plasmides, les bactériophages, les îlots de pathogénicité ou les cassettes chromosomiques peuvent être partagés par transferts horizontaux et être à l'origine de l'émergence de souches pathogènes ou résistantes aux antibiotiques. En effet, les EGM peuvent porter des gènes qui vont coder des facteurs de virulence ou des gènes de résistance, mais aussi des gènes qui vont assurer leur propre transfert, réplication et intégration dans le chromosome bactérien¹³⁴.

Chez *S. aureus*, de nombreux facteurs de virulence ont été acquis par des transferts horizontaux d'EGM comme de nombreuses toxines (entérotoxines, leucocydine, *toxic shock syndrome* toxin...) codées par des gènes portés par des phages ou des îlots de pathogénicité¹³⁴. Ces EGM peuvent représenter jusqu'à 25% du génome de cette espèce et peuvent être transférés à d'autres espèces de staphylocoques¹³⁵. Néanmoins au sein des SCN, peu de facteurs de virulence portés par des EGM ont été identifiés ; seuls quelques phages porteurs de gènes de résistance ont été décrits chez des isolats cliniques, principalement chez *S. epidermidis* et *S. hominis*, présentant des relations phylogénétiques étroites avec les phages de *S. aureus*¹³⁶.

Chez *S. lugdunensis*, l'analyse du génome de la souche N920143 a permis d'identifier une région phagique Φ SL1⁷³. Le séquençage récent de sept génomes de *S. lugdunensis* a aussi permis d'identifier quatre régions phagiques nommées Φ SL2 à Φ SL5, d'une taille de 44,4 à 53,5 kb⁹⁶. La comparaison de ces phages par l'outil Blastn a montré que 30% des CDS de ces phages sont partagées avec des phages appartenant à *S. hominis* (Φ Stb12) et *S. epidermidis* (Φ PH15). Malgré la présence de nombreuses CDS au sein de ces phages, aucun gène codant des facteurs de virulence n'a été identifié⁹⁶.

Par ailleurs, l'analyse des sept génomes séquencés de *S. lugdunensis* a permis l'assemblage de *contigs* supplémentaires extra-chromosomiques correspondant à des séquences plasmidiques. Ainsi, cinq plasmides, pVISLISI1 à pVISLISI5, ont été identifiés, d'une longueur de 3,1 à 12,6 kb et présentant de trois à 13 CDS dont certaines impliquées dans la répllication plasmidique⁹⁶. Aucun gène codant des facteurs de virulence n'a été identifié, mais le plasmide pVISLISI1 correspond à un plasmide déjà décrit dans la littérature, pLUG10 porteur d'un gène de résistance au cadmium *cadD*²⁸. Le plasmide pVISLISI3 présente quant à lui une homologie parfaite avec le plasmide pRIVM de *S. aureus*. En revanche, seules 6,3 kb du plasmide pVISLISI5 présentent une homologie parfaite avec le plasmide VRSAP de *S. aureus*⁹⁶.

Ainsi, le séquençage haut débit de plusieurs génomes de *S. lugdunensis* a permis d'identifier plusieurs EGM, mais aucun îlot de pathogénicité. Ces EGM ne portent pas de gènes codant de facteurs de virulence ni de gènes de résistance. Ils suggèrent un transfert horizontal de gènes de *S. aureus* à *S. lugdunensis*.

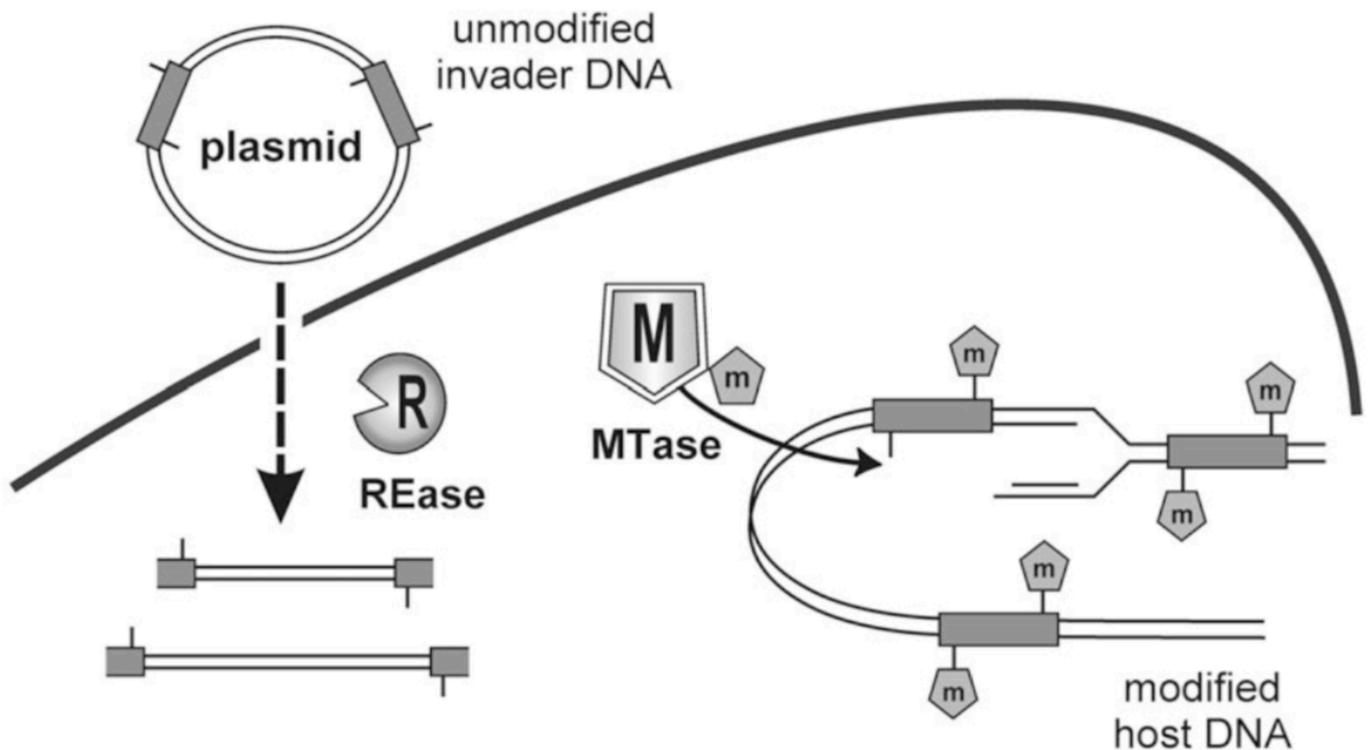


Figure 31 : Mécanisme général des Systèmes de Restriction-Modification (SRM)¹³⁸.

Les SRM reconnaissent si les adénines ou cytosines des séquences cibles sont méthylées ou non. L'ADN méthylé est reconnu par le SRM comme appartenant au génome de l'hôte grâce aux méthyltransférases (MTase) qui maintiennent le niveau de méthylation au niveau de séquences spécifiques. Les ADN exogènes, ici un plasmide, sont reconnus comme étrangers et sont clivés par les endonucléases de restriction (REase).

2.2.4. Identification de facteurs de restriction

2.2.4.1. Système de restriction-modification

Il a été identifié chez les staphylocoques des systèmes pour contrôler la stabilité du génome bactérien. Chez *S. aureus* et *S. epidermidis* ont été identifiés quatre types de Systèmes de Restriction-Modification (SRM) qui les protègent des transferts horizontaux et notamment vis-à-vis de l'intégration d'ADN phagique¹³⁷.

Les SRM sont constitués de protéines aux activités opposées, une endonucléase de restriction, tandis que l'activité de méthylation est attribuée à une méthyltransférase, reconnaissant une séquence spécifique (**Figure 31**)¹³⁸. Les endonucléases de restriction reconnaissent et clivent les ADN exogènes du fait de l'absence de méthylation des séquences cibles spécifiques. A l'inverse, les méthyltransférases jouent un rôle protecteur de l'ADN génomique en méthylant des cytosines ou des adénines de séquences spécifiques. Les SRM sont largement répandus parmi les procaryotes et plusieurs espèces ou souches possèdent leur propre combinaison de type de SRM. Chacun des types de SRM est composé d'enzymes spécifiques reconnaissant des séquences cibles différentes et des sites de clivage variables¹³⁸.

Chez *S. aureus*, le SRM de type I, Saul, est aussi composé des gènes *hsdR* (impliqué dans la restriction de l'ADN), *hsdM* (impliqué dans la méthylation de l'ADN) et *hsdS* (impliqué dans la reconnaissance des séquences cibles) qui codent deux complexes protéiques M₂S et R₂M₂S. Le complexe M₂S catalyse le transfert d'un groupement méthyle d'une S-adénylméthionine à une adénine d'une séquence cible. Cette méthylation protège cette séquence ADN de la restriction¹³⁹. Le complexe R₂M₂S, quant à lui, est une endonucléase de restriction qui reconnaît les adénines non méthylées d'une séquence cible et les clive, empêchant l'intégration d'ADN exogène.

Au sein du génome complet de la souche N920143 a été identifié un SRM de type I, appelé Slul, composé des gènes *hsdR*, *hsdM* et *hsdS*. Ce SRM rend difficile la transformation des souches de *S. lugdunensis* par de l'ADN exogène^{137,140}.

De façon intéressante, Heilbronner *et al.*¹⁴⁰ ont montré qu'une souche de *S. lugdunensis* était plus aisément transformable par un plasmide cloné chez des souches d'*Escherichia coli* exprimant les gènes *hsdM* et *hsdS* de *S. lugdunensis*.

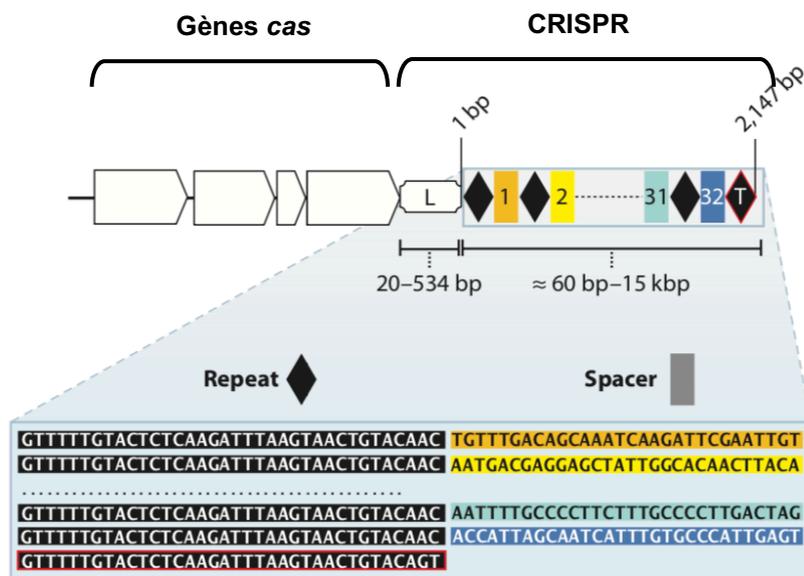


Figure 32 : Organisation générale du *locus Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR) associé aux protéines Cas (CRISPR/Cas)¹⁴³.

Les cadres blancs en amont du *locus* CRISPR correspondent aux gènes *cas*. La séquence en aval des gènes *cas* correspond à la séquence *leader* (L) dont la taille varie selon les *loci* entre 20 et 534 pb. Les losanges noirs représentent les *repeats* qui sont des séquences conservées au sein d'un même *locus*. Les rectangles de couleurs représentent les *spacers* de 30 pb environ.

L'analyse comparative des 15 génomes complets de *S. lugdunensis* réalisée par Argemi *et al.*¹⁴¹ a montré que le SRM de type I n'était présent que chez six souches. Chez ces dernières, les trois gènes codant le SRM étaient très conservés avec une homologie de 100%. Le gène codant la méthyltransférase présentait selon les souches, une homologie nucléotidique de 90 à 98% avec celui de la méthyltransférase Sau18 de la souche C18 de *S. aureus*. Par ailleurs, il a été identifié pour les sept autres souches, un SRM de type II dont le gène codant la méthyltransférase présentait une homologie variable selon les souches avec les méthyltransférases ShaJ de *S. haemolyticus* (82%), de *S. epidermidis* (66%) et *Enterococcus faecium* (60%)¹⁴¹.

Chez *S. aureus*, les SRM, et notamment celui de type I, sont un des mécanismes majeurs pour prévenir des échanges d'ADN étranger. Chez cette espèce, le transfert de plasmide et d'ADN phagique est strictement contrôlé par le SRM de type I entre les souches de différentes lignées. Ainsi, la présence de différents types de SRM chez *S. lugdunensis* pourrait constituer une première barrière contre les transferts horizontaux.

2.2.4.2. Système *Clustered Regularly Interspaced Short Palindromic Repeat*

Heilbronner *et al.*⁷³ ont identifié dans le génome de la souche N920143 un *locus Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR) associé aux protéines Cas (CRISPR/Cas) de 12,5 kb.

Le système de CRISPR/Cas est connu chez *S. aureus* et *S. epidermidis* pour protéger la cellule bactérienne face à l'invasion d'ADN exogène, comme celui des phages ou des plasmides¹⁴². Le *locus* CRISPR/Cas est composé des gènes *cas* suivis d'une région *leader* riche en adénine et thymine située en amont du *locus* CRISPR (**Figure 32**)¹⁴³. Ce dernier comprend des séquences répétées appelées *repeats* de 27 à 50 pb, très conservées au sein d'un même CRISPR. Entre les *repeats* sont présents des *protospacers*, séquences d'environ 30 pb provenant d'EGM. Il existe plusieurs classes, types et sous-types de système CRISPR/Cas en fonction de l'organisation génétique du *locus* et de la présence des gènes *cas* pour le clivage de l'ADN exogène^{143,144}.

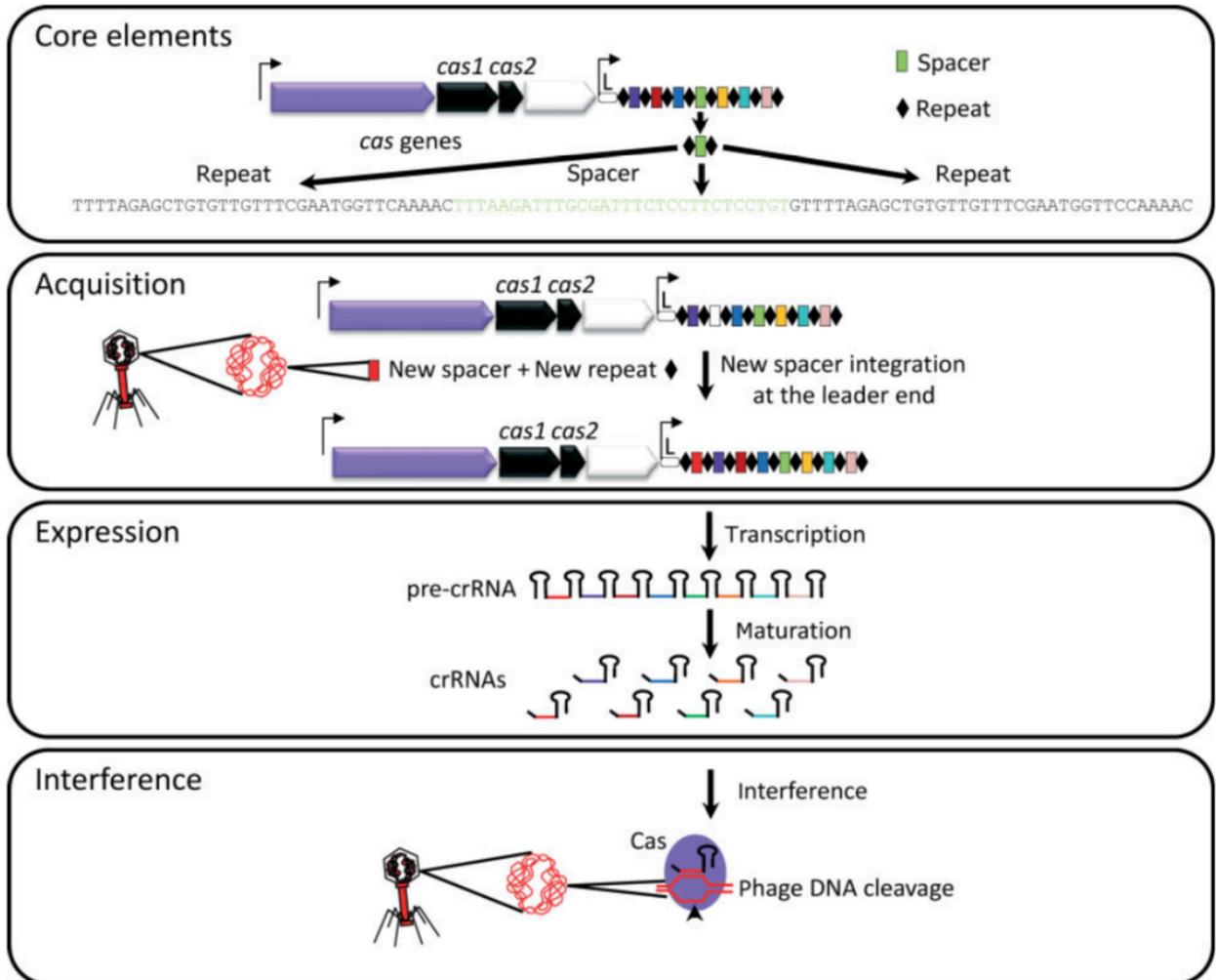


Figure 33 : Description des étapes d'adaptation, d'expression et d'interférence du *Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR) associé aux protéines Cas (CRISPR/Cas)¹⁴⁵.

Le fonctionnement du système CRISPR se décompose en trois étapes^{143,145} (**Figure 33**) :

- La phase d'adaptation qui consiste en la détection et l'intégration d'un fragment d'acide nucléique étranger d'environ 30 pb. La sélection de cette séquence est rendue possible par une protéine Cas, capable de reconnaître un motif de quelques nucléotides appelé *protospacer* adjacent motif (PAM). Le fragment d'ADN étranger est ensuite reconnu par un dimère de protéines Cas1 et Cas2, conservées entre tous les types CRISPR/Cas. Une fois liée au *protospacer*, ces protéines vont permettre l'intégration de cette séquence dans le locus CRISPR/Cas entre des repeats.
- Lors d'une seconde introduction d'un ADN exogène, la cellule bactérienne se protège par l'expression des ARN du CRISPR (crRNA). Les crRNA sont de courtes séquences d'ARN qui contiennent un *protospacer* unique permettant de s'hybrider de façon spécifique et complémentaire avec l'ADN étranger. Cette synthèse des crRNA varie en fonction des différentes classes et types de CRISPR/Cas.
- La phase d'interférence avec la séquence cible consiste en la formation d'un duplex entre le crRNA et l'ADN exogène permettant le ciblage par le complexe Cas comprenant des nucléases pour cliver de façon spécifiques les séquences homologues.

Les systèmes CRISPR/Cas ont été classés en trois types principaux et 12 sous-types en fonction de leur contenu génétique et de leur différences structurelles et fonctionnelles¹⁴⁶. La caractéristique fondamentale de cette classification repose sur le contenu du *locus* en gènes *cas* qui sont très divers génétiquement et en termes de fonction. Les gènes *cas1* et *cas2* sont universels à tous les types. A l'inverse, les gènes *cas6* et *cas10* ne sont retrouvés que dans les types I et III alors que le gène *cas9* ne sera présents que dans le type II¹⁴⁵.

L'analyse comparative des 15 génomes complets de *S. lugdunensis* a montré que le système CRISPR/Cas était présent chez six souches. Parmi ces six souches, cinq présentaient un système CRIPR/Cas de type IIIA et une souche présentait un système de type IIC. Toutefois, chez cette dernière, le gène *cas9* présentait plusieurs codons stop qui rendaient probablement le système CRISPR inopérant. Au sein du *locus* CRISPR des cinq souches, 19 *protospacers* distincts ont été identifiés. Pour sept d'entre eux, une homologie a été identifiée par Blast avec des séquences provenant d'EGM de différentes espèces

bactériennes tels que des plasmides de *Lactobacillus salivarius*, de *Clostridium botulinum* et des phages de *Streptococcus sp*, de *Campylobacter sp* et de *Staphylococcus sp*¹⁴¹.

Ainsi, le système CRISPR/Cas consiste en une autre barrière potentielle contre les transferts d'ADN horizontaux chez certaines souches de *S. lugdunensis*.

2.2.4.3. Système de toxine/antitoxine

L'analyse génomique comparative de 15 génomes de *S. lugdunensis* a permis d'identifier un système toxine/antitoxine (T/AT) MazEF très conservé chez toutes les souches de *S. lugdunensis*¹⁴¹. Ce système a été identifié dans le génome de plusieurs espèces à Gram positif, mais parmi les SCN, il n'a été mis en évidence que chez *S. equorum*¹⁴⁷. Il a été montré que la protéine MazF, la toxine, est une RNase qui dégrade les transcrits présentant une séquence UACAU dont l'activité peut être inhibée par l'antitoxine MazE¹⁴⁸. Dans la littérature, les systèmes T/AT sont décrits comme impliqués dans de multiples processus comme la réponse au stress, la mort cellulaire, la persistance, la formation de biofilm mais aussi dans la défense contre les phages^{149,150}. Ainsi, ce système T/AT pourrait être chez *S. lugdunensis* un système de protection putatif contre les EGM. Il est néanmoins nécessaire de vérifier la fonction de ces protéines et si elles sont opérantes par la construction de mutants de délétion.

En plus de tous ces facteurs de restriction, les staphylocoques présentent une paroi épaisse qui constitue une barrière physique à l'entrée d'ADN exogène, tel que des plasmides. C'est pourquoi, il est parfois nécessaire de recourir à la méthode des protoplastes pour transformer *S. lugdunensis* et construire des mutants⁷².

Ainsi, cette paroi épaisse et ces facteurs de restriction rendent en pratique difficile et même parfois impossible la transformation de certaines souches cliniques de *S. lugdunensis*¹⁴⁰.

2.2.5. Analyse du *pan-genome* de *S. lugdunensis*

Au sein d'une espèce bactérienne donnée, le *pan-genome* correspond à l'ensemble des gènes décrits chez l'espèce. Le *pan-genome* comprend plusieurs sous-ensembles¹⁵¹ :

- **le *core-genome*** : ensemble des gènes communs à toutes les souches d'une même espèce. Cet ensemble comprend le plus souvent les gènes qui participent au maintien du métabolisme et des processus biologiques essentiels comme la réplication, la transcription, l'homéostasie ...
- **le *génomme accessoire*** : ensemble des gènes présents chez quelques souches. Ce sous-ensemble est le plus souvent constitué de gènes acquis par des EGM intégrés au chromosome de la bactérie. Le *génomme accessoire* comprend également des gènes codant des fonctions spécifiques comme des facteurs de virulence ou de résistance à un antibiotique.
- **les gènes spécifiques** d'une souche ou *singletons*

L'étude du *pan-genome* permet de connaître le répertoire génique complet d'une espèce bactérienne. Pour ce faire, il est nécessaire de déterminer combien de nouveaux gènes sont identifiés pour chaque nouveau génome séquencé. Il s'agit du concept de *pan-genome* ouvert ou fermé.

La loi de Heap permet de préciser si le *pan-genome* d'une espèce est ouvert ou fermé. Par des extrapolations mathématiques, cette loi permet en génomique de connaître le nombre de gènes du *pan-genome* en fonction du nombre de génomes analysés¹⁵¹. Selon la loi de Heap, un *pan-genome* est considéré comme fermé si l'ajout de nouveaux génomes n'augmente pas significativement le nombre de gènes. A l'inverse, il est ouvert si le nombre de gènes augmente significativement à la suite du séquençage de nouveaux génomes. Par exemple, il a été montré que le *pan-genome* de *S. agalactiae* est ouvert puisque le nombre de nouveaux gènes augmente même après le séquençage de centaines de génomes¹⁵². Cela signifie également que les espèces présentant un *pan-genome* ouvert sont plus facilement transformables et sont capables d'intégrer de l'ADN exogène¹⁵². A l'inverse, les espèces qui présentent une niche écologique restreinte et/ou qui ont une capacité réduite à intégrer de l'ADN exogène par échange horizontal ou par recombinaison ont un *pan-genome* fermé.

Argemi *et al.*¹⁴¹ ont étudié le *pan-genome* et le *core-genome* de *S. lugdunensis*, *S. aureus* et *S. epidermidis*. Pour ce faire, ils ont analysé les génomes de 15 souches de *S. lugdunensis*, 13 de *S. epidermidis* et 15 de *S. aureus*. Ils ont montré que *S. aureus* et *S. epidermidis* ont un *pan-genome* ouvert puisque le répertoire génique augmente pour les deux espèces à chaque nouveau génome analysé. Le nombre total atteint était de 3828 gènes après l'analyse de 15 de *S. aureus* génomes et de 3864 gènes après l'analyse de 13 génomes de *S. epidermidis*. *A contrario*, *S. lugdunensis* possède un *pan-genome* fermé qui stagne en dessous des 3000 gènes après l'analyse de 15 génomes. Cependant, le *core-genome* de ces trois espèces de staphylocoques présente une évolution similaire stagnant autour des 2000 gènes¹⁴¹.

Le fait que *S. lugdunensis* présente un *pan-genome* fermé est surprenant pour une espèce pathogène chez l'Homme, et ce d'autant plus que *S. aureus* et *S. epidermidis* ont, eux, un génome ouvert. Ce caractère coïncide avec la présence de plusieurs mécanismes qui pourraient empêcher les transferts horizontaux *via* les EGM. Toutefois, la présence de tels systèmes de restriction chez *S. aureus* et, plus rarement, chez d'autres SCN présentant un génome ouvert, laisse penser qu'il pourrait exister d'autres mécanismes limitant l'acquisition de nouveaux gènes. Par ailleurs, ce caractère fermé de son génome et cette faible capacité de transformation de l'espèce peut expliquer que *S. lugdunensis* reste sensible à de très nombreuses molécules antibiotiques, ayant une capacité réduite à intégrer des EGM portant des gènes de résistance, contrairement à *S. aureus* et aux autres SCN dont *S. epidermidis*.

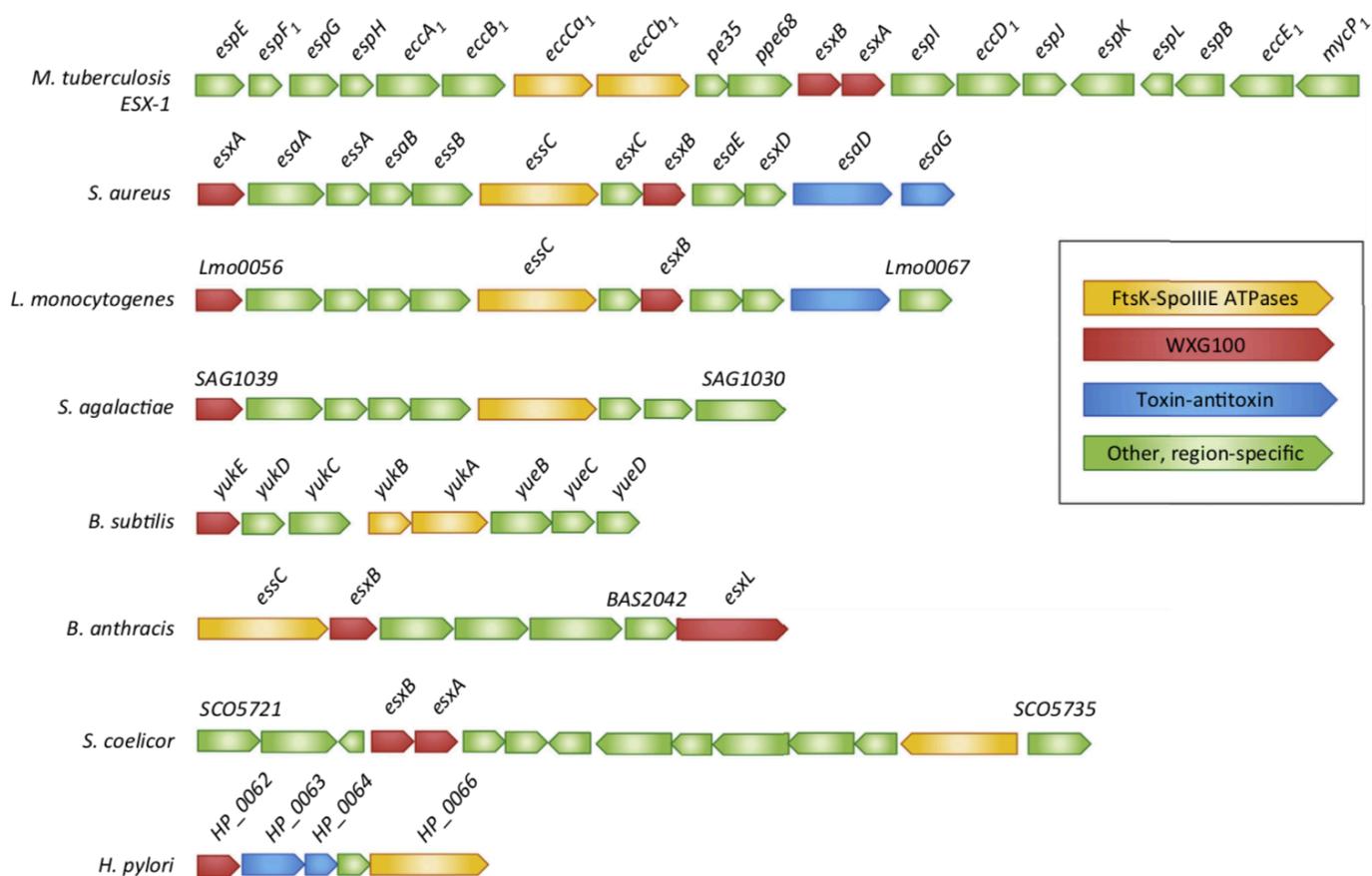


Figure 34 : Organisation génomique du système de sécrétion de type VII (T7SS) chez différentes espèces bactériennes¹⁶³.

L'organisation génomique du *locus* codant le T7SS chez certaines espèces des *phyla* des Actinobactéries (*Mycobacterium tuberculosis* et *Streptomyces coelicor*), des Firmicutes (*Staphylococcus aureus*, *Listeria monocytogenes*, *Streptococcus agalactiae*, *Bacillus subtilis* et *Bacillus anthracis*) et des Gammaprotéobactéries (*Helicobacter pylori*) est représentée. Les gènes codant des protéines de la famille FtsK-SpoIIIE sont représentées en jaune, ceux codant les protéines associées au système toxine-antitoxine sont en bleu. Les gènes sont représentés sous forme de rectangles orientés selon le sens de la transcription et dont la longueur est proportionnelle à celle du gène.

3. Le système de sécrétion Ess/type VII

3.1. Origine et identification du système de sécrétion Ess/type VII

Les bactéries possèdent plusieurs mécanismes qui exportent des protéines à travers la membrane cytoplasmique. Les systèmes de sécrétion bactériens qui interviennent dans la translocation des protéines à travers la bicouche lipidique des bactéries à Gram négatif sont classés en six systèmes, les types I à VI¹⁵³. Chez les bactéries à Gram positif, les systèmes de sécrétion Sec et Tat étaient initialement considérés comme impliqués dans toutes les sécrétions protéiques, jusqu'à la découverte d'un nouveau système de sécrétion spécialisé, ESX-1, chez *Mycobacterium tuberculosis*¹⁵⁴. Comme la paroi des mycobactéries possède une structure didermique analogue à la paroi des bactéries à Gram négatif, la nomenclature a été étendue au système ESX-1, renommé ainsi système de sécrétion de type VII (SST7)¹⁵⁵.

Le système de sécrétion ESX-1 est codé par un *locus* comprenant au moins 20 gènes, porté par la région RD1 (**Figure 34**). Il a été montré que l'atténuation de la virulence de la souche vaccinale vivante de *Mycobacterium bovis*, le bacille de Calmette et Guérin (BCG), est liée à la perte de la région RD1 contenant le *locus* codant le système ESX-1. En effet, dans un modèle d'infection chez la souris, la souche H37Rv de *M. tuberculosis* délétée de la région RD1 (Δ RD1) montrait une virulence atténuée, semblable à la souche BCG, par rapport à la souche sauvage et à la souche complémentée. Avec ces dernières, après 15 jours, aucune des souris infectées n'avait survécu, alors que les souris infectées avec la souche BCG ou avec le mutant Δ RD1 avaient toutes survécu¹⁵⁴.

Dans des modèles d'infection chez l'animal, des différences majeures ont été observées entre les mutants de délétion du *locus* ESX-1 et leurs souches parentales. L'infection par des mutants a résulté en une augmentation de la survie des animaux infectés, une diminution de la réplication bactérienne et une réponse inflammatoire réduite avec un nombre plus faible de granulomes¹⁵⁶.

Ce système a été très étudié chez *M. tuberculosis* car il est responsable de l'exportation de deux protéines immunogènes, l'*Early Secreted Antigen Target-6 kDa* (ESAT-6) et la *Culture Filtrate Protein of 10 kDa* (CFP-10) aussi appelées EsxA et EsxB, respectivement¹⁵⁷. Ces protéines sécrétées seraient impliquées dans le développement de

la tuberculose en interférant avec la réponse immunitaire de l'hôte^{154,156,158,159}. L'infection tuberculeuse se caractérise par la formation de granulomes, constitués de cellules immunitaires regroupées autour des foyers d'infection¹⁵⁶. Les cellules bactériennes survivent et persistent dans les cellules de l'immunité constitutives des granulomes. Les protéines EsxA et EsxB seraient impliquées dans les premiers stades de l'infection en favorisant la croissance intracellulaire dans les macrophages et en interférant avec la maturation du phagolysosome. Ces protéines se sont montrées essentielles dans la survie intracellulaire de souches de *M. tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum* et d'une souche recombinante de BCG exprimant un système ESX-1 fonctionnel^{160,161}. En effet, ces espèces bactériennes sont capables d'échapper au phagolysosome et d'atteindre le cytoplasme des macrophages contrairement à des souches avec un système ESX-1 inactivé ou délété. Cette capacité est imputée aux protéines sécrétées qui joueraient le rôle de *pore-forming toxins* en s'insérant dans la membrane du phagolysosome pour la rompre^{156,162}. Après la phase de survie initiale, les cellules infectées recrutent de nouvelles cellules immunitaires telles que les macrophages et les polynucléaires neutrophiles (PNN) sur le site de l'infection, qui sont exploités dans la formation du granulome¹⁵⁶.

Après la découverte du système ESX-1 chez *M. tuberculosis*, les analyses génomiques ont permis d'identifier des systèmes de sécrétion homologues chez les Actinobactéries et majoritairement dans le phylum des *Firmicutes*, en particulier chez de nombreuses espèces bactériennes pathogènes telles que *S. aureus*, *Bacillus anthracis*, *Streptococcus agalactiae* ou *Listeria monocytogenes*¹⁶³ (**Figure 34**). Le fait que le SST7 soit conservé au sein de nombreuses espèces bactériennes suggèrent qu'il pourrait jouer plusieurs rôles au sein des bactéries. Chez les espèces de mycobactéries, en plus de son rôle dans la virulence, le SST7 a été décrit comme impliqué dans les transferts d'ADN conjugatif, dans l'acquisition du fer et du zinc, dans la perméabilité et l'intégrité de la paroi...¹⁶³

Malgré la prévalence de ce système de sécrétion au sein d'un grand nombre d'espèces bactériennes, il n'a été étudié expérimentalement que pour quelques-unes d'entre elles et notamment chez *Bacillus sp.*¹⁶⁴ et *S. aureus*¹⁶⁵.

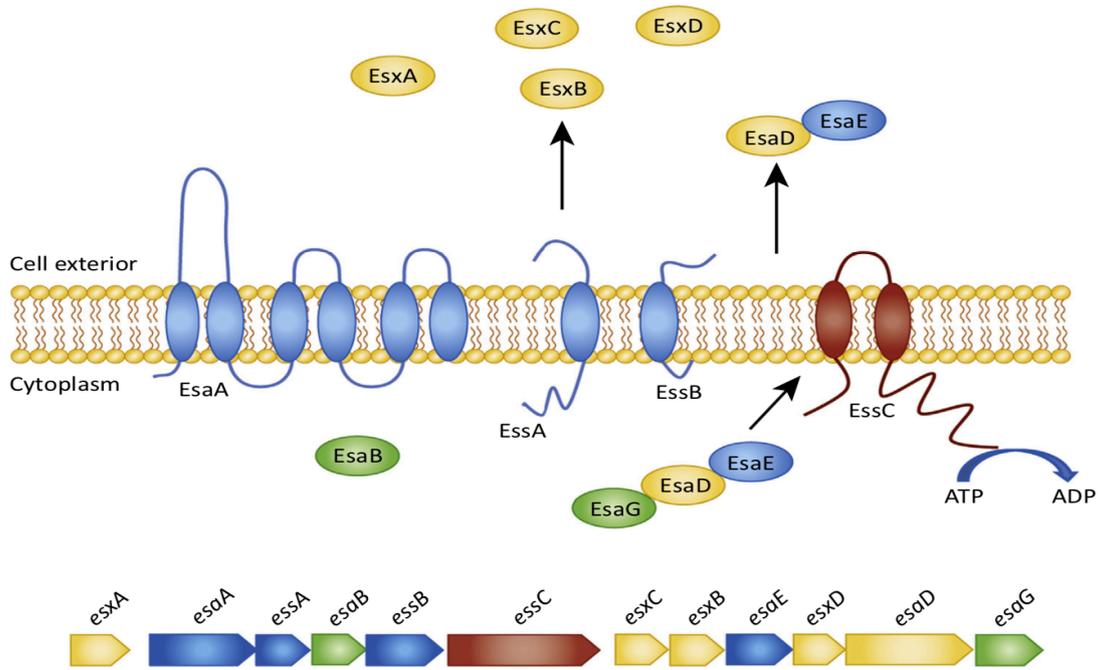


Figure 35 : Le système de sécrétion de type VII (T7SS) et le *locus* *ess* de *S. aureus*¹⁶³.

Représentation de la machinerie du T7SS avec les gènes codant les protéines transmembranaires ou associées à la membrane en bleu, les gènes codant les protéines sécrétées en jaune, les gènes codant les protéines cytoplasmiques en vert et le gène codant la protéine ATPase de la famille SpoIIIE-Fstk est en rouge.

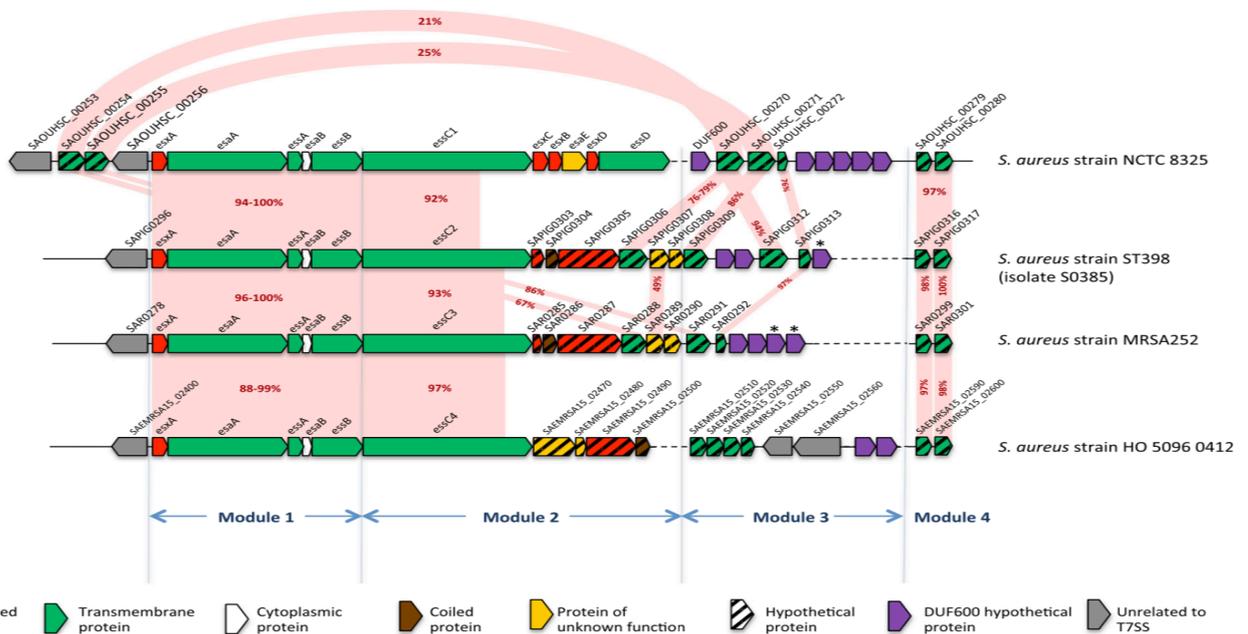


Figure 36 : Représentation de l'organisation génétique du *locus* *ess* codant le système de sécrétion de type VII (T7SS) chez quatre souches de *S. aureus*¹⁶⁶.

Représentation du *locus* *ess* chez quatre souches de *S. aureus* sous la forme de quatre modules. Le module 1 comprend cinq gènes conservés (*esxA* à *essB*). Le module 2 est une région variable comprenant le gène *essC* pour lequel quatre variants ont été décrits (*essC1* à *essC4*), chacun associé en aval à un cluster de gènes différents. Le module 3 est une région hautement variable comprenant des gènes codant des protéines transmembranaires putatives et des protéines avec un domaine DUF600. Le module 4 est quant à lui très conservé. Les zones roses montrent les régions conservées entre les isolats et le pourcentage d'identité protéique est écrit en rouge.

3.2. Le système de sécrétion Ess/type VII de *S. aureus*

Hormis celui de *M. tuberculosis*, le SST7 le mieux caractérisé aujourd'hui est celui de *S. aureus*. Chez cette espèce, le SST7 est codé par un *locus* appelé *ess* (pour *ESAT-6 Secretion System*)¹⁶³. Plusieurs équipes de recherche se sont attachées à déterminer son organisation génétique et transcriptionnelle, et à décrire le rôle de chacune des protéines impliquées dans le mécanisme de sécrétion.

3.2.1. Organisation génétique et transcriptionnelle

Le fonctionnement du SST7 est assuré par six protéines principales (**Figure 35**) : la protéine sécrétée EsxA, la protéine cytoplasmique EsaB et quatre protéines transmembranaires EsaA, EssA, EssB et EssC^{166,167}. Ainsi, la protéine EsxA peut être exportée hors de la cellule bactérienne, par l'intermédiaire des quatre protéines transmembranaires et de la protéine cytoplasmique. Toutes ces protéines sont codées par le *locus ess* qui est très conservé chez *S. aureus*. Cependant, le SST7 ne se limite pas à ces protéines, puisque de nombreux autres gènes ont été décrits au sein du *locus ess*.

Organisation génétique

L'étude des génomes complets d'une collection de 153 isolats de *S. aureus* la plus représentative possible de la diversité génétique de l'espèce (95 *Sequence Types* [ST] et 24 Complexes Clonaux [CC], définis par *MultiLocus Sequence Typing* [MLST]) a permis de décrire l'organisation et la variabilité du *locus ess*¹⁶⁶. Pour faciliter la description des différentes organisations du *locus*, les auteurs ont divisé le *locus* en quatre modules distincts (**Figure 36**)¹⁶⁶. Les gènes des **modules 1 et 4** étaient largement conservés au sein des 153 génomes (de 88 à 100% d'identité) alors que les **modules 2 et 3** étaient beaucoup plus variables (0 à 97% d'identité).

- **Le module 1** comprend cinq gènes, *esxA*, *esaA*, *esaB*, *essA* et *essB*, très conservés au sein de tous isolats. L'organisation génétique observée était la même que celle observée dans des études précédentes avec peu de variations dans la séquence prédite en acides aminés^{168,169}. La nature hautement conservée de ces gènes suggère qu'ils codent des composants essentiels du SST7.

- **Le module 2** comprend plusieurs gènes, notamment le gène *essC* dont la région 5' est conservée. Néanmoins, il existe quatre variants de la région 3' de ce gène (*essC1* à *essC4*), chacun associé à un *cluster* de quatre à six gènes différent en aval. Le variant *essC1* est le variant le plus représenté (90/153 isolats). Il est présent chez plusieurs souches de *S. aureus* de référence (*Newman*, USA300, MW2...) qui ont été utilisées pour décrire la fonction biologique de la protéine EssC¹⁶⁵. Par ailleurs, le module 2 comportant le variant *essC1* se distingue des autres variants par la présence des gènes codant les protéines sécrétées EsxB, EsxC, EsxD, EsaD et EsaE. A la place, chacun des trois autres variants possède un ensemble distinct de gènes codant des protéines hypothétiques non caractérisées à ce jour. Il a également été observé une association entre les variants du gène *essC* et certains ST et CC. En effet, le variant *essC2* n'a été identifié que chez des isolats de CC2 et de ST398, alors que le variant *essC4* a été retrouvé uniquement chez des isolats de CC22.
- **Le module 3** est le plus variable de tous les modules du *locus ess*. Il comprend, selon les variants du gène *essC*, un arrangement complexe de gènes dont certains codent des protéines membranaires putatives tandis que d'autres présents en nombre variable codent des protéines d'environ 166 aa avec un domaine de fonction inconnue, nommé DUF600 (*Domain of Unknown Function* [DUF]). Par ailleurs, le nombre de gènes codant des protéines hypothétiques DUF600 est différent selon les CC. Par exemple, ces gènes ont été retrouvés au nombre de deux chez des isolats de CC22 alors que 13 ont été identifiés chez des souches du CC8. Toutefois, le nombre de gènes codant ces protéines hypothétiques peut également varier entre les isolats d'un même CC. Le rôle et la relation des gènes codant ces protéines DUF600 avec ceux codant les principales protéines du SST7 reste à ce jour à élucider. Pourtant, ces gènes semblent impliqués dans le SST7 puisque tous les gènes du *locus ess* sont sur le même cadre de lecture avec de faibles espaces inter-géniques, suggérant qu'ils soient soumis à un même contrôle transcriptionnel¹⁶⁶.
- **Le module 4** comporte deux gènes codant des protéines membranaires putatives conservées chez tous les isolats de *S. aureus*.

Ainsi, le *locus* codant le SST7 de *S. aureus* présente une importante diversité génétique et varie entre des isolats de différents CC. Par ailleurs, il existe également une organisation transcriptionnelle variable du *locus ess* entre les souches de *S. aureus*.

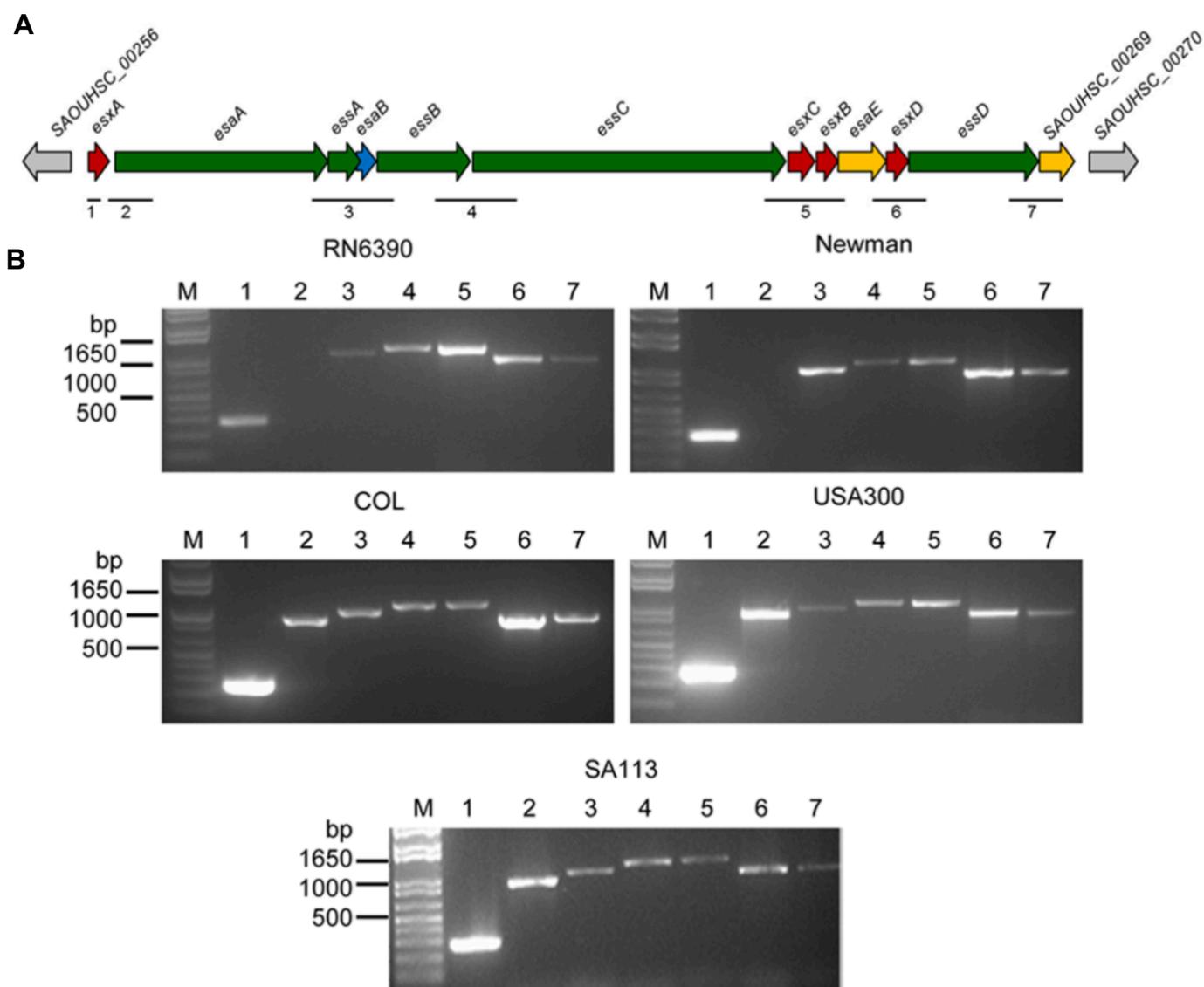


Figure 37 : Étude de l'organisation transcriptionnelle du *locus ess* de cinq souches de *S. aureus* par *Reverse Transcription* couplée à la PCR (RT-PCR)¹⁶⁷.

A, Représentation du *locus ess* chez la souche de référence de *S. aureus* NCTC 8325. Les gènes codant les protéines sécrétées sont en rouge, les gènes codant les protéines transmembranaires sont en vert et le gène codant la protéine cytoplasmique est en bleu. En jaune, sont représentés les gènes codant des protéines aux fonctions inconnues. Les numéros représentent les fragments amplifiés par RT-PCR (cf B).

B, Analyse par RT-PCR des transcrits de cinq souches différentes de *S. aureus*. La taille attendue des fragments 1 à 7 est de 272, 953, 1023, 1153, 1168, 959 and 946 pb respectivement. La PCR n°2, positive pour les souches COL, USA300 et SA113, indique que le gène *esxA* est co-transcrit avec tous les autres gènes du *locus ess*, à l'inverse des souches RN6390 et Newman.

Organisation transcriptionnelle

Kneuper *et al.*¹⁶⁷ ont étudié l'organisation transcriptionnelle des gènes du *locus ess* chez cinq souches de *S. aureus* de référence proches phylogénétiquement, appartenant au même CC (CC8) : *Newman* (*S. aureus* Sensible à la Méricilline [SASM] hospitalier), RN6390 (*S. aureus* Résistant à la Méricilline [SARM] hospitalier), COL (SARM hospitalier), USA300 (SARM communautaire) et SA113 (SARM hospitalier). L'analyse des ARN de ces souches par *reverse transcription* (RT) couplée à la PCR a montré deux organisations différentes (**Figure 37**). En effet, une seule unité transcriptionnelle a été identifiée pour les souches COL, USA300 et SA113. Cela signifie que le gène *esxA* est co-transcrit avec tous les autres gènes du *locus*, contrairement aux souches *Newman* et RN6390 chez lesquelles le gène *esxA* est sur un transcrit différent. En outre, des travaux de 5'-RACE (pour *rapid amplification of cDNA-ends*) PCR ont permis d'identifier un promoteur entre le gène *esxA* et le gène *esaA* chez la souche RN6390 mais pas chez la souche USA300, confirmant ainsi l'existence de deux organisations transcriptionnelles différentes¹⁶⁷. Cela signifie aussi qu'il existe une séquence potentielle de terminaison de la transcription *rho*-indépendante dans la région inter-génique *esxA-esaA* de 83 pb, chevauchant le site d'initiation de la transcription et le promoteur putatif du gène *esaA*. Or, la séquence de la région inter-génique est identique pour les cinq souches. Ainsi, même si cette séquence agit probablement comme un terminateur de transcription efficace chez les souches *Newman* et RN6390, elle permettrait une lecture partielle pour les trois autres souches¹⁶⁷.

Des travaux de séquençage des ARN (RNA-seq) ont été réalisés chez les quatre souches de *S. aureus* représentatives des variants du gène *essC*¹⁶⁶. Cette analyse a permis de confirmer chez les quatre souches (HO50960412, S0385, NCTC8325 MRSA252), dont la souche NCTC8325 dérivée de la souche de *S. aureus* RN6390, que le transcrit du gène *esxA* est monocistronique, alors que les gènes *esaA*, *essA*, *esaB*, *essB*, *essC* et les gènes du module 2 sont sur un transcrit polycistronique¹⁶⁶. Par ailleurs, ces travaux de RNA-seq ont permis de quantifier l'expression des gènes du *locus ess*. Le plus haut niveau d'expression observé était celui du gène *esxA* chez les quatre souches étudiées. Les autres gènes du module 1, *esaA*, *essA*, *esaB* et *essB*, étaient eux aussi transcrits mais à des taux bien plus faibles, de l'ordre de 60 à 70 fois inférieurs¹⁶⁶. Kneuper *et al.*¹⁶⁷ avaient eux aussi montré une différence d'expression de même ordre entre le gène *esxA* et le gène *esaA*.

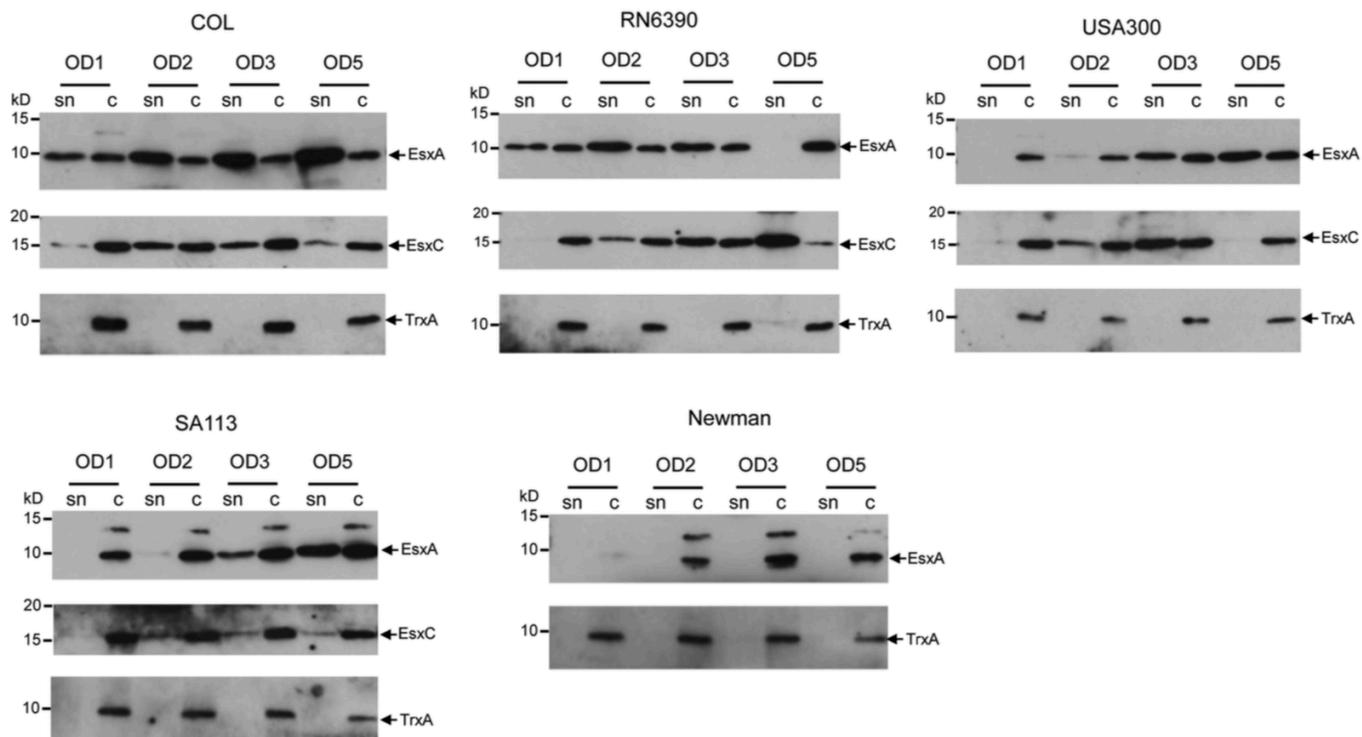


Figure 38 : Profils de sécrétion des protéines EsxA et EsxC de cinq souches de *S. aureus* obtenus par *western blot*¹⁶⁷.

Les cinq souches de *S. aureus* ont été cultivées en milieu liquide trypticase soja jusqu'à atteindre une densité optique (DO) de 1, 2, 3 ou 5. A ces DO, un aliquot de la culture a été prélevé et centrifugé pour récupérer le surnageant (sn) et la fraction cellulaire (c). Les échantillons ont été analysés par *western blot* et les protéines EsxA, EsxC et TrxA ont été révélés à l'aide d'anticorps spécifiques. La protéine TrxA est une protéine cytoplasmique servant de témoin positif. La présence d'une bande de taille attendue indique que la protéine est présente dans la fraction testée.

Ces résultats confirment la présence de deux promoteurs, dont un promoteur fort pour la transcription du gène *esxA*^{166,167}.

D'autre part, la comparaison de l'expression relative des gènes du *locus ess* a permis d'évaluer des différences dépendantes des souches étudiées. Le plus fort taux d'expression était observé chez les souches S0385 et NCTC8325, en particulier pour les gènes du module 1, par rapport aux souches HO 5096 0412 et MRSA252. Par ailleurs, il a été observé des différences d'expression entre les modules du *locus ess* ; tous les gènes des modules 1, 2 et 4 étaient exprimés, à la différence des gènes du module 3¹⁶⁶. Parmi ce dernier, les gènes codant les protéines DUF600 présentaient le niveau d'expression le plus variable. Chacune des souches possédant plusieurs copies de gènes codant des protéines DUF600, il pourrait alors y avoir une redondance fonctionnelle de ces protéines avec une potentielle régulation de l'expression différente de celles des autres gènes du *locus ess*¹⁶⁶.

Profils de sécrétion

Enfin, en plus de la diversité génétique du *locus ess* et de la variabilité transcriptionnelle, différents profils de sécrétions des protéines du SST7 ont été observés. Kneuper *et al.*¹⁶⁷ ont analysé par *western blot* la synthèse dans le cytoplasme et la sécrétion dans le milieu extérieur des protéines EsxA et EsxC à différentes densités optiques (DO) correspondant aux phases de croissance exponentielle (DO 1 et 2) et stationnaire (DO 3 et 5) (**Figure 38**). La sécrétion des deux protéines EsxA et EsxC a pu être détectée pour les souches RN6390, USA300, COL et SA113. Pour les souches COL et RN6390, elles ont été détectées dès la phase exponentielle, alors que pour les souches USA300 et SA113, ces protéines ont été détectées principalement lors de la phase de croissance stationnaire. Chez ces dernières, même si les protéines EsxA et EsxC n'étaient sécrétées que tardivement, elles étaient tout de même produites puisqu'elles ont été détectées par *western blot* dans le cytoplasme dès la phase de croissance exponentielle.

Contrairement aux quatre autres souches, chez la souche *Newman*, la protéine EsxA n'a été détectée dans aucun des surnageants étudiés aux différentes DO, alors qu'elle était détectée dans le cytoplasme. Les auteurs rapportent aussi qu'ils n'ont pu détecter ni la sécrétion, ni la production de la protéine EsxC. Ces résultats sont cohérents avec une étude

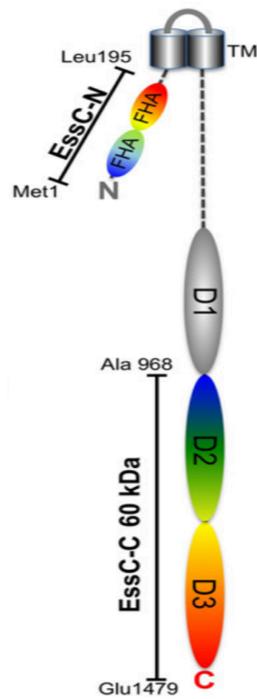


Figure 39 : Structure de la protéine transmembranaire EssC¹⁷³.

La protéine transmembranaire EssC est composée en région N-terminale de deux domaines *forkhead-associated* (FHA), suivis de deux segments transmembranaires (TM) et de trois domaines de la famille des ATPases SpoIIIE-Fstk (D1, D2 et D3)

précédente qui a montré que la production de la protéine EsxC était réprimée lorsque la souche *Newman* était cultivée en milieu liquide trypticase soja¹⁶⁸. Cette analyse démontre des profils de sécrétion variables selon les souches de *S. aureus* et suggère que les protéines du SST7 sont soumises à de possibles modifications post-traductionnelles pour leur sécrétion dans le milieu extérieur¹⁶⁷.

En conclusion, le SST7 de *S. aureus* présente non seulement une organisation génétique et transcriptionnelle variable selon les souches mais aussi des variations en termes de sécrétion protéique. Il est surprenant que de telles différences d'expression aient pu être observées chez cinq souches phylogénétiquement proches, appartenant toutes au même CC, alors que la séquence nucléotidique de leur *locus ess* est quasi identique (seulement deux mutations sur 14 kb). L'impact de cette diversité reste encore à élucider.

3.2.2. Structure et fonction des protéines codées

A ce jour, le *locus ess* de *S. aureus* comprend au moins 12 gènes codant :

- Cinq protéines membranaires EssA, EssB, EssC et EsaA et EsaE (aussi appelée EssE¹⁷⁰)
- Cinq protéines sécrétées EsxA, EsxB, EsxC (aussi appelée EsaC¹⁶⁸), EsxD et EsaD (aussi appelée EssD¹⁷¹)
- Deux protéines cytoplasmiques EsaB et EsaG

Le rôle de certaines d'entre elles a pu être caractérisé ; il est décrit ci-dessous.

3.2.2.1. Les protéines transmembranaires

Pour assurer la synthèse, la translocation et la sécrétion des protéines du SST7, le *locus ess* code des protéines transmembranaires EssA, EssB, EssC et EsaA. Il a été montré que certaines protéines transmembranaires du SST7 étaient essentielles pour la production et/ou la sécrétion selon les souches de *S. aureus* étudiées. Plusieurs études ont consisté à construire des mutants de délétion ou d'inactivation des gènes du *locus ess* chez les souches *Newman*, USA300 et RN6390, et à caractériser leur capacité à sécréter les protéines du SST7¹⁶⁷⁻¹⁶⁹. Ces travaux ont mis en évidence que les protéines

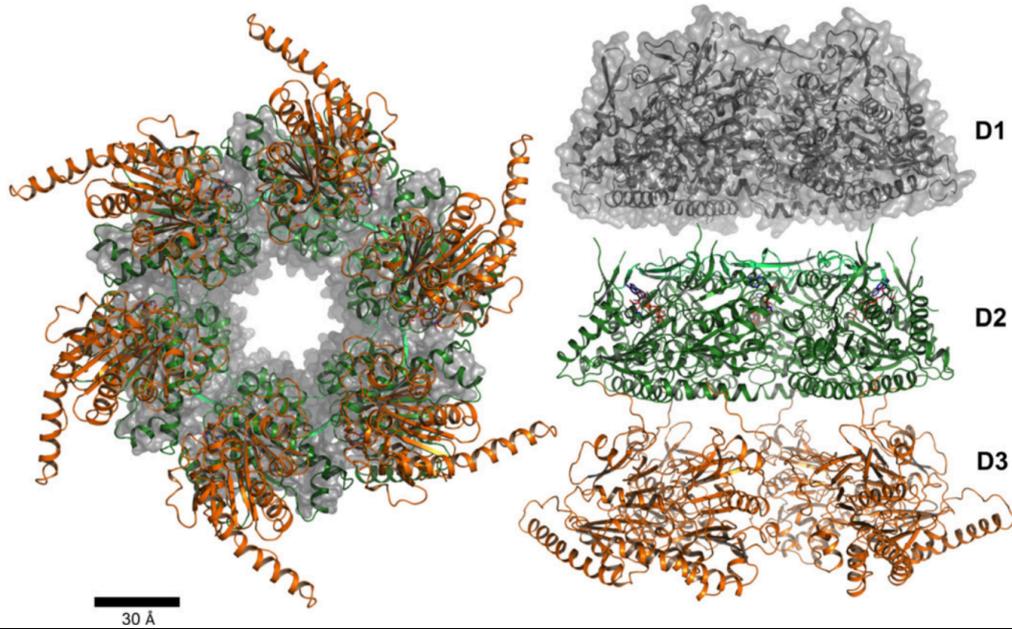
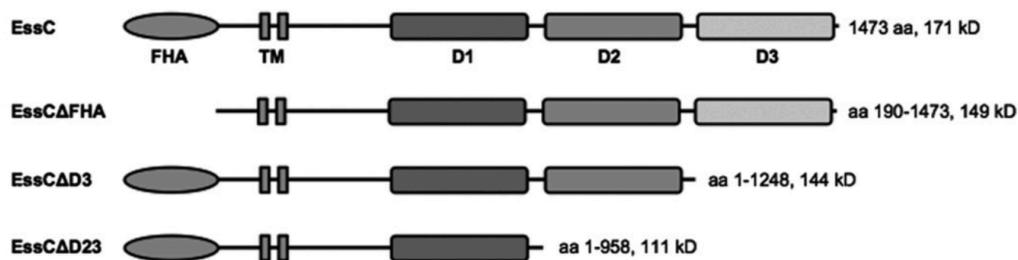


Figure 40 : Modèle hexamérique des domaines ATPases SpoIIIE-Ftsk de la protéine transmembranaire EssC¹⁷³.

Le modèle hexamérique de la protéine EssC a été généré par superposition des domaines D1, D2 et D3. Cette structure forme un pore d'un diamètre de 30 Ångström.

A



B

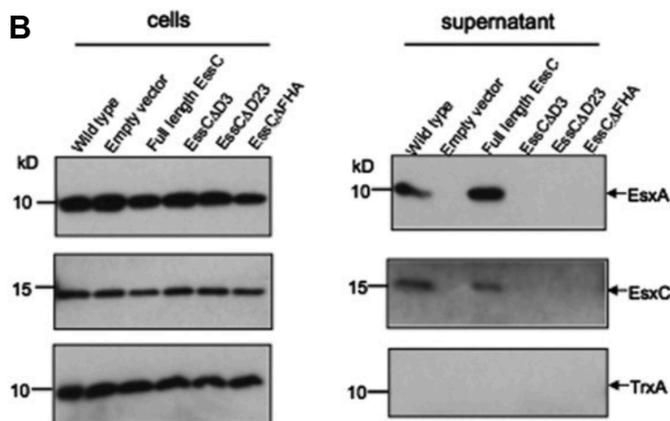


Figure 41 : Analyse des rôles des domaines de la protéine transmembranaire EssC¹⁷³.

A, Représentation schématique de la protéine EssC de la souche sauvage, du mutant de délétion des domaines *forkhead-associated* (FHA) EssCΔFHA, du domaine ATPase SpoIIIE-Ftsk D3 EssCΔD3 et des domaines ATPase SpoIIIE-Ftsk D2 et D3 EssCΔD2D3. Le nombre d'acides aminés et le poids moléculaire des protéines sont indiqués à droite.

B, Les échantillons de fraction cellulaire ou de surnageant ont été obtenus à partir de culture de la souche de *S. aureus* RN6390 sauvage ou du mutant de délétion Δ*essC* avec le plasmide exprimant un des variants de la protéine EssC. Les échantillons ont été analysés par *western blot* et les protéines EsxA, EsxC et TrxA ont été révélés à l'aide d'anticorps spécifiques. La protéine TrxA est une protéine cytoplasmique servant de témoin positif. La présence d'une bande de taille attendue indique que la protéine est présente dans la fraction testée.

transmembranaires EssA, EssB et EssC sont essentielles pour l'export des protéines sécrétées, voire pour leur production. *A contrario*, selon les souches testées, la protéine EsaA n'est pas une protéine déterminante pour le fonctionnement du SST7 de *S. aureus*. Chez *Bacillus subtilis*, l'homologue de la protéine membranaire EsaA appelée YueB est par contre essentielle pour la sécrétion de la protéine YukE, homologue de la protéine EsxA¹⁶⁴.

Parmi les protéines transmembranaires du SST7, la protéine EssC a été la plus étudiée. Bien qu'aucune structure de la protéine EssC n'ait encore été obtenue chez *S. aureus*, un modèle a été proposé à partir de la structure de son homologue chez l'espèce de *Firmicutes*, *Geobacillus thermodenitrificans*¹⁷². La protéine EssC présente au niveau de la région N-terminale deux domaines *forkhead-associated* (FHA) dont le rôle est méconnu, suivis de deux domaines transmembranaires et de trois domaines ATPase de la famille SpoIIIE-FtsK (D1, D2 et D3) (**Figure 39**)¹⁷³. Les protéines SpoIIIE-FtsK sont conservées au sein du monde bactérien et fournissent l'énergie nécessaire à la translocation de l'ADN et de protéines à travers des pores membranaires^{172,173}. Elles sont connues pour s'oligomériser sous forme d'hexamères. Une modélisation de la structure hexamérique des domaines D1, D2 et D3 suggère que la protéine EssC forme un pore d'un diamètre de 30 Ångström, suffisamment grand pour la sécrétion des protéines du SST7^{163,173} (**Figure 40**).

Une délétion du gène codant la protéine EssC chez la souche RN6390 de *S. aureus* a été réalisée (**Figure 41**)¹⁷³. Ainsi, les mutants Δ essC ont été complétés avec des plasmides exprimant des variants de la protéine EssC tels que le variant EssC Δ FHA délété de la séquence codant les domaines FHA, le variant EssC Δ D3 délété de la séquence codant le troisième domaine ATPase SpoIIIE-FtsK et le variant EssC Δ D2D3 délété de la séquence codant le deuxième et le troisième domaine ATPase SpoIIIE-FtsK. La seule perte du domaine D3 suffisait à l'abolition complète de la sécrétion des protéines EsxA et EsxC. Le même résultat a été observé avec les variants *essC* Δ D2D3 et *essC* Δ FHA. Par ailleurs, Burts *et al.*¹⁶⁵ ont également construits chez la souche *Newman* des mutants exprimant des variants de la protéine EssC par insertion d'un transposon au niveau des séquences codant les différents domaines protéiques. De la même manière, la perte du domaine D1 empêche la sécrétion des protéines EsxA et EsxB, mais curieusement, les mutants avec l'insertion du transposon dans les deux derniers domaines D2 et D3 restent capables de sécréter les

protéines EsxA et EsxB. Cette différence de résultats très divergents reste encore aujourd'hui à élucider.

Quoiqu'il en soit, la protéine EssC semble être un composant majeur pour la translocation des protéines dans le milieu extérieur. Les domaines FHA et transmembranaires sont nécessaires à la stabilité de la protéine et à son enchâssement dans la membrane cytoplasmique. D'autre part, les domaines ATPase SpoIIIE-FtsK D1, D2 et D3 sont essentiels structurellement pour former un pore membranaire et permettent la sécrétion des protéines du SST7^{163,173}.

3.2.2.2. Les protéines cytoplasmiques

Le *locus ess* code deux protéines cytoplasmiques. L'une d'entre elles est la protéine EsaG impliquée dans un système T/AT en interagissant avec les protéines sécrétées EsaD et EsaE (cf paragraphe 3.2.2.3).

La seconde protéine cytoplasmique du SST7 est la protéine EsaB de 80 aa structurellement proche d'une ubiquitine¹⁷⁴. Chez les souches de *S. aureus Newman* et USA300, la construction d'un mutant par l'insertion d'un transposon dans le gène *esaB* n'a pas aboli la sécrétion des protéines EsxA, EsxB et EsxC, au contraire elle a entraîné une augmentation de la transcription du gène *esxC*^{165,168}. Ces résultats ont alors laissé penser que cette protéine cytoplasmique pourrait jouer un rôle dans la régulation des transcrits du gène *esxC*, voire de tous les gènes du *locus ess*. Casabona *et al.*¹⁷⁴ ont alors étudié par RNA-seq, le niveau d'expression des gènes du SST7 chez la souche RN6390 mutante et délétée du gène *esaB*. Ils ont montré que la protéine EsaB ne régule pas la transcription du gène *esxC*, ni d'aucun autre gène du *locus ess*, puisqu'aucune différence d'expression n'a été notée entre les deux souches. Ainsi, le rôle de cette protéine n'est pas clairement défini, mais elle ne semble ni impliquée dans les mécanismes de sécrétion, ni dans la régulation de l'expression du *locus ess*.

Une analyse complète des transcrits de la souche mutante Δ *esaB* a révélé une surexpression de 17 gènes, tous codant des protéines impliquées dans l'acquisition du fer et notamment les gènes du *locus isd*. La protéine EsaB et le SST7 pourraient ainsi être impliquée dans la régulation et l'homéostasie du fer¹⁷⁴.

3.2.2.3. Les protéines sécrétées

Les protéines sécrétées EsxA, EsxB, EsxC et EsxD

Les protéines sécrétées EsxA, EsxB, EsxC et EsxD sont de petits peptides d'une taille d'environ 100 aa¹⁶³. Les protéines EsxA et EsxB appartiennent à la famille des protéines WXG100. Ces protéines sont très conservées chez les mycobactéries et chez les Firmicutes¹⁶³. La principale caractéristique de cette famille de protéines est une structure en double hélice reliée par un coude β (structure en forme de boucle en épingle à cheveux) avec comme motif central W – X – G (Tryptophane – X – Glycine). Une autre caractéristique est la présence dans la partie C-terminale d'une séquence consensus HxxxD/ExhxxxxH où H représente un acide aminé hydrophobe très conservé et h n'importe quel acide aminé hydrophobe¹⁷⁵. Chez *M. tuberculosis*, comme chez *B. subtilis* ou *S. aureus*, ce type de motif constitue un signal de sécrétion et d'adressage au SST7.

Le transcrit le plus abondant du SST7 est celui de la protéine EsxA¹⁶⁷. Cette protéine possède un fort degré de conservation au sein des souches de *S. aureus* et est nécessaire pour la sécrétion d'autres substrats^{166,167,169}. En effet, les différents travaux de mutagenèse visant à déléter ou à inactiver les gènes du *locus ess* ont montré que l'absence de la protéine EsxA stoppe la sécrétion, mais pas la synthèse des protéines EsxB, EsxC et EsxD chez les souches RN6390 et USA300^{167,169}. Chez la souche *Newman*, la délétion du gène *esxA* stoppe la sécrétion et même la synthèse de la protéine EsxB ; l'inverse est également vrai puisque la délétion du gène *esxB* abolit aussi la production de la protéine EsxA¹⁶⁵. Par ailleurs, chez la souche RN6390, la délétion du gène *esxB* stoppe la sécrétion de la protéine EsxA, alors que chez la souche USA300, elle abolit sa synthèse^{167,169}.

Les protéines EsxC et EsxD n'appartiennent pas à la famille des protéines WXG100, mais elles peuvent s'associer avec les protéines EsxA et EsxB pour former des dimères. Il a été montré chez la souche USA300 de *S. aureus* que les protéines EsxA et EsxC pouvaient former un hétéro-dimère, mais qu'elles pouvaient également s'auto-assembler afin de former des homo-dimères (EsxA-EsxA ou EsxC-EsxC)¹⁶⁹. En outre, la protéine EsxB est capable de former un hétéro-dimère uniquement avec la protéine EsxD. Toutefois, la raison pour laquelle ces protéines s'associent entre elles pour former des hétéro- ou des homo-dimères n'est pas encore claire. Cette dimérisation pourrait être une condition préalable à la sécrétion des protéines dans le milieu extérieur. Chez *M. tuberculosis*, les

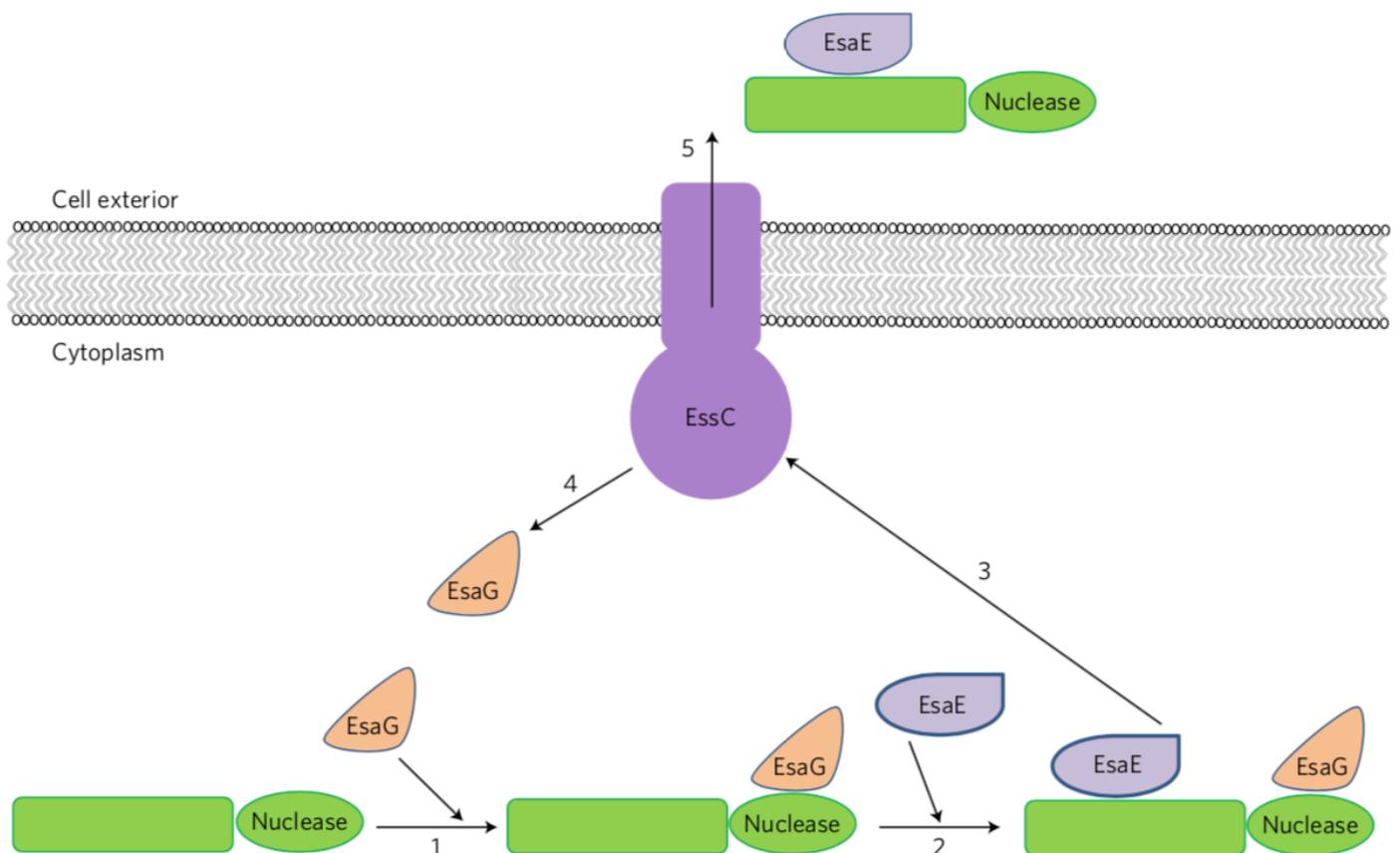


Figure 42 : Modèle de la synthèse et de la sécrétion de la toxine EsaD de *S. aureus*¹⁸⁰.

Après la synthèse de la protéine EsaD (en vert), la protéine cytoplasmique EsaG (orange) se lie au domaine nucléase de la toxine EsaD (1). La protéine EsaE (violet clair) se lie à la région C-terminale de la protéine EsaD (2). Le complexe est ensuite pris en charge par le système de sécrétion de type VII pour l'export de la toxine EsaD *via* la protéine transmembranaire EssC (violet foncé) (3). La protéine EsaG libère la protéine EsaD et reste intra-cytoplasmique (4). Le complexe EsaD-EsaE est libéré dans le milieu extérieur (5).

protéines ESAT-6 et CFP-10 pourraient s'associer afin de former un hétéro-dimère reconnu par des récepteurs membranaires de la cellule de l'hôte et conduire à la modulation du comportement cellulaire à l'avantage de la bactérie¹⁷⁶. Chez *M. tuberculosis* et *M. marinum*, la protéine ESAT-6 peut s'insérer dans les membranes cellulaires de l'hôte et former des pores pour échapper aux compartiments intracellulaires, comme les phagolysosomes¹⁷⁷.

Malgré ces différentes études, les fonctions de ces protéines sécrétées chez *S. aureus* restent vagues. Les protéines EsxB et EsxC sont nécessaires à la virulence de *S. aureus* dans un modèle murin de formation d'abcès^{165,168} et la protéine EsxA a été montrée essentielle dans la virulence en interférant avec les voies de signalisation de l'apoptose dans les cellules hôtes (**cf paragraphe 3.2.3**)¹⁷⁸. Cependant, comme la délétion de ces gènes affecte la sécrétion et/ou la synthèse des autres protéines sécrétées, leur implication directe ou indirecte dans la virulence n'est pas clairement démontrée.

Les protéines EsaD, EsaE et EsaG du complexe T/AT

Le SST7 est également impliqué dans la sécrétion d'un système T/AT qui nécessite la synthèse de trois protéines dont deux protéines sécrétées et une protéine cytoplasmique. Ainsi, la protéine EsaD, décrite initialement comme transmembranaire (d'où son ancien nom EssD), est également sécrétée par le SST7¹⁷⁹. La protéine EsaD est une toxine anti-bactérienne avec un domaine LXG en position N-terminale, couplée à un domaine à activité nucléase en région C-terminale¹⁸⁰ (**Figure 42**). Deux autres protéines codées par le *locus ess* sont essentielles à la synthèse de la protéine EsaD : les protéines EsaE et EsaG. La protéine cytoplasmique EsaG est une antitoxine qui se lie au domaine nucléase de la protéine EsaD, protégeant la bactérie productrice de l'auto-intoxication. En effet, l'introduction de l'antitoxine EsaG dans une souche sensible de *S. aureus* confère une protection contre la lyse provoquée par la toxine EsaD. Une deuxième protéine, EsaE, se lie à la partie N-terminale de la toxine EsaD et permet la translocation du complexe EsaD-EsaE *via* la protéine EssC¹⁸⁰.

Chez *S. aureus*, 50% des souches ne portent pas les gènes codant la toxine EsaD et l'antitoxine EsaG^{166,180}. Cependant, ces souches possèdent dans le *locus ess* au moins deux gènes codant des protéines homologues de l'antitoxine EsaG, les protéines

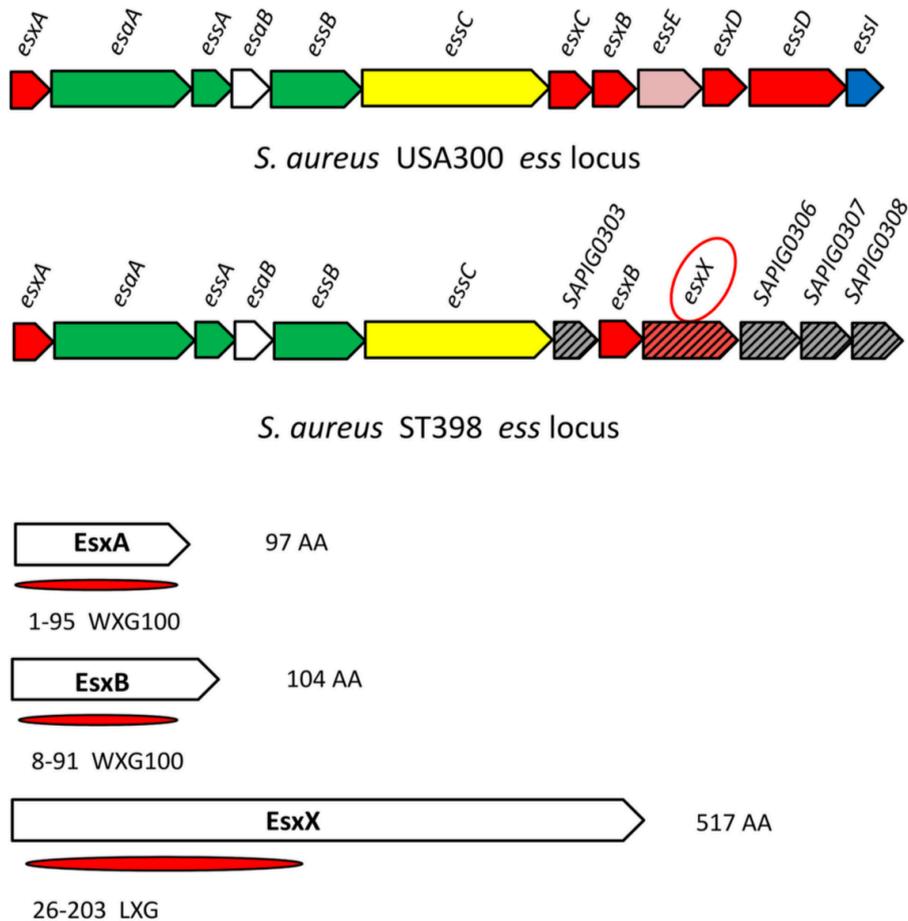


Figure 43 : Localisation du gène *esxX* dans le *locus ess* de la souches RJ-ST398 de *S. aureus* de *Sequence Type* (ST) 398¹⁸².

En haut, est représenté le *locus ess* des souches USA300 (ST8) et RJ-ST398 (ST398). Les gènes codant des protéines ATPases de la famille SpoIIIE-Fstk sont en jaune, les gènes codant des protéines sécrétées sont en rouge, les gènes codant des protéines transmembranaires sont en vert et le gène codant la protéine cytoplasmique est en blanc. Le gène codant la protéine EssE (EsaE) est en rose et le gène codant la protéine EssI (EsaG) est en bleu.

En bas, sont représentés le nombre d'acides aminés des motifs WXG100 et des protéines EsxA et EsxB. La protéine EsxX d'une taille de 517 acides aminés présente un domaine LXG.

DUF600¹⁶⁶. Il a été observé que les protéines DUF600 étaient capables d'interagir avec la toxine EsaD, suggérant que ces souches produisent les protéines DUF600 pour se protéger des souches productrices de la toxine EsaD¹⁸⁰. Ainsi, le système T/AT du SST7 jouerait également un rôle dans la compétition bactérienne¹⁶³.

La protéine EsxX des souches de ST398 de *S. aureus*

Dans le surnageant des souches de ST398, une protéine sécrétée, appelée EsxX a été détectée. La lignée phylogénétique des souches de SARM de ST398 a été décrite initialement comme principalement responsable d'infection chez le bétail et pouvant provoquer des cas sporadiques d'infection chez l'Homme travaillant au contact des animaux¹⁸¹.

Les protéines du SST7 ont principalement été caractérisées chez des souches possédant un *locus ess* avec le variant *essC1*¹⁶⁶. Chez les souches communautaires de ST393 présentant le variant *essC2*, les gènes codant les protéines sécrétées EsxC, EsxD, EsaE et EsaD sont absents¹⁸² (**Figure 43**). L'analyse génomique des souches de ST398 a révélé, en aval du gène *esxB*, un gène codant une protéine de 517 aa correspondant à la protéine EsxX¹⁸². Cette protéine présente un domaine LXG à son extrémité N-terminale, soulignant un potentiel de sécrétion. Le séquençage complet du génome de 76 isolats de ST398 dont 61 responsables d'infection chez l'Homme et 15 responsables de mammites chez le bovin, a été effectué¹⁸². Les données de séquençage ont montré que toutes les souches de ST398 possédaient la protéine EsxX ; 86,8% (66/76) d'entre elles présentaient une identité de séquence de 100% et 13,2% ne présentaient une unique mutation ponctuelle¹⁸².

Les travaux de délétion du gène *esxX* n'ont montré aucun d'impact sur la sécrétion des protéines EsxA et EsxB ; elles étaient toujours présentes dans le surnageant des souches RJ-ST398 mutante et sauvage. Dai *et al.*¹⁸² se sont intéressés à la régulation des gènes codant ces protéines sécrétées par RT-PCR quantitative chez un mutant de délétion du *locus agr*. Ainsi, il a été observé que les gènes *esxA* et *esxB* étaient régulés par *agr* puisque leur expression relative était divisée par deux chez la souche mutante par rapport à la souche parentale. A l'inverse, aucun impact de la délétion du *locus agr* sur l'expression du gène *esxX* n'a été mesuré, suggérant qu'il n'est pas sous son contrôle¹⁸².

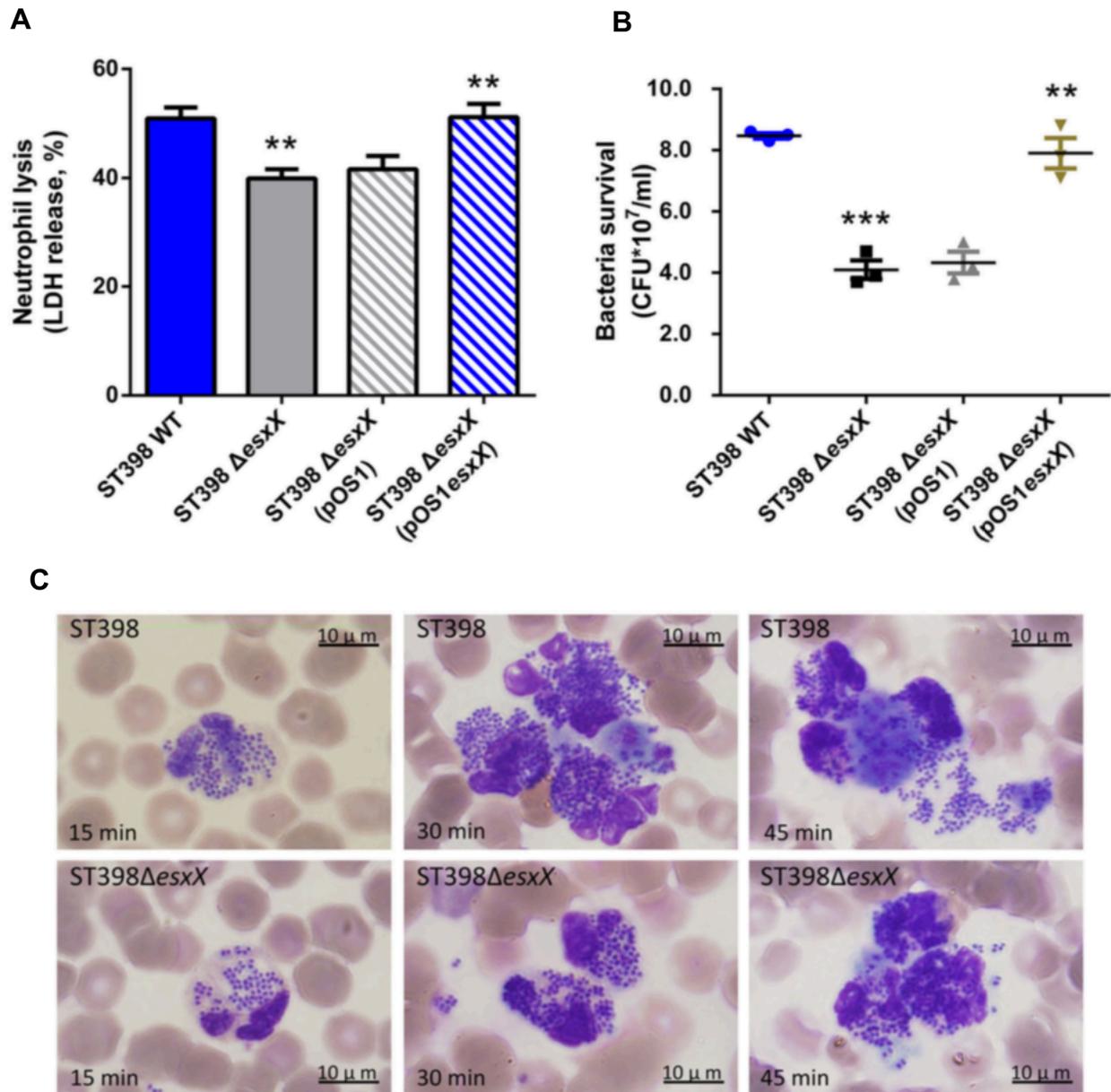


Figure 44 : La protéine EsxX induit la lyse des polynucléaires neutrophiles (PNN) et participe à l'échappement immunitaire de la souche de *S. aureus* RJ-ST398¹⁸².

A, Test de lyse des PNN humains par la souche RJ-ST398 sauvage, délétée du gène Δ esxX ou complémentée. La lyse des PNN a été évaluée par mesure du taux de lactate déshydrogénase, une protéine cytoplasmique. La souche mutante avec le plasmide vide (pOS1) a été testée comme contrôle négatif. Les valeurs statistiques indiquées ont été obtenues par le test de Student avec $**p < 0,01$.

B, Mesure de la survie bactérienne des souches RJ-ST398 sauvage, délétée du gène *esxX* ou complémentée après incubation pendant trois heures avec les PNN humains. Les valeurs statistiques indiquées ont été obtenues par le test de Student avec $**p < 0,01$ et $***p < 0,001$.

C, Observation microscopique de l'internalisation de la souche sauvage et de la souche mutée Δ esxX par les PNN après 15, 30 et 45 min d'incubation.

Dai *et al.*¹⁸² ont voulu déterminer si la protéine EsxX pouvait contribuer à l'échappement de *S. aureus* aux premières lignes de défense du système immunitaire, comme les PNN. La souche sauvage RJ-ST398, la souche délétée du gène *esxX* et la souche complémentée ont été utilisées dans un modèle d'infection *in vitro* de PNN humains. Après trois heures d'incubation, la lyse des PNN a été évaluée par mesure de la lactate déshydrogénase et la survie bactérienne a été comparée entre les souches. Une diminution significative de la lyse des PNN infectés par le mutant de délétion a été observée (**Figure 44A**). Par ailleurs, la survie de la souche mutante $\Delta esxX$ était deux fois moins élevée que la souche parentale (**Figure 44B**). Enfin, lors du test de lyse des PNN, des observations microscopiques ont permis de décrire un taux d'internalisation comparable entre la souche sauvage et la souche mutante (**Figure 44C**). Ces résultats suggèrent donc que la protéine EsxX des souches de ST398 favorise la lyse des PNN après internalisation de la cellule bactérienne et contribue ainsi à l'échappement des défenses immunitaires innées¹⁸².

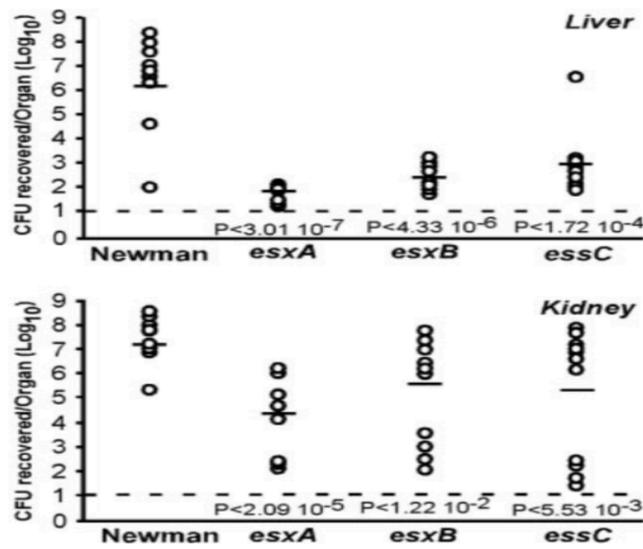


Figure 45 : Modèle de formation d'abcès hépatiques et rénaux chez des souris infectées par la souche de *S. aureus* Newman sauvage ou les mutants dont les gènes *esxA*, *esxB* et *essC* sont inactivés¹⁶⁵.

Les souris ont été infectées par injection rétro-orbitale avec 10^6 UFC pour chacune des quatre souches. Le foie et les reins des souris ont été récupérés quatre jours après l'infection pour dénombrer les bactéries viables dans les abcès.

Les valeurs de significativité statistique ont été obtenues avec le test de Student.

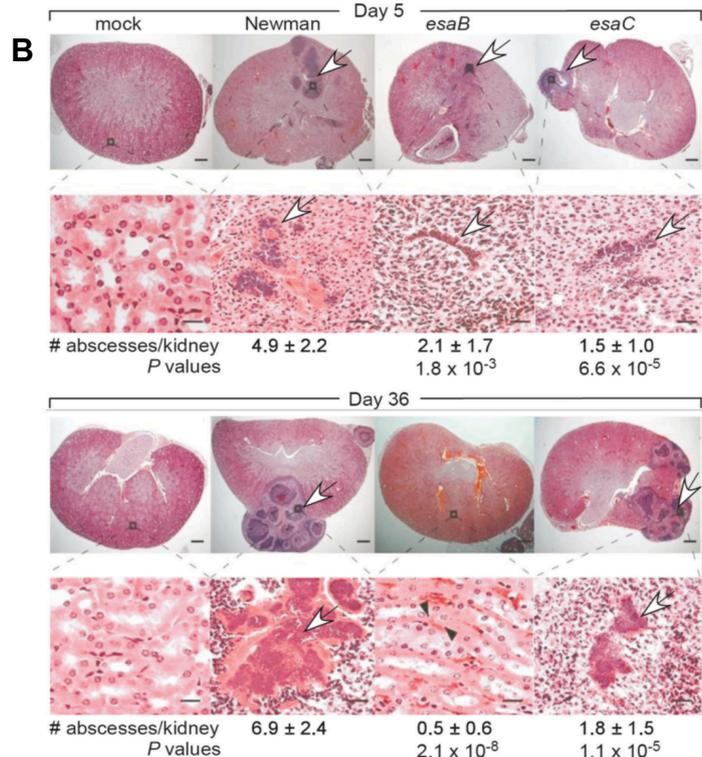
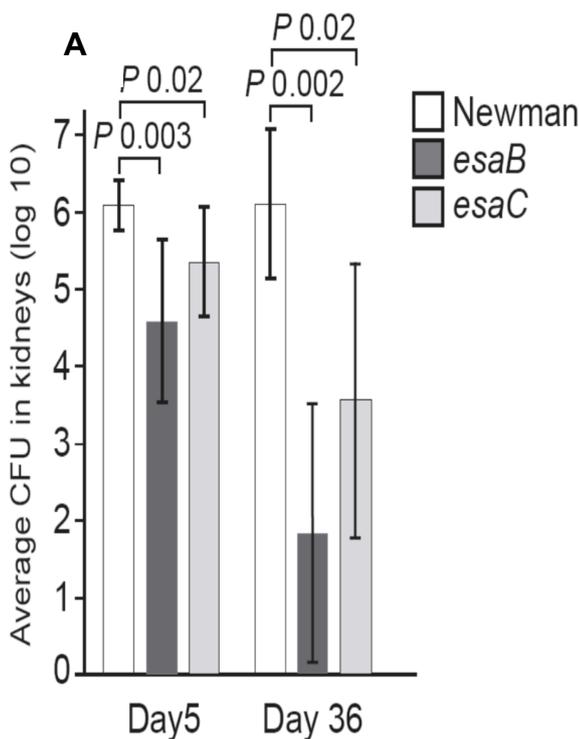


Figure 46 : Modèle de formation d'abcès rénaux chez des souris infectées par injection rétro-orbitale par la souche de *S. aureus* Newman sauvage ou par les mutants dont les gènes *esaB* et *esaC* (*esxC*) sont inactivés¹⁶⁸.

A, Les reins des souris ont été récupérés cinq jours et 36 jours après l'infection pour dénombrer les bactéries viables dans les abcès.

B, Les tissus rénaux ont été fixés dans du formol et colorés avec hématoxyline/éosine. L'observation microscopique des tissus rénaux montre un nombre significativement moins important d'abcès chez les souris infectées par les mutants de délétion *esaB* et *esaC*. Les flèches blanches montrent des abcès avec une forte concentration de bactéries et de cellules immunitaires. Les nombres sous chaque photo correspondent au nombre d'abcès moyens par rein.

Les valeurs statistiques ont été obtenues par le test de Student.

3.2.3. Rôle du système de sécrétion Ess/type VII dans la virulence de *S. aureus*

SST7 & abcès chroniques

Le rôle du SST7 dans la virulence de *S. aureus* a été mis en évidence pour la première fois par l'équipe de Burts *et al.*¹⁶⁵ dans un modèle murin de formation d'abcès après injection rétro-orbitale. La souche de *S. aureus Newman* a été utilisée pour construire des mutants dont les gènes *esxA*, *esxB* et *essC* ont été inactivés par l'insertion d'un transposon. La souche sauvage et les mutants ont été inoculés chez la souris à une concentration de 10⁶ UFC/ml. Après quatre jours, les bactéries vivantes dans les abcès formés dans le foie et les reins ont été dénombrées. Une diminution significative de la capacité de la souche *Newman* à former des abcès rénaux et hépatiques a été observée après inactivation des gènes *esxA*, *esxB* et *essC* (**Figure 45**). En outre, la moitié des animaux infectés avec les mutants avait éliminé la majorité des bactéries dans les reins après quatre jours d'infection. La capacité à former des abcès dans le foie a été encore plus affectée que celle dans les reins¹⁶⁸.

La même équipe a également cherché à déterminer si la protéine sécrétée EsxC et la protéine cytoplasmique EsaB étaient nécessaires à la réplication bactérienne chez l'hôte infecté en utilisant le même modèle d'infection chez la souris et des mutants de délétion des gènes codant les protéines EsxC et EsaB¹⁶⁸. La formation d'abcès rénaux a cette fois été analysée cinq et 36 jours après l'inoculation. Le rein droit des souris a servi à quantifier le nombre de bactéries (**Figure 46A**) et le rein gauche à évaluer la formation d'abcès (**Figure 46B**). Après cinq jours et 36 jours d'infection, la charge bactérienne et le nombre d'abcès intra-rénaux étaient significativement réduits chez les souris infectées avec les mutants *esxC* et *esaB*. Ainsi, le SST7, et notamment les protéines EsxC et EsaB, auraient un rôle clé dans la formation d'abcès chroniques et permettraient à *S. aureus* de persister chez l'hôte infecté¹⁶⁸.

Le rôle dans la virulence de la protéine EsxX identifiée chez les souches de ST398 a également été évalué dans un modèle murin de formation d'abcès cutanés. Pour ce faire, les souris ont été infectées avec la souche *S. aureus* RJ-ST398 sauvage ou mutante Δ *esxX* par une injection sous-cutanée dorsale. Après deux jours d'infection, la taille des abcès sous-cutanés a été mesurée. Les abcès formés présentaient une taille deux fois supérieure chez les souris infectées avec la souche sauvage. L'analyse histologique des tissus infectés

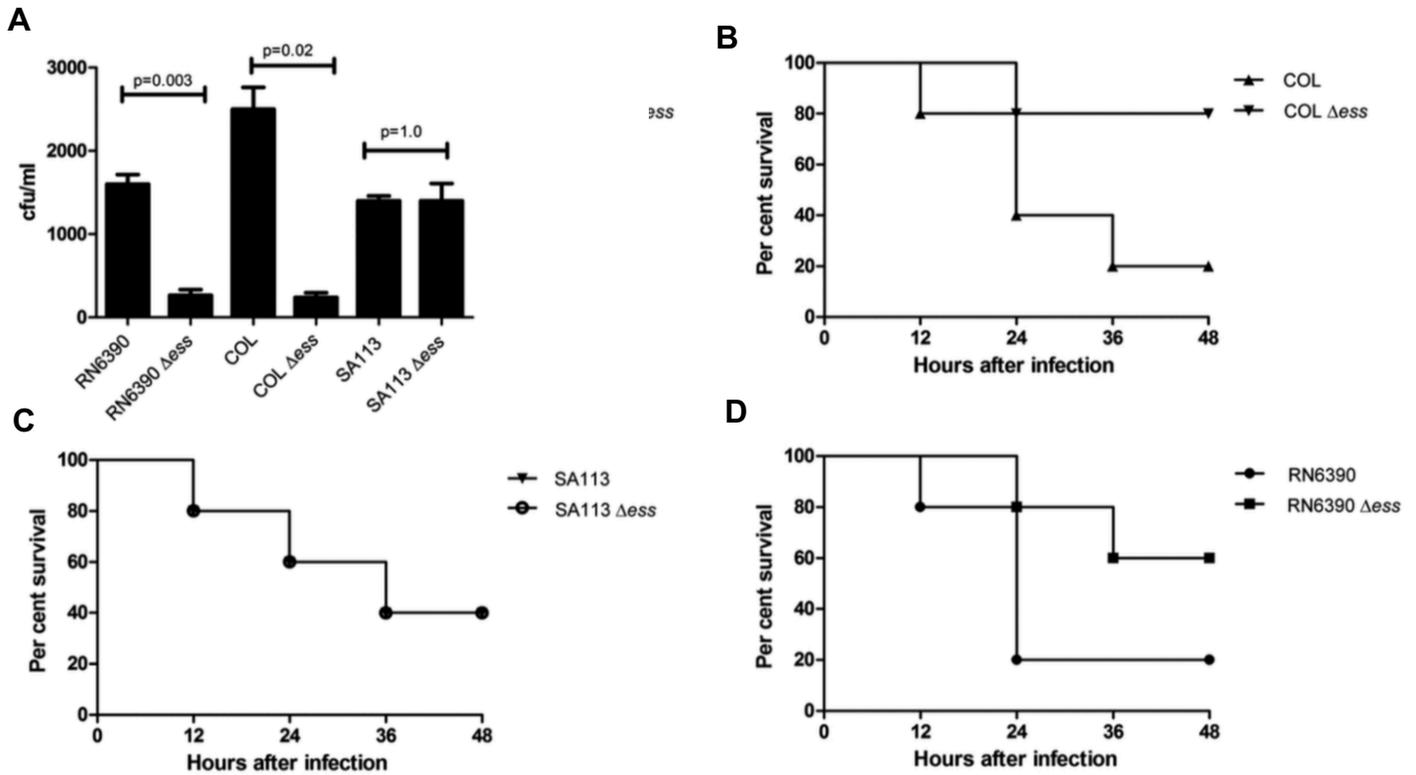


Figure 47 : Contribution du système de sécrétion de type VII des souches de *S. aureus* COL, SA113 et RN6390 à la colonisation nasale et à la virulence dans des modèles de colonisation (A) et de pneumonie (B-D) chez la souris¹⁶⁷.

A, Les souris du modèle de colonisation ont été infectées par les souches sauvage et mutantes Δ_{ess} de *S. aureus* COL, SA113 et RN6390. Après trois jours, le nombre de bactéries intra-nasales a été évalué.

B-D, Les souris du modèle de pneumonie ont été infectées par les souches sauvage de *S. aureus* (B) COL, (C) SA113 et (D) RN6390 ou avec les souches mutantes ayant une délétion du *locus ess*. La survie des souris a été évaluée et comparée jusqu'à 48 heures après l'infection par la méthode Kaplan-Meier.

par la souche sauvage montrait une infiltration des bactéries dans le tissu épithélial, une importante inflammation et une désorganisation cellulaire, absentes dans les tissus infectés par la souche mutante¹⁸².

Ainsi, le SST7 est responsable de la sécrétion de protéines impliquées dans la virulence bactérienne et dans la formation d'abcès. Ces protéines sécrétées pourraient être une des causes de la persistance et de la chronicité des infections à *S. aureus*.

SST7 & colonisation/infection pulmonaire

S. aureus est une espèce qui colonise fréquemment les fosses nasales chez les individus sains. Cependant, c'est également l'un des premiers pathogènes bactériens à coloniser et à infecter les voies respiratoires des patients atteints de mucoviscidose, mais il est aussi responsable de pneumopathie hors contexte de mucoviscidose¹⁸³. Le rôle du SST7 a donc été évalué dans un modèle de colonisation nasale et dans un modèle de pneumonie chez la souris pour déterminer s'il est nécessaire à l'implantation de *S. aureus* dans les fosses nasales puis dans le développement d'une infection pulmonaire (**Figure 47A**). Des souris ont été inoculées par voie intra-nasale soit avec les souches sauvages RN6390, COL ou SA113, soit avec leur mutant de délétion du *locus ess*. Trois jours après, les mutants de délétion des souches RN6390 et COL étaient en quantité moindre par rapport à la souche sauvage dans les fosses nasales, suggérant que le SST7 aurait un rôle dans la colonisation de l'arbre respiratoire par ces deux souches de *S. aureus* (**Figure 47**). Par contre, aucune différence dans la capacité à coloniser n'a été observée entre le mutant et sa souche parentale SA113. Cette différence pourrait être due à la production d'un plus grand nombre de MSCRAMM par cette souche, palliant l'absence du SST7 fonctionnel¹⁸⁴. Dans le modèle de pneumonie, après 48 heures d'infection, l'absence d'un SST7 fonctionnel chez les souches RN6390 et COL a entraîné une réduction significative de la mortalité alors que la souche SA113 mutante et parentale présentaient le même niveau de virulence. Ainsi, le SST7 est important pour la colonisation des fosses nasales et dans le développement de pneumopathie, mais pas pour la souche SA113¹⁶⁷. Les raisons de ces différences souches-dépendantes n'ont pas été corrélées à l'activité de sécrétion du SST7¹⁶⁷. Il a aussi été rapporté que l'expression du gène *essC* de la souche *Newman* de *S. aureus* était induite en présence de surfactant pulmonaire et que la délétion

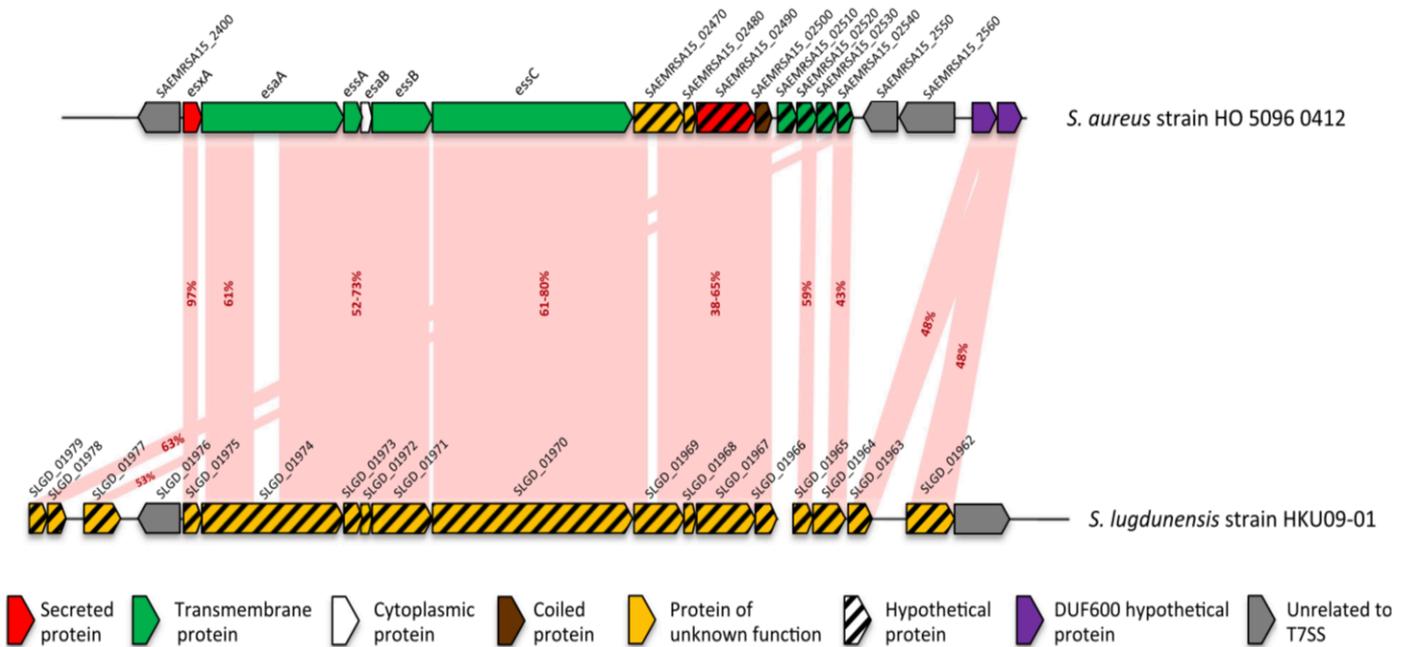


Figure 48 : Comparaison de l'organisation du locus *ess* de *S. aureus* et de *S. lugdunensis*¹⁶⁶.

La comparaison des deux locus met en évidence des régions conservées chez les deux espèces (du gène *esxA* au gène *essB*). Un homologue du gène *essC4* et des gènes codant des protéines avec un domaine DUF600 ont également été identifiés. Les zones roses montrent les régions conservées entre les isolats et le pourcentage d'identité protéique est écrit en rouge.

du gène *essC* était associée à une virulence atténuée dans un modèle murin de pneumonie¹⁸⁵.

Il a été mis en évidence que les protéines EsxA et EsxB étaient capables de former un hétéro-dimère. Cependant, le rôle de ces protéines sur les cellules de l'hôte n'a pas été élucidé. Pour ces raisons, Korea *et al.*¹⁷⁸ ont infecté des cellules épithéliales pulmonaires A549 avec des mutants de délétion des gènes *esxA* et *esxB* ainsi qu'un double mutant de délétion des deux gènes (Δ *esxAB*). Ces mutants ont montré un taux d'invasion des cellules épithéliales pulmonaires équivalent à celui de la souche sauvage. Toutefois, le nombre de bactéries intracellulaires et le taux d'apoptose était plus important avec le double mutant de délétion par rapport aux mutants simples et à la souche sauvage.

Afin de déterminer si la protéine EsxA pourrait interférer avec les voies apoptotiques cellulaires, les cellules A549 ont été transfectées avec un plasmide pour qu'elles expriment soit la protéine EsxA sauvage, soit une protéine EsxA tronquée. Il a été montré par cytométrie en flux que les cellules A549 exprimant la protéine EsxA sauvage entraient 2,5 fois plus en apoptose que les cellules exprimant la protéine EsxA tronquée. Ainsi, la protéine EsxA interférerait avec les voies de signalisation apoptotique des cellules de l'hôte et avec la protéine EsxB, pour médier la libération de *S. aureus*¹⁷⁸.

3.3. Le système de sécrétion Ess/type VII de *S. lugdunensis*

Une analyse comparative du génome de *S. aureus* avec des génomes de référence d'autres espèces de staphylocoques a été réalisée pour étudier la distribution et l'organisation génétique du *locus ess* au sein du genre *Staphylococcus*¹⁶⁶. Le *locus* codant le SST7 a été identifié dans le génome de la souche de *S. lugdunensis* HKU09-01¹³³ (**Figure 48**). Cette région génomique présente des gènes homologues de la partie conservée du module 1 du *locus ess* chez *S. aureus*, allant du gène *esxA* au gène *essB*. Dans le module 2, il a été identifié un gène homologue au variant *essC4* et un gène codant une protéine sécrétée putative. Au sein du module 3, des gènes présentent une homologie de séquence avec les gènes codant les protéines DUF600, mais aucun homologue des gènes du module 4 n'a été identifié.

L'analyse comparative avec le génome de plusieurs SCN tels que : *S. epidermidis*, *S. haemolyticus*, *Staphylococcus pasteurii*, *Staphylococcus pseudintermedius*,

Staphylococcus saprophyticus et *Staphylococcus warneri* a permis de mettre en évidence qu'aucun autre SCN ne possédait le *locus* *ess* en entier. *S. pasteurii* possède un gène homologue à ceux codant les protéines DUF600 et quatre gènes homologues au *cluster* de gènes codant des protéines transmembranaires putatives chez *S. aureus* (SAEMRSA15_02510, 02520, 02530 et 02540). Des homologues de ces quatre gènes ont également été identifiés chez *S. warneri*. Toutefois, aucun des gènes codant les principales protéines du SST7 n'a été identifié chez des SCN autres que *S. lugdunensis*.

S. lugdunensis étant le seul SCN à posséder le *locus* complet codant le SST7, cela suggère une recombinaison spécifique avec *S. aureus*¹⁶⁶. Ceci est un nouvel élément en faveur de sa proximité avec *S. aureus*, plutôt qu'avec les autres SCN. Le SST7 n'a pas encore été caractérisé chez *S. lugdunensis*; il pourrait se révéler être un facteur de virulence chez cette espèce.

4. Génotypage au sein du genre *Staphylococcus*

Pendant longtemps, la classification et le typage des espèces bactériennes ont reposé sur des caractères phénotypiques comme des tests biochimiques ou antigéniques, la croissance ou l'inhibition en présence de substrats ou d'antibiotiques spécifiques, permettant d'identifier des profils distincts. Ces techniques ont ensuite laissé place à des méthodes de typage basées sur des techniques moléculaires utilisant la PCR (amplification aléatoire, ribotypage ...) ou l'analyse de fragments d'ADN.

Une de ces méthodes autrefois largement utilisées pour le typage bactérien est la technique d'électrophorèse en champ pulsé ou *Pulsed Field Gel Electrophoresis* (PFGE)¹⁸⁶. Cette méthode permet de distinguer des souches bactériennes sur la base du nombre de fragments séparés par électrophorèse en champ pulsé après digestion du chromosome bactérien par restriction enzymatique. Bien que cette méthode soit discriminante, elle ne permet pas de produire des données facilement échangeables et comparables entre laboratoires.

Avec l'essor du séquençage, des méthodes basées sur l'analyse des séquences nucléotidiques ont été développées afin de produire des données non-ambiguës échangeables permettant d'effectuer des comparaisons inter-laboratoires et des études épidémiologiques à différentes échelles¹⁸⁷. Le typage bactérien est alors devenu un outil épidémiologique puissant et utile pour évaluer la diversité génétique de populations bactériennes à différents niveaux :

- Les analyses micro-épidémiologiques sont des études locales et à court-terme. Il s'agit alors de vérifier si des isolats collectés dans une même aire géographique (par exemple un service hospitalier ou un hôpital) sont une seule et même souche. A ce niveau, le typage bactérien répond à une simple question : y a-t-il une transmission croisée entre patients ou une persistance environnementale d'une souche bactérienne ?
- Les analyses macro-épidémiologiques sont des études globales à long-terme de populations bactériennes. Il s'agit d'établir les relations phylogénétiques entre des isolats collectés à l'échelle nationale voire mondiale. A cette échelle, le typage bactérien a pour objectif l'étude de la structure génétique des populations bactériennes

par l'analyse du polymorphisme de gènes conservés et les modalités de cette variabilité au cours de l'évolution.

La macro-épidémiologie permet d'évaluer la prévalence de souches ou de génotypes au sein d'un pays ou d'une région du monde ou d'associer un génotype à un tableau clinique donné à un hôte ou à des résistances particulières aux antibiotiques, définissant ainsi des lignées hypervirulentes ou des clones multi-résistants.

Afin de définir les liens phylogéniques entre les souches, les méthodes de typage se doivent d'être discriminantes. Pour ce faire, elles sont basées sur la variabilité de différents marqueurs bactériens¹⁸⁷ :

- Des *loci* bien caractérisés et conservés au cours de l'évolution adaptés à un suivi macro-épidémiologique sur de longues périodes (comme les gènes domestiques lors des études par MLST).
- Des *loci* hypervariables, évoluant rapidement, qui permettent d'identifier des variations entre des souches, même sur une courte période. Les analyses de variations fondées sur ces loci sont plus adaptées à des études micro-épidémiologiques (comme la *Multiple-Locus Variable Number Tandem Repeat [VNTR] Analysis [MLVA]*).

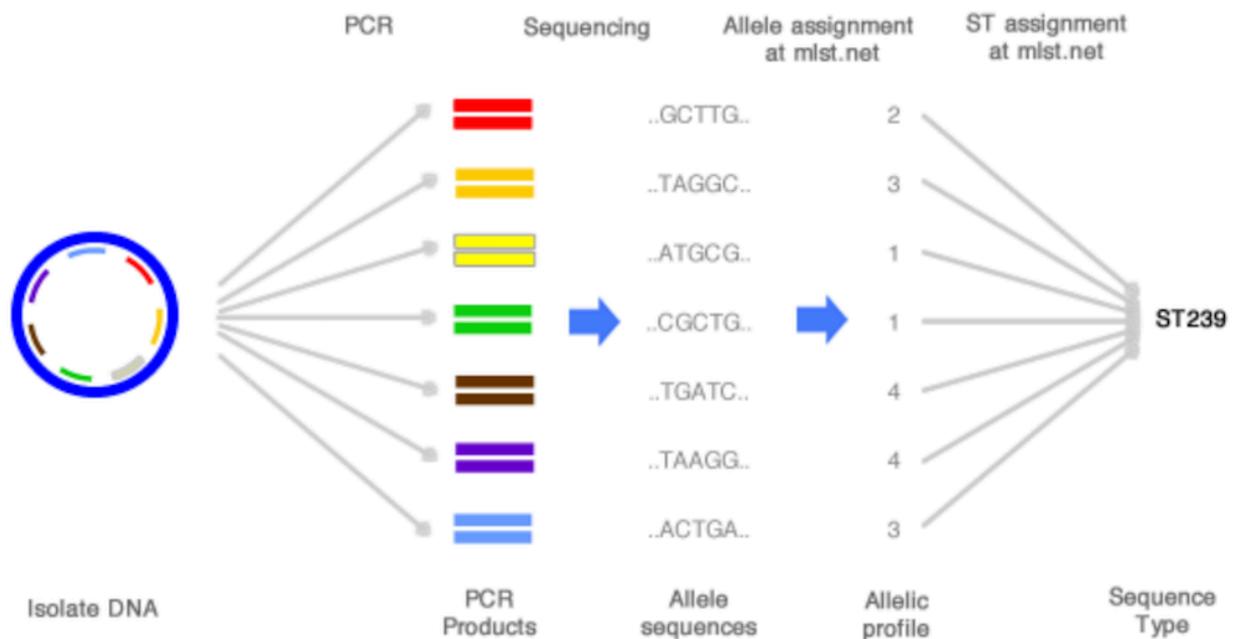


Figure 49 : Principe général du typage d'une souche par *MultiLocus Sequence Typing* (MLST).

Les différentes étapes de la méthode de MLST sont les suivants :

- Extraction de l'ADN génomique
- Amplification et séquençage d'un fragment interne de sept gènes domestiques
- Comparaison des séquences et identification des allèles
- Identification d'un profil allélique unique à sept chiffres pour chaque *Sequence Type*

Source : <http://beta.mlst.net/Instructions/default.html>

4.1. L'analyse par MultiLocus Sequence Typing

4.1.1. Principe de la méthode

Le typage par MLST a été développé sur le principe d'une ancienne technique appelée *MultiLocus Enzyme Electrophoresis* (MLEE). Cette dernière est fondée sur la caractérisation du polymorphisme de gènes codant des enzymes métaboliques sous la forme de variants de migration électrophorétique¹⁸⁷.

Bien que discriminant et capable d'identifier des *clusters* de souches responsables d'infections chez plusieurs espèces bactériennes, le typage par MLEE présentait néanmoins quelques limites¹⁸⁷. Tout d'abord, cette méthode ne permettait pas d'identifier tous les polymorphismes puisque les mutations silencieuses n'étaient pas prises en compte. Par ailleurs, comme le typage par PFGE, cette technique ne produisait pas de données et de profils facilement comparables entre laboratoires. C'est pour ces raisons qu'a été développée la MLST, dont le principe ne repose pas sur l'analyse de migrations électrophorétiques mais sur des données de séquençage non-ambiguës.

Le génotypage par MLST consiste en l'analyse du polymorphisme de séquence de gènes codant des enzymes impliquées dans le métabolisme bactérien. Généralement, il s'agit de six à huit gènes domestiques bien répartis sur le chromosome dont une région d'environ 500 pb est amplifiée et séquencée sur les deux brins pour tous les *loci*, et ce pour chaque isolat^{188,189}. Les séquences obtenues sont alignées et pour chaque nouvelle mutation mise en évidence, un numéro d'allèle arbitraire est attribué. L'analyse des *loci* génère un profil numérique à sept chiffres, ou profil *multilocus* résumé par un numéro de ST (**Figure 49**).

Avec l'algorithme *eBURST* (<http://eburst.mlst.net>)¹⁹⁰, il est possible de regrouper les isolats de ST proches en des CC. Ces *clusters* rassemblent, dans le cas d'un schéma à sept gènes, les isolats de ST partageant au moins cinq allèles identiques avec au moins un autre ST. Au sein d'un CC, les ST qui se distinguent par un seul *locus* sont appelés *Single Locus Variant* (SLV) et ceux qui diffèrent par deux *loci* sont des *Double Locus Variant* (DLV). Il est également possible d'identifier, au sein d'un CC, le génotype ancestral (ancêtre commun). Ce dernier est défini comme le ST qui diffère du plus grand nombre de ST par un seul *locus*, c'est-à-dire le ST qui présente le plus de SLV au sein d'un CC¹⁹⁰.

La MLST est considérée comme une méthode de référence en génétique des populations bactériennes, standardisée et discriminante, avec une bonne reproductibilité intra- et inter-laboratoires¹⁸⁸. Ainsi, les données de séquençage des allèles des différents *loci* et les profils alléliques sont déposés sur des bases de données comme celle appelée "PubMLST" hébergée par l'université d'Oxford (<https://pubmlst.org>) ou celle hébergée par l'Institut Pasteur (<http://bigsd.b.pasteur.fr>). Ces bases de données sont régulièrement enrichies avec de nouveaux allèles et ST et font de la MLST un outil puissant pour les études épidémiologiques. En effet, cette méthode permet de confronter des données de différents laboratoires et de rechercher la diffusion de souches bactériennes virulentes, d'identifier des clones résistants aux antibiotiques, de réaliser des surveillances épidémiologiques¹⁸⁸... Depuis sa mise au point, il existe aujourd'hui de nombreux schémas de MLST adaptés à plusieurs espèces bactériennes, notamment des espèces posant des problèmes majeurs de santé publique : *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Streptococcus agalactiae*, *S. aureus*¹⁸⁸...

Les analyses de MLST permettent également d'étudier la structure et la dynamique d'évolution de la population d'une espèce donnée¹⁸⁹. En effet, il existe deux types de mécanismes d'évolution génomique menant à deux grands types de populations bactériennes. La première est appelée population panmictique, c'est-à-dire évoluant par recombinaisons *via* des échanges horizontaux d'ADN, par exemple *N. meningitidis*. La seconde est une population dite clonale, évoluant principalement par mutations, et pour laquelle le taux de recombinaison est faible. Dans ce type de population, la pression de l'environnement sélectionne des mutations présentant un avantage adaptatif. Toute mutation altérant profondément les fonctions vitales de la bactérie, menant à une cellule non viable, ne sont pas conservées. Toutefois, les mutations non délétères pour la bactérie sont conservées au fil des générations ; on peut ainsi parler de clones bactériens. Ce mode d'évolution se caractérise par une association non aléatoire ou un déséquilibre de liaison entre les allèles des différents gènes étudiés. Le calcul de l'index de clonalité ou d'association (IA) reflète la probabilité que des recombinaisons génétiques se produisent¹⁹¹. Lorsqu'il est supérieur à zéro, cela signifie qu'il existe un déséquilibre de liaison entre les allèles, c'est-à-dire que leur répartition n'est pas aléatoire au sein des ST et donc que la population est clonale¹⁹².

Décrite initialement chez *Neisseria meningitidis*¹⁸⁷, la MLST est une méthode de typage largement utilisée chez *S. aureus* et chez les SCN.

4.1.2. MLST & *S. aureus*

Le schéma de MLST dédié à l'étude de l'espèce *S. aureus* a été développé en 2000 à partir de 155 isolats collectés d'infections communautaires et associées aux soins entre 1997 et 1998 dans la région d'Oxford¹⁹³. Cette étude s'est intéressée au polymorphisme de 14 gènes de ménage, mais sept d'entre eux n'étaient pas assez discriminants. Ainsi, les sept gènes domestiques qui présentaient le plus grand nombre d'allèles ont été conservés :

- *arcC* : carbamate kinase
- *aroE* : shikimate deshydrogenase
- *glpF* : glycerol kinase
- *gmk* : guanylate kinase
- *pta* : phosphate acetyltransferase
- *tpi* : triosephosphate isomerase
- *yqiL* : acetyl coenzyme A acetyltransferase

Le polymorphisme relevé à partir de ces sept gènes a permis d'identifier entre 11 et 17 allèles par *locus* menant à l'identification de 53 ST pour 155 isolats¹⁹³. Pour juger du pouvoir résolutif d'un schéma de typage, le calcul de l'index de diversité de *Simpson* (DI) peut être réalisé. Plus ce dernier est proche de 1, plus le schéma est discriminant. Ainsi, ce schéma est très résolutif puisque le DI calculé est égal à 0,996.

A ce jour, le typage par MLST de l'espèce *S. aureus* a permis de répertorier 4926 ST sur la base de données *PubMLST* par l'étude de 33735 isolats.

Depuis le développement de cette méthode, les données de MLST ont été largement utilisées pour caractériser la structure de la population de *S. aureus* et effectuer un suivi épidémiologique mondial des infections à *S. aureus*¹⁹⁴. De nombreuses études ont révélé que *S. aureus* possède une population clonale, avec un nombre limité de CC majeurs prédominants dans le monde. Les CC en question, principalement le CC1, CC5, CC8, CC15, CC22, CC30, CC45, CC59, CC80 et CC121, sont largement répandus dans le monde, bien que leur fréquence puisse varier en fonction de l'origine géographique, de l'hôte (Homme ou animal), du lieu d'isolement de la souche (hôpital ou communauté) et de

Location	Standard name	Common name	Molecular characteristics					
			PFGE	MLST	spa type	SCCmec type	agr type	
USA/Canada	ST5-MRSA II	USA100	USA100/CMRSA2	ST5	t002/t242/t311	II	2	
	ST45-MRSA II	USA600	USA600/CMRSA1	ST45		II	1/4	
	ST36-MRSA II	USA200	USA200/CMRSA4,8,9	ST36	t018	II	3	
	ST8-MRSA II			ST8	t064/t068	II	1	
	ST241-MRSA II			ST241	t037/t138	II	1	
	ST239-MRSA III			PFGE type B	ST239	t037/t138	III	1
Mexico/South America	ST5-MRSA I	Cordobes/Chilean clone	PFGE type C	ST5	t002/t003	I	2	
	ST239-MRSA III	Brazilian clone	PFGE type B	ST239	t037	III	1	
	ST5-MRSA II	New York/Japan clone	USA100/CMRSA2	ST5	t002/t003/t311	II	2	
Europe	ST36-MRSA II	EMRSA-16	USA200/CMRSA4,8,9	ST36	t018	II	3	
	ST5-MRSA I	EMRSA-3	PFGE type C	ST5	t002/t003	I	2	
	ST239-MRSA III	EMRSA-1,4,11, Vienna/ Brazilian/Hungarian clone	PFGE type B	ST239	t037/t138	III	1	
	ST247-MRSA I/II	EMRSA-5,7 Iberian clone	PFGE type A	ST247	t051/t052	I/II	1	
	ST241-MRSA III	UK-Finland clone		ST241	t037/t138	III	1	
	ST8-MRSA II			ST8	t064/t068	II	1	
	ST105-MRSA II			ST105		II	2	
	ST228-MRSA I			ST228		I	2	
	ST5-MRSA II	New York/Japan clone	USA100/CMRSA2	ST5	t002/t003/t311	II	2	
	ST239-MRSA III variant	Portuguese clone	PFGE type B	ST239		III variant	1	
	ST225-MRSA II			ST225		II		
	Africa	ST239-MRSA III	Brazilian/Hungarian clone	PFGE type B	ST239	t037	III/III mercury	1
		ST241-MRSA III			ST241		III	1
ST250-MRSA I				ST250	t194/t292	I	1/4	
ST5-MRSA I			PFGE type C	ST5	t002/t003	I	2	
ST36-MRSA II			USA200/CMRSA4,8,9	ST36	t018	II	3	
ST5-MRSA III			USA100/CMRSA2	ST5		III	2	
ST247-MRSA I				ST247	t051/t052	I	1	
ST1819-MRSA I				ST1819		I	1	
Middle East		ST239-MRSA III		PFGE type B	ST239	t037/t030	III	1
Asia	ST239-MRSA III		PFGE type B	ST239	t030/t037	III/III A	1	
	ST241-MRSA III			ST241		III/III A	1	
	ST5-MRSA II	New York/Japan clone	USA100	ST5	t002/t003/t242	II	2	
	ST254-MRSA I			ST254		I	1	
	ST30-MRSA I			ST30	t019/t318	I	3	
	ST2590-MRSA II			ST2590	t002	II	2	
Australia/New Zealand	ST239-MRSA III		PFGE type B	ST239	t030/t037	III	1	

^a—, absence of PVL. Blank entries indicate a lack of available information.

Tableau V : Sequence types des clones épidémiques hospitaliers de *S. aureus* résistants à la méticilline et leur caractéristiques moléculaires¹⁹⁹.

la prévalence des souches de SASM ou de SARM^{194,195}. En outre, le typage par MLST a permis de suivre l'émergence et la diffusion de clones résistants à la méticilline, en particulier lorsqu'il était associé au typage de la cassette *SCCmec*.

Les données de typage ont montré que les SARM hospitaliers étaient fréquemment associés à cinq CC (CC5, CC8, CC22, CC30 et CC45)¹⁹⁴. La répartition et la prévalence de ces CC peuvent toutefois varier selon les régions du monde (**Tableau V**). Par ailleurs, il a été montré que les lignées majeures de SARM hospitaliers ont émergé à partir de clones épidémiques de SASM communautaires qui ont acquis la cassette *SCCmec*, comme par exemple les clones de SARM USA100 de CC5 ou ESARM-16 de CC30 identifiés aux États-Unis et au Royaume-Uni^{193,196}. Ces résultats ne sont pas surprenants puisque des souches de SASM ont déjà été montrées comme appartenant aux mêmes CC que les SARM (CC5, CC8, CC30 et CC45) suggérant que les souches de SASM présents dans ces CC ont une plus grande capacité à accepter la cassette *SCCmec*¹⁹⁶. Les souches de SASM ont toutefois la caractéristique d'avoir une répartition plus hétérogène, dans des lignées avec un fond génétique différent des principaux clones de SARM, tels que les CC7, CC9, CC12, CC15, CC25, CC51 et CC101¹⁹⁶.

Contrairement aux SARM hospitaliers, les SARM communautaires montrent eux une diversité génétique plus importante¹⁹⁴. Parmi les clones de SARM communautaires décrits, figure le clone épidémique isolé aux États-Unis, nommé USA300, appartenant au CC8^{195,197}. En effet, une étude américaine a montré parmi une collection de 249 SARM isolés dans 11 villes américaines que 98% des isolats appartenaient au clone USA300¹⁹⁸. Ce clone a la particularité de porter le gène codant la leucocidine de Pantone-Valentine et de causer des ITPM chez l'enfant¹⁹⁵.

A côté de ces souches de SARM hospitaliers et communautaires, ont également été identifiées des souches de SARM associées à des infections chez les animaux d'élevage, en particulier chez le porc et le veau¹⁹⁹. Ces souches sont génétiquement distinctes des isolats humains (CC1, CC9, CC97 et CC398) et constituent un important réservoir de SARM en dehors de l'hôpital. Le clone de SARM le plus répandu dans le monde est celui de CC398, surtout en Europe et en Amérique du Nord. Les souches de CC398 sont également capables de coloniser de provoquer des infections chez l'Homme travaillant au contact des animaux¹⁹⁹.

Des souches de SASM de CC398 ont également été identifiées chez l'Homme dont certaines sont responsables d'infections sévères telles que des bactériémies, des endocardites infectieuses et des pneumonies nécrosantes^{199,200}. L'analyse du génome de souches de SASM de CC398 d'origine humaine a montré qu'elles présentaient des gènes de phage et d'échappement au système immunitaire, absents des isolats issus de chez l'animal, et qui permettrait une adhésion accrue aux cellules épithéliales humaines¹⁹⁷.

4.1.3. MLST & SCN

4.1.3.1. Génotypage de l'espèce *S. epidermidis* par MLST

Chez *S. epidermidis*, il existe quatre schémas différents de MLST²⁰¹⁻²⁰³. Les trois premiers schémas utilisent principalement les mêmes gènes domestiques que le schéma décrit chez *S. aureus*^{201,202}. Cependant, il n'a jamais été possible de distinguer plus de deux CC malgré l'utilisation de plusieurs collections d'isolats de *S. epidermidis*. C'est pourquoi, en 2007, Thomas *et al.*²⁰³ ont évalué le pouvoir discriminant des *loci* des trois schémas de MLST et ont sélectionné les plus discriminants afin d'en définir un nouveau.

Les trois différents schémas ont été testés sur une collection de 66 isolats d'origines géographiques diverses, dont 46 provenant d'endocardites infectieuses sur valve prothétique et 20 de bactériémies. Tous schémas confondus, quatre à huit allèles différents par gène ont été identifiés avec des DI par *locus* de 0,172 à 0,744 respectivement. Le pouvoir résolutif de chacun des schémas a alors été défini par le calcul des DI. Pour les anciens schémas, des DI de respectivement, 0,871, 0,889 et 0,922 ont été obtenus permettant l'individualisation de 15, 23 et 25 ST regroupés entre un et trois CC. Le DI du schéma utilisant les *loci* les plus discriminants était de 0,956, associé à l'identification de 31 ST et cinq CC²⁰³.

Avec ce nouveau schéma, tous les ST, à l'exception de deux, regroupaient des isolats provenant de patients atteints soit d'endocardite soit de bactériémie. Le CC2 rassemblait les isolats de trois ST distincts : le ST2 comprenant des isolats responsables d'endocardites et de bactériémies, le ST28 comprenant un isolat responsable d'endocardite et le ST22 comprenant un isolat responsable d'une bactériémie. Au sein du CC2, le ST2 s'est révélé être le génotype ancestral du ST28 et du ST22. Ainsi, le ST2 pourrait être l'ancêtre commun

de deux lignées distinctes, une responsable d'endocardites infectieuses et l'autre responsable de bactériémies²⁰³.

Par ailleurs, une autre étude a montré que le ST2 était significativement plus associé à des souches responsables d'infections (49,2%) par rapport aux autres ST. En outre, il a été observé que les souches de ST2 étaient plus fréquemment identifiées à l'hôpital, suggérant que ce clone serait très adapté à l'environnement hospitalier et qu'il diffère des souches de *S. epidermidis* communautaires²⁰⁴.

A l'heure actuelle, 780 ST ont été référencés sur la base de données *PubMLST* par l'étude de 1229 isolats de *S. epidermidis*.

Des schémas de typage de MLST ont été mis au point pour d'autres espèces de SCN : *S. haemolyticus*²⁰⁵, *Staphylococcus hominis*²⁰⁶ et *Staphylococcus pseudintermedius*²⁰⁷, mais ils ne seront pas décrits ici.

4.1.3.2. Génotypage de l'espèce *S. lugdunensis* par MLST

Pour *S. lugdunensis*, le schéma a été mis au point plus récemment, en 2012, par Chassain *et al.*²⁰⁸ à partir de 87 isolats d'origines clinique (IPTM, IOA, bactériémie et infections sur matériel) et géographique diverses. Ce schéma porte sur l'analyse du polymorphisme de sept gènes domestiques dont trois sont communs à celui de *S. aureus*¹⁹³ : *aroE*, *gmk* et *yqiL*.

- *aroE* : shikimate deshydrogenase
- *dat* : D-amino acid aminotransferase
- *ddl* : D-alanine :D-alanine ligase
- *gmk* : guanylate kinase
- *ldh* : L-lactate dehydrogenase
- *recA* : recombinaise
- *yqiL* : acetyl coenzyme A acetyltransferase

Le séquençage de ces marqueurs a permis d'identifier entre quatre et neuf allèles distincts par *locus*, permettant d'individualiser 20 ST regroupés en cinq CC (**Figure 50**)²⁰⁸. Le CC le plus représenté était le CC1 (29/87), suivi du CC3 (20/87) et du CC2 (17/87). Le

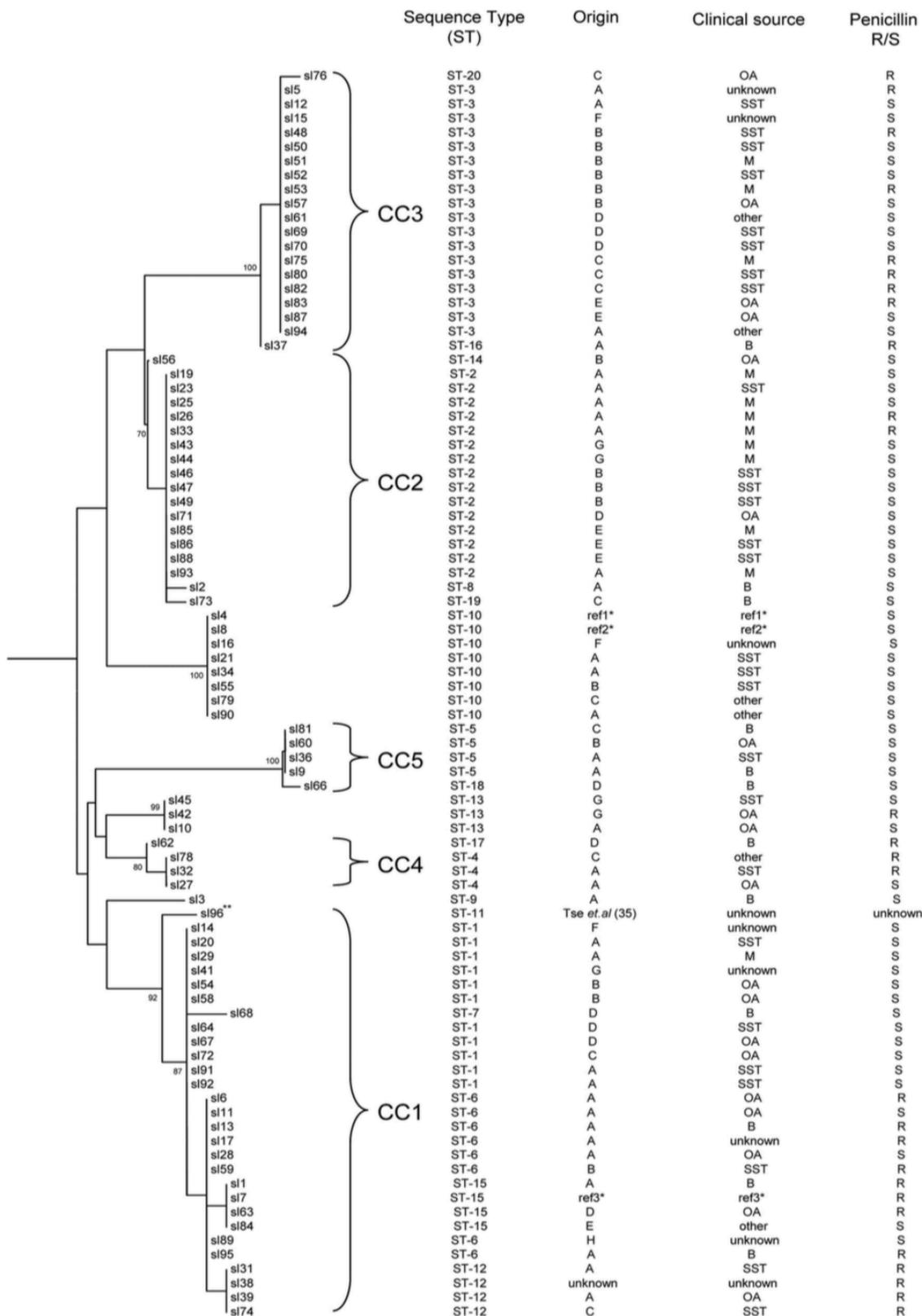


Figure 50 : Dendrogramme construit par la méthode *neighbor-joining* représentant les relations phylogénétiques entre les 87 isolats de *S. lugdunensis* caractérisés par *MultiLocus Sequence Typing*²⁰⁸.

Les complexes clonaux (CC) sont représentés par des accolades. L'origine clinique des souches est représentée par une abréviation : OA, infection ostéoarticulaire ; SST, infection de la peau et des tissus mous ; M, infection sur matériel ; B, bactériémie.

L'origine géographique des souches est indiquée par une lettre majuscule : A, Rouen ; B, Nantes ; C, Nancy ; D, Bordeaux ; E, Montpellier ; F, Versailles ; G, Louvain (Belgique) ; H, Maribor (Slovénie)

nombre plus faible de génotypes identifiés par rapport aux autres espèces de staphylocoques suggère une moins grande diversité génétique de l'espèce ou un pouvoir discriminant de la MLST inférieur à celui des schémas décrits chez *S. aureus* et *S. epidermidis*. Par ailleurs, aucune corrélation entre les ST ou les CC avec l'origine clinique ou géographique des isolats n'a été identifiée. Parmi ces 87 souches, 27 isolats étaient producteurs d'une pénicillinase, dont 13 répartis dans le CC1 (**Figure 50**)²⁰⁸. Par ailleurs, un déséquilibre de liaison significatif entre les allèles des *loci* étudiés a été mis en évidence par le calcul de l'IA (égal à 2,91), suggérant une structure de population clonale de l'espèce²⁰⁸.

Par la suite, une étude rétrospective a été conduite à Taïwan sur 118 isolats collectés de sites normalement stériles entre 2003 et 2013 : 105 isolés à partir d'hémocultures et 13 de liquides de ponction d'épanchement pleural, liquide d'ascite, liquide synovial, liquide céphalo-rachidien et liquide amniotique²³. En parallèle d'un typage de la cassette *SCCmec*, un typage par PFGE a été réalisé et a permis d'identifier trois groupes majeurs. Dans l'un de ces groupes, 14 souches étaient résistantes à la méticilline et possédaient la cassette de *SSCmec* de type II. De plus, ces 14 souches ont été isolées sur une période de plus de 40 mois dont dix dans un service d'urgence, trois dans des unités de soins intensifs et une dans un service de médecine générale. Un typage par MLST sur huit de ces 14 isolats a montré que tous appartenaient au ST6 du CC1. Cette étude est la première à montrer la diffusion et la persistance d'un clone endémique résistant à la méticilline de *S. lugdunensis* en milieu hospitalier à Taïwan²³.

En 2016, une autre étude rétrospective a été effectuée par la même équipe²⁴ à Taïwan. Elle a caractérisé cette fois par MLST 129 isolats de *S. lugdunensis* isolés entre 2003 et 2014. Sur ces 129 isolats, 81 étaient issus d'infections associées aux soins (IAS) et 48 d'infections communautaires. Ces derniers étaient plus fréquemment isolés de cas d'IPTM (85,4% vs 19,8%) alors que les isolats d'IAS étaient surtout responsables de bactériémies (34,6% vs 4,2%). Une résistance à l'oxacilline a été observée pour 27 isolats et une résistance à la pénicilline pour 87 des 129 isolats.

Le typage par MLST de ces isolats a permis d'identifier 12 ST répartis dans six CC distincts. Le CC1 était prédominant (43%), suivi par le CC3 (25,6%). De plus, tous les ST avaient une distribution similaire entre les IAS et les infections communautaires, excepté le ST38 (correspondant aujourd'hui au ST3) montré comme significativement plus

fréquemment associé à des IAS (33.3% vs 12.5%). Par ailleurs, parmi les 27 souches résistantes à l'oxacilline, la majorité appartenait au ST38 (77,8%), suivi du ST6 (22,2%). Le typage de la cassette *SCCmec* a montré que 17 des isolats de ST38 résistants à l'oxacilline avaient une cassette *SCCmec* de type V et cinq isolats de ST6 résistants à l'oxacilline avaient une cassette *SCCmec* de type II. Ainsi, cette étude montre l'émergence et la diffusion de clones résistants à la méticilline de ST38-*SCCmec* de type V et ST6-*SCCmec* de type II²⁴.

A l'heure actuelle, 42 ST ont été référencés par l'étude de 109 isolats de *S. lugdunensis* sur la base de données hébergée par l'Institut Pasteur.

Les études de MLST chez *S. lugdunensis* ne permettent donc pas de mettre en évidence d'association entre génotype et présentation clinique probablement du fait d'une évolution clonale, mais elles permettent toutefois d'observer une prévalence en France et en Asie du CC1 et du CC3. Ces travaux permettent également de suivre l'émergence et la diffusion de clones de *S. lugdunensis* endémiques résistants aux antibiotiques en Asie, où il a été montré une résistance à la méticilline plus fréquente (20% à 32%)^{21,26} que dans le reste du monde (8%)²⁰⁹.

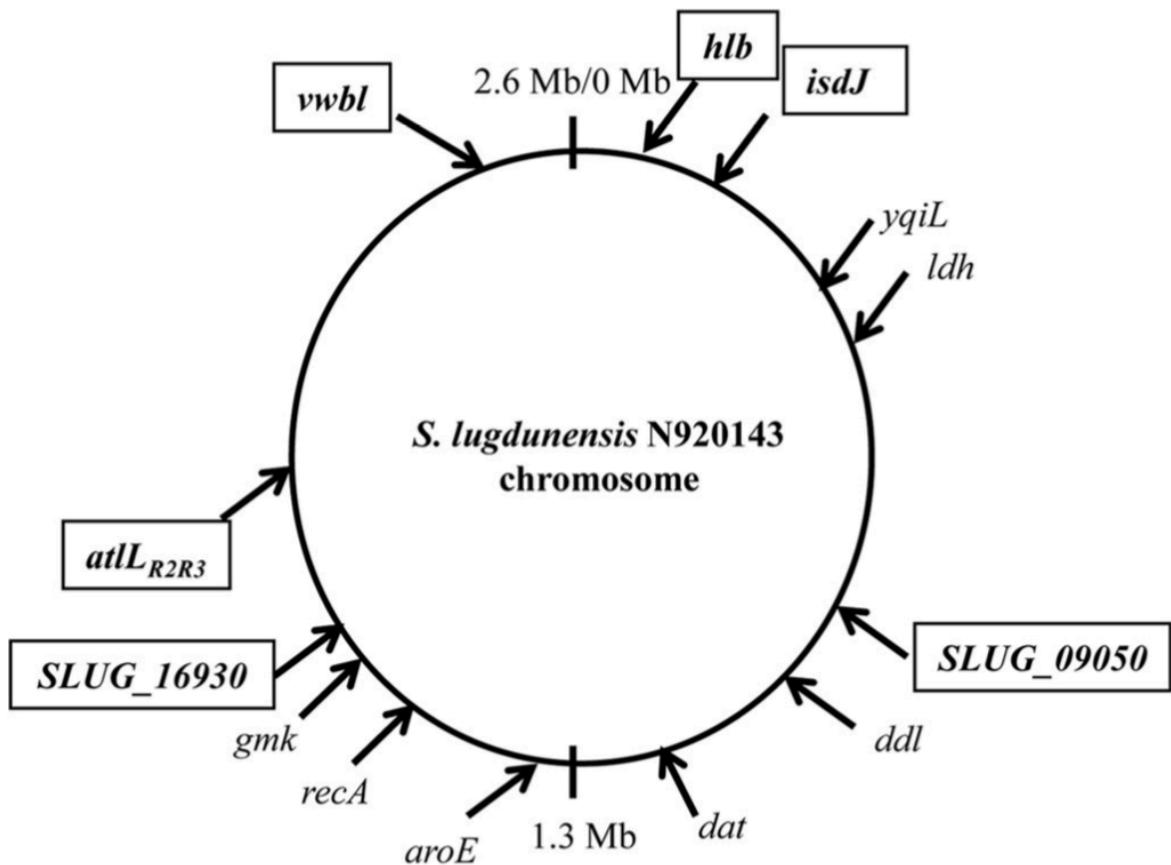


Figure 51 : Localisation des gènes inclus dans les schémas de *MultiLocus Sequence Typing* (MLST) et de *Multi-Virulence Locus Sequence Typing* (MVLST) sur le génome de la souche de *S. lugdunensis* N920143⁹⁸.

Les sept gènes de ménage de MLST sont en italique. Les sept gènes codant des protéines reconnues comme des facteurs de virulence ou des protéines potentiellement associées à la virulence sont encadrés.

4.2. L'analyse par Multi-Virulence Locus Sequence Typing (MVLST)

L'objectif du typage par MVLST est de mettre en évidence des lignées hypervirulentes associées à un contexte clinique sévère. Son principe repose sur celui de la MLST, excepté que son analyse est basée sur le polymorphisme de six à huit *loci* codant des facteurs de virulence reconnus ou des protéines potentiellement associées à la virulence. Le pouvoir résolutif de cette méthode repose sur le fait que ces gènes, plus soumis à la pression de l'environnement que les gènes domestiques, évoluent plus rapidement. Ici, les profils alléliques sont appelés des *Virulence Types* (VT) et peuvent être regroupés en Complexes Clonaux associés à la virulence (CC_v) s'ils partagent au moins cinq *loci* identiques⁹⁸.

Cette méthode de génotypage a été mise au point pour plusieurs espèces bactériennes à Gram positif, comme *Clostridium difficile*²¹⁰ ou *Listeria monocytogenes*²¹¹. Chez *S. aureus*, un schéma de MVLST a été développé par Verghese *et al.*²¹² en 2012 sur une collection de SARM préalablement caractérisés par typage de la cassette SCC_{mec}. Les gènes associés à la virulence choisis pour ce schéma sont décrits ci-dessous :

- *atl* : autolysine bifonctionnelle
- *essC* : protéine transmembranaire du SST7
- *geh* : lipase
- *hlgA* : γ -hémolysine
- *htrA* : sérine protéase
- *sdrC* : adhésine riche en sérine – aspartate (SD)

L'analyse par MVLST s'est montrée plus discriminante que la MLST puisqu'à partir des 107 isolats, 13 ST et 38 VT ont été identifiés. Elle a aussi permis de subdiviser les SARM au sein des CC définis par MLST et de distinguer les isolats responsables d'infections communautaires de ceux responsables d'IAS²¹².

Chez *S. lugdunensis*, le schéma de MVLST a été développé par Didi *et al.*⁹⁸ en 2014 sur la collection de 87 isolats cliniques utilisée pour la mise au point de la MLST²⁰⁸. A ces derniers, ont été ajoutées six souches isolées en simple condition de portage. Le typage par MVLST a consisté en l'analyse des sept gènes codant des facteurs de virulence ou des protéines potentiellement associées à la virulence suivants (**Figure 51**) :

- *atlLR2* et *atlLR3* : domaines putatifs d'ancrage de l'autolysine AtIL à la paroi

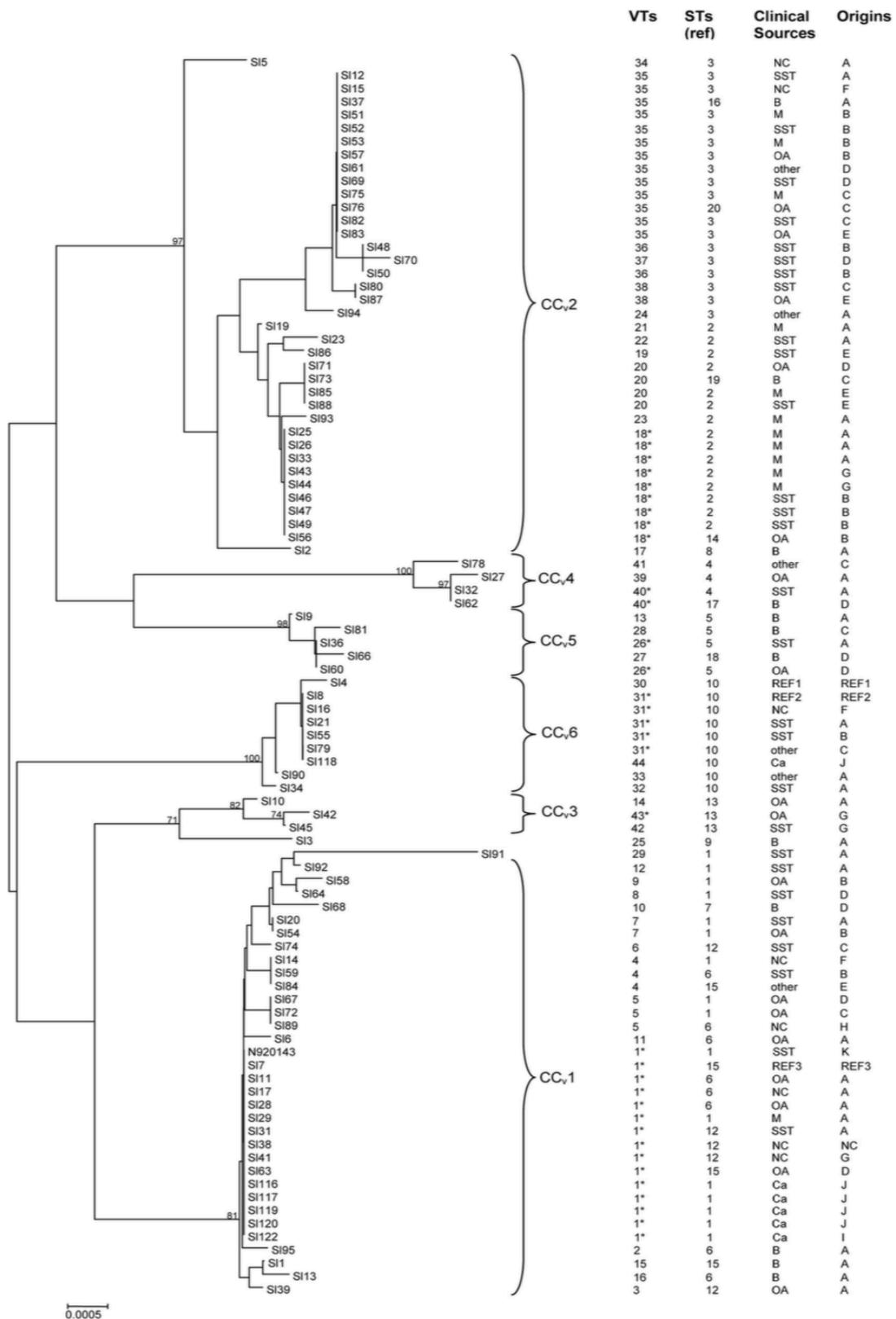


Figure 52 : Dendrogramme construit par la méthode *neighbor-joining* représentant les relations phylogénétiques entre les 94 isolats de *S. lugdunensis* caractérisés par *Multi-Virulence Locus Sequence Typing*⁹⁸.

Les complexes clonaux associés à la virulence (CC_v) sont représentés par des accolades. L'origine clinique est représentée par une abréviation : OA, infection ostéoarticulaire ; SST, infection de la peau et des tissus mous ; M, infection sur matériel ; B, bactériémie ; Ca, portage. L'origine géographique des souches est indiquée par une lettre majuscule : A, Rouen ; B, Nantes ; C, Nancy ; D, Bordeaux ; E, Montpellier ; F, Versailles ; G, Louvain (Belgique) ; H, Maribor (Slovénie) ; I, Tours ; J, Kronoberg (Suède) ; K, Lyon.

- *hly* : hémolysine putative
- *isdJ* : protéine putative impliquée dans la captation du fer
- *SLUG_09050* : hémolysine III putative
- *SLUG_16930* : protéine putative de liaison au fibrinogène et à la fibronectine
- *vwbl* : protéine de liaison au vWF

Pour les 93 isolats d'origines clinique et géographique différentes, 43 VT répartis en six CC_vs ont été décrits. Le schéma de MVLST est donc plus discriminant que celui de MLST, qui avait mené à l'identification de 20 ST (**Figure 52**). Les quatre VT majeurs étaient partagés par cinq à 14 isolats, et 32 des 43 VT étaient représentés par un seul isolat, soulignant la diversité génétique de la population étudiée. De plus, tout comme pour la MLST, le calcul de l'index d'association (IA = 3,740) confirmait le caractère clonal de la population de *S. lugdunensis*. Cet index suggère une faible fréquence voire une absence de recombinaison des gènes associés à la virulence. Au total, la méthode de MVLST confirme la structure clonale de la population de *S. lugdunensis* et suggère une co-évolution des gènes de virulence avec les gènes domestiques⁹⁸.

Un schéma *trilocus*, simplifié, se basant uniquement sur l'analyse de trois *loci*, (*atlL_{R3}*, *isdJ* et *SLUG_16930*) s'est révélé aussi discriminant que le schéma de MVLST à 7 gènes avec un DI de 0,922 pour le schéma *trilocus* contre 0,943 pour le schéma complet⁹⁸.

Tout comme avec le schéma de MLST, aucune corrélation entre l'origine clinique et géographique des isolats et les VT n'a été observée. Enfin, aucune lignée hypervirulente, ni aucun génotype spécifique des souches de portage n'a été identifié. Toutefois, l'analyse de MVLST portait seulement sur l'étude de six souches de portage⁹⁸.

Ainsi, il existe à ce jour deux méthodes de typage et de suivi macro-épidémiologique des souches de *S. lugdunensis*, la MLST et la MVLST, basées sur l'analyse de sept *loci* pour chacune d'elles, avec la possibilité pour la MVLST d'utiliser un schéma *trilocus* résolutif. Néanmoins, aucun schéma de typage n'a permis d'associer un génotype à un tableau clinique, et ce d'autant plus que les souches invasives et celles isolées en simple condition de portage (mais seulement au nombre de six) étaient associées au sein des mêmes complexes clonaux.



Figure 53 : Organisation générale d'un *Variable Number Tandem Repeat* (VNTR)²¹⁴.

Les rectangles bleus représentent les motifs répétés qui composent le VNTR. Au niveau des régions 5' et 3' sont représentés les régions flanquantes conservées. Les flèches au niveau de ces régions représentent les sites d'appariement des amorces pour l'amplification par PCR.

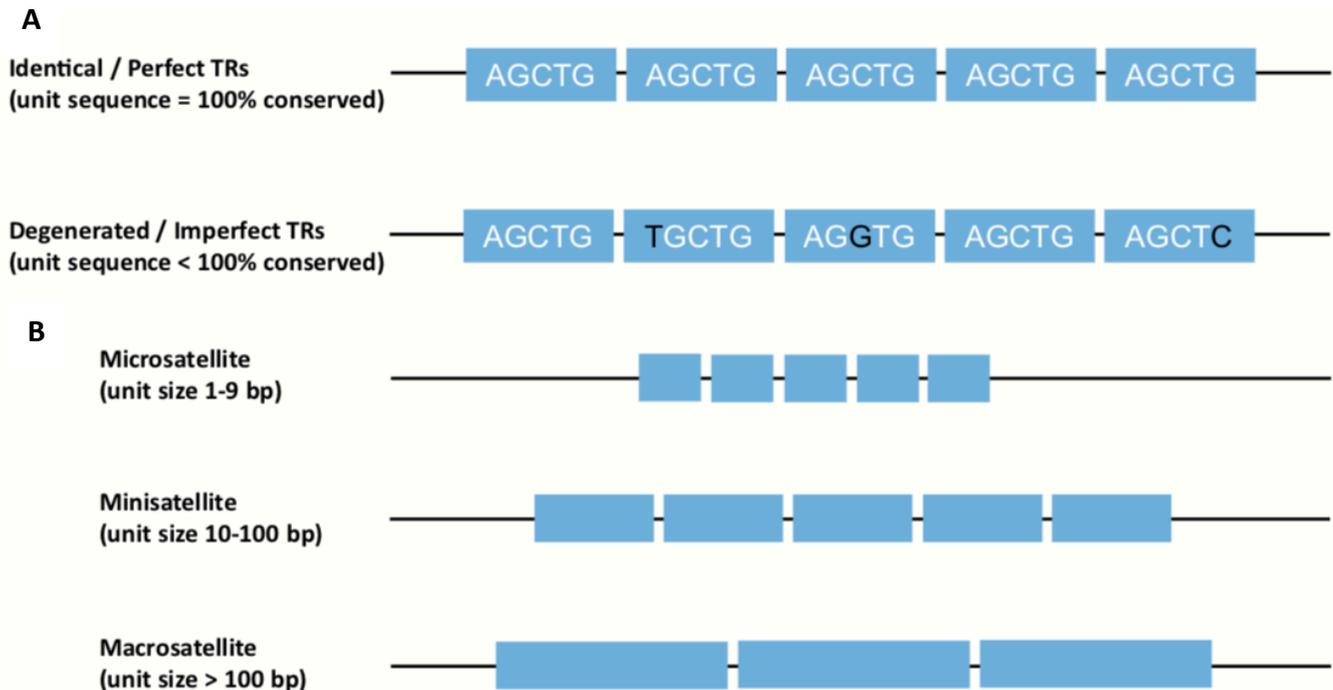


Figure 54 : Représentation des différents types de répétitions en tandem (TR) (A) et des différents types de *Variable Number Tandem Repeat* (VNTR) (B)²¹³.

A, Le premier VNTR présente des TR conservés (parfaits) alors que le second présente des TR dégénérés (imparfaits) avec des variations.

B, Taille des TR en fonction des différents types de VNTR.

4.3. Le génotypage basé sur l'étude des *Variable Number Tandem Repeats*

Dans le génome bactérien, une source intéressante de polymorphisme (de taille et de séquence) est fournie par des séquences répétées en tandem, appelées VNTR. Les VNTR sont des séquences nucléotidiques de taille variable et de localisation inter- ou intragénique²¹³. Les répétitions en tandem (TR) qui composent les VNTR sont des séquences nucléotidiques qui sont directement répétées les unes à la suite des autres. En amont de la région 5' et en aval de la région 3', sont présentes des régions conservées, appelées régions flanquantes qui encadrent le VNTR (**Figure 53**). Selon le degré d'homologie des séquences répétées, les TR sont classés en deux catégories : les TR conservées à 100%, sont dites TR parfaites ou identiques ; les TR présentant des variations de séquence sont appelées TR imparfaites ou dégénérées (**Figure 54A**)²¹³. Par ailleurs, les VNTR sont classés en trois catégories selon leur taille (**Figure 54B**)²¹³ :

- Les micro-satellites présentant des TR d'une taille comprise entre 1 et 9 pb
- Les mini-satellites présentant des TR d'une taille allant de 10 à 100 pb
- Les macro-satellites présentant des TR d'une taille supérieure à 100 pb.

Le polymorphisme de ces régions hypervariables est lié d'une part à un polymorphisme de taille et d'autre part à un polymorphisme de séquence. Cette variabilité est due à l'instabilité des VNTR, conséquence du glissement de l'ADN polymérase (*strand-slippage mispairing*) au niveau des TR de petite taille lors de la réplication. Ce phénomène conduit à l'augmentation ou à la réduction d'un nombre de copies de TR. Pour les TR de grande taille, le polymorphisme des VNTR serait le fruit de phénomènes de recombinaisons homologues²¹³.

Cette variabilité génétique fait des VNTR des marqueurs de choix pour discriminer des individus au sein d'une même espèce bactérienne. Ce polymorphisme est exploité *via* une méthode de typage nommé *Multiple-Loci Variable Number Tandem Repeat (VNTR) Analysis*.

4.3.1. Le génotypage par Multiple-Loci VNTR Analysis

4.3.1.1. Principe de la méthode

La MLVA est une méthode de typage qui exploite le polymorphisme de taille des VNTR, c'est-à-dire la variation du nombre de répétitions de ces marqueurs entre les isolats d'une même espèce. Le principe de cette technique repose sur l'amplification par PCR de plusieurs VNTR situés à différentes localisations sur le génome. Il est possible d'amplifier un VNTR par l'utilisation d'amorces positionnées au niveau des régions flanquantes. Dès lors, la taille totale de l'amplicon peut être déterminée par électrophorèse capillaire ou sur gel d'agarose. Ainsi, connaissant la taille du motif répété et la taille des régions flanquantes en amont et en aval du VNTR, il est possible de déduire le nombre de motifs répétés selon la formule suivante :

$$\text{Nombre de TR} = \frac{\text{Taille de l'amplicon (pb)} - \text{Taille des régions flanquantes 5' et 3' (pb)}}{\text{Taille des motifs répétés (pb)}}$$

Une fois le nombre de TR déterminé pour chaque VNTR, il est possible de définir des profils *multilocus*, dont les numéros correspondent au nombre de TR de chaque *locus*. Chaque code numérique correspond à un profil unique appelé *MLVA-type* (MT). Des MT peuvent être regroupés en *MLVA-Complexes* (MC), s'ils ont le même nombre de TR pour six des sept *loci* étudiés.

La mise au point d'un schéma de MLVA commence par l'identification *in silico* de VNTR. Elle nécessite l'utilisation du logiciel *Tandem Repeat Finder* (TRF) qui permet d'identifier des séquences répétées en tandem au sein de génomes complets²¹⁴. Il est alors nécessaire d'évaluer *in silico* le polymorphisme des VNTR par comparaison de plusieurs génomes de la même espèce bactérienne. Les marqueurs présentant des variations suffisamment importantes sont alors testés *in vitro* sur un panel de souches bien caractérisées et représentatifs de la diversité de la population de l'espèce étudiée.

Ainsi, un VNTR est validé comme un bon candidat pour le schéma de MLVA (i) s'il est présent chez tous les isolats testés, (ii) s'il présente une taille supérieure à 5 pb pour un motif répété, (iii) s'il ne présente pas d'insertion ou de délétion dans les motifs répétés, (iv) si les régions flanquantes sont conservées entre tous les isolats testés. De plus, les amorces nécessaires à son amplification doivent être situées le plus proche possible du VNTR²¹⁴.

La MLVA est une méthode qui fait preuve d'un important pouvoir résolutif chez plusieurs espèces bactériennes. Chez les bactéries à Gram négatif comme *Salmonella* par exemple, la MLVA est utilisée comme un outil puissant de surveillance épidémiologique au niveau mondial^{215,216}. Chez les bactéries à Gram positif, des schémas de MLVA ont été adaptés à plusieurs espèces bactériennes : *L. monocytogenes*²¹⁷, *S. agalactiae*²¹⁸, *S. aureus*²¹⁹, *S. epidermidis*²²⁰ ...

4.3.1.2. MLVA & *S. aureus*

Le premier schéma de MLVA dédié à l'espèce *S. aureus* a été développé par Sabat *et al.*²¹⁹ en 2003. La méthode a été adaptée d'une technique de typage moléculaire rapide basée sur le polymorphisme de taille de fragments amplifiés après digestion enzymatique des gènes codant la protéine A, la coagulase libre et les régions hypervariables en amont du gène *mecA*^{221,222}.

Par ailleurs, le séquençage de génomes de *S. aureus* a permis l'identification de la présence de VNTR intra-géniques. C'est pourquoi, Sabat *et al.*²¹⁹ ont développé un schéma de MLVA basé sur l'étude de cinq VNTR localisés dans les gènes suivant :

- *clfA* : *clumping factor A*
- *clfB* : *clumping factor B*
- *sdrCDE* : locus codant les adhésines riches en répétitions SD, SdrC, D et E
- *sspA* : sérine protéase V8
- *spa* : protéine A de liaison au fibrinogène

L'analyse des VNTR a montré des TR d'une taille allant de neuf à 24 pb par électrophorèse sur gel d'agarose à partir d'une collection de 34 isolats de SARM et a permis l'identification de 26 profils de MLVA. L'analyse des profils générés a montré que le MLVA était capable de distinguer les isolats d'origine géographique différente tout en regroupant les isolats épidémiologiquement liés dans des mêmes génotypes ou des génotypes proches. Cette étude a en outre prouvé que le pouvoir discriminant de la MLVA était comparable à celui de la technique de typage par électrophorèse en champ pulsé²¹⁹. Toutefois la MLVA présente aussi l'avantage d'être une méthode sensible, simple et permettant l'analyse rapide d'un

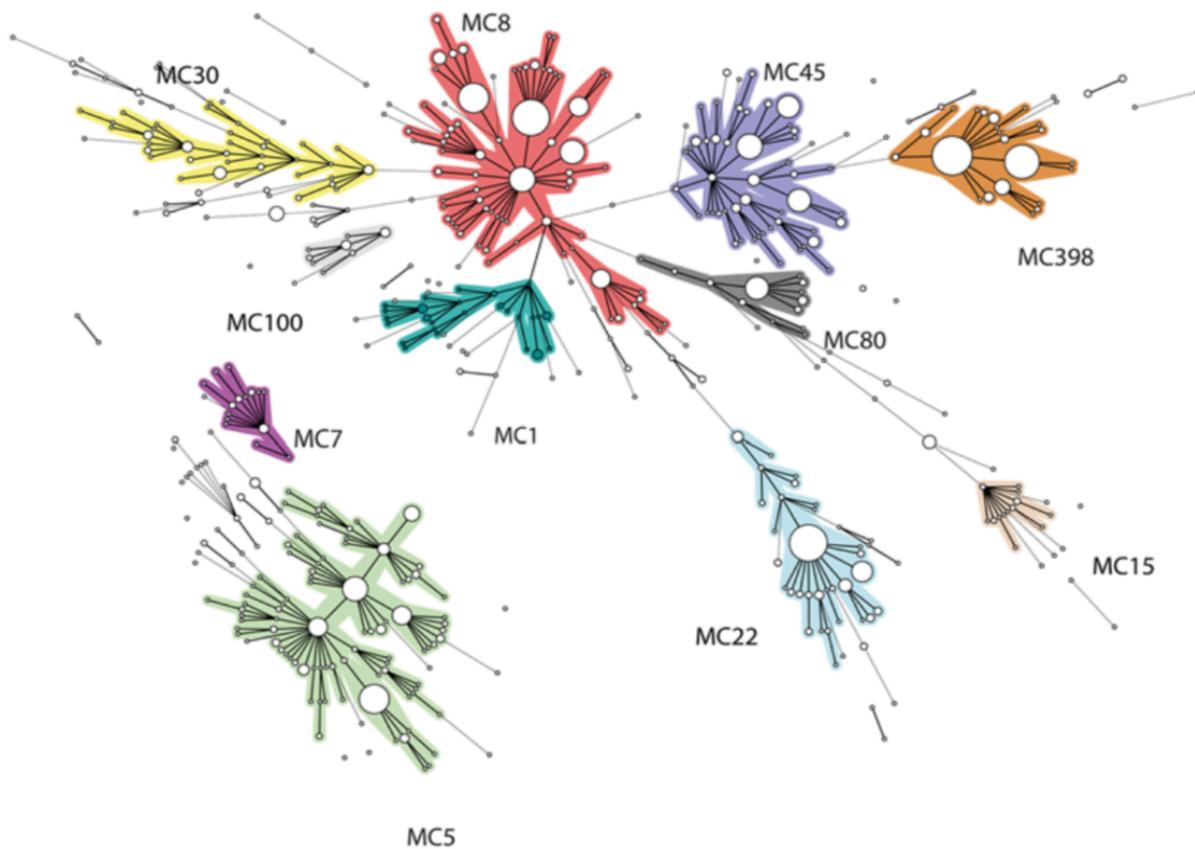


Figure 55 : *Minimum Spanning Tree* des 1681 isolats de *S. aureus* typés par *Multiple Locus Variable Number Tandem Repeat (VNTR) Analysis (MLVA)*²²⁷.

Chaque cercle représente un MLVA *Type* (MT) et la taille du cercle est proportionnelle au nombre d'isolats de ce MT. Les zones de couleur autour des cercles délimitent les MLVA complexes. Les lignes épaisses relient deux MT qui diffèrent par un seul VNTR alors que les lignes fines relient deux MT qui diffèrent de deux VNTR.

plus grand nombre d'isolats avec la possibilité de produire des données exportables et échangeables depuis l'utilisation de l'électrophorèse capillaire.

Il existe plusieurs schémas de MLVA chez *S. aureus* utilisant de six à 12 VNTR différents²²³⁻²²⁷. L'inconvénient de certains de ces schémas est qu'ils utilisent des VNTR avec des TR d'une taille trop faible pour une détermination précise du nombre de motifs répétés par électrophorèse sur gel d'agarose. Par ailleurs, des amplifications aspécifiques ont été obtenues pour certains *loci* comme le *locus sdr* rendant l'assignation de profils difficile pour certaines souches. Ces difficultés rendent alors la méthode peu reproductible et peu fiable en produisant des données de typage ambiguës²²⁷.

C'est pourquoi en 2009, Schouls *et al.*²²⁷ ont utilisé les VNTR présents dans neuf génomes de *S. aureus* pour mettre au point un nouveau schéma de MLVA. L'analyse *in silico* à l'aide du logiciel TRF a mené à l'identification de plus de 20 VNTR, dont la majorité présentaient des TR imparfaits. Ces VNTR ont été testés sur un panel comprenant des souches de *S. aureus* de référence dont le génome a été séquencé. Dès lors, n'ont été conservés que les VNTR (i) amplifiables par PCR chez toutes les souches, (ii) ne présentant pas d'amplification aspécifique et (iii) présentant au moins trois, mais pas plus de dix allèles différents²²⁷. Au final, huit VNTR ont été conservés dont cinq localisés dans des régions non-codantes (VNTR61_01, VNTR61_02, VNTR67_01, VNTR21_01 et VNTR63_01) et trois localisés dans le gène *sspA* codant la cystéine protéine SspA, le gène *spa* codant la protéine A et le gène *coa* codant la coagulase libre.

Ce schéma a alors été appliqué à une collection de 1681 isolats de *S. aureus* collectées chez des patients néerlandais. Le nombre de répétitions entre les différents *loci* allait de cinq (VNTR21_01) à 25 répétitions (*sspA*), générant un pouvoir discriminant variable selon les VNTR, souligné par des valeurs de DI très hétérogènes variant de 0,342 (VNTR21_01) à 0,830 (*spa*)²²⁷. Ainsi, les 1681 isolats ont été individualisés en 511 MT répartis en 11 MC (**Figure 55**). Ces résultats ont été comparés à ceux des méthodes de typage de PFGE et de *spa-typing* (**cf paragraphe 4.3.2**). Il a été montré que la MLVA était aussi discriminante que la méthode de PFGE avec des DI similaires de 0,985 et de 0,977 respectivement. Le DI de la méthode de *spa-typing* était légèrement inférieur aux deux autres méthodes puisque égal à 0,962.

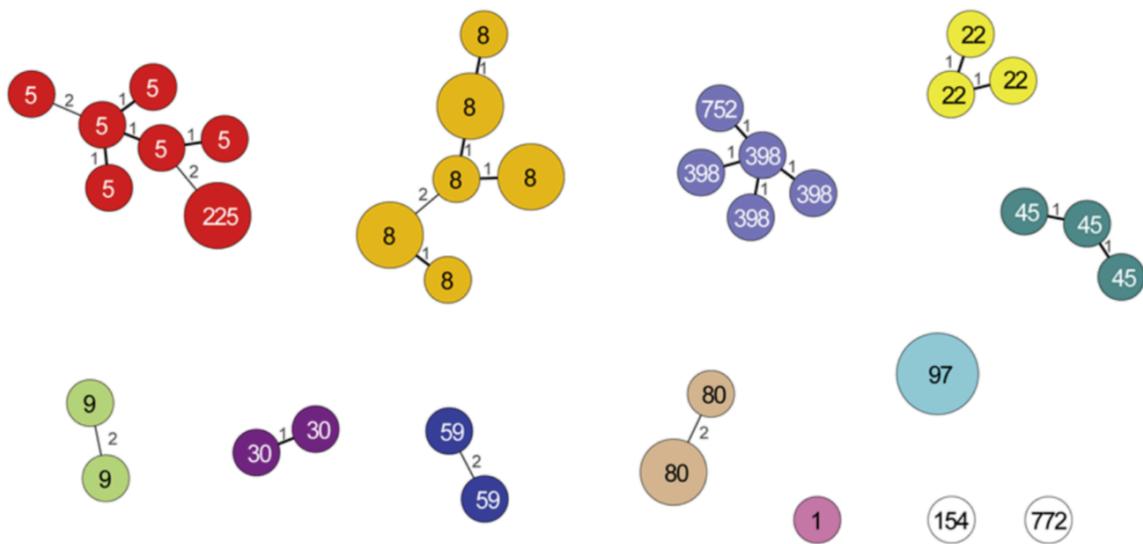


Figure 56 : *Minimum Spanning Tree (MST)* basé sur les données de typage de *Multiple Locus Variable Number Tandem Repeat (VNTR) Analysis* pour 40 isolats de *S. aureus*²²⁷.

Le MST illustre la concordance des génotypes obtenus par MLVA et *MultiLocus Sequence Typing (MLST)*. Chaque cercle représente un MLVA *Type (MT)* et la taille du cercle est proportionnelle au nombre d'isolats de ce MT. Le nombre inscrit dans chaque cercle représente le *Sequence Type (ST)* et les couleurs représentent les complexes clonaux (CC) obtenus par MLST. Les cercles blancs représentent des ST qui ne font pas partis d'un CC. Seuls les MT qui ne varient que par un ou deux *loci*, sont reliés par un trait.

Bien que discriminante, la MLVA était congruente avec les autres méthodes de typage. Le niveau de congruence peut être défini par le coefficient de corrélation de Pearson. Ce coefficient définit la probabilité que deux isolats de même génotype, défini par une première méthode, soient aussi de même génotype avec une seconde méthode de typage. Dans cette étude, les résultats de la méthode de MLVA étaient congruents avec ceux de PFGE à 77%, c'est-à-dire que deux isolats avaient une probabilité de 77% d'être de même génotype avec les deux méthodes. Les auteurs ont alors évalué la congruence des résultats de typage de MLVA et de MLST. Pour ce faire, les résultats de typage de 40 isolats ont été comparés. A partir d'un arbre phylogénétique de type *Minimum Spanning Tree* (MST) basé sur les données de MLVA, la répartition des isolats a montré un haut degré de congruence avec les CC. En effet, les ST appartenant à un même CC étaient aussi regroupés par MLVA (**Figure 56**). Par ailleurs, le coefficient de corrélation de Pearson a montré une congruence de 86,5% entre les résultats de MLST et de MLVA. Ainsi, bien qu'elle soit congruente avec la MLST, la MLVA reste plus discriminante et permet de subdiviser les CC²²⁷.

Schouls *et al.*²²⁷ ont aussi évalué la concordance épidémiologique de la méthode de MLVA sur une collection de 658 d'isolats de SARM provenant de différents hôpitaux sur une période de 10 mois. Le typage par MLVA s'est montré capable de regrouper des isolats d'un même hôpital en un même MT ou en plusieurs MT proches. Ainsi, cette méthode de typage est fiable pour identifier des isolats épidémiologiquement liés et constitue un outil de micro-épidémiologie pour étudier la dissémination des souches de SARM dans des contextes d'épidémies.

Aujourd'hui, c'est le schéma de Schouls *et al.*²²⁷ qui prévaut grâce à une standardisation de la méthode permettant une reproductibilité des résultats entre laboratoires. En effet, jusqu'en 2009 la méthode de MLVA utilisait l'électrophorèse sur gel d'agarose pour déterminer la taille totale du fragment amplifié et déduire le nombre de TR. Ainsi, cette équipe a mis au point un schéma utilisant deux PCR multiplex permettant chacune l'amplification de quatre VNTR à l'aide de quatre sondes fluorescentes différentes. Ces marqueurs fluorescents sont utilisés pour déterminer la taille précise de chaque VNTR par électrophorèse capillaire. Les protocoles et la procédure détaillée de la méthode sont disponibles sur le site (<https://www.mlva.net/saureus>) où il est également possible de déterminer les MT des isolats testés en fonction des profils obtenus.

4.3.1.3. MLVA & SCN

Comme pour la méthode de MLST, des schémas de MLVA existent pour plusieurs espèces de SCN comme *S. epidermidis*²²⁰, *S. haemolyticus*²⁰⁵ et *S. hominis*²²⁸.

Pour *S. epidermidis*, la technique de MLVA a été développée par Johansson *et al.*²²⁰ pour étudier une population de 30 souches multi-résistantes aux antibiotiques isolées à l'hôpital universitaire d'Umeå en Suède.

L'analyse du génome de la souche ATCC12228 avec le logiciel TRF a permis d'identifier 137 VNTR. Au final, seulement cinq ont été conservés, présentant au moins deux motifs répétés et une taille minimum du motif répété de 18 pb. Un seul VNTR était inter-génique, les quatre autres étaient localisés dans des gènes dont trois codant des protéines de liaison au fibrinogène de la famille des adhésines riches en répétition SD²²⁰.

Afin de juger du pouvoir discriminant des cinq VNTR, un typage par électrophorèse en champ pulsé a également été réalisé. La congruence entre les deux méthodes était remarquable puisque 16 génotypes ont été identifiés pour chacune d'elles avec une répartition identique des isolats.

L'analyse de la dissémination de génotypes multi-résistants de *S. epidermidis* à l'hôpital d'Umeå a montré que de nombreux isolats liés épidémiologiquement étaient répandus dans plusieurs services hospitaliers. En effet, lors de cette étude, il a été observé une association entre le MT prédominant et des isolats responsables d'infections profondes de cicatrices post-opératoires après implantation d'un dispositif médical. Par ailleurs, le génotype majoritaire a été identifié à partir d'isolats prélevés dans plusieurs services de chirurgie thoracique, de chirurgie-ophtalmologie et de chirurgie orthopédique⁷⁵. Ces résultats suggèrent qu'un clone de *S. epidermidis* a persisté à l'hôpital et que ce clone a diffusé au sein des services pour coloniser les patients.

En conclusion, la MLVA est une méthode de génotypage qui produit des données non-ambiguës et échangeables pouvant être comparées entre les laboratoires, contrairement à la méthode de typage par électrophorèse en champ pulsé. Par ailleurs, elle a l'avantage, par rapport à la MLST, d'être rapide d'utilisation et moins coûteuse, ne nécessitant pas l'utilisation du séquençage. En outre, à nombre de *loci* comparable, la MLVA permet d'augmenter de façon significative le nombre de génotypes obtenus par son haut pouvoir

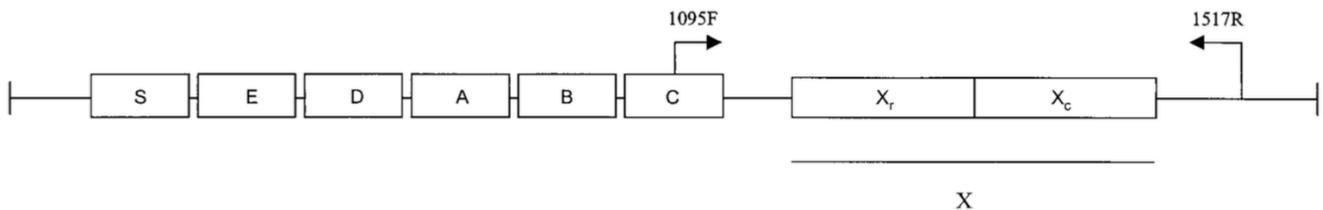


Figure 57 : Représentation des domaines composant le gène codant la protéine A de *S. aureus*²³⁴.

Les rectangles représentent les portions du gène *spa* qui codent le peptide signal (S), le domaine de liaison à l'immunoglobuline G (A-D), une région homologue à la région codant le domaine de liaison à l'immunoglobuline G (E), la région C-terminale (X) qui comprend la région composée de motifs répétés de 24 pb (X_r) et le motif d'ancrage à la paroi (X_c). Les flèches 1095F et 1517R représentent les sites d'appariement des amorces *forward* et *reverse* utilisées pour le *spa*-typing.

résolutif. Par ailleurs, l'utilisation de l'électrophorèse capillaire fait de cette technique une méthode très fiable et reproductible, surtout pour des études micro-épidémiologiques.

Il existe une autre méthode de typage bactérien basée sur le polymorphisme des VNTR et notamment celui des TR dégénérés. Cette technique appelée *Tandem Repeat Sequence Typing* (TRST) consiste, en plus de la détermination du nombre de motifs répétés, à déterminer la séquence nucléotidique par séquençage Sanger de chacun des motifs répétés du VNTR. Cette méthode a été appliquée chez différentes espèces bactériennes telles que *Salmonella sp*²¹⁵, *Streptococcus pneumoniae*²²⁹, *C. difficile*²³⁰, *Legionella pneumophila*²³¹ et *Francisella noatunensis*²³², par séquençage d'un ou plusieurs *loci* et a révélé un haut pouvoir discriminant.

4.3.2. Typage de *S. aureus* par *spa*-typing

Une méthode de typage basée sur le séquençage du gène *spa*, nommée *spa*-typing, est très utilisée chez *S. aureus*. C'est en 1996 que Frénay *et al.*²³³ proposent le séquençage du VNTR localisé dans le gène *spa*, codant la protéine A, pour le typage d'isolats de *S. aureus* résistants à la pénicilline.

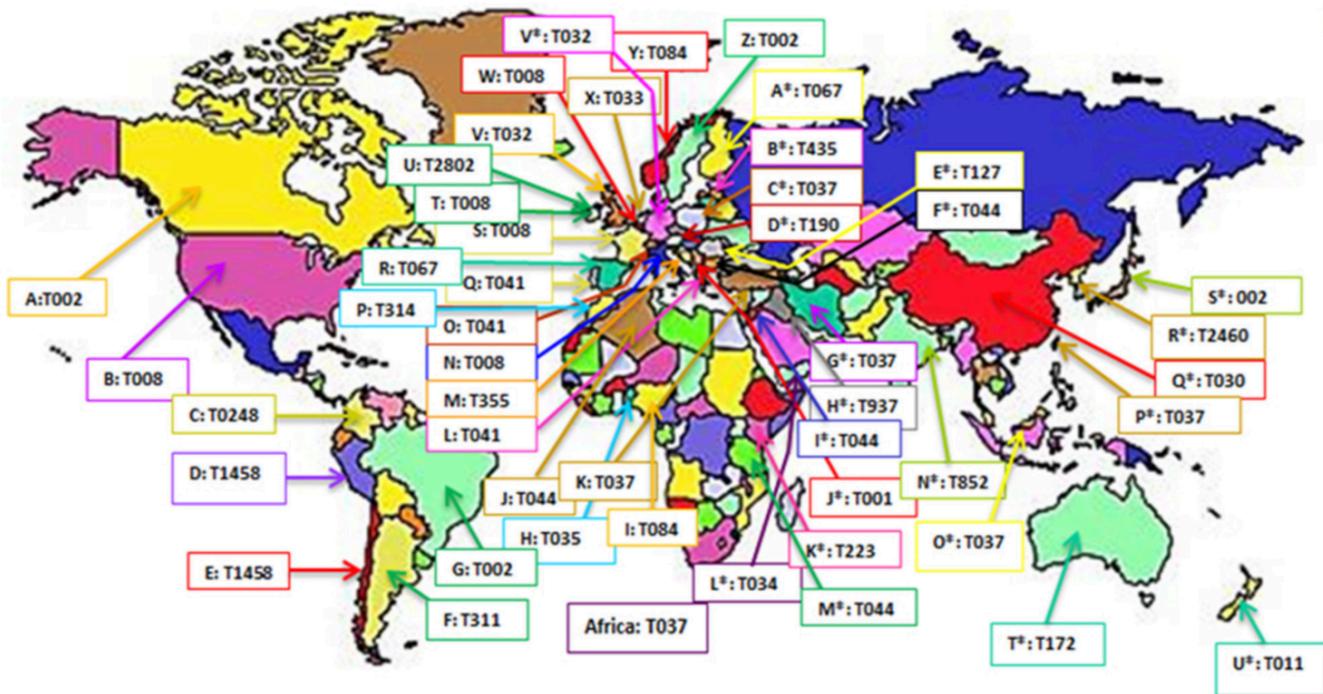
Le gène *spa*, de taille variable, n'est présent que chez *S. aureus*. La protéine A de *S. aureus* est une MSCRAMM qui se lie aux immunoglobulines pour échapper à la phagocytose. Le gène *spa* se compose d'une région codant un domaine de liaison au fragment constant de la chaîne lourde de l'immunoglobuline G, suivi de la région X constituée de motifs répétés en nombre variable et enfin un motif d'ancrage à la paroi (**Figure 57**)²³⁴. Le séquençage de la région X a révélé la présence de motifs répétés hautement polymorphes dont la grande majorité font une taille de 24 pb. La diversité de cette région est le produit de duplications et de délétions de ces motifs répétés, de recombinaison intra-génique, mais aussi de SNP²³⁴.

Malgré cette diversité, la région X semble suffisamment stable dans le temps pour être utilisée comme un marqueur épidémiologique. En effet, le séquençage de la région X a montré que des souches cultivées *in vitro* pendant dix jours consécutifs sur gélose au sang de mouton conservaient la même séquence et le même nombre de répétitions. Par ailleurs, des résultats identiques ont été observés à partir de plusieurs isollements, sur une période de cinq ans, d'une souche de SARM persistant chez un patient²³³.

Pour réaliser une amplification par PCR et un séquençage Sanger de ce VNTR, des amorces positionnées au niveau des régions flanquantes conservées de la région X ont été utilisées. Le séquençage permet de déterminer le nombre de motifs répétés mais aussi de déterminer la séquence de chacun de ces motifs de 24 pb. Pour chaque séquence unique d'un motif, un numéro est assigné (01, 02, 03 ...). Ainsi, à chaque combinaison unique de motifs répétés (26-23-17-34-17-20-17-12-17-16) est attribué un numéro de génotype appelé *spa-type* (t002)²³⁵. Une fois le séquençage effectué, il est possible de déterminer le *spa-type* d'un isolat, mais également le CC de MLST correspondant, à partir de la base de données "Ridom StaphType" (<http://www.ridom.de/staphtype>). Cette base de données regroupe à ce jour 762 motifs de répétitions différents et 17897 *spa-types*. De plus, cette méthode standardisée produit des résultats reproductibles facilement interprétables, non-ambigus et exportables.

Cette méthode permet de prédire le CC voire le ST auquel un isolat appartient car il existe une excellente congruence avec la méthode de MLST²³⁶. Il a été calculé un pourcentage de congruence de 72,9 à 96,8% entre les génotypes identifiés selon les études^{236,237}. Bien que le *spa-typing* se base sur le polymorphisme d'un unique *locus*, il permet tout de même d'obtenir un excellent pouvoir discriminant par rapport aux méthodes utilisant plusieurs *loci*²³⁸. Le *spa-typing* permet d'obtenir un pouvoir discriminant (DI = 0,913) supérieur à celui de la méthode de MLST (DI = 0,887)²³⁹. Ce qui fait du *spa-typing* un outil de première ligne pour la surveillance mondiale des infections à *S. aureus* mais aussi pour les études micro-épidémiologiques^{194,236,240}.

L'inconvénient de cette méthode est qu'elle puisse entraîner une classification erronée des isolats du fait de phénomènes de recombinaison et/ou de l'homoplasie, c'est-à-dire la similitude pour un caractère donné entre deux isolats, indépendante d'un ancêtre commun²⁴¹. Les allèles définis par l'étude de micro-satellites se réfèrent à des fragments d'ADN de taille différente entre les isolats. Des isolats qui ont une taille identique pour un *locus* ne sont pas nécessairement liés phylogénétiquement. Cet état identique peut être le résultat de mutations convergentes au niveau d'ancêtres communs ; il s'agit de l'homoplasie de taille²⁴¹. Ainsi, des isolats présentant une taille identique pour le gène *spa* pourraient présenter le même nombre de motifs répétés sans avoir de lien phylogénétique. Cependant, l'analyse des polymorphismes nucléotidiques des motifs répétés tend à surpasser cette problématique, les SNP étant moins sujets à l'homoplasie²⁴¹.



A: Canada, B: USA, C: Colombia, D: Peru, E: Chile, F: Argentina, G: Brazil, H: Ghana, I: Nigeria, J: Algeria, K: Turkey, L: Serbia, M: Croatia, N: Italy, O: Austria, P: Morocco, Q: Portugal, R: Spain, S: France, T: Ireland, U: Iceland, V: UK, W: Belgium, X: Netherlands, Y: Norway, Z: Sweden, A*: Finland, B*: Latvia, C*: Poland, D*: Austria, E*: Romania, F*: Greece, G*: Iran, H*: Iraq, I*: Lebanon, J*: Bosnia and Herzegovina, K*: Kenya, L*: Oman, M*: Tanzania, N*: India, O*: Malaysia, P*: Taiwan, Q*: China, R*: Korea, S*: Japan, T*: Australia, U*: New Zealand, V*: Germany

Figure 58 : Distribution des *spa*-types de *S. aureus* les plus prévalents à travers le monde²³⁸.

Continent	The most common <i>spa</i> types (No.): No. of MRSA and/or MSSA isolates		
Europe	t032 (1,250): 1250 MRSA	t008 (899): 510 MRSA, 229 MSSA, 90 uncertified	t002 (794): 450 MRSA, 100 MSSA, 244 uncertified
Asia	t030 (1,748): 1686 MRSA, 51 MSSA, 11 uncertified	t037 (1,467): 1415 MRSA, 51 MSSA, 1 uncertified	T002 (1,285): 8064 MRSA, 9 MSSA, 340 uncertified
America	t008 (2,100): 2151 MRSA, 56 MSSA, 107 uncertified	t002 (1,525): 855 MRSA, 80 MSSA, 857 uncertified	t242 (752): 478 MRSA, 274 uncertified
Africa	t037 (381): 381 MRSA	t084 (267): 217MSSA, 50 uncertified	t064 (123): 256 MRSA, 11 uncertified
Australia	t202 (50): 50 uncertified	t037 (32): 32 MRSA	t437 (19): 19 MRSA

Tableau VI : Distribution des *spa*-types les plus prévalents à travers le monde en termes de nombre de souches de *S. aureus* sensibles (SASM) et résistants à la méticilline (SARM)²³⁸.

Du fait de sa facilité d'exécution et de sa nomenclature standardisée, le *spa-typing* est devenu un outil de typage très utile et développé pour l'épidémiologie mondiale. L'utilisation de cette méthode par de nombreux laboratoires a permis d'évaluer la prévalence des *spa-types* des isolats de *S. aureus* sur chaque continent et de distinguer des clones endémiques (**Figure 58**). Il a été montré que les *spa-types* les plus prévalents en Europe étaient les *spa-types* t032, t008 et t002. En effet, le t032 particulièrement prévalent au Royaume-Uni et en Allemagne, suivi du *spa-type* t008 qui est particulièrement présent en France et en Italie. Quant au t002, il est également le second *spa-type* le plus répandu aux États-Unis, au Canada et au Brésil²³⁸.

En parallèle du *spa-typing*, des études de la prévalence de la cassette SCCmec au sein de la population de *S. aureus* ont été réalisées (**Tableau VI**). Ces travaux ont permis d'identifier des *spa-types* associés à des souches de SARM²³⁸. Par exemple, les *spa-types* les plus représentés dans le monde sont associés à un plus grand nombre de SARM que de SASM. En effet, le t008 (2629 SARM vs 258 SASM), suivi du t002 (9364 SARM vs 189 SASM), du t037 (SARM 1971 vs 51 SASM) et du t044 (590 SARM vs 0 SASM) ont été identifiés dans 12 pays à travers le monde. Par exemple, en Europe, tous les isolats de t032 sont des SARM et les isolats de t037 et t437 en Afrique et en Australie sont également des SARM. Toutefois, ces résultats dépendent des collections d'échantillons utilisées dans la littérature. De nombreuses études cliniques s'attachent à caractériser en priorité des souches de SARM ; une conclusion spécifique à chaque étude doit alors être tirée pour savoir si les isolats de SARM sont associés à des *spa-types* particuliers et inversement²³⁸.

La méthode de *spa-typing* est une méthode de typage qui présente de nombreux avantages lorsqu'on la compare aux autres méthodes de typage en termes de rapidité, de reproductibilité, de simplicité d'application et d'interprétation, mais aussi de standardisation et de comparaison de données. De plus, son haut pouvoir discriminant et sa congruence avec les autres méthodes de typage, font du *spa-typing* un outil épidémiologique important pour le typage de première intention afin de suivre la diffusion et la répartition mondiale des isolats de *S. aureus*²³⁸.

Tout comme l'automatisation et la démocratisation du séquençage Sanger a permis la mise au point de méthodes de typage, les progrès considérables dans les technologies de séquençage à haut débit ont conduit au développement de nouvelles méthodes de typage exploitant la totalité du génome bactérien. En effet, le séquençage haut débit permet une recherche exhaustive des mutations (SNP, insertion, délétion) au sein des génomes bactériens. Des méthodes de typage sont basées sur la comparaison des SNP sur le génome complet appelées *whole genome* SNP. En termes de pouvoir discriminant, ces méthodes supplantent toutes les méthodes de typage précédentes^{194,242}. Ainsi, ces applications du séquençage haut débit sont des outils d'une grande pertinence pour les études micro-épidémiologiques^{194,242,243-245}.

TRAVAUX PERSONNELS

1. Analyse génomique comparative de souches pathogènes et de portage de *S. lugdunensis*

1.1. Objectifs et principaux résultats de l'étude

S. lugdunensis est une espèce commensale appartenant à la flore cutanée de l'Homme. Cependant, il possède un pouvoir pathogène plus proche de celui de *S. aureus* en termes d'agressivité et d'évolution clinique, que de celui des autres SCN². En effet, les infections à *S. lugdunensis* présentent un tel caractère agressif et destructeur qu'il peut être souvent nécessaire de recourir à des procédures de drainage ou de chirurgie, en plus de l'antibiothérapie, pour en limiter la mortalité³⁵. Même si les souches de *S. lugdunensis* restent sensibles à la quasi-totalité des antibiotiques anti-staphylococciques, il a été observé chez cette espèce, l'émergence de souches résistantes à l'oxacilline^{26,246} ; antibiotique utilisé en première intention dans le traitement des infections sévères à SCN.

Malgré la description de quelques facteurs de virulence, aucune donnée moléculaire n'explique à ce jour une telle pathogénicité chez cette espèce. Plusieurs méthodes de typage moléculaire ont été développées comme la MLST²⁰⁸ et la MVLST⁹⁸ pour rechercher un éventuel lien entre lignées phylogénétiques et virulence. Ces travaux de typage ont permis de décrire le caractère clonal de la population de *S. lugdunensis* et d'identifier sept CC (<http://bigsdbs.web.pasteur.fr/staphlugdunensis>), mais aucune lignée hypervirulente associée à un contexte clinique sévère. De nouvelles méthodes de typage basées sur l'étude du polymorphisme de motifs répétés en tandem ont confirmé cette absence de lien entre groupements phylogénétiques et les contextes d'isolement des souches (**cf partie travaux associés**).

Le séquençage de génomes complets est aujourd'hui un outil puissant en microbiologie pour les études épidémiologiques, l'identification de gènes de résistance (résistome), ou d'EGM (mobilome) et la caractérisation de gènes codant des facteurs de virulence (virulome). Néanmoins, seulement 15 génomes complets de *S. lugdunensis* sont actuellement disponibles sur NCBI et seuls deux études ont consisté à la comparaison de génomes entier chez cette espèce^{96,141}. Ces études ont révélé que (i) le génome de *S. lugdunensis* est plus proche de celui de *S. aureus* que de celui des autres SCN, (ii) qu'il peut porter des EGM tels que des prophages et des plasmides et (iii) que cette espèce, à

la différence de *S. aureus* et *S. epidermidis*, présente un *pan-genome* fermé avec plusieurs systèmes limitant les transferts de gènes horizontaux.

Dans ce contexte, le premier volet de nos travaux a consisté à séquencer le génome complet de trois souches pathogènes (SL13, SL29 et SL55) et de trois souches de portage (SL117, SL118 et SL122) pour les comparer aux 15 génomes de *S. lugdunensis* disponibles sur la base de données du NCBI. La comparaison génomique a été réalisée dans le but de caractériser le *pan-genome*, le *core-genome* et le génome accessoire des 21 souches de *S. lugdunensis*. D'autre part, cette comparaison a été effectuée pour identifier de potentiels déterminants génétiques spécifiques des souches pathogènes ou de portage, ou spécifiques de souches responsables d'un contexte clinique sévère. Cette étude s'est également attachée à rechercher des gènes de résistance et de nouveaux facteurs de virulence, portés ou non par des EGM, afin d'expliquer la virulence particulière de cette espèce. Parallèlement, nous avons évalué le pouvoir pathogène des six souches séquencées dans un modèle d'infection de la larve de *Galleria mellonella* jusqu'alors non mis en œuvre chez cette espèce bactérienne.

Le séquençage des six souches de *S. lugdunensis* a été réalisé à l'aide de la technologie *Pacific Biosciences*. Il a généré pour chaque souche un unique contig correspondant au génome complet et circularisé d'une taille comprise entre 2,53 et 2,61 Mb. L'annotation des 21 génomes nous a ensuite permis de déterminer le répertoire total de gènes de l'espèce par la caractérisation du *pan-genome*. A l'aide du logiciel *orthologsorter*, il a alors été possible de former des groupes de séquences protéiques orthologues afin d'identifier ceux communs ou non à toutes les souches de *S. lugdunensis*. De cette façon, le nombre total de gènes pour l'espèce s'élève à 3077 gènes dont 2185 composant le *core-genome* (communs aux 21 souches) et 892 pour le génome accessoire.

A partir de cette analyse, nous avons recherché si, au sein des gènes composant le génome accessoire, certains étaient spécifiques des souches pathogènes ou des souches de portage. Aucun gène spécifique d'un des deux groupes ou d'un contexte clinique donné n'a été identifié (endocardite vs IPTM par exemple). En revanche, l'analyse du génome accessoire a révélé la présence de gènes spécifiques de certains CC définis par MLST, comme la présence de gènes codant pour le système CRISPR/Cas uniquement chez les souches de CC1 ou la présence d'un SRM de type II seulement chez les souches appartenant au CC6.

L'analyse génomique a également permis de rechercher la présence d'EGM tels que de l'ADN de phage intégré au sein du génome et de plasmides potentiellement porteurs de gènes de résistance aux antibiotiques ou de gènes codant des facteurs de virulence. Ainsi, nous avons identifié cinq nouvelles régions phagiques nommées Φ SL6 à Φ SL10 après comparaison de leurs séquences avec celles des régions phagiques précédemment identifiées chez *S. lugdunensis*^{73,96}. Parmi celles-ci aucun gène de virulence, ni de résistance n'a été identifié. Toutefois, des séquences codant une autolysine bifonctionnelle, une hydrolase du peptidoglycane, une protéine avec une triple hélice collagène et un répresseur LexA ont été identifiées.

La comparaison des génomes des 21 souches de *S. lugdunensis* a également permis d'identifier une région particulière de 13 kb chez seulement deux souches de CC1, la souche SL13 responsable d'une endocardite infectieuse et la souche VISLISI_33 isolée d'un abcès hépatique. Environ 5 kb de cette région présentait une homologie de séquence de 94% avec des îlots de pathogénicité de *S. aureus* (SaPI), SaPITokyo11212 et SaPIlivm60. Il s'est avéré que cette région portait des gènes codant une protéine de liaison à la pénicilline Ear, un transporteur de sidérophores FhuD et une protéine hypothétique avec un domaine SaPI.

Par ailleurs, chez les six souches dont le génome a été séquencé par *Pacbio*, des contigs supplémentaires extra-chromosomiques ont été générés. La comparaison de leur séquence avec les bases de données sur NCBI et leur annotation a permis d'identifier trois plasmides, dont un jamais identifié jusqu'alors, chez la souche SL13 (CC1) nommé pSL1 et deux chez la souche SL117 (CC1), nommés pSL2 et pSL3. Aucun gène de virulence n'était porté par ces plasmides, mais le plasmide pSL1 était porteur d'un gène de résistance à la tétracycline *tetK* et le plasmide pSL2 portait un gène de résistance au cadmium *cadC*. Nous avons alors recherché des gènes de résistance au sein du chromosome des 21 souches de *S. lugdunensis*. Plusieurs gènes ont été identifiés dont un gène de résistance à la fosfomycine, *fosB*, présent chez toutes les souches. Un gène de résistance à la pénicilline *blaZ* porté par un transposon de 6 kb a également été retrouvé chez sept souches. Enfin, nous avons identifié un gène de résistance à l'acide fusidique *fusB* uniquement chez la souche VISLISI_27 (CC1). La séquence du gène *fusB* possède 91% d'identité avec le gène *fusB* porté par un îlot génomique de *S. aureus* nommé SaRI*fusB*. Ainsi, une région de 16 kb correspondant à un îlot génomique putatif a été identifiée chez la souche VISLISI_27 ;

cette région porte également le gène *vapE* (*virulence-associated protein E*) codant un facteur de virulence putatif.

Cette étude a également fait l'objet d'une caractérisation du virulome de *S. lugdunensis*, c'est-à-dire d'une recherche de différences entre les souches au niveau du répertoire génique codant pour des facteurs de virulence. Malgré l'identification de plusieurs gènes codant des facteurs impliqués dans la virulence de *S. lugdunensis* ou de plusieurs espèces de staphylocoques, aucune variation spécifique des souches pathogènes ou des souches de portage n'a été identifiée. Cependant, nous avons mis en évidence plusieurs variations génétiques spécifiques des souches appartenant au CC3 pour le *locus* codant le système *agr* et celui codant la capsule polysaccharidique. De plus, nous avons observé plusieurs décalages du cadre de lecture et mutations non-sens dans les gènes codant les adhésines putatives SIs.

Lors de la comparaison des gènes potentiellement impliqués dans la virulence, nous avons observé au sein du *locus* *ess* codant le SST7 une importante diversité génétique et différentes organisations spécifiques des CC de *S. lugdunensis*. En effet, deux variants de séquence du gène *essC* ont été identifiés ; le premier variant du gène était associé à un *cluster* de cinq gènes retrouvé chez toutes les souches appartenant aux CC1, CC5, CC6 et une souche de CC7. Le second variant du gène *essC* était associé à un *cluster* de six gènes uniquement retrouvé chez les souches de CC3 et une souche de CC7. En aval de ces gènes, un nombre variable de gènes codant des protéines hypothétiques a été observé. D'importantes variations ont également été observées entre les souches appartenant au CC1 comme la présence d'un transposon inséré dans le *locus* *ess* ou un nombre variable de gènes codant des protéines avec un domaine DUF600.

Enfin, dans le modèle d'infection chez la chenille *G. mellonella*, trois souches de portage (SL117, SL118 et SL122) mais aussi la souche isolée d'une infection sur prothèse vasculaire (SL29) ont montré une virulence modérée. A l'inverse les souches isolées d'une infection de la peau et des tissus mous (SL55) et d'une endocardite infectieuse (SL13) se sont avérées significativement plus virulente que les quatre autres souches.

Ces travaux font l'objet d'un article scientifique, actuellement en cours d'écriture.

1.2. Publication n°1

Comparative genome analysis of Staphylococcus lugdunensis shows complex clonal-dependent diversity of a putative virulence factor, the *ess/type VII* locus

Jérémie Lebeurre, Sandrine Dahyot, Seydina Diene, Amandine Paulay, Marion Aubourg, Xavier Argemi, Jean-Christophe Giard, Isabelle Tournier, Patrice François, Martine Pestel-Caron. Article en cours d'écriture.

1 **Comparative genome analysis of *Staphylococcus lugdunensis* shows complex clonal-**
2 **dependent diversity of a putative virulence factor, the *ess*/type VII locus**

3
4
5
6 Lebeurre J.^{1*}, Dahyot S.², Diene S.³, Paulay A.¹, Aubourg M.⁴, Argemi X.⁵, Giard JC.⁴, Tournier
7 I⁶, François P.³, Pestel-Caron M.².

8
9 ¹Normandie Univ, UNIROUEN, GRAM EA2656, F-76000 Rouen, France

10 ²Normandie Univ, UNIROUEN, GRAM EA2656, Rouen University Hospital, F-76000 Rouen,

11 ³University of Geneva Hospitals, Genomic Research Laboratory, Service of Infectious
12 Diseases, Geneva, Switzerland

13 ⁴Université de Caen Basse-Normandie, EA4655 U2RM (Équipe Antibio-Résistance), Caen,
14 France.

15 ⁵Université de Strasbourg, CHRU de Strasbourg, VBP EA7290, Fédération de Médecine
16 Translationnelle de Strasbourg (FMTS), Institut de Bactériologie, 3 rue Koeberlé, F-67000
17 Strasbourg, France

18 ⁶Normandie Univ, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and
19 Personalized Medicine, Rouen University Hospital, F-76000 Rouen, France

20 *Corresponding author:

21 E-mail: jeremie.lebeurre@univ-rouen.fr

22 Phone number : +33235148299 ; fax number : +33235148203

23 Groupe de Recherche sur l'Adaptation Microbienne (EA2656), Normandie Univ, UNIROUEN,
24 F-76000 Rouen, France

25

26 **Abstract**

27 *Staphylococcus lugdunensis* is a commensal bacterium of human skin that has emerged as a
28 virulent Coagulase-Negative Staphylococci (CoNS) in both community-acquired and
29 healthcare associated infections. Genotyping methods showed a clonal population structure of
30 this pathogen but failed to identify hypervirulent lineages. Here, complete genomes of three
31 pathogenic and three carriage *S. lugdunensis* strains were obtained by Single-Molecule
32 sequencing (Pacbio) and compared to 15 complete genomes available in GenBank database.
33 The goal was to identify (i) genetic determinants specific of pathogenic or carriage strains or
34 specific to clonal complexes (CC) defined by MultiLocus Sequence Typing, and (ii) antibiotic
35 resistance genes and new putative virulence factors worn or not by mobile genetic elements
36 (MGE). Comparative genomic analysis did not show strict correlation between gene content
37 and the ability of the six strains to cause infections in humans and in a *Galleria mellonella*
38 infection model. However, this study identified new mobile genetic elements (five prophages,
39 two genomic islands and one plasmid) and genetic variations of some putative virulence-
40 associated *loci*, especially in CC3 strains. For a so clonal population, high variability and eight
41 CC-dependent genetic organizations were observed for the *ess* locus, which encodes a putative
42 type VII secretion system (T7SS) homologous to that of *S. aureus*. Further phenotypic and
43 functional studies are needed to characterize this particular CC3 and to evaluate the role of the
44 T7SS in the virulence of *S. lugdunensis*.

45 **Introduction**

46 *Staphylococcus lugdunensis* is a coagulase-negative staphylococcal species (CoNS) that is a
47 part of the normal human skin flora. However, unlike other CoNS, *S. lugdunensis* exhibits a
48 high degree of virulence similar, in some respects, to *Staphylococcus aureus*¹.
49 *S. lugdunensis* is responsible for infections ranging from minor skin and soft tissue infections^{2,3}
50 to invasive infections such as infective endocarditis⁴, bone and joint infections⁵⁻⁷, prosthetic
51 device-infections^{8,9} and vascular catheter-related infections¹⁰, both in healthy host and immune-
52 compromised patients. One of the characteristics of *S. lugdunensis* infections is the
53 development of deep-seated abscess lesions that infected patients can not clear without surgical
54 intervention associated to an effective antibiotic therapy². Interestingly, few virulence factors
55 have been described to date to explain the high capacity of *S. lugdunensis* to cause acute and
56 suppurative infections, and particularly the tissue damage and the elevated mortality observed
57 for infections such as infective endocarditis^{9,11}. Furthermore, oxacillin-resistant strains have
58 recently emerged^{12,13} which are worrying as beta-lactams are first-line drugs for the treatment
59 of severe infections.

60 Several molecular typing methods have been described to search for a potential link between
61 within-species genetic variations and characteristics such as pathogenic potential and virulence.
62 Multi-Locus Sequence typing (MLST) showed a clonal population structure of this pathogen¹⁴,
63 and allowed to describe seven clonal complexes (CC)
64 (<http://bigsdbs.web.pasteur.fr/staphlugdunensis>) but failed to identify hypervirulent lineages or
65 clusters specific to carriage strains. Multi-Virulence Locus Sequence Typing¹⁵ as well as
66 Multiple Locus Variable Number Tandem Repeat Analysis (MLVA) and Tandem Repeat
67 Sequence Typing (TRST) recently developed, confirmed this absence of link between
68 clustering and clinical settings¹⁶.

69 Whole-genome sequencing (WGS) has become a powerful tool in microbiology but to date
70 only 15 complete genomes of *S. lugdunensis* are available on the National Center for
71 Biotechnology Information (NCBI). The only two studies which have compared *S. lugdunensis*
72 whole genomes revealed (i) that its genome is closer to that of *S. aureus* than other CoNS, (ii)
73 that it can harbor mobile genetic elements (MGE) such as prophages and plasmids but (iii) lastly
74 that this species displays a closed pan-genome and several systems (restriction-modification,
75 toxin/antitoxin and CRISPR/Cas systems) to prevent horizontal gene transfer^{17,18}.

76 In this context, we first sequenced the whole genome of three pathogenic and three carriage
77 strains, previously characterized by MLST¹⁵ and collected from 4 geographic origins in France
78 and Sweden. Then, these whole-genome sequences were combined with the data from the 15

79 genomes available in the GenBank database. This study was conducted to identify genetic
80 variations specific to pathogenic or carriage strains or to CC, antibiotic resistance genes and
81 new putative virulence factors worn or not by MGE.

82 Comparative genomic analysis did not show correlation between gene content and the ability
83 of the six strains to cause infections in humans and in a *Galleria mellonella* infection model but
84 genetic variations of some putative virulence-associated loci among the 21 genomes analyzed.

85 Surprisingly for a so clonal population, a great diversity was observed for the *ess* locus, which
86 putatively encodes a type VII secretion system (T7SS) homologous to that of *S. aureus*. The
87 T7SS of *S. aureus* is a multiprotein complex dedicated to the export of several virulence factors
88 during host infection which plays a key role in pathogenesis. Experiments using mouse models
89 of infection have indeed shown that T7SS is required for colonization, intraspecies competition
90 and the establishment of persistent abscess lesions¹⁹⁻²².

91

92 **Results**

93 **Pacific biosciences genome sequences and functional genome annotation**

94 PacBio sequencing of the six *S. lugdunensis* strains SL13, SL29, SL55, SL117, SL118 and
95 SL122 resulted in full genome recovery on a single contig with sequence length ranging from
96 2.53 (SL118) to 2.61 Mb (SL55). Between 67457 (SL55) and 79450 (SL118) post-filter
97 polymerase reads were generated with an average read length ranging from 13.7 (SL55) to 14.8
98 kb (SL13). The N50 read length was comprised between 19.2 and 22.7 kb, and the average
99 coverage between 272X and 367X. In addition to the contigs corresponding to the genomes,
100 zero (SL29 and SL118) to three (SL117) extra-chromosomal contigs were generated.

101 The complete genome sequences of 15 different *S. lugdunensis* strains available in the NCBI
102 database (Table S1) were compared to the six genomes sequenced by PacBio. The size of these
103 21 complete genomes ranged between 2.53 (SL118) and 2.69 Mb (N920143). All strains
104 presented similar %GC-content (33.6 to 33.9%). The number of genes automatically predicted
105 by Prokka software was ranging from 2443 (FDAARGOS381) to 2675 (N920143) genes
106 encoding between 2338 and 2541 proteins. Genomic data for the 21 strains are shown in
107 Table 1.

108 **Pan-, core- and accessory-genome analysis**

109 Pan-genome analysis lead to identify the total gene repertoire of the species and sets of core
110 and accessory genes available in each *S. lugdunensis* genome. The pan-genome of the 21
111 *S. lugdunensis* strains was found to be composed of a total repertoire of 3077 genes. Of these,
112 2185 genes (71.01%) were part of the core-genome, and 892 genes (28.99%) were part of the
113 accessory genome. Based on this analysis, no gene from the accessory genome was retrieved to
114 belong only to the pathogenic strains or only to the carriage strains. In the same way, no gene
115 from the accessory genome was associated only to strains isolated from a severe clinical
116 infection such as infective endocarditis, skin and soft tissues infection, deep-seated abscess or
117 device-associated infection.

118 However, CC-dependent genetic variations were observed. Accessory-genome analysis of the
119 CC1 strains revealed effectively that only 18 genes were unique to these nine strains. Among
120 them, nine genes corresponded to the type III-A clustered, regularly interspaced, short,
121 palindromic repeat (CRISPR) loci-associated Cas1, Cas2, Cas10, Csm2, Csm3, Csm4, Csm5,
122 Csm6 and Cas6 proteins. The *cas* locus from the CC1 strains showed a sequence homology of

123 76% with that of the *S. epidermidis* RP62a strain. In all the CC1 strains except the strains SL29
124 and SL122, a potential frameshift is located in the gene encoding the Csm6 protein and resulting
125 in a probable non-functional CRISPR/Cas system. The remaining nine genes unique to the CC1
126 strains were carried by a transposon and encode arsenical resistance proteins.

127 Eight genes were unique to the CC6 strains. Six of them encode hypothetical proteins but two
128 encode a type II restriction-modification system (RMS). Indeed, we identified a C-5 cytosine-
129 specific DNA methylase and a Type-2 restriction enzyme Sau3AI sharing 82% of identity with
130 the RMS type II from *S. haemolyticus* JCSJ1435 strain.

131 Only two genes were found specific of the CC7 strains: one encoding a DUF600 protein and
132 one encoding a hypothetical protein. No gene was identified as specific to the strains of CC3.

133 **Mobile genetic elements identification**

134 Five new intact prophage regions named Φ SL6 to Φ SL10 were identified by a PHASTER
135 analysis for 12 of the 21 strains (Table 2). The prophage region Φ SL6, Φ SL7 and Φ SL8 were
136 identified in the strains of CC1 (SL13, SL29, SL117 and SL122) while the prophage regions
137 Φ SL9 and Φ SL10 were found in strains of CC6 (SL55 and SL118) and CC7 (FDAARGOS143)
138 respectively. The total length of these prophages was ranging from 46.6 kb (Φ SL9) to 49.9 kb
139 (Φ SL7); these prophages encode between 26 to 69 proteins. Among these proteins, no virulence
140 factors or antibiotic resistance-associated proteins were identified. However, we identified a
141 putative bifunctional autolysin, a cell wall hydrolase lytN precursor, a collagen triple-helix
142 protein and a LexA repressor which might be involved in the remodeling of the bacterial cell
143 wall.

144 Functional annotation by using Prokka on additional contigs lead to identify three plasmids:
145 pSL1 in the strain SL13, and pSL2 and pSL3 in the strain SL117 (Table 3). The total length of
146 the plasmids identified was ranging from 1.9 (pSL3) to 3.9 kb (pSL1); they encode between 4
147 and 5 proteins. Plasmids pSL2 and pSL3 respectively correspond to the cadmium resistance-
148 associated plasmid pLUG10²³ and the plasmid pVISLISI_2¹⁷ previously described. Of note, no
149 virulence factor was carried by these plasmids but tetracyclin and cadmium resistance genes
150 were found.

151 In addition, we identified a region of 13 kb length restricted to the CC1 strains SL13 and
152 VISLISI_33. Around 5 kb of this region shared a sequence homology of 94% with the

153 pathogenicity island SaPITokyo11212 of the *S. aureus* strain Tokyo11212 and with the
154 pathogenicity island SaPIvm60 of the *S. aureus* strains IVM60. This region encodes an
155 integrase, an excisionase, a primase and a terminase. Other interesting genes were carried by this
156 potential pathogenicity island as those encoding a Ear penicillin binding protein, a FhuD
157 ferrichrome ABC transporter and a protein with a *S. aureus* Pathogenicity Island (SAPI)
158 domain.

159 **Identification of antibiotic resistance-associated genes in *S. lugdunensis***

160 Several antibiotic resistance-associated genes were identified by using the web tool ResFinder
161 v3.0 (Table 4). A gene homologous (83% of identity) to the *fosB* gene of *Bacillus cereus* AS4-
162 12 was identified in the chromosome of all *S. lugdunensis* strains. A penicillin resistance-
163 associated gene, *blaZ* was also identified. This gene was carried by a transposon and found in
164 seven of the 21 *S. lugdunensis* strains. The *blaZ* gene showed a homology ranging from 92%
165 to 100% with the *blaZ* gene carried by the plasmid pSAP106A of *S. epidermidis*. Otherwise,
166 the annotation of extra-chromosomal contig of SL13 strain lead to identify a *tetK* tetracyclin
167 resistance gene which shared an identity of 99% with the *tetK* gene from the plasmid pT181 of
168 *S. aureus*. Of note, no gene of methicillin resistance, within a composite staphylococcal
169 chromosome cassette (SCC) *mec* element, was found.

170 Finally, we also identified a fusidic acid resistance gene, but only in strain VISLISI_27. This
171 gene had a 91% homology with the *fusB* gene carried by a genomic island (SaRIfusB) of the
172 *S. aureus* CS6-EEFIC strain. The analysis of the location of the *fusB* gene in the genome of the
173 strain VISLISI_27 revealed that it is located on a genomic island of 16 kb. In addition, by
174 analyzing this latter region we identified a gene encoding a potential virulence factor, *vapE* (for
175 virulence-associated protein E). The gene *vapE* was also found in a genomic island of 40 kb in
176 the genome of the strain HKU09-01 with an identity of 85%. However, Blastn comparison of
177 these two genomic islands showed 80% of identity for only 3 kb, suggesting that strains
178 VISLISI_27 and HKU09-01 have acquired these regions independently and from different
179 sources.

180 **Comparison of virulence factors genes in *S. lugdunensis***

181 Analysis of the virulence gene composition based on the Virulence Factor Database (VFDB)
182 showed a number of putative virulence factors that were detected in most of the *S. lugdunensis*
183 strains analyzed. To improve the identification of notable features, tblastn searches of well-

184 characterized staphylococcal virulence factors and transcriptional regulators were performed
185 (Table 5).

186 Of the genes belonging to the accessory gene regulator (*agr*) locus, *agrA*, *agrB*, *agrC* and *agrD*
187 were found in all the *S. lugdunensis* strains. However, we have identified a strong variability
188 among the *agr* locus as described by Dufour *et al.* (Dufour 2002), who identified two genetic
189 groups of the hypervariable *agr* region (*agrB* to *agrC*). Phylogenetic analysis based on the
190 alignment of the hypervariable *agr* region confirmed the repartition into two groups (Figure
191 S1A). Indeed, all the CC3 strains are clustered with the allele *agrSI*-type II whereas all other
192 strains were closer to the allele *agrSI*-type I.

193 In a region of 16 kb in the genome of the 21 *S. lugdunensis* strains, 16 genes homologous to
194 those encoding polysaccharidic capsule type 5 or 8 in *S. aureus* were detected (Figure S1B). In
195 *S. lugdunensis*, only the central region composed of the *capH*, *capI*, *capJ* and *capK*, was found
196 different between strains; the 12 other genes of the *cap* locus were conserved. Based on the
197 phylogenetic analysis, the *cap* locus of CC3 and VISLISI_25 strains (CC5) was closer to the
198 *cap8* locus of *S. aureus* while other strains were related to the *cap5* locus.

199 **Sls genes showed a highly diversity in *S. lugdunensis***

200 Comparative analysis of the Sls nucleotide and predicted protein sequences revealed an
201 important diversity between the 21 *S. lugdunensis* strains. The comparison of the SlsA
202 sequences showed that their size varied from 1634 to 5228 amino acids between *S. lugdunensis*
203 strains, correlated with the number of IgG-like fold ranging from eight to 45. Sequence
204 comparison of the SlsA proteins by using Blastp showed 51 to 52% of identity with the Bhp
205 protein from the strain *S. epidermidis* RP62a and 63 to 64% with a cell wall associated biofilm
206 Bhp/SesD protein from *Staphylococcus caprae* JMUB590. The pfam analysis of these two
207 staphylococcal proteins showed also the presence of similar IgG-like fold, with 18 and 50
208 repeats respectively. Interestingly, the *slsA* gene was absent from the genomes of all the CC3
209 strains and from the FDAARGOS377 strain (CC5). Of note, a nonsense mutation was detected
210 in the *slsA* gene of the strain VISLISI_22 (CC6).

211 Furthermore, in the *slsD* gene, a nonsense mutation located upstream the region coding for the
212 LPXTG motif was found in 14 of 21 *S. lugdunensis* strains but was not present for the CC3
213 strains and the strain VISLISI_33 (CC1). Sequence comparison of the SlsD proteins by using

214 Blastp showed an identity of 40% and 32% with the adhesins SdrF and SdrG from
215 *S. epidermidis* respectively.

216 Analysis of the SlsE protein sequences showed that their size varied from 2818 to 4166 amino
217 acids among the 21 *S. lugdunensis* strains. Interestingly, in the *slsE* gene a frameshift was
218 located upstream the region encoding the LPXTG motif and was recovered for all the CC1
219 strains except the strains HKU09-01 and VISLISI_37. The best match obtained by using Blastp
220 for SlsE (39% of identity) was obtained for a non-characterized “YSIRK-type signal protein”
221 of *Staphylococcus simulans*.

222 The comparison of the SlsG protein sequences showed that their size varied from 1604 to 2286
223 amino acids between *S. lugdunensis* strains, correlated with the Rib repeat number ranging from
224 seven to 21. In the strains VISLISI_21 and VISLISI_25 a frameshift was located within the Rib
225 repeats.

226 Finally, the SlsB, SlsC and SlsF proteins were highly conserved in the 21 *S. lugdunensis* strains,
227 between 99 and 100% of identity.

228

229 **Description of the genetic diversity of the *S. lugdunensis* *ess* locus**

230 Comparative genomic analysis showed variations in the size of a nucleotidic region of the *ess*
231 locus from the 21 strains of *S. lugdunensis*. Indeed, the *ess* locus size ranged from 22134 bp
232 (HKU09-01) to 38834 bp (SL29 and SL122). This difference is the result of an important
233 polymorphism identified in the strains of CC1. For a better understanding and to facilitate the
234 description of these variations, we have divided the *ess* locus into six distinct modules, each
235 with a particular coding sequences based-content and a relative conservation (Figure 1). Eight
236 genetic organizations were observed in the 21 genomes of *S. lugdunensis*.

237 The module 1 contained five genes present in all sequences analyzed and homologous to genes
238 encoding proteins of the T7SS in *S. aureus*: EsxA, EsaA, EssA, EsaB and EssB. This module
239 was highly conserved among all the *S. lugdunensis* strains, with slight modifications in the
240 amino acid sequences of these five proteins (99 to 100% of identity depending on the proteins).

241 The module 2 was variable between *S. lugdunensis* clonal complexes. Indeed, two distinct
242 organizations of the *ess* locus were identified; one shared by all the CC3 strains and one CC7
243 strain (C33), and the other one for the strains the other CC (CC1, CC5 and CC6). The first
244 important variation concerned the *essC* gene for which two *essC* sequences variants (*essC1* and

245 *essC2*) were observed. The 5' region of the *essC* gene was conserved within all *S. lugdunensis*
246 isolates from any CC (3762 bp) whereas the 3' region differed (around 681 bp). Thus, the *essC1*
247 variant encode a protein of 1476 residues and was identified in 15 of the 21 *S. lugdunensis*
248 strains. Of note, *S. lugdunensis* N920143 contained a frameshift in the gene encoding the
249 membrane associated protein EssC. On the other hand, the *essC2* sequence variant encodes a
250 protein of 1481 amino acids which was only identified in the five CC3 strains and in the C33
251 strain (CC7).

252 Otherwise, each *essC* gene variant was associated with a proper cluster of genes. The *essC1*
253 gene was associated with five genes highly conserved among the 15 *S. lugdunensis* strains. The
254 *essC2* gene variant was associated with six genes conserved within the five CC3 strains and the
255 C33 strain. Conversely, these six genes shared no identity with the five genes associated with
256 the *essC1* variant (Figure 2).

257 In the module 3, 4 and 6, it was observed a complex arrangement of coding sequences that
258 encode a variable number of uncharacterized proteins. Effectively, an important number of
259 genes which encode proteins with Domains of Unknow Function (DUFs) related to the T7SS
260 were identified (DUF5079, DUF5080, DUF5020, DUF5084 and DUF5085). The number of
261 genes encoding these proteins ranged from 6 to 10 among the CC1 strains. The highest number
262 observed was 11 in the CC3 strains and the strain C33. In addition, the number of genes
263 encoding DUF600 proteins was variable among isolates; the minimum number identified was
264 two in CC6 strains while the highest number was five in the CC3 strains.

265 With regard to the module 5, two distinct genetic organizations were observed. Indeed, there
266 was a particular genetic arrangement that included nine coding sequences unrelated to the T7SS
267 carried by a transposon found only in four CC1 strains (SL29, SL177, SL122 and N920143).
268 In these strains, the transposon was associated with three coding sequences which encode
269 hypothetical DUF600 proteins of 152, 159 and 165 amino acids. In the strains which have no
270 transposon, two genes encoding hypothetical DUF600 proteins were identified. One of them
271 encodes a DUF600 protein of 165 amino acids conserved among the 21 *S. lugdunensis* strains
272 and the other one encoded a protein of 339 residues with two DUF600.

273

274 **Comparison of the *ess* modules 1 and 2 from *S. lugdunensis* and *S. aureus***

275 Here, the comparison will focus on the module 1 and 2 which are the only ones to have a strong
276 homology with those of *S. aureus* (48 to 92% of identity). The comparison of the module 1 of
277 the *ess* locus was performed between the genome of the *S. aureus* HO 5096 0412 and RJ-ST398
278 strains and the genome of the *S. lugdunensis* HKU09-01 and VISLISI_21 strains, both chosen
279 as representative of the diversity observed among the 21 strains (Figure 2).

280 This showed that the *S. lugdunensis* genes *SLGD_RS09680* and *B7466_RS09445* which encode
281 a putative secreted protein shared 92% of identity with the EsxA protein of *S. aureus* strains.
282 The genes *SLGD_RS09680* and *B7466_RS09445* encode a protein of 97 amino acids with a
283 WXG100 domain and conserved C-terminal hydrophobic patterns for the formation of specific
284 α -helical surface. In the same way, the transmembrane proteins EsaA, EssA and EssB of
285 *S. aureus* strains showed high homologies (51 to 64% of identity) with several putative
286 transmembrane proteins in *S. lugdunensis* HKU09-01 strain (*SLGD_RS09675*,
287 *SLGD_RS09670* and *SLGD_RS09660*) and VISLISI_21 (*B7466_RS09440*, *B7466_RS09435*
288 and *B7466_RS09425*). These homologous proteins exhibited the same family domains in
289 *S. aureus* and *S. lugdunensis*.

290 Finally, the cytoplasmic protein EsaB of *S. aureus* strains exhibited an identity of 64% with the
291 proteins *SLGD_RS09065* and *B7466_RS09420* which contained a YukD family domain.
292 Functional analysis of the T7SS core components using pfam database allowed the prediction
293 and the identification of the same protein domains for *S. aureus* and *S. lugdunensis*.

294 The genes *SLGD_RS09655* (*essC1*) and *B7466_RS09420* (*essC2*) of *S. lugdunensis* strains
295 which encode the putative EssC protein contained a N-terminal FtsK-SpoIIIE domain for DNA
296 transportation and two domains of the SpoIIIE-FtsK-like ATPase family. Otherwise, the *essC1*
297 gene encoded a protein sharing 77% of identity with the EssC protein of *S. aureus* strain HO
298 5096 0412, while the *essC2* gene encoded a protein sharing 77% of identity with the EssC
299 protein of *S. aureus* strain RJ-ST398.

300 Then, a comparison of the module 2 of the *ess* locus from the *S. lugdunensis* strain HKU09-01
301 with that of the *S. aureus* strain HO 5096 0412 was performed (Figure 2, Table 6). In the strain
302 HKU09-01, the *essC1* gene is associated to five genes encoding hypothetical proteins
303 (*SLGD_RS09650* and *SLGD_RS09645*), a LXG domain protein (*SLGD_RS09640*), a
304 DUF5082 protein (*SLGD_RS09635*) and a cystatin-like (*DUF4467*) protein

305 (SLGD_RS09630). The DUF5082 protein of 137 amino acids encoded by the gene
306 *SLGD_RS09635* showed a central WXG motif but no C-terminal hydrophobic conserved
307 pattern was observed. The percentage of identity of the different staphylococcal proteins is
308 described in the Table 6. The comparison of the module 2 of the *ess* locus from the strain
309 HKU09-01 with that of the *S. aureus* strain HO 5096 0412 showed that they shared the same
310 organization.

311 A comparison of the module 2 from the *S. lugdunensis* strain VISLISI_21 with the *S. aureus*
312 strain *aureus* RJ-ST398 was also performed (Figure 2, Table 7). The *essC2* sequence variant,
313 is associated to six genes. Five of them encode proteins with a DUF5344, DUF5082, DUF5079,
314 and two with a DUF5085 (B7466_RS09415, B7466_RS09410, B7466_RS09400,
315 B7466_RS09395 and B7466_RS09390, respectively). The protein B7466_RS09410 with a
316 DUF5344 contained a conserved C-terminal hydrophobic domain and the protein
317 B7466_RS09405 exhibit a N-terminal LXG domain. These proteins had respectively an identity
318 of 50 and 56% with the proteins EsxB and EsxX of the strain *S. aureus* RJ-ST398. The
319 comparison of the module 2 of the *ess* locus between the strain VISLISI_21 and the *S. aureus*
320 strain RJ-ST398 showed the same organization. Otherwise, in the *S. lugdunensis* strain C33, a
321 frameshift is located in the gene encoding the homolog of the EsxX protein.

322

323 ***G. mellonella* infection experiments**

324 In order to evaluate and compare the virulence of the six *S. lugdunensis* strains in the same
325 conditions *in vivo*, the *G. mellonella* model of infection has been developed. The growth
326 kinetics of *S. lugdunensis* strains (SL13, SL29, SL55, SL117, SL118, and SL122) were first
327 carried out in BHI medium at 37°C with agitation. The six strains showed similar growth rates
328 with optical densities (OD_{600nm}) after 7h culture ranging from 1.2 to 1.6 (data not shown).
329 Strains of *S. lugdunensis* from clinical isolates (SL13, SL29, SL55) and carriage (SL117,
330 SL118, SL122) were then injected into *G. mellonella* larvae ($3(\pm 2) \times 10^6$ UFC per larvae) and
331 their survival was monitored during three days. As shown in Figure 3, the carriage strains
332 (SL117, SL118, SL122) as well as the pathogenic isolates SL29 (vascular prosthesis infection)
333 revealed a moderate virulence in this model of infection. After 72 hours, 43%, 31%, 87% and
334 72% of the larvae infected with these 4 strains were still alive, respectively. However, they
335 could not be clustered because statistical analysis displayed significant differences of virulence
336 between them (*p* values < 0.05) (Table 8). In contrast, the strains SL13 (isolated from infective
337 endocarditis) and SL55 (isolated from a skin and soft tissue infection) were significantly more

338 virulent than the others since only 4% and 8% of the infected worms survived 3 days post
339 infection, respectively (Figure 3, Table 8). Of note, no mortality was observed for the larvae for
340 whom sterile saline buffer was injected (data not shown). These results revealed that, according
341 to their origin, the virulence of *S. lugdunensis* were strain-specific and that two clinical isolates
342 obviously appeared more pathogenic.
343

344 **Discussion**

345 *S. lugdunensis* is associated with a wide range of diseases including skin and soft tissue
346 infections, infective endocarditis and bone and joint infections¹. This organism is considered as
347 an opportunistic pathogen which can cause infections with unusual severity compared to other
348 CoNS. It is even considered as a “wolf in sheep’s clothing”¹. This particular virulence is
349 currently not explained by the rare virulence factors described. No typing method^{14,15} has yet
350 been able to identify invasive clonal complexes as described for *Neisseria meningitidis*²⁴ or
351 *Streptococcus pneumoniae*²⁵.

352 In this study, we present the first genomic comparison of three carriage and three pathogenic
353 *S. lugdunensis* strains representative of the major lineages within *S. lugdunensis* species to
354 explore whether virulence can be linked to carriage of specific genes or genetic variations.
355 Furthermore, we compared these genomic data to already-existing genomic data to characterize
356 the pan-, core- and accessory-genome of the species.

357 This sequencing brought the number of available genomes to 21 and therefore increased the
358 potential genetic diversity of the *S. lugdunensis* genomic data as the six strains sequenced in
359 this study were collected from one Swedish and three different French cities, whilst seven of
360 the 15 genomes in GenBank were originated from a unique French location¹⁸.

361 Pan-genome analysis confirmed that the *S. lugdunensis* pan-genome could be considered
362 closed, as supported by the Heaps' law extrapolation used by Argemi *et al.*¹⁸. The *S. lugdunensis*
363 pan-genome appears in fact to be poorly expanding with the inclusion of six new complete
364 genomes and reaches only a total of 3077 genes. As described previously¹⁸, *S. lugdunensis* pan-
365 genome cognates to approximately 3000 genes based on the analysis of 15 genomes whereas
366 the *S. aureus* and *S. epidermidis* pan-genomes reach more 3800 genes.

367 In this context, it is not surprising that a large number of *S. lugdunensis* strains have barriers to
368 horizontal gene transfer CRISPR/Cas system or RMS¹⁸. Our study shows that 48% of the strains
369 possess a CRISPR/Cas system. These findings confirm the particular place of this species
370 among CoNS as Rossi *et al.*²⁶ showed that only 9% (i.e. 3/34) and 3% (i.e. 1/32) of
371 *S. epidermidis* and *Staphylococcus haemolyticus* strains respectively presented characteristic
372 CRISPR/Cas systems. Of note, the CRISPR/Cas system was only recovered in the CC1 strains.
373 This high percentage may be explained by the fact that most of the sequenced genomes
374 belonged to CC1 strains (9/21, i.e. 43%). This CC is currently over-represented in NCBI in
375 comparison with other CCs (CC3 [5/21, i.e. 24%], CC5 [1/21, i.e. 4%], CC6 [4/21, i.e. 19%])

376 and CC7 [2/21, i.e. 10%]). It would be interesting to sequence the whole genome of strains of
377 CC2 and CC4 as none is available among the 21 genomes in the GenBank database.

378 Despite CRISPR/Cas systems, the CC1 strains contained more prophage regions and plasmids
379 than others, especially than CC3 strains which contained only one plasmid. We identified in the
380 CC1 *cas* locus, nine genes required to form a ribonucleotide complex and for crRNA
381 maturation²⁷. However, a frameshift was observed within the *csm6* gene that might impact the
382 initial role of the CRISPR/Cas system. Indeed, in *S. epidermidis*, the protein Csm6 was found
383 essential for preventing plasmid transformation^{28,29}. Further investigations on a larger collection
384 of *S. lugdunensis* genomes and a functional characterization are needed to understand the
385 prevention capacity of the CRISPR/Cas against MGE acquisition. Another hypothesis would
386 be that MGE may have been present before the acquisition of the CRISPR/Cas region.

387 While the CC3 strains do not exhibit CRISPR/Cas systems or RMS, no foreign DNA was
388 detected within their genome. This can be compared with the great difficulties encountered to
389 genetically manipulate *S. lugdunensis*. The rare mutant strains published have been constructed
390 only following introduction of plasmids by protoplast transformation^{30,31} or by electroporation
391 of *Escherichia coli* strains expressing *hsdM/hsdS* RMS genes of *S. lugdunensis* and
392 interestingly, despite a new electroporation protocol, Heilbronner *et al.*³² highlighted that CC3
393 and CC4 isolates were harder to transform than CC1, CC2 and CC5 strains.

394 This suggests that other mechanisms may play a role of barrier against the acquisition of MGE
395 in *S. lugdunensis*. Bacteria can display the equivalent of physical barriers such as extracellular
396 polysaccharidic substances, including capsules to prevent phage infection^{33,34}. Thus, the
397 hypothesis could be that *cap8* capsule of *S. lugdunensis* CC3 strains could act as a protective
398 barrier. Lambe *et al.*³⁵ have described the presence of a glycocalyx surrounding *S. lugdunensis*
399 cells, but to date very little is known about extracellular polysaccharidic and proteinaceous
400 structures of *S. lugdunensis*^{1,36}.

401 Despite the presence of CRISPR/Cas immunity systems and several type I and type II RMS,
402 most (15/21, i.e. 71%) of the *S. lugdunensis* strains have acquired MGE. In this study, five
403 prophage regions and one plasmid were newly identified. Gene content analysis of these MGE
404 revealed the presence of a tetracyclin resistance-encoding plasmid in one strain. Fosfomycin,
405 penicillin and fusidic acid resistance-associated genes were also identified with a chromosomal
406 location. Fosfomycin resistance-associated gene was highly conserved among the 21 strains
407 while the distribution of penicillin (33%, i.e. 7/21) and fusidic acid (< 1%, i.e. 1/21) resistance
408 genes was variable between isolates. This finding is in accordance with the results of McHardy

409 *et al.*³⁷ who showed that the *blaZ* gene was detected in 39.1% (36/92) of oxacillin-susceptible
410 *S. lugdunensis* isolates. The presence of the *blaZ* gene does not follow a clonal distribution as
411 it was found in several CC1 and CC3 strains. On the other hand, several studies reported the
412 emergence of methicillin resistance mostly for Asian *S. lugdunensis* isolates^{38–40}, but from the
413 21 genomes of *S. lugdunensis*, no methicillin resistance gene carried by a SCC*mec* element was
414 identified. The low efficiency of transformation of this species seems to be the most likely
415 hypothesis to explain the highly conserved antibiotic susceptibility to most anti-staphylococcal
416 antibiotics^{37,41}. CRISPR/Cas, RMS, thick cell wall and capsules are all factors which may act
417 as barriers against integrations of foreign DNA carrying antibiotic resistance determinants.

418 Given the homologies found, the prophage regions identified here probably originated from
419 *S. aureus*, *S. epidermidis* and *S. hominis*, and therefore suggest a probable horizontal gene
420 transfer between these staphylococcal species. No gene already associated with staphylococcal
421 virulence has been identified within these regions, but several ones were potentially interesting
422 because they encode a bifunctional autolysin (with CHAP amidase and
423 N-acetylglucosaminidase domains) or a LytN cell wall hydrolase (with a N-acetylmuramoyl-
424 L-alanine amidase domain) that might be involved in the regulation of cell wall remodeling and
425 growth, autolysis and virulence like AtIL in *S. lugdunensis*^{42,43}.

426 Prophage regions identified also harbored a gene encoding a LexA repressor protein. LexA is
427 an accessory-regulator of the bacterial DNA repair SOS response but also a factor encoded for
428 and by the prophage itself. Thus, LexA can regulate the expression of genes carried by MGE
429 and lead to genetic rearrangements, movement of MGE and prophage induction⁴⁴. It was in
430 particular shown that antibiotic-induced SOS response promoted horizontal dissemination of
431 MGE carrying virulence factors in staphylococci⁴⁵.

432 A genomic island of 16 kb which carried fusidic acid resistance *fusB* gene and *vapE* gene was
433 identified in two genomes of *S. lugdunensis*, which might contribute to bacterial adaptation.
434 This genomic island is homologous to a genomic island of 16.6 kb including *fusB* gene and
435 named *Staphylococcus aureus* resistance island SaRIfusB⁴⁶. This MGE was described in an
436 epidemic clone CS6 of *S. aureus* causing impetigo⁴⁷. In *S. epidermidis*, phage-related islands
437 (SEPIfusB-857 and 14.1.R1SE) carrying putative virulence gene *vapE* and resistance gene *fusB*
438 were identified in fusidic acid-resistant isolates⁴⁸. The *vapE* gene was initially found in virulent
439 *Dichelobacter nodosus* causing severe ovine infection⁴⁹. Thereby, identification of these genes
440 in the genomes of *S. lugdunensis* suggest that genomic island, other than MGE as bacteriophage

441 or plasmids, could also contribute to the acquisition of antibiotic resistance and virulence genes
442 in this species even if the phenomenon is probably rare¹⁸.

443 In this study, no specific virulence determinant was identified in *S. lugdunensis* pathogenic
444 strains compared to the carriage strains. Nevertheless, a region of 13 kb partially homologous
445 to pathogenicity islands SaPITokyo11212 and SaPIvm60 of *S. aureus* was found in the
446 *S. lugdunensis* strains SL13 (causing infective endocarditis) and VISLISI_33 (causing liver
447 abscess). In *S. aureus*, SaPIs play a important role in the genomic evolution, leading to an
448 increase in its ability to cause severe infections by acquisition of numerous virulence
449 determinants such as pore-forming toxins and superantigens^{50,51}. An Ear penicillin binding
450 protein and a FhuD ferrichrome ABC transporter were carried by this potential pathogenicity
451 island. These genes have already been described by Novick *et al.*⁵² in several SaPIs from
452 *S. aureus* such as SaPImw2, SaPIm4 and SaPI1 where they are generally associated to genes
453 encoding Seq, Sek and TSST-1 toxins⁵¹. However, no enterotoxin, superantigen or toxic shock
454 syndrome toxin 1 was found in this region of 13 kb in *S. lugdunensis*. In *S. aureus*, the Ear
455 protein may act as a superantigenic exoprotein which could play roles in virulence⁵³ and was
456 found to be up-regulated only in human blood or serum⁵⁴. Otherwise, the ferric hydroxamate
457 uptake D (FhuD) protein could be implicated in the iron acquisition by *S. lugdunensis* as it can
458 promote ferrisiderophore import in *S. aureus*⁵⁵ and facilitate heme uptake in blood and serum⁵⁴.

459 The development of bacterial infections is determined by the presence and the expression of
460 particular genes (so-called virulence factors) but also relies on host factors. To better
461 characterize the virulence of the three carriage strains from those isolated from invasive or
462 severe infections, we have for the first time implemented a *G. mellonella* infection model. This
463 model has been used to characterize the virulence and pathogenesis of a wide range of
464 microorganisms^{56,57}. Interestingly, our results showed that this is a relevant model for assessing
465 the *in vivo* *S. lugdunensis* virulence as the SSTI-associated strain (SL55) and especially the
466 causing infective endocarditis strain (SL13) were significantly more lethal than the four others.
467 Although insect lack an adaptative immune response, their innate immune response still retains
468 remarkable similarities with the immune response in vertebrates; this allows to obtain relevant
469 information about the infection process. In addition, the infected larvae can be incubated at
470 37°C, which favors the study of human pathogenic bacteria⁵⁷. Of course, these results must be
471 confirmed by the test of a larger number of isolates. Although it will not replace mammalian
472 models, the larvae of *G. mellonella* provide a rapid and cost-effective alternative to collect
473 initial data.

474 The role of the different virulence factors on the clinical characteristics (the pathogenicity) of
475 *S. lugdunensis* is to date not fully understood (the precise relationship between the few virulence
476 factors described for *S. lugdunensis* and pathogenicity is unclear). Based on genomic
477 comparison using VFDB and previously characterized staphylococcal virulence factors, no
478 significant difference in virulence gene content was observed from carriage and pathogenic
479 strains. Notwithstanding, we identified CC-dependent variations in several loci potentially
480 associated to the virulence of *S. lugdunensis*.

481 No link between *S. lugdunensis* *agr* type and specific clinical infection entities has been shown
482 to date but our study shows that all CC3 strains belonged to the *agr* type II whereas all other
483 strains were associated to the allele *agr* type I. Tseng *et al.*⁵⁸ showed that the seven of 11 isolates
484 collected from bacteremia which belonged to *agr* type II presented a higher hemolytic and
485 protease activity than the four *agr* type I isolates, all isolates forming a similar level of biofilm.
486 On the other hand, many studies have attempted to associate *agr* types with one or more of
487 biological characteristics of *S. aureus*. It was shown that *agr* groups vary by clonal lineages
488 distribution, antibiotic resistance profile, biofilm formation, expression of virulence factors, and
489 autoinducing peptide structures⁵⁹. In this way, *agr* group I is usually found in clonal lineages
490 CC8, CC25, CC22, CC45, and CC395. CC5, CC12, and CC15 isolates usually harbor *agr* group
491 II, CC30 is often characterized by *agr* group III, and CC121 harbors *agr* group IV⁶⁰. In addition,
492 Jarraud *et al.* have shown that *S. aureus* endocarditis strains were mainly associated to the
493 phylogenetic group AF2 (*agr* groups I or II) while the phylogenetic group AF1 (*agr* group IV)
494 strains are mainly responsible for generalized exfoliative syndromes and bullous impetigo⁶¹.
495 Similarly, *agr*-specificity groups in *S. epidermidis* are also associated with particular clinical
496 infections^{62,63}. This putative correlation between *agr* types and *S. lugdunensis* lineages must be
497 confirmed by the study of a larger collection of isolates from various CC described as well as
498 the potential impact of this association on the regulation of virulence could be explored by a
499 functional analysis of the *agr* locus.

500 Interestingly, the CC3 strains are also distinguished from all the others by the allele of the *cap*
501 gene (*cap8*). Outside the study of Lambe *et al.*³⁵ that reported that the transmission electron
502 microscopy of pure cultures of selected strains showed that *S. lugdunensis* possessed
503 glycocalyx, no work has described the prevalence, composition, diversity and role of the
504 capsule in this species. In *S. aureus*, several studies have shown that serotype 5 or 8 capsular
505 isolates account for approximately 80% of the isolates recovered from humans, the remaining
506 isolates being not encapsulated or nontypeable⁶⁴⁻⁶⁶. Like *S. aureus*⁶⁷, the genes specifying

507 capsular polysaccharides in *S. lugdunensis* are encoded by a region with highly conserved
508 *capABCDEFGHIJKLMN*OP and four genes, *capHIJK*, specifying chemical diversity among
509 serotypes.

510 Heilbronner *et al.*⁶⁸ described seven genes encoding “*S. lugdunensis* surface proteins” (SIs)
511 named SIsA to SIsG within the genome of *S. lugdunensis* N920143 strain. Here, we have
512 identified that several SIs proteins were absent or potentially nonfunctional in some *S.*
513 *lugdunensis* strains. Indeed, the SIsD and SIsE proteins are probably nonfunctional in most of
514 the CC1 strains due to the presence of nonsense mutation and frameshift in the gene sequences.
515 Conversely, all the CC1 strains presented the gene encoding the SIsA proteins whereas it was
516 absent from all the CC3 strains. These CC-dependent differences, if confirmed on a larger
517 number of strains, could result in variation for the biofilm-forming and adhesion capacities
518 between isolates. Indeed, SIsA protein contain protein domains and structural organization
519 homologous to a known cell wall-associated biofilm Bhp⁶⁹. Bhp has been shown to promote
520 primary attachment to abiotic surfaces as well as intercellular adhesion during biofilm
521 formation^{70,71}. In addition, the SIsD protein has sequence and putative structural similarity to
522 the members of the Sdr family of adhesins such as SdrF and SdrG necessary and sufficient for
523 the attachment of *S. epidermidis* to surfaces coated with human fibrinogen or type I
524 collagen^{70,72}. A functional characterization of these SIs proteins could be helpful to explore
525 whether they are involved in the ability of *S. lugdunensis* to bind medical devices or biotic
526 surfaces, such as host tissue^{5,36}.

527 To our knowledge, this study provides the first description of the genetic diversity of the *ess*
528 locus from *S. lugdunensis*. Surprisingly, this locus showed a considerable diversity among only
529 21 strains for a species whose population structure is clonal^{15,16}.

530 Homologous genes of those encoding the six core components (EsxA, EssA, EssB, EssC, EsaA,
531 and EsaB) of the *S. aureus* T7SS machinery⁷³ were identified and conserved among the
532 genomes of the 21 strains of *S. lugdunensis*. Moreover, the same protein domains were
533 identified between the six core proteins of *S. aureus* and *S. lugdunensis*. At the molecular level,
534 the *S. aureus* T7SS is composed of four membrane-associated proteins (EsaA, EssA, EssB, and
535 EssC), three soluble cytosolic proteins (EsaB, EsaE, and EsaG), and five secreted virulence
536 factors (EsxA, EsxC, EsxB, EsxD, and EsaD)⁷⁴. Through a poorly understood mechanism, the
537 membrane and cytoplasmic proteins cooperate to export the effectors proteins during host
538 infection. This system secretes in particular two small protein ESAT-6 (also termed EsxA) and
539 Culture Filtrate Protein-10 (CFP-10) (also termed EsxB). These two proteins are members of

540 the WXG100 protein superfamily characterized by two small helical hairpins linked with a
541 central WXG motif (Tryptophan – X – Glycine). A total of four membrane proteins, EsaA,
542 EssA, EssB, and the Ftsk-SpoIIIE ATPase EssC, and a cytoplasmic ubiquitin-like fold protein
543 EsaB⁷⁵, has been demonstrated to be necessary for translocation of EsxA and EsxB across the
544 bacterial membrane⁷³.

545 Otherwise, the secreted EsxA protein of both species contain conserved C-terminal
546 hydrophobic patterns HxxxD/ExhxxxH where H stands for highly conserved hydrophobic and
547 h for less conserved hydrophobic residues, x for any amino acid and D/E for either aspartic or
548 glutamic acids, respectively. This motif constitutes a key component of *S. aureus*
549 T7SS-substrate recognition⁷⁶. In addition, *S. lugdunensis* possess a gene encoding a protein
550 with LXG domain which is found in a group of polymorphic toxins predicted to use T7SS
551 pathway^{77,78}. Thereby, these data highlight a possible extracellular translocation of these
552 proteins. In *S. aureus*, proteins with domain LXG are involved in immune evasion and virulence
553 as the protein EsxX (Dai 2017) or in the bacterial competition as the EsaD nuclease^{19,79}. Of
554 note, no gene homologous to the *esaD* gene encoding the anti-bacterial toxin with a C-terminal
555 nuclease domain was identified within the genome of *S. lugdunensis* strains.

556 A great diversity of the genetic organization of the *ess* locus was described by Warne *et al.* in
557 153 *S. aureus* isolates belonging to multiple CC⁷⁴. They showed that the *S. aureus* *essC* gene
558 exhibited four sequence variants, each associated to a specific cluster of four to six different
559 genes. In *S. aureus*, the *essC1* variant was the most prevalent among the 153 isolates and was
560 strongly associated to several CCs (CC1, 5, 7, 8, 9, 25, 41, 51 and 88). The second most
561 prevalent variant was the variant *essC3* associated to CC30 and the ST239 while the *essC4*
562 variant was only found in CC22 isolates including the reference *S. aureus* strain HO 5096 0412.
563 The *essC2* variant was identified in the CC15 and ST398 strains as the reference strain
564 RJ-ST398. In *S. lugdunensis*, two *essC* variants were also observed, for which a strong
565 association with MLST CCs was noted. The *essC2* variant was identified in all the CC3 strains
566 and in the C33 strain of CC7 while the *essC1* variant was the most prevalent and found in 15
567 strains belonging to the CC1, CC5, CC6 and CC7. We also noticed that the *ess* locus of the
568 CC3 and the C33 strains were closer to the *ess* locus of *S. aureus* strain RJ-ST398 (*S. aureus*
569 *essC2* variant) whereas the other strains were more related to the *ess* locus of
570 *S. aureus* strain HO 5096 0412 (*S. aureus* *essC4* variant). These results suggest that
571 *S. lugdunensis* strains may have acquired these regions independently and from difference
572 sources.

573 Genetic diversity of the *ess* locus of *S. lugdunensis* between CCs and within a same lineage
574 concerned also the organization of the locus itself downstream the modules 1 and 2. Effectively,
575 eight genetic organization CC-dependent were observed, based on (i) a variable number of
576 genes encoding uncharacterized proteins, (ii) the presence of a cluster of seven genes specific
577 to the CC1 and CC3, (iii) the insertion of a transposon in CC1 strains, and (iv) the variable copy
578 number of genes encoding DUF600 proteins. The mechanisms of occurrence of these variations
579 inter- and intra-CC are unclear and need further investigations in a larger collection of *S.*
580 *lugdunensis* strains, and especially in strains belonging to the CC2 and CC4 which could present
581 several others organizations of this locus.

582 We describe here homologs of a large number of genes encoding the T7SS in *S. lugdunensis*
583 which shares a high number of clinical features with *S. aureus*, including ability to form
584 abscesses. Thus, *S. lugdunensis* possesses all the components required to a functional Ess
585 machinery which could play a key role in the virulence of this bacterial species. In *S. aureus*,
586 the T7SS promotes chronically long-term persistence infections with abscesses lesions and
587 lethal outcomes in mouse model^{20,21}. The mechanisms of *S. lugdunensis* virulence are poorly
588 understood but, as it possesses all the components required to a functional Ess machinery, the
589 T7SS could play a key role in abscess development and thereby a role in the virulence of this
590 bacterial species.

591 Taken together, these results confirm the “closed” state of the pan-genome of *S. lugdunensis*,
592 but identification of MGE carrying putative virulence and resistance genes suggests some
593 possible horizontal gene transfers between *S. lugdunensis* and other staphylococcal species. No
594 specific variation or virulence determinant was associated to the pathogenic strains but the
595 distribution of potential genomic islands constitute an interesting option to explore on a broader
596 range of strains. Here, we found several CC-dependent variations especially for the CC3 strains
597 which showed specific variations in several loci potentially associated with virulence. Further
598 phenotypic and functional studies are needed to fully characterize this particular CC and to
599 evaluate the role of the T7SS in the virulence of *S. lugdunensis*. Finally, whole genome
600 sequencing of CC2 and CC4 strains are needed for a complete picture of the genetic diversity
601 of *S. lugdunensis*.

602
603
604

605 **Methods**

606 **Pacific biosciences sequencing**

607 The complete genome of three pathogenic and three carriage strains of *S. lugdunensis*
608 previously characterized by Multi-Locus Sequence Typing (MLST) and Multi-Virulence Locus
609 Sequence Typing^{14,15} was sequenced using the Single Molecule sequencing Pacific Biosciences
610 (Pacbio) technology⁸⁰. The strains SL13, SL29 and SL55 were respectively isolated from
611 infective endocarditis, vascular prosthesis infection and skin and soft tissues infection. The
612 carriage strains were isolated from groin (SL118 and 122) and toe (SL117)^{81,82}. Four strains
613 belonged to the clonal complexe (CC) 1 (SL13, SL29, SL117, SL122) and two to the CC6, as
614 defined by MLST¹⁴ (Table1).

615

616 **Computational analysis**

617 In addition to the six genomes sequenced by Pacbio, 15 complete genomes published or
618 available were obtained from GenBank for comparative analyses. All complete genomes were
619 annotated using the software Prokka v1.13⁸³. Genomes were compared using orthologsorter
620 (www.git.facom.ufms.br/bioinfo/orthologsorter.git) to generate ortholog clustering sequences
621 among the 21 strains and to define the pan-genome, the core-, and the accessory-genome of
622 *S. lugdunensis*. A similarity matrix of complete genomes and a phylogenetic tree were created
623 based on the neighbor-joining method thanks to the software BioNumerics v7.6. Sequence Type
624 (ST) and Clonal Complexe (CC) defined by MLST were identified by extracting sequences of
625 the genes *aroE*, *dat*, *ddl*, *gmk*, *ldh*, *recA* and *yqiL*¹⁴. ST and CC were obtained by submission
626 of the gene sequences into the international MLST database
627 (<http://bigsdbs.web.pasteur.fr/staphlugdunensis/>). The presence of genes encoding restriction
628 enzymes among was predicted by comparison and Blastn with datasets on Restriction Enzyme
629 Database (REBASE)⁸⁴.

630 Integrated prophage regions were identified and annotated by using the online PHAge Search
631 Tool Enhanced Release (PHASTER)⁸⁵ (<http://phaster.ca>) with default parameters for the six
632 genomes sequenced by Pacbio, the complete genome sequences of the five strains FDAARGOS
633 and the strain Klug, available on the NCBI, and for which no phage was described to date.
634 PHASTER provided the region length and position of the prophages on the chromosome, their
635 GC content and the most common related phages. The known prophage sequences Φ SL1⁶⁸,
636 Φ SL2, Φ SL3, Φ SL4 and Φ SL5¹⁷ were compared to the prophages regions identified and Mauve
637 alignments were used to confirm PHASTER results.

638 Plasmids identification was done by using Prokka v1.13 for annotation of extra-chromosomal
639 contigs obtained by Pacbio sequencing and was compared to the plasmid sequences available
640 on the NCBI (pVISLISI_1, pVISLISI_2, pVISLISI_3, pVISLISI_4 pVISLISI_5¹⁷ and to
641 unnamed plasmid from the strain FDAARGOS381). Nucleotide sequence similarities were
642 investigated by Blastn to identify the closest related plasmid.

643 Virulence factors were identified using the Virulence Factors database (VFDB)
644 (<http://www.mgc.ac.cn/VFs>)⁸⁶ and tBlastn searches were made to compare the presence or
645 absence of known virulence genes of staphylococci.

646 Comparisons of the *agr* and *cap* hypervariable regions were performed by extracting and
647 aligning the *agr* and *cap* loci with MUSCLE. The allelic *agr* type-I (*agr-1Sl*, accession number:
648 AF173933) and type-II (*agr-2Sl*, accession number:AF346728) described by Dufour *et al.*⁸⁷
649 were used as reference. The hypervariable regions (*capH* to *capK*) from the
650 *S. aureus cap8* and *cap5* loci were downloaded from the NCBI database. Phylogenetic trees
651 were constructed using the maximum likelihood method implemented in the PhyML program
652 v3.0.

653 Resistance genes were identified by using ResFinder v3.1
654 (<https://cge.cbs.dtu.dk/services/ResFinder>)⁸⁸. A cutoff of 80% minimum length and 80%
655 identity were used to search for resistance genes in the 21 genomes and plasmids of
656 *S. lugdunensis*.

657 Comparisons of individual sequences of the *ess* locus were performed on the 21 genomes of
658 *S. lugdunensis* by using Blastn and Blastp on the NCBI, and were facilitated by using the
659 Artemis Comparison Tool⁸⁹. T7SS genes and proteins were identified from the available
660 genomes of reference strains HO 5096 0412 (Accession number HE861097)⁹⁰ and RJ-ST398
661 (Accession number AM990992)⁹¹. Pfam database was also used to identify protein domains.

662 ***G. mellonella* infection experiments**

663 Infection of *G. mellonella* larvae with *S. lugdunensis* was performed as previously described
664 by Michaux *et al.*⁹². Briefly, using a syringe pump (KD Scientific, Holliston, MA), larvae (about
665 0.3 g and 3 cm in length) were infected subcutaneously with washed cells of *S. lugdunensis*
666 from an overnight culture in BHI, with 3(±2)×10⁶ CFU per larvae administered in 10 µl of
667 sterile saline buffer. In each test, 10 insects were infected, and the experiments were repeated
668 at least four times. Larval killing was then monitored at 1, 2 and 3 days post-infection. Sterile
669 saline buffer was also tested under the same conditions as control. Results were analyzed using

670 one-way analysis of variance with a Bonferroni correction with the free available “R” software.
671 For all comparisons, a p value of less than 0.05 was considered significant.

672

673 **Acknowledgements**

674 This study was supported by an educational grant for Jérémie Lebeurre from the Charles Nicolle
675 Foundation.

676

677 **Author contributions statement**

678 JL, SD, PF and MPC designed the study. JL, SD and AP analyzed the data. JC and MA realized
679 the *Galleria mellonella* infection experiments. JL and MPC wrote the manuscript.

680

681 **Additional information**

682 Competing financial interests: the authors declare that they have no competing interests.

683

684 **References**

- 685 1. Frank, K. L., del Pozo, J. L. & Patel, R. From Clinical Microbiology to Infection
686 Pathogenesis: How Daring To Be Different Works for *Staphylococcus lugdunensis*. *Clin.*
687 *Microbiol. Rev.* **21**, 111–133 (2008).
- 688 2. Arias, M. *et al.* Skin and soft tissue infections caused by *Staphylococcus lugdunensis*:
689 Report of 20 cases. *Scand. J. Infect. Dis.* **42**, 879–884 (2010).
- 690 3. Bocher, S., Tonning, B., Skov, R. L. & Prag, J. *Staphylococcus lugdunensis*, a Common
691 Cause of Skin and Soft Tissue Infections in the Community. *J. Clin. Microbiol.* **47**, 946–950
692 (2009).
- 693 4. Non, L. R. & Santos, C. A. Q. The occurrence of infective endocarditis with
694 *Staphylococcus lugdunensis* bacteremia: A retrospective cohort study and systematic review. *J.*
695 *Infect.* **74**, 179–186 (2017).
- 696 5. Argemi, X. *et al.* *Staphylococcus lugdunensis* small colony variant conversion resulting
697 in chronic prosthetic joint infection. *Médecine Mal. Infect.* **47**, 498–501 (2017).
- 698 6. Douiri, N. *et al.* *Staphylococcus lugdunensis* : a virulent pathogen causing bone and joint
699 infections. *Clin. Microbiol. Infect.* **22**, 747–748 (2016).
- 700 7. Lourtet-Hascoët, J., Bicart-See, A., Félicé, M. P., Giordano, G. & Bonnet, E.
701 *Staphylococcus lugdunensis*, a serious pathogen in periprosthetic joint infections: comparison
702 to *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Int. J. Infect. Dis.* **51**, 56–61 (2016).
- 703 8. Shah, N. B. *et al.* Laboratory and clinical characteristics of *Staphylococcus lugdunensis*
704 prosthetic joint infections. *J. Clin. Microbiol.* **48**, 1600–1603 (2010).
- 705 9. Anguera, I. *Staphylococcus lugdunensis* infective endocarditis: description of 10 cases
706 and analysis of native valve, prosthetic valve, and pacemaker lead endocarditis clinical profiles.
707 *Heart* **91**, e10–e10 (2005).
- 708 10. Ebright, J. R., Penugonda, N. & Brown, W. Clinical experience with *Staphylococcus*
709 *lugdunensis* bacteremia: a retrospective analysis. *Diagn. Microbiol. Infect. Dis.* **48**, 17–21
710 (2004).
- 711 11. Sabe, M. A., Shrestha, N. K., Gordon, S. & Menon, V. *Staphylococcus lugdunensis* : a
712 rare but destructive cause of coagulase-negative staphylococcus infective endocarditis. *Eur.*
713 *Heart J. Acute Cardiovasc. Care* **3**, 275–280 (2014).
- 714 12. Ho, P.-L. *et al.* Carriage niches and molecular epidemiology of *Staphylococcus*
715 *lugdunensis* and methicillin-resistant *S. lugdunensis* among patients undergoing long-term renal
716 replacement therapy. *Diagn. Microbiol. Infect. Dis.* **81**, 141–144 (2015).
- 717 13. Yeh, C.-F. *et al.* Molecular Characteristics of Disease-Causing and Commensal
718 *Staphylococcus lugdunensis* Isolates from 2003 to 2013 at a Tertiary Hospital in Taiwan. *PLOS*
719 *ONE* **10**, e0134859 (2015).
- 720 14. Chassain, B. *et al.* Multilocus Sequence Typing Analysis of *Staphylococcus lugdunensis*
721 Implies a Clonal Population Structure. *J. Clin. Microbiol.* **50**, 3003–3009 (2012).
- 722 15. Didi, J., Lemee, L., Gibert, L., Pons, J.-L. & Pestel-Caron, M. Multi-Virulence-Locus
723 Sequence Typing of *Staphylococcus lugdunensis* Generates Results Consistent with a Clonal
724 Population Structure and Is Reliable for Epidemiological Typing. *J. Clin. Microbiol.* **52**, 3624–
725 3632 (2014).
- 726 16. Dahyot, S. *et al.* Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA)
727 and Tandem Repeat Sequence Typing (TRST), helpful tools for subtyping *Staphylococcus*
728 *lugdunensis*. *Sci. Rep.* **8**, (2018).
- 729 17. Argemi, X. *et al.* Whole genome sequencing of 7 strains of *Staphylococcus lugdunensis*
730 allows identification of mobile genetic elements. *Genome Biol. Evol.* (2017).
731 doi:10.1093/gbe/evx077
- 732 18. Argemi, X. *et al.* Comparative genomic analysis of *Staphylococcus lugdunensis* shows
733 a closed pan-genome and multiple barriers to horizontal gene transfer. *BMC Genomics* **19**,

- 734 (2018).
- 735 19. Cao, Z., Casabona, M. G., Kneuper, H., Chalmers, J. D. & Palmer, T. The type VII
736 secretion system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor
737 bacteria. *Nat. Microbiol.* **2**, 16183 (2016).
- 738 20. Burts, M. L., Williams, W. A., DeBord, K. & Missiakas, D. M. EsxA and EsxB are
739 secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus*
740 *aureus* infections. *Proc. Natl. Acad. Sci.* **102**, 1169–1174 (2005).
- 741 21. Burts, M. L., DeDent, A. C. & Missiakas, D. M. EsaC substrate for the ESAT-6
742 secretion pathway and its role in persistent infections of *Staphylococcus aureus*. *Mol.*
743 *Microbiol.* **69**, 736–746 (2008).
- 744 22. Kneuper, H. *et al.* Heterogeneity in *ess* transcriptional organization and variable
745 contribution of the Ess/Type VII protein secretion system to virulence across closely related *S*
746 *taphylococcus aureus* strains: Type VII secretion in *S. aureus*. *Mol. Microbiol.* **93**, 928–943
747 (2014).
- 748 23. Poitevin-Later, F., Vandenesch, F., Dyke, K., Fleurette, J. & Etienne, J. Cadmium-
749 resistance plasmid in *Staphylococcus lugdunensis*. *FEMS Microbiol. Lett.* **78**, 59–63 (1992).
- 750 24. Bratcher, H. B., Bennett, J. S. & Maiden, M. C. J. Evolutionary and genomic insights
751 into meningococcal biology. *Future Microbiol.* **7**, 873–885 (2012).
- 752 25. Dagerhamn, J. *et al.* Determination of accessory gene patterns predicts the same
753 relatedness among strains of *Streptococcus pneumoniae* as sequencing of housekeeping genes
754 does and represents a novel approach in molecular epidemiology. *J. Clin. Microbiol.* **46**, 863–
755 868 (2008).
- 756 26. Rossi, C. C., Souza-Silva, T., Araújo-Alves, A. V. & Giambiagi-deMarval, M. CRISPR-
757 Cas Systems Features and the Gene-Reservoir Role of Coagulase-Negative Staphylococci.
758 *Front. Microbiol.* **8**, (2017).
- 759 27. Hatoum-Aslan, A., Maniv, I. & Marraffini, L. A. Mature clustered, regularly
760 interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism
761 anchored at the precursor processing site. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 21218–21222
762 (2011).
- 763 28. Hatoum-Aslan, A., Maniv, I., Samai, P. & Marraffini, L. A. Genetic Characterization of
764 Antiplasmid Immunity through a Type III-A CRISPR-Cas System. *J. Bacteriol.* **196**, 310–317
765 (2014).
- 766 29. Jiang, W., Samai, P. & Marraffini, L. A. Degradation of Phage Transcripts by CRISPR-
767 Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**, 710–721 (2016).
- 768 30. Marlinghaus, L. *et al.* Construction and characterization of three knockout mutants of
769 the *fbl* gene in *Staphylococcus lugdunensis*: characterization of isogenic mutants of *fbl*. *APMIS*
770 **120**, 108–116 (2012).
- 771 31. Szabados, F. *et al.* Occurrence of genes of putative fibrinogen binding proteins and
772 hemolysins, as well as of their phenotypic correlates in isolates of *S. lugdunensis* of different
773 origins. *BMC Res. Notes* **4**, 113 (2011).
- 774 32. Heilbronner, S., Hanses, F., Monk, I. R., Speziale, P. & Foster, T. J. Sortase A Promotes
775 Virulence in Experimental *Staphylococcus lugdunensis* Endocarditis. *Microbiology* (2013).
776 doi:10.1099/mic.0.070292-0
- 777 33. Wilkinsont, B. J. & Holmes', K. M. *Staphylococcus aureus* Cell Surface: Capsule as a
778 Barrier to Bacteriophage Adsorption. *Infect Immun* **4**
- 779 34. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat.*
780 *Rev. Microbiol.* **8**, 317–327 (2010).
- 781 35. Lambe, D. W., Jeffery, C., Ferguson, K. P. & Cooper, M. D. Examination of the
782 glycocalyx of four species of *Staphylococcus* by transmission electron microscopy and image
783 analysis. *Microbios* **78**, 133–143 (1994).

- 784 36. Frank, K. L., Reichert, E. J., Piper, K. E. & Patel, R. In Vitro Effects of Antimicrobial
785 Agents on Planktonic and Biofilm Forms of *Staphylococcus lugdunensis* Clinical Isolates.
786 *Antimicrob. Agents Chemother.* **51**, 888–895 (2007).
- 787 37. McHardy, I. H. *et al.* Clinical and Microbiological Aspects of β -Lactam Resistance in
788 *Staphylococcus lugdunensis*. *J. Clin. Microbiol.* **55**, 585–595 (2017).
- 789 38. Cheng, C. *et al.* Persistence of a major endemic clone of oxacillin-resistant
790 *Staphylococcus lugdunensis* sequence type 6 at a tertiary medical centre in northern Taiwan.
791 *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **36**, 72–77 (2015).
- 792 39. Yeh, C.-F. *et al.* Clinical Features, Outcomes, and Molecular Characteristics of
793 Community- and Health Care-Associated *Staphylococcus lugdunensis* Infections. *J. Clin.*
794 *Microbiol.* **54**, 2051–2057 (2016).
- 795 40. Chang, S.-C. *et al.* Characterization of two novel variants of staphylococcal cassette
796 chromosome mec elements in oxacillin-resistant *Staphylococcus lugdunensis*. *J. Antimicrob.*
797 *Chemother.* **72**, 3258–3262 (2017).
- 798 41. Higaki, S., Kitagawa, T., Morohashi, M. & Yamagishi, T. Distribution and antimicrobial
799 susceptibility of coagulase-negative staphylococci from skin lesions. *J. Int. Med. Res.* **27**, 191–
800 195 (1999).
- 801 42. Gibert, L. *et al.* The major autolysin of *Staphylococcus lugdunensis*, AtlL, is involved
802 in cell separation, stress-induced autolysis and contributes to bacterial pathogenesis. *FEMS*
803 *Microbiol. Lett.* **352**, 78–86 (2014).
- 804 43. Diene, S. M., Corvaglia, A. R., François, P., van der Mee-Marquet, N. & Regional
805 Infection Control Group of the Centre Region. Prophages and adaptation of *Staphylococcus*
806 *aureus* ST398 to the human clinic. *BMC Genomics* **18**, 133 (2017).
- 807 44. Fornelos, N., Browning, D. F. & Butala, M. The Use and Abuse of LexA by Mobile
808 Genetic Elements. *Trends Microbiol.* **24**, 391–401 (2016).
- 809 45. Ubeda, C. *et al.* Antibiotic-induced SOS response promotes horizontal dissemination of
810 pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.* **56**, 836–844
811 (2005).
- 812 46. O’Neill, A. J., Larsen, A. R., Skov, R., Henriksen, A. S. & Chopra, I. Characterization
813 of the epidemic European fusidic acid-resistant impetigo clone of *Staphylococcus aureus*. *J.*
814 *Clin. Microbiol.* **45**, 1505–1510 (2007).
- 815 47. Novick, R. P., Christie, G. E. & Penadés, J. R. The phage-related chromosomal islands
816 of Gram-positive bacteria. *Nat. Rev. Microbiol.* **8**, 541–551 (2010).
- 817 48. Chen, H.-J. *et al.* New Structure of Phage-Related Islands Carrying *fusB* and a Virulence
818 Gene in Fusidic Acid-Resistant *Staphylococcus epidermidis*. *Antimicrob. Agents Chemother.*
819 **57**, 5737–5739 (2013).
- 820 49. Bloomfield, G. A., Whittle, G., McDonagh, M. B., Katz, M. E. & Cheetham, B. F.
821 Analysis of sequences flanking the vap regions of *Dichelobacter nodosus*: evidence for multiple
822 integration events, a killer system, and a new genetic element. *Microbiol. Read. Engl.* **143 (Pt**
823 **2)**, 553–562 (1997).
- 824 50. Malachowa, N. & DeLeo, F. R. Mobile genetic elements of *Staphylococcus aureus*.
825 *Cell. Mol. Life Sci.* **67**, 3057–3071 (2010).
- 826 51. Novick, R. P. & Ram, G. Staphylococcal pathogenicity islands — movers and shakers
827 in the genomic firmament. *Curr. Opin. Microbiol.* **38**, 197–204 (2017).
- 828 52. Novick, R. P. & Subedi, A. The SaPIs: Mobile Pathogenicity Islands of *Staphylococcus*.
829 in *Chemical Immunology and Allergy* (ed. Marone, G.) 42–57 (KARGER, 2007).
830 doi:10.1159/000100857
- 831 53. Singh, V. K. *et al.* Phylogenetic distribution and expression of a penicillin-binding
832 protein homologue, Ear and its significance in virulence of *Staphylococcus aureus*. *J. Med.*
833 *Microbiol.* **66**, 1811–1821 (2017).

- 834 54. Malachowa, N. *et al.* Global Changes in *Staphylococcus aureus* Gene Expression in
835 Human Blood. *PLoS ONE* **6**, e18617 (2011).
- 836 55. Beasley, F. C. *et al.* Characterization of staphyloferrin A biosynthetic and transport
837 mutants in *Staphylococcus aureus*. *Mol. Microbiol.* **72**, 947–963 (2009).
- 838 56. Kavanagh, K. & Reeves, E. P. Exploiting the potential of insects for in vivo
839 pathogenicity testing of microbial pathogens. *FEMS Microbiol. Rev.* **28**, 101–112 (2004).
- 840 57. Ramarao, N., Nielsen-Leroux, C. & Lereclus, D. The insect *Galleria mellonella* as a
841 powerful infection model to investigate bacterial pathogenesis. *J. Vis. Exp. JoVE* e4392 (2012).
842 doi:10.3791/4392
- 843 58. Tseng, S.-P. *et al.* Genotypes and phenotypes of *Staphylococcus lugdunensis* isolates
844 recovered from bacteremia. *J. Microbiol. Immunol. Infect.* **48**, 397–405 (2015).
- 845 59. Tan, L., Li, S. R., Jiang, B., Hu, X. M. & Li, S. Therapeutic Targeting of the
846 *Staphylococcus aureus* Accessory Gene Regulator (*agr*) System. *Front. Microbiol.* **9**, (2018).
- 847 60. Holtfreter, S. *et al.* Clonal distribution of superantigen genes in clinical *Staphylococcus*
848 *aureus* isolates. *J. Clin. Microbiol.* **45**, 2669–2680 (2007).
- 849 61. JARRAUD, S. *et al.* Exfoliatin-Producing Strains Define a Fourth *agr* Specificity Group
850 in *Staphylococcus aureus*. *J BACTERIOL* **182**, 6 (2000).
- 851 62. Harris, L. G. *et al.* Limitations in the use of PSM γ , *agr*, RNIII, and biofilm formation
852 as biomarkers to define invasive *Staphylococcus epidermidis* from chronic biomedical device-
853 associated infections. *Int. J. Med. Microbiol. IJMM* **307**, 382–387 (2017).
- 854 63. Hellmark, B., Söderquist, B., Unemo, M. & Nilsson-Augustinsson, Å. Comparison of
855 *Staphylococcus epidermidis* isolated from prosthetic joint infections and commensal isolates in
856 regard to antibiotic susceptibility, *agr* type, biofilm production, and epidemiology. *Int. J. Med.*
857 *Microbiol. IJMM* **303**, 32–39 (2013).
- 858 64. Arbeit, R. D., Karakawa, W. W., Vann, W. F. & Robbins, J. B. Predominance of two
859 newly described capsular polysaccharide types among clinical isolates of *Staphylococcus*
860 *aureus*. *Diagn. Microbiol. Infect. Dis.* **2**, 85–91 (1984).
- 861 65. Cocchiari, J. L. *et al.* Molecular characterization of the capsule locus from non-typeable
862 *Staphylococcus aureus*. *Mol. Microbiol.* **59**, 948–960 (2006).
- 863 66. Roghmann, M. *et al.* Epidemiology of capsular and surface polysaccharide in
864 *Staphylococcus aureus* infections complicated by bacteraemia. *J. Hosp. Infect.* **59**, 27–32
865 (2005).
- 866 67. O’Riordan, K. & Lee, J. C. *Staphylococcus aureus* capsular polysaccharides. *Clin.*
867 *Microbiol. Rev.* **17**, 218–234 (2004).
- 868 68. Heilbronner, S. *et al.* Genome sequence of *Staphylococcus lugdunensis* N920143 allows
869 identification of putative colonization and virulence factors: *Staphylococcus lugdunensis*
870 genome sequence. *FEMS Microbiol. Lett.* **322**, 60–67 (2011).
- 871 69. Post, V. *et al.* Comparative Genomics Study of *Staphylococcus epidermidis* Isolates
872 from Orthopedic-Device-Related Infections Correlated with Patient Outcome. *J. Clin.*
873 *Microbiol.* **55**, 3089–3103 (2017).
- 874 70. Bowden, M. G. *et al.* Identification and preliminary characterization of cell-wall-
875 anchored proteins of *Staphylococcus epidermidis*. *Microbiol. Read. Engl.* **151**, 1453–1464
876 (2005).
- 877 71. Cucarella, C. *et al.* Bap, a *Staphylococcus aureus* surface protein involved in biofilm
878 formation. *J. Bacteriol.* **183**, 2888–2896 (2001).
- 879 72. Arrecubieta, C., Lee, M.-H., Macey, A., Foster, T. J. & Lowy, F. D. SdrF, a
880 *Staphylococcus epidermidis* surface protein, binds type I collagen. *J. Biol. Chem.* **282**, 18767–
881 18776 (2007).
- 882 73. Unnikrishnan, M., Constantinidou, C., Palmer, T. & Pallen, M. J. The Enigmatic Exs
883 Proteins: Looking Beyond Mycobacteria. *Trends Microbiol.* **25**, 192–204 (2017).

884 74. Warne, B. *et al.* The Ess/Type VII secretion system of *Staphylococcus aureus* shows
885 unexpected genetic diversity. *BMC Genomics* **17**, (2016).

886 75. van den Ent, F. & Löwe, J. Crystal structure of the ubiquitin-like protein YukD from
887 *Bacillus subtilis*. *FEBS Lett.* **579**, 3837–3841 (2005).

888 76. Poulsen, C., Panjikar, S., Holton, S. J., Wilmanns, M. & Song, Y.-H. WXG100 Protein
889 Superfamily Consists of Three Subfamilies and Exhibits an α -Helical C-Terminal Conserved
890 Residue Pattern. *PLOS ONE* **9**, 12 (2014).

891 77. Whitney, J. C. *et al.* A broadly distributed toxin family mediates contact-dependent
892 antagonism between gram-positive bacteria. *eLife* **6**, (2017).

893 78. Zhang, D., Iyer, L. M. & Aravind, L. A novel immunity system for bacterial nucleic
894 acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic
895 Acids Res.* **39**, 4532–4552 (2011).

896 79. Anderson, M., Aly, K. A., Chen, Y.-H. & Missiakas, D. Secretion of atypical protein
897 substrates by the ESAT-6 Secretion System of *S. taphylococcus aureus*: ESAT-6 secretion in
898 *S. aureus*. *Mol. Microbiol.* **90**, 734–743 (2013).

899 80. Gupta, P. K. Single-molecule DNA sequencing technologies for future genomics
900 research. *Trends Biotechnol.* **26**, 602–611 (2008).

901 81. Bieber, L. & Kahlmeter, G. *Staphylococcus lugdunensis* in several niches of the normal
902 skin flora. *Clin. Microbiol. Infect.* **16**, 385–388 (2010).

903 82. van der Mee-Marquet, N. *et al.* *Staphylococcus lugdunensis* Infections: High Frequency
904 of Inguinal Area Carriage. *J. Clin. Microbiol.* **41**, 1404–1409 (2003).

905 83. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**,
906 2068–2069 (2014).

907 84. Roberts, R. J., Carneiro, M. O. & Schatz, M. C. The advantages of SMRT sequencing.
908 *Genome Biol.* **14**, 405 (2013).

909 85. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool.
910 *Nucleic Acids Res.* **44**, W16-21 (2016).

911 86. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: hierarchical and refined
912 dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).

913 87. Dufour, P. *et al.* High Genetic Variability of the *agr* Locus in *Staphylococcus* Species.
914 *J. Bacteriol.* **184**, 1180–1186 (2002).

915 88. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J.
916 Antimicrob. Chemother.* **67**, 2640–2644 (2012).

917 89. Carver, T. J. *et al.* ACT: the Artemis Comparison Tool. *Bioinforma. Oxf. Engl.* **21**,
918 3422–3423 (2005).

919 90. Holden, M. T. G. *et al.* A genomic portrait of the emergence, evolution, and global
920 spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**, 653–664
921 (2013).

922 91. Wang, Y. *et al.* Role of the ESAT-6 secretion system in virulence of the emerging
923 community-associated *Staphylococcus aureus* lineage ST398. *Sci. Rep.* **6**, (2016).

924 92. Michaux, C. *et al.* SlyA Is a Transcriptional Regulator Involved in the Virulence of
925 *Enterococcus faecalis*. *Infect. Immun.* **79**, 2638–2645 (2011).

926

927

928 **Figure legends**

929 **Figure 1.** Schematic representation of the *ess* locus from genomes of the 21 strains of
930 *S. lugdunensis*. For each strain, clonal complexes (CC) are indicated to the right. The *ess* locus
931 was divided into six modules depending on the gene-based content and the relative conservation
932 defined with the Artemis comparison tool. The module 1 contained five conserved genes, *esxA*
933 to *essB*, at the 5' end of the locus one gene encodes a putative secreted protein EsxA and
934 putative other core components of the T7SS. The module 2 included the *essC* gene, of which
935 there are two variants, each associated to a distinct cluster of five to six genes. In the module 3,
936 4 and 6 contained a complex arrangement of coding sequences that encode a variable number
937 of uncharacterized proteins. The module 5 is composed of two distinct genetic organizations
938 were observed, one of them contained nine coding sequences unrelated to the T7SS carried by
939 a transposon with three coding sequences which encode hypothetical DUF600 proteins. The
940 other one comprised only two genes encoding DUF600 proteins. Orange areas highlight
941 conserved coding sequences between strains with the percentage of amino acid identity in red
942 between *S. lugdunensis* strains and in blue between *S. aureus* RJ-ST398.

943

944 **Figure 2.** Comparison of the *ess* module 1 and 2 of *S. aureus* HO 5096 0412 and RJ-ST398
945 strains with HKU09-01 and VISLISI_21 *S. lugdunensis* strains. The gene-based content and
946 the relative conservation defined with the Artemis comparison tool. Orange areas highlight
947 conserved coding sequences between strains with the percentage of amino acid identity in red.

948

949

950 **Figure 3.** Kaplan-Meier survival curves of infected larvae of *G. mellonella* after infection by
951 *S. lugdunensis* strains (SL13, SL29, SL55, SL117, SL118, and SL122).

952

953

Figure 1.

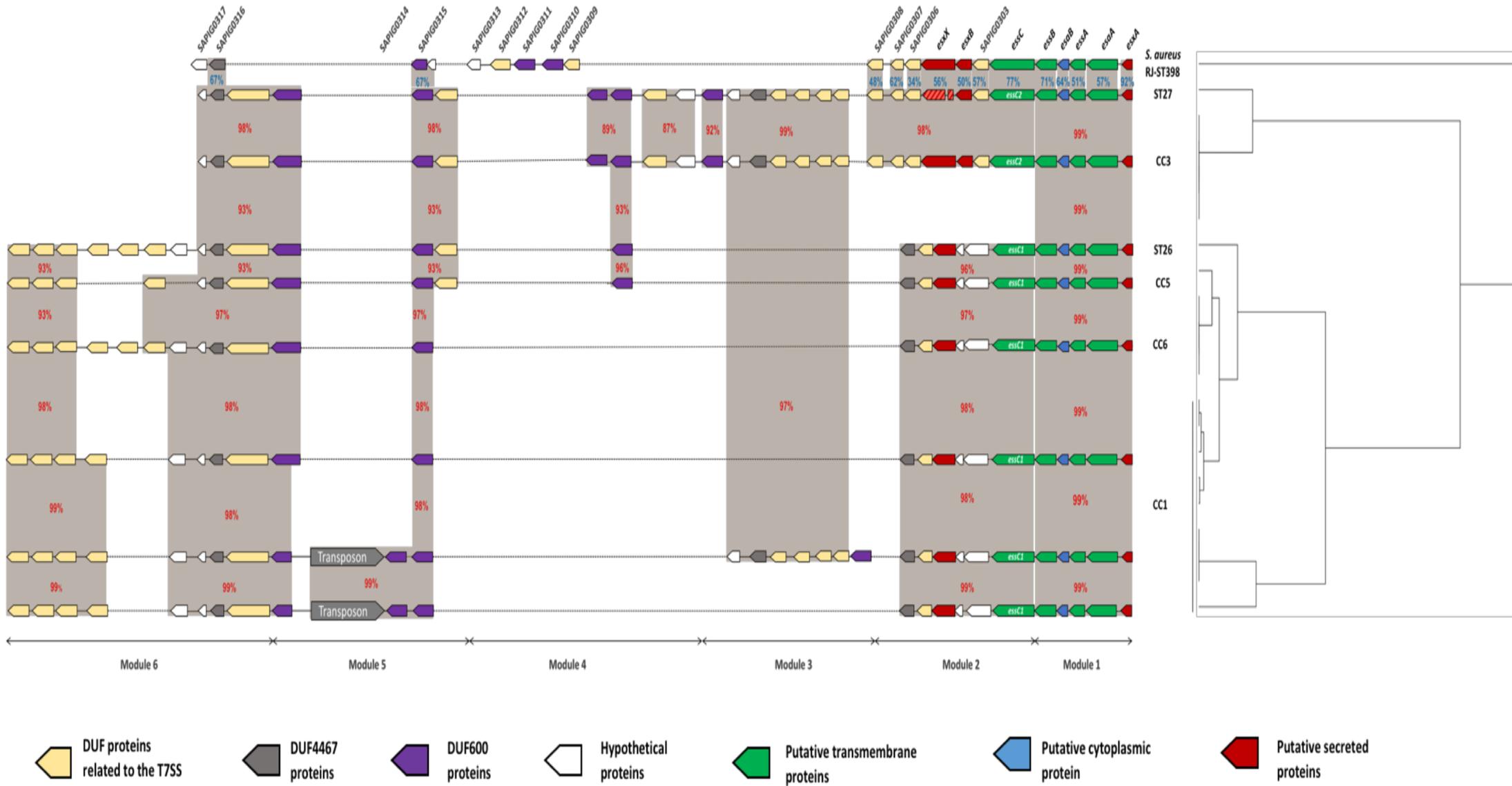


Figure 2.



Figure 3.

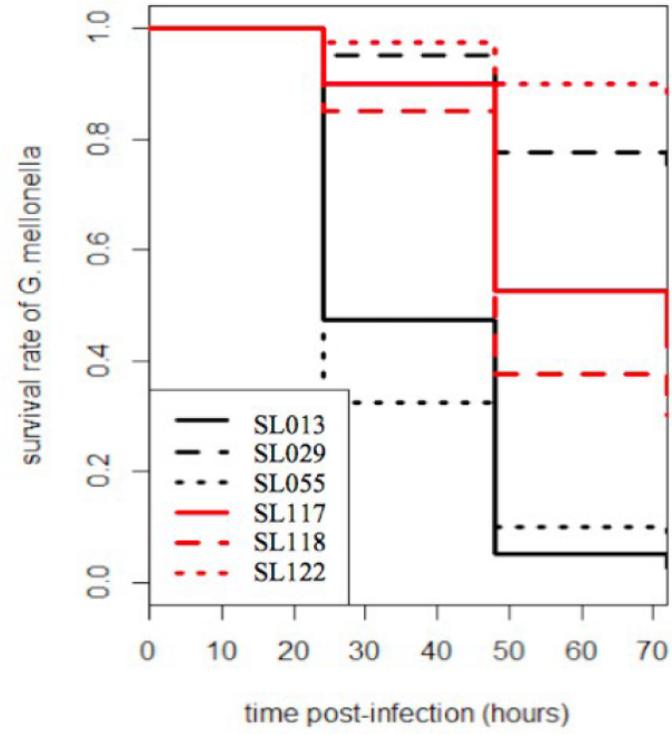


Table 1. Whole genome sequence content of the six *S. lugdunensis* strains sequenced with Pacific Biosciences technology and the 15 *S. lugdunensis* genomes available on NCBI.



Strain	Isolement	ST ^a	CC ^b	Genome size (Mb)	%GC content	Genes	Proteins	tRNA	rRNA	tmRNA	N920143 (%)	Phage	Plasmid	CRISPR/CAS ^c	MRS ^d
FDAARGOS222	SSTI	ST24	CC6	2.54	33.7	2479	2373	47	6	1	96.44122%	0	0	-	Type I
VISLISI_22	Infective endocarditis	ST24	CC6	2.58	33.8	2544	2400	70	20	1	97.40251%	1	1	-	Type II
N920143	SSTI	ST1	CC1	2.69	33.8	2675	2541	61	19	1		1	0	+	Type I
VISLISI_37	Infective endocarditis	ST1	CC1	2.66	33.8	2615	2480	61	19	1	98.76250%	1	0	+	Type I
VISLISI_27	Material	ST12	CC1	2.65	33.8	2591	2457	91	19	1	96.83107%	0	1	+	Type I
FDAARGOS141	Bursitis	ST3	CC3	2.62	33.9	2596	2465	59	19	1	95.40951%	0	0	-	-
FDAARGOS377	Abcess	ST3	CC3	2.61	33.8	2579	2452	59	15	1	95.40504%	0	0	-	-
FDAARGOS381	Abcess	ST3	CC3	2.49	33.8	2443	2338	48	4	1	95.44507%	0	1	-	-
Klug	Carriage	ST3	CC3	2.54	33.8	2468	2340	59	16	1	95.40951%	0	0	-	-
VISLISI_21	Bacteremia	ST3	CC3	2.56	33.7	2515	2396	59	7	1	94.61263%	0	0	-	-
SL55*	SSTI	ST10	CC6	2.61	33.9	2527	2394	61	19	1	96.56137%	1	0	-	Type II
SL118*	Carriage	ST10	CC6	2.53	33.8	2425	2315	52	5	1	96.59604%	1	0	-	Type II
C33	Carriage	ST27	CC7	2.66	33.9	2604	2471	61	19	1	93.97828%	0	2	+	Type II
VISLISI_25	Material	ST5	CC5	2.57	33.8	2514	2381	61	19	1	93.55531%	0	0	-	Type II
HKU09-01	SSTI	ST1	CC1	2.66	33.7	2637	2523	56	5	1	97.86123%	1	0	+	Type I
VISLISI_33	Liver abcess	ST6	CC1	2.65	33.9	2601	2469	59	19	1	97.16748%	2	1	+	Type I
SL29*	Prosthesis infection	ST1	CC1	2.59	33.6	2547	2427	60	8	1	99.32308%	1	0	+	Type I
SL122*	Carriage	ST1	CC1	2.56	33.8	2514	2381	61	19	1	99.32308%	1	0	+	Type I
SL117*	Carriage	ST1	CC1	2.6	33.8	2545	2420	56	16	1	99.34146%	2	2	+	Type I
FDAARGOS143	SSTI	ST26	CC7	2.57	33.8	2512	2379	61	19	1	96.61349%	1	0	-	Type I
SL13*	Infective endocarditis	ST6	CC1	2.58	33.7	2527	2415	52	6	1	98.21626%	1	1	+	Type I

^aSequence Type, ^bClonal Complex, ^cClustered, Regularly Interspaced, Short, Palindromic Repeat-associated (cas) genes, ^dModification-Restriction System

*Genomes sequenced in this study

SSTI: Skin and Soft Tissue Infection

Table 2. Prophage regions identified among the 21 genomes of *S. lugdunensis*.

	Prophages				
	Φ SL6	Φ SL7	Φ SL8	Φ SL9	Φ SL10
Host strain	SL13	SL29*, SL117 and SL122	SL117	SL55, SL118	F143
Length (kb)	45,9	49,9	49,1	46,6	57
GC%	34,94	35,33	35,3	34,45	35,24
Coding sequences	51	26	62	69	61
Virulence factor gene	0	0	0	0	0
Resistance gene	0	0	0	0	0
Related phage	187	Stb12	CNPx	CNPx	CNPH82
Original host	<i>S. aureus</i>	<i>S. hominis</i>	<i>S. epidermidis</i>	<i>S. epidermidis</i>	<i>S. epidermidis</i>

* phage duplication

Table 3. Plasmids identified among the 21 genomes of *S. lugdunensis*.

	Plasmids		
	pSL1	pSL2	pSL3
Host strain	SL13	SL117	SL117
Plasmid length (kb)	2,7	3,9	1,9
GC%	30,59	28,99	31,3
Coding sequences	4	5	4
Virulence factor gene	0	0	0
Resistance gene	<i>tet(k)/tet(L)</i>	<i>cadC</i>	0
Related plasmid	pSSTET1	pLUG10	pVISLISI_2
Identities with the related plasmid (%)	99%	98%	100%
Coverage (%)	100%	48%	100%

Table 4. Antibiotic resistance-associated genes identified among the 21 genomes of *S. lugdunensis* and their genomic location.

Antibiotic	Gene	Strains																			Genomic location	
		SL13	SL29	SL55	SL117	SL118	SL122	HKU09-01	N920143	V21	V22	V25	V27	V33	V37	C33	F141	F143	F222	F377		F381
Fosfomycin	<i>fosB</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Chromosome
Penicillin	<i>blaZ</i>	+	-	-	-	-	-	+	-	+	-	+	+	-	-	+	-	-	+	-	-	Transposon
Fusidic acid	<i>fusD</i>	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	Genomic island
Tetracyclin	<i>tetK</i>	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Plasmid

Table 5. Comparison of virulence factor from the Virulence Factors Database and from known virulence factor of staphylococci among the 21 *S. lugdunensis* strains

Function	gene	CC1										CC3					CC6				CC7		CC5
		ST1					ST6		ST12	ST3					ST10		ST24		ST26	ST27	ST5		
		SL29	SL117	SL122	V37	N920	HKU	SL13	V33	V27	V21	FD141	FD381	FD377	Klug93	SL55	SL118	V22	FD222	FD143	C33	V25	
Exoenzymes	<i>hly</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	FS	+	+	0	+	
	<i>hlyIII</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sagB</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sagD</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>nuc</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
HTH transcription factors	<i>sarA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sarR</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sarV</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sarX</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sarZ</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>mgrA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>rot</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>locus agr</i>	type I	type II	type I	type I	type I	type I	type II	type II	type I													
<i>saeS</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
<i>saeR</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
Proteases	<i>clpB</i>	+	+	+	FS	+	+	+	+	+	+	+	+	+	+	+	FS	+	+	+	+	+	
	<i>clpC</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>clpP</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>clpL</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	S	+	+	+	+	+	
	<i>clpX</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Iron metabolism	<i>locus sir</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>locus hts</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>locus sst</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sfaA</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	<i>sfaB</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sfaC</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>sfaD</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	<i>locus isd</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>srtB</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>hrtA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>hrtB</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>hssS</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
<i>hssR</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
Host immune evasion	<i>cap PS</i>	cap5	cap8	cap8	cap8	cap8	0	cap5	cap5	cap5	cap5	cap5	cap5	cap8									
	<i>cap PG</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>oatA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>dlt</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>mprF</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Adhesion MSCRAMM	<i>locus icaA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	FS icaA	FS icaA	FS icaA	FS icaA	+	+	+	
	<i>enolase</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>atlL</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>ebpS</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>fnbA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
<i>srtA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		

The presence of a putative virulence gene is indicated by a “+” while the absence is mentioned by “0”. PS, Polysaccharidic; PG, Polyglutamic; FS, Frameshift; S, non-sens mutation; *, duplication.

Table 6. Comparison of the proteins encoded by the *ess* module 1 and 2 of *S. aureus* HO 5096 0412 with HKU09-01 *S. lugdunensis* strains.

<i>S. aureus</i> HO 5096 0412 strain			<i>S. lugdunensis</i> HKU09-01 strain			
Proteins	length (amino acids)	pfam family domain	Proteins	length (amino acids)	pfam family domain	% identity
EsxA	97	WXG100	SLGD_RS09680	97	WXG100	92%
EsaA	1009	N/A	SLGD_RS09675	1006	N/A	57%
EssA	153	EssA	SLGD_RS09670	144	EssA	64%
EsaB	80	YukD	SLGD_RS09665	80	YukD	70%
EssB	444	YukC	SLGD_RS09660	437	YukC	70%
EssC	1476	FtsK SpoIIIE N two FtsK SpoIIIE	SLGD_RS09655	1476	FtsK SpoIIIE N two FtsK SpoIIIE	77%
SAEMRSA15_02470	359	N/A	SLGD_RS09650	360	N/A	47%
SAEMRSA15_02480	96	N/A	SLGD_RS09645	96	N/A	64%
SAEMRSA15_02490	420	LXG	SLGD_RS09640	420	LXG	58%
SAEMRSA15_02500	137	DUF5082	SLGD_RS09635	137	DUF5082	66%
SAEMRSA15_02510	157	DUF5083	N/A	N/A	N/A	N/A
SAEMRSA15_02520	124	DUF4467	SLGD_RS09630	127	DUF4467	58%

Table 7. Comparison of the proteins encoded by the *ess* module 1 and 2 of *S. aureus* RJ-ST398 strains with VISLISI_21 *S. lugdunensis* strains.

<i>S. aureus</i> RJ-ST398 strain			<i>S. lugdunensis</i> VISLISI_21 strain			
Proteins	length (amino acids)	pfam family domain	Proteins	length (amino acids)	pfam family domain	% identity
EsxA	97	WXG100	B7466_RS09445	97	WXG100	92%
EsaA	1009	N/A	B7466_RS09440	1006	N/A	57%
EssA	153	EssA	B7466_RS09435	144	EssA	51%
EsaB	80	YukD	B7466_RS09430	80	YukD	64%
EssB	444	YukC	B7466_RS09425	437	YukC	71%
EssC	1481	FtsK SpoIIIE N two FtsK SpoIIIE	B7466_RS09420	1481	FtsK SpoIIIE N two FtsK SpoIIIE	77%
SAPIG0303	98	DUF5344	B7466_RS09415	95	DUF5344	57%
EsxB	134	DUF5082	B7466_RS09410	107	DUF5082	50%
EsxX	517	LXG	B7466_RS09405	523	LXG	56%
SAPIG0306	228	DUF5079	B7466_RS09400	223	DUF5079	34%
SAPIG0307	149	DUF5085	B7466_RS09395	150	DUF5085	62%
SAPIG0308	149	DUF5085	B7466_RS09390	151	DUF5085	48%

Table 8. Pair test statistical comparison (*p* values) of the virulence of the 6 strains of *S. lugdunensis* in the *Galleria mellonella* infection model.

Strains	Strains				
	SL13	SL29	SL55	SL117	SL118
SL029	2.2e-11 ^a				
SL055	1	1.7e-10			
SL117	1.1e-06	1	4.3e-06		
SL118	0.00016	0.0021	0.00029	1	
SL122	1.5e-14	1	1.5e-13	0.00043	2.7e-9

^a in bold are *p* values < 0.05

Supplementary data

Figure S1. Phylogenetic trees based on (A) the *agr* hypervariable region (*agrB* to *agrD*) and (B) the *cap* hypervariable region (*capH* to *capK*) of the 21 genomes of *S. lugdunensis*.

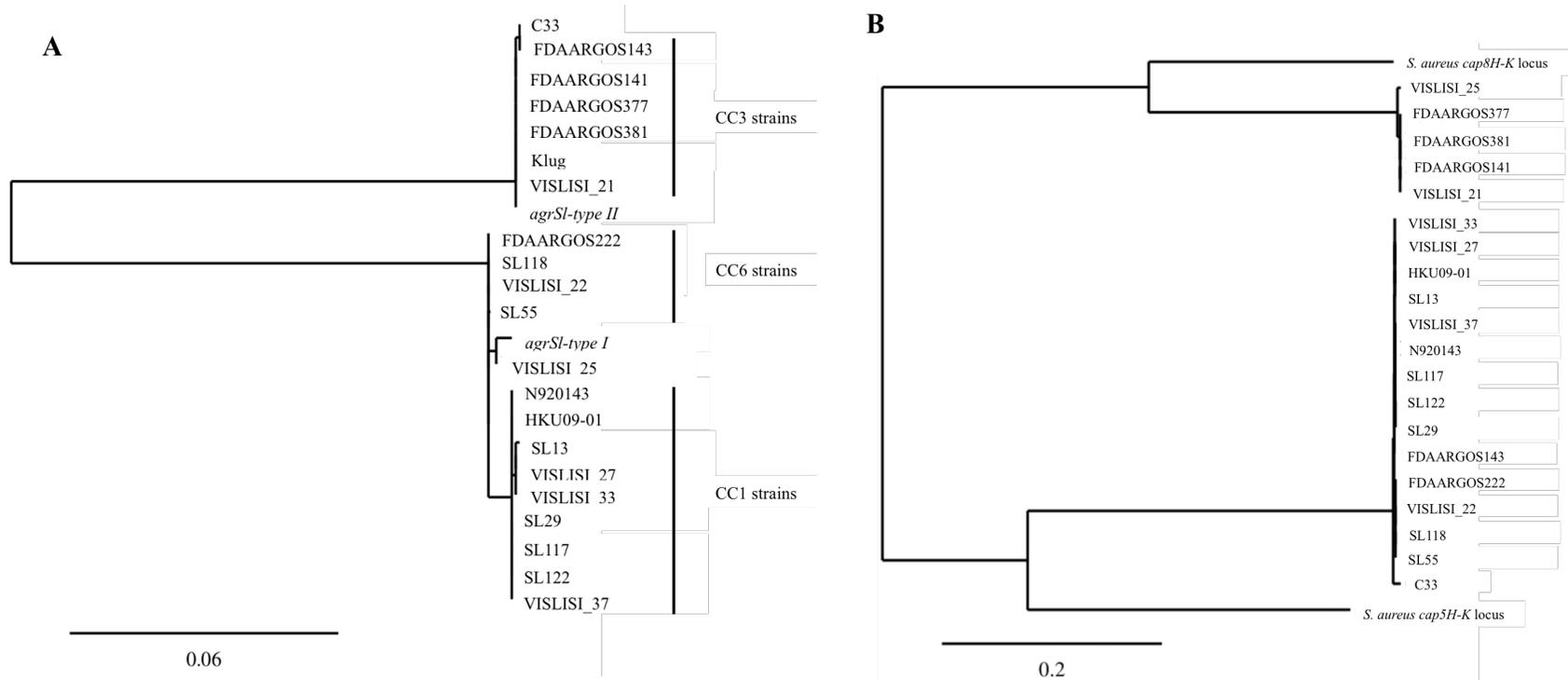


Table S1. Clinical and geographical origins of complete genomes of *S. lugdunensis* strains from NCBI.

<i>S. lugdunensis</i> strains	NCBI accession numbers	Isolation source	Country
HKU0901	NC_013893	Skin and soft tissue infection	Hong Kong
N920143	NC_017353	Skin and soft tissue infection	France
FDAARGOS141	NZ_CP014022	Bursitis	USA
FDAARGOS143	NZ_CP014023	Skin and soft tissue infection	USA
FDAARGOS222	NZ_CP020406	Skin and soft tissue infection	Strasbourg
Klug93G4	NZ_CP017069	Carriage	Hong Kong
FDAARGOS377	NZ_CP023539	Abscess	USA
FDAARGOS381	NZ_CP023970	Abscess	USA
VISLISI_21	NZ_CP020762	Bacteraemia	France
VISLISI_22	NZ_CP020764	Infective endocarditis	France
VISLISI_25	NZ_CP020763	Material infection	France
VISLISI_27	NZ_CP020735	Material infection	France
VISLISI_33	NZ_CP020769	Liver abscess	France
VISLISI_37	NZ_CP020761	Infective endocarditis	France
C33	NZ_CP020768	Carriage	France

2. Caractérisation du système de sécrétion Ess/type VII

2.1. Objectifs et principaux résultats de l'étude

Chez les bactéries, les systèmes de sécrétion jouent un rôle clé dans l'export de facteurs de virulence hors du cytoplasme, dans la communication avec l'environnement et avec les cellules de l'hôte²⁴⁷. Outre les systèmes de sécrétion Sec et Tat, de nombreuses bactéries à Gram positif possèdent un système de sécrétion de protéines spécialisé, appelé SST7, codé par le *locus ess*. Chez *S. aureus*, ce système s'est avéré essentiel pour la virulence et la sécrétion des protéines ESAT-6 (également appelé EsxA) et le CFP-10 (également appelé EsxB)¹⁶⁵. Ces deux protéines font partie de la famille des protéines WXG100 caractérisée par une structure en double hélice reliée par un motif central WXG (Tryptophane-X-Glycine). Chez *S. aureus*, la sécrétion de ces protéines est assurée par une protéine cytoplasmique EsaB et par quatre protéines transmembranaires EsaA, EssA, EssB et EssC¹⁶³. La protéine EssC est composée de domaines protéiques qui lui permettent de former un pore et de jouer ainsi un rôle central dans la translocation des protéines telles que EsxA et EsxB, mais aussi des protéines EsxC, EsxD et EsaD^{167,169}.

Par ailleurs, l'étude de 153 génomes complets de *S. aureus* a permis de décrire la variabilité génétique et différentes organisations du *locus ess*¹⁶⁶. En plus de cette variabilité génétique, ont été décrites plusieurs organisations transcriptionnelles et différents profils de sécrétion des protéines EsxA, EsxB et EsxC selon les souches de *S. aureus* étudiées¹⁶⁷.

Un *locus* homologue au *locus ess* de *S. aureus* a été identifié par Warne *et al.*¹⁶⁶ chez *S. lugdunensis*. Compte-tenu du pouvoir pathogène de cette espèce, en particulier de son implication dans des abcès cutanés de sujets sains en tous points comparables à ceux générés par *S. aureus*³⁵ et d'infections invasives particulièrement sévères et inhabituelles pour un SCN^{2,52}, là encore semblables à celles produites par *S. aureus*, il nous a paru pertinent de contribuer à la caractérisation moléculaire et phénotypique de ce système chez *S. lugdunensis*, seul SCN à posséder tous les gènes nécessaires au fonctionnement du SST7.

A la faveur de l'analyse comparée de 21 génomes, nous avons identifié au sein du *locus ess* de *S. lugdunensis* une région conservée qui contient tous les gènes codant les protéines considérées comme essentielles au fonctionnement du SST7 chez

S. aureus^{163,166}, telles que les protéines membranaires EssA, EssB, EssC et EsaA, la protéine cytoplasmique EsaB et la protéine sécrétée EsxA. Par ailleurs, nous avons identifié au sein de ce *locus* un gène, nommé pour cette étude *lxg*, codant une protéine présentant un domaine LXG, prédit pour être sécrété par le SST7²⁴⁸. Alors que cette espèce possède une structure de population clonale, nous avons également identifié au sein du *locus* *ess* de nombreuses variations entre les souches de différents CC, mais aussi entre des souches appartenant à un même CC, comme la présence d'un transposon et d'un *cluster* de sept gènes additionnels.

Dans ce contexte, le premier objectif de ces travaux était de confirmer la diversité génétique du *locus* *ess* en évaluant la distribution du transposon et du *cluster* de sept gènes au sein d'une collection de 94 isolats de *S. lugdunensis* préalablement caractérisés par MLST⁹⁸. L'étude de la distribution de ces polymorphismes majeurs a montré que le transposon était principalement associé à des souches appartenant au CC1 alors que le *cluster* de sept gènes était principalement retrouvé chez des souches de CC2 et de CC3.

Notre second objectif était de caractériser l'organisation transcriptionnelle du *locus* *ess* de *S. lugdunensis* à partir des six souches dont le génome a été séquencé par *Pacbio* et de la souche DSM 4804. Aucune différence dans l'organisation transcriptionnelle du *locus* *ess* n'a été observée entre les sept souches. En effet, les gènes étudiés étaient tous co-transcrits.

Afin d'étudier le rôle du SST7 dans la virulence de *S. lugdunensis*, nous avons construit un mutant de délétion du gène codant la protéine EssC et évalué l'impact de la délétion (i) sur la virulence *in vitro* et *in vivo* et (ii) sur le sécrétome.

Dans un premier temps, nous avons comparé *in vitro*, l'activité β -hémolytique et δ -hémolytique synergique de la souche mutée par rapport à la souche sauvage sur des milieux solides supplémentés avec du sang de cheval ou de mouton. Une différence très significative a été observée dans l'activité hémolytique des deux souches. En effet, la délétion du gène *essC* a entraîné la perte de la capacité de la souche à lyser les érythrocytes de cheval.

En collaboration avec le Pr Olivier Lesouhaitier du Laboratoire de Microbiologie – Signaux et Micro-environnement (LMSM EA 4312), nous avons alors comparé la virulence

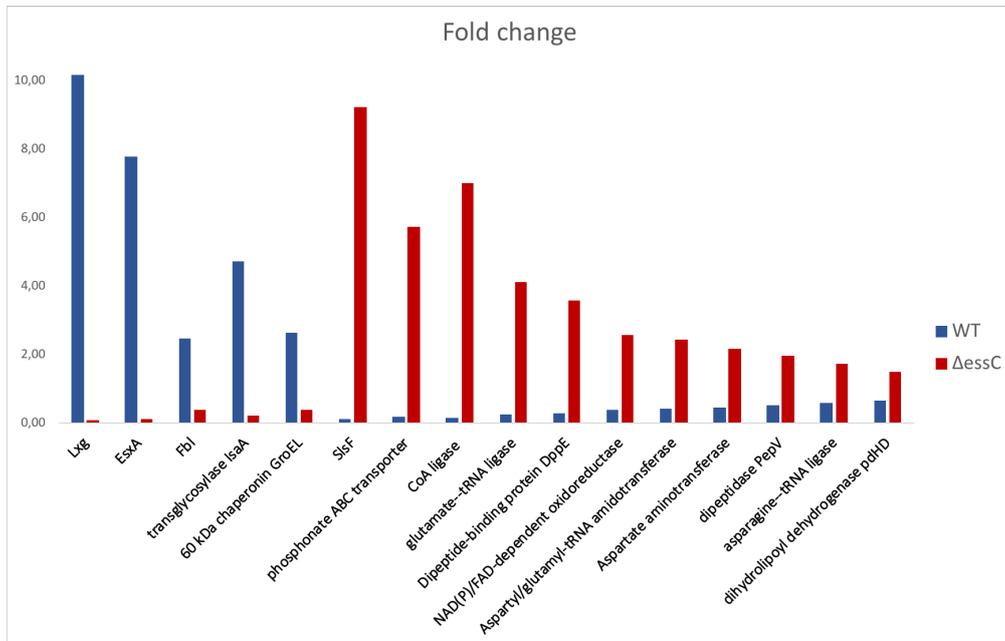


Figure 59 : Expression différentielle des protéines identifiées comme surexprimées dans le surnageant de culture de la souche sauvage DSM WT et de la souche délétée du gène Δ essC.

Highest mean condition	Protein	Maximum fold change	Anova (P)
WT	Lxg	10,18	0,00009
	EsxA	7,78	0,02068
	Transglycosylase IsaA	4,71	0,04950
	60 kDa chaperonin GroEL	2,65	0,02549
	Fbl	2,47	0,04797
Δ essC	SIsF	9,24	0,03983
	2-amino-3-ketobutyrate CoA ligase	7,00	0,00062
	Phosphonate ABC transporter	5,73	0,04507
	Glutamate--tRNA ligase	4,11	0,03424
	Dipeptide-binding protein DppE	3,58	0,01751
	NAD(P)/FAD-dependent oxidoreductase	2,58	0,02959
	Aspartyl/glutamyl-tRNA amidotransferase	2,43	0,01864
	Aspartate aminotransferase	2,17	0,00024
	Dipeptidase PepV	1,96	0,02643
	Asparagine--tRNA ligase	1,73	0,00688
	Dihydrolipoyl dehydrogenase pdHD	1,50	0,02402

Tableau VII : Valeurs de *fold change* obtenues pour les protéines identifiées comme surexprimées dans le surnageant de culture de la souche sauvage DSM WT ou chez la souche délétée du gène Δ essC.

Les valeurs de significativité statistique ont été obtenues avec le test de Anova ($P < 0,05$).

de la souche sauvage et de la souche mutée (i) dans un modèle *in vitro* de cytotoxicité cellulaire sur une lignée de kératinocytes (HaCaT) et (ii) *in vivo* dans un modèle de survie du nématode *C. elegans* en milieu liquide (*fast killing*) et solide (*slow killing*). Toutefois, dans ces modèles, aucune différence n'a été observée entre la souche mutée et la souche sauvage.

Enfin, en collaboration avec le Pr Pascal Cosette et le Dr Julie Hardouin de la plateforme de protéomique PISSARO (Plateforme Instrumentale en Sciences Séparatives et Analytiques de Rouen), nous avons initié une analyse comparative du sécrétome de la souche mutée et de la souche parentale pour évaluer si la souche mutée présentait une diminution du nombre de protéines sécrétées dans le milieu extérieur et notamment les protéines sécrétées par le SST7. Les résultats de l'étude du sécrétome ont été obtenus par analyse des protéines contenues dans le surnageant de culture des deux souches en milieu minimum supplémenté avec 5% de glucose. Ces protéines ont été extraites par précipitation avec de l'acide trichloracétique et ensuite analysées par nano-chromatographie liquide couplée à une analyse par spectrométrie de masse en tandem.

Les résultats présentés ici ne sont que préliminaires du fait de problème de répétabilité et de reproductibilité entre les réplicats biologiques (**Figure 59, Tableau VII**). Toutefois, il a été observé que la souche sauvage exprimerait 10 fois plus la protéine EsxA et 7 fois la protéine Lxg. De plus, la souche sauvage exprimerait plus la protéine Fbl, une transglycolase IsaA et une protéine de stress GroEL. A l'inverse, l'adhésine SIsF, une dipeptidase PepV et plusieurs protéines membranaires ou cytosoliques seraient plus exprimées chez la souche mutante que chez la souche sauvage.

Ces premiers résultats doivent être confirmés par de nouvelles expérimentations.

Le chapitre suivant présente l'article scientifique en cours de rédaction. Cet article comporte tous les matériels et méthodes utilisés ainsi que tous les résultats obtenus à ce jour sauf ceux qui concernent le sécrétome des souches sauvage et mutante qui demandent à être confirmés. En l'absence de ces derniers, aucune discussion n'a été rédigée en anglais. Ce travail est discuté en français dans la partie 3. Discussion et perspectives de ce manuscrit.

2.2. Publication n°2

Secretome analysis defines the role of EssC in the Staphylococcus lugdunensis Ess/Type VII secretion system

Jérémie Lebeurre, Charline Dubos, Clément Guillou, Thomas Clamens, Julie Hardouin, Olivier Lesouhaitier, Stéphanie Legris, Pascal Cosette, Patrice François, Sandrine Dahyot, Martine Pestel-Caron. Article en cours de préparation.

1 **Secretome analysis defines the role of EssC**
2 **in the *Staphylococcus lugdunensis* Ess/Type VII secretion system**

3
4 Lebeurre J.^{1*}, Dubos C.², Guillou C.^{3,4}, Clamens T.⁵, Hardouin J.^{3,4}, Lesouhaitier O.⁵, Legris
5 S.¹, Cosette P.^{3,4}, François P.⁶, Dahyot S.², Pestel-Caron M.²

6
7 ¹Normandie Univ, UNIROUEN, GRAM EA2656, F-76000 Rouen, France

8 ²Normandie Univ, UNIROUEN, GRAM EA2656, Rouen University Hospital, F-76000 Rouen,
9 France

10 ³Normandie Univ, UNIROUEN, INSA Rouen, CNRS, PBS, F-76000 Rouen, France

11 ⁴PISSARO Proteomic Facility, IRIB, F-78821 Mont-Saint-Aignan, France

12 ⁵ Normandie Univ, UNIROUEN, Laboratory of Microbiology Signals and Microenvironment
13 LMSM EA 4312, 27000 Evreux, France

14 ⁶Geneva University Hospitals, Genomic Research Laboratory, Division of Infectious Diseases,
15 Geneva, Switzerland

16
17
18 *Corresponding author:

19 E-mail: jeremie.lebeurre@univ-rouen.fr

20 Phone number : +33235148299 ; fax number : +33235148203

21 Groupe de Recherche sur l'Adaptation Microbienne (EA2656), Normandie Univ, UNIROUEN,
22 F-76000 Rouen, France

23
24
25

26 **Introduction**

27 Protein secretion is critical for bacteria to interact with and manipulate their environments, and
28 play critical roles in virulence. Bacteria possess a number of different secretion machineries to
29 transport proteins (in particular toxins) across their cell envelopes. Besides the Sec, Tat, and
30 flagellar secretion systems, many Gram-positive bacteria also produce a specialized protein
31 secretion machinery termed the Type VII secretion system (T7SS)¹. First identified in
32 pathogenic and commensal mycobacteria^{2,3}, T7SS was found essential for virulence and
33 secretion of, amongst others, two small proteinaceous T-cell antigens, the Early Secreted
34 Antigen Target 6 (ESAT-6, also termed EsxA) and Culture Filtrate Protein-10 (CFP-10) (also
35 termed EsxB)^{4,5}. These two proteins are members of the WXG100 protein superfamily
36 characterized by two small helical hairpins linked with a central WXG motif (Tryptophan – X
37 – Glycine).

38 T7SS are also encoded in the genome of diverse bacteria among the *Actinobacteria* and
39 *Firmicutes* phyla^{3,6,7}. The T7SS of *Staphylococcus aureus* is encoded by the *ess* locus. Genetic
40 analyses and functional characterization of the *ess* locus allowed to identify genes needed to
41 the secretion machinery⁷⁻⁹. The Ess machinery comprises six essential core components. Four
42 membrane proteins (EsaA, EssA, EssB, and the Ftsk-SpoIIIE ATPase EssC) and a cytoplasmic
43 ubiquitin-like fold protein EsaB, have been demonstrated to be necessary for translocation of
44 EsxA across the bacterial membrane³. Besides EsxA, four other proteins are secreted: the
45 WXG-like EsxB, EsxC, EsxD and EsaD^{3,9}. Of note, EsaD is a secreted anti-bacterial toxin
46 involved in the interbacterial competition¹⁰. This T7SS nuclease interacts during its
47 biosynthesis with two Ess accessory proteins, EsaG, a cytoplasmic antitoxin, and EsaE a
48 putative chaperone secreted.

49 T7SS have been demonstrated to be required in the virulence of *S. aureus*. Experiments using
50 mouse models of infection have indicated that the T7SS is required for colonization and
51 virulence, in particular for the pneumonia development, skin soft infections and for the
52 establishment of persistent abscesses^{7,8,11}. T7SS has also been involved in the modulation of
53 immune responses during staphylococcal bloodstream infections¹² and in iron homeostasis¹³.

54 Recently, Warne *et al.*¹⁴ have described a highly genetic diversity and variable organization
55 within the *ess* locus across *S. aureus* isolates from a broad range of clonal complexes (CCs).
56 This study showed that the *ess* locus comprises four distinct clusters, three of which do not
57 encode EsxB, EsxC, EsxD or EsaD. Instead, each cluster is associated with a distinct set of
58 genes downstream of the core *ess* genes. Warne et al could not show evidence of intra- and

59 inter-species recombination to explain these highly diverse modules but identified a homolog
60 region of the *ess* locus in the genome of *Staphylococcus lugdunensis*, only Coagulase-negative
61 staphylococci (CoNS) species to possess the entire *ess* locus.

62 *S. lugdunensis* represents a regular part of the normal skin flora that has emerged as a virulent
63 CoNS in both community-acquired and healthcare associated infections¹⁵⁻¹⁷. This bacterium is
64 an opportunistic pathogen, mainly responsible for skin and soft tissue infections, particularly
65 abscesses which are one of the most common clinical presentation in the breast, perianal region
66 and abdomen¹⁸⁻²⁰. These infections are clinically similar to infections caused by *S. aureus* and
67 should thereby be considered as serious as *S. aureus* infections^{15,18}. *S. lugdunensis* has also
68 involved in native valve endocarditis²¹, bacteremia²² and bone and joint infections^{23,24}. To date,
69 little is known about the potential disease-causing capacity of *S. lugdunensis* even if some
70 virulence factors were reported as haemolysins, fibrinogen binding protein Fbl, von Willebrand
71 binding protein vWbl and the lugdulysine metalloprotease¹⁵.

72 In a recent genomic comparative analysis, we have shown that the *ess* locus of *S. lugdunensis*
73 is also subject to an important genetic diversity, in particular for the CC3 strains which have
74 their own genetic organization compared to strains from other CCs. These two genetic
75 organizations are linked to the existence of two *essC* gene variants (*essC1* and *essC2*), each
76 associated with a specific cluster of downstream genes. Furthermore, great polymorphism was
77 observed within the module 3 (presence of cluster of seven genes encoding hypothetical
78 proteins) and the module 5 (presence of a 7 kb-transposon) of the locus.

79 While the mechanisms of the virulence of this pathogen characterized by its ability to cause
80 abscesses of various localizations²⁵⁻²⁸ remain poorly understood, it seemed appropriate to
81 explore the role of this *ess* locus in *S. lugdunensis* virulence.

82 Therefore, the aims of the present study was (i) to explore the polymorphisms described within
83 the modules 3 and 5 of the locus on a large collection of isolates, (ii) to determine the *ess*
84 transcriptional organization, and (iii) to characterize the role of EssC protein in the protein
85 secretion of T7SS and in the pathogenicity of *S. lugdunensis* by comparison of the secretome
86 and the *in vivo* and *in vitro* virulence of an in-frame deletion *essC* gene mutant and its parental
87 strain.

88

89

90 **Methods**

91 **Strains, plasmid construction and protoplasts formation**

92 If not mentioned otherwise, *S. lugdunensis* strains were grown on trypticase soy broth (TSB).
93 *Escherichia coli* Stb14 strain was cultured aerobically in Luria-Bertani medium. When
94 necessary, ampicillin (100 µg/mL) or erythromycin (1 µg/mL) were added to the medium for
95 plasmid selection.

96 All DNA extractions were performed by using the InstaGene Matrix kit (Bio-Rad) according
97 to the manufacturer's recommendations. In-frame deletion of *essC* gene (Δ essC) by
98 homologous recombination was performed by using the plasmid pMAD³¹. Two up and
99 downstream fragments flanking the *essC* gene with the first two and last five codons were
100 amplified and overlapped by using the following primers (BamHI-F1
101 aaaaatggatccaaagaagcgttatacatctatgcaa, R1 aataatttaatcatcattctgccctcaatcc and F2
102 cagaatgatgattaaattttgaatagggaggca, EcoRI-R2 ctatatgaattctaattgtaatgatgcagtttcgta). Then, the
103 amplified fragment was digested by BamHI and EcoRI enzymes (New England BioLabs), and
104 ligated with T4 DNA ligase (New England BioLabs) into the pMAD vector following
105 manufacturer's instructions. Clones containing the amplified flanking regions in the pMAD
106 plasmid were selected in *E. coli* Stb14 and the recombinant plasmid was extracted (NucleoBond
107 Xtra Midi endotoxin free, Macherey Nagel). *S. lugdunensis* DSM 4808 (DSM WT) was
108 transformed with the plasmid construction Δ essC by protoplasts formation as previously
109 described^{32,33}. After incubation at 30°C for 5 days, *S. lugdunensis* clones containing the plasmid
110 were selected, and the double-crossing over was performed³¹.

111

112 **Investigation of the genetic diversity of the *ess* locus**

113 A collection of 94 isolates, previously characterized by MLST^{29,30} (stored at -80°C before use)
114 were grown overnight at 37°C on tryptic soy agar with 5% horse blood agar plate. The
115 collection of strains was composed of 34 isolates belonging to the CC1, 17 to the CC2, 20 to
116 the CC3, 5 to the CC4, 5 to the CC5, 9 to the CC6 and 4 singletons.
117 All DNA extractions were performed by using the InstaGene Matrix kit (Bio-Rad) according
118 to the manufacturer's recommendations. PCR was performed on a Veriti Thermal Cycler
119 (Applied Biosystems) by using GoTaq G2 Green Master Mix (Promega) and specific primer
120 pairs targeting either the transposon or the cluster of seven genes (Table S1).

121

122 **Total RNA extraction, Reverse-Transcription PCR (RT-PCR) and 5'-RACE experiments**

123 Fresh TSB was inoculated with an overnight (ON) culture to an initial OD_{600nm} of 0.01 and
124 grown at 37°C in a shaking incubator set at 150 rpm. Cells were grown to mid-exponential
125 phase (OD_{600nm} = 1), and 1 ml was centrifuged (5min at 8,000xg, 4°C). Supernatants were
126 carefully removed and the pellets were resuspended in 200 µl TE Buffer (10 mM Tris-HCl pH
127 8 and 1 mM EDTA) containing 2,1 mg lysozyme (Sigma-Aldrich, St. Louis, MO, USA) and 10
128 µg lysostaphin (Sigma-Aldrich). Samples were vortexed for 30s, incubated at 37°C for 10min,
129 and lysed using the RA1 lysis buffer provided in the NucleoSpin® RNA kit (Macherey Nagel)
130 supplemented with β-mercaptoethanol (Sigma-Aldrich). Next, a bead-beating procedure
131 alternating 1-min cycles of beating with incubation on ice was performed. RNA were isolated
132 using the NucleoSpin® RNA kit (Macherey Nagel). The amount of RNA yielded was assessed
133 with a NanoDrop spectrophotometer (Nanodrop Technologies). RNA samples were then treated
134 by TURBO DNA-free™ kit (Ambion). The cDNA synthesis was performed using the
135 Omniscript® RT kit (Qiagen). A total of 2 µg of RNA from respective samples was reverse
136 transcribed in a 20 µL volume reaction using random primers (Promega, Madison, USA) and
137 RNase inhibitor (40 U/µl) (Invitrogen, Carlsbad, USA). PCR was performed on a Veriti
138 Thermal Cycler (Applied Biosystems) by using GoTaq G2 Green Master Mix (Promega) and
139 region-specific primers that either primed within each gene of the *ess* locus or that spanned
140 intergenic regions (region-x-f and region x-r) (Table S2). PCR products visualized on 1%
141 agarose gels. All kits were used according to the manufacturer's instructions.

142 5'-RACE PCR was performed with the protocol from the second generation 5'/3'-RACE kit
143 (Roche) using RNA samples prepared as described above, gene-specific primer pairs for *esxA*
144 or *esaA*.

145

146 **Gene expression analysis by RT-qPCR**

147 RT-qPCR was performed on the strain DSM WT at OD₆₀₀ of 1.0, 2.0, 3.0 and 5.0. PCR
148 amplification was carried out with the SYBR Green (Bio-Rad). Reactions were performed in
149 triplicate using a CFX96 real-time PCR thermal cycler (Bio-Rad) with primers listed in Table
150 S2. The 16S rRNA housekeeping gene was used to normalize the results. The comparative cycle
151 threshold method ($2^{-\Delta Ct}$ method) was used to analyze the mRNA levels.

152

153

154 **Hemolysis assay and CAMP test**

155 Columbia agar with 5% horse blood was inoculated with 2 μ l of ON culture of strains DSM
156 WT and DSM Δ essC at OD₆₀₀ of 1. After incubation at 37°C for 2 days, the hemolysis diameter
157 was determined subtracting the diameter of the colony from the total diameter of hemolysis.
158 *S. aureus* strain ATCC29123 was used as positive control. Synergic hemolysis was investigated
159 by CAMP test on tryptic soy agar with 5% sheep blood by inoculation of a single central straight
160 line of *S. aureus* strain ATCC29123. Strains DSM WT and DSM Δ essC were inoculated by
161 perpendicular straight lines to that of *S. aureus*. Synergic hemolysis was observed after 24h at
162 37°C and then 24h at room temperature.

163

164 **HaCaT cells cytotoxicity tests**

165 The direct cytotoxicity activity of the bacteria was studied using the human keratinocyte cell
166 line HaCaT (Eppelhein, Germany) as previously described³⁴. Briefly, HaCaT cells were grown
167 at 37 °C under a 5% CO₂ atmosphere, in Dulbecco's modified Eagle's medium (DMEM, Lonza,
168 Levallois-Perret, France) supplemented with 10% fetal calf serum and 1% antibiotic cocktail
169 (HyClone Thermo Scientific, Illkirch, France). Cells were used between passages 41 and 65.
170 One day before use, the HaCaT cells were starved of antibiotic and fetal calf serum. The cells
171 were incubated for 24 hours with bacteria at a multiplicity of infection (MOI) of 10:1. The lethal
172 effect of bacteria was determined by assaying LDH which is released by keratinocytes, by using
173 a Cytotox 96 assay (Promega)

174

175 ***Caenorhabditis elegans* synchronization and virulence assays**

176 The *Caenorhabditis elegans* wild-type Bristol strain N2 was obtained from the *Caenorhabditis*
177 Genetics Center (Minneapolis, MN, USA). *C. elegans* were maintained under standard
178 culturing conditions at 23°C on nematode growth medium (NGM: 3 g NaCl, 2.5 g peptone,
179 17 g agar, 5 mg cholesterol, 1 ml 1 M CaCl₂, 1 ml 1 M MgSO₄, 25 ml 1 M KH₂PO₄, H₂O to 1
180 liter) agar plates with *E. coli* OP50 as a food source. Synchronous cultures of worms were
181 generated as previously described³⁵, and used for both slow-killing assay and liquid killing
182 assay.

183 Bacterial lawns used for *C. elegans* survival slow killing assays were prepared by spreading 50
184 μ l of DSM WT and DSM Δ essC strains on 35 mm NGM conditioned Petri dishes supplemented
185 with both 0.05 mg/ml of 5-fluoro-2'-deoxyuridine and 5 μ mol/l of tryptophan. The plates were
186 incubated ON at 37 °C and then placed at room temperature for 4 h. Fifteen to twenty L4 (48

187 hours old) synchronized worms were harvested with M9 solution (3 g KH₂PO₄, 6 g NaHPO₄, 5
188 g NaCl, 1 ml 1 M MgSO₄, H₂O to 1 litre), placed on the 35 mm assay Petri dishes and incubated
189 at 23 °C. Worm survival was scored at 1 hour, 24 hours and each subsequent day, using an
190 Axiovert S100 optical microscope (Zeiss, Oberkochen, Germany) equipped with a Nikon
191 digital Camera DXM 1200F (Nikon Instruments, Melville, NY, USA). The worms were
192 considered dead when they remained static without grinder movements for 20 s. The results are
193 expressed as the percentage of living worms. The results are the average of two independent
194 assays.

195 Concerning the liquid killing assay³⁶, bacteria from ON cultures were centrifuged for 10 min at
196 7,500 x g. The supernatants were collected and 80 µl were distributed onto microtiter plate
197 wells. Immediately after, fifteen to twenty L4 (48 hours old) synchronized worms, harvested
198 with M9 solution, were added to the bacterial supernatants and incubated at 22 °C. Worm
199 survival was scored at 1 hour and each following hours using an Axiovert S100 optical
200 microscope (Zeiss, Oberkochen, Germany).

201 For the killing assay, nematode survival was calculated by the Kaplan-Meier method, and
202 survival differences were tested for significance by using the log rank test (GraphPad Prism
203 version 4.0; GraphPad Software, San Diego, California, USA).

204

205 **Sample preparation for secretome analysis**

206 Strains DSM WT and DSMΔ*essC* were cultured in Minimal Glucose Medium at 37°C until
207 mid-exponential phase which corresponded to OD₆₀₀ 1 and centrifuged at 4°C for X min. The
208 supernatants were frozen until use. To perform TCA (TriChloroacetic Acid) (Sigma)
209 precipitation, sample was thawed slowly at 4°C. For each sample (WT and Mutant), 20% of
210 TCA was added for 3 hours at 4°C. After incubation, samples were centrifugated for 20 minutes
211 at 10000g (4°C) and supernatants were removed. Pellets were washed twice with cold acetone
212 and centrifugated for 15 minutes at 12000g (4°C). Then, pellets were dried and solubilized with
213 50µL of homemade specific buffer (Urea 7M [Sigma], Thiourea 2M [Sigma], DTT 20mM
214 [Sigma], Chaps 2% [Sigma], C7BzO 0.5% [Sigma], TriButylPhosphin 5µL [Fluka Chemika]).
215 To estimate protein concentration for each sample, Bradford was performed according to the
216 manufacturer's instructions (Biorad). The linear range for BSA standard was 0.1 to 1 mg/mL.
217 The dilution of all samples was 1:1. MMG buffer was used as negative control.

218 Then, protein samples (from 3 to 25µg depending on samples) were mixed with 2X Laemmli
219 buffer (SDS 4% [Biorad], Tris HCl 0.15M [Sigma], Glycerol 17% [GE Healthcare],
220 Bromophenol blue 0.008% [Sigma] and DTT 0.1M [Sigma]), were heated for 10 minutes at

221 95°C and were stacked into a 7% polyacrylamide gel (Acrylamide/Bis-Acrylamide 30% [29 :1]
222 [Sigma]). After migration (10mA/gel for 45 minutes then 20mA/gel for 90 minutes), gel was
223 stained with Coomassie blue (30 sec) to highlight proteins. Thereafter, protein gel bands were
224 excised with scalpels and were discolored using specific buffer (H₂O, EtOH and Acetic Acid).
225 Then, « in gel » digestion was performed. Firstly, a reduction/alkylation step (DTT
226 50mM/Iodoacetamide 125mM [Sigma]) was needed before tryptic digestion. Secondly, gel
227 bands were dried with acetonitrile (ACN) (Fischer) and incubated with 1µg of trypsin
228 (Promega) in 10mM carbonate buffer (Ammonium Hydrogen Carbonate [Fluka Chemika]) to
229 obtain a peptidic fraction. After ON incubation at 37°C, digests were recovered with ACN
230 baths, dried and solubilized in 0.1% formic acid (Fischer).

231

232 **NanoLC–MS/MS Analyses**

233 All experiments were performed on an LTQ-Orbitrap Elite apparatus coupled to an Easy nLC
234 II system (both from Thermo Scientific). Samples were injected onto an enrichment column
235 (Acclaim PepMap100, Thermo Scientific). The separation was achieved with an analytical
236 column needle (NTCC-360/100-5-153, NikkyoTechnos). The mobile phase consisted of
237 H₂O/FA 0.1% (buffer A) and ACN/FA 0.1% (buffer B). Tryptic peptides were eluted at a flow
238 rate of 300 nL/min using a three-step linear gradient: from 2 to 40% B over 75 min, from 40 to
239 80% B in 4 min, and at 80% B for 11 min. The mass spectrometer was operated in positive
240 ionization mode with capillary voltage and source temperature set at 1.5 kV and 275 °C,
241 respectively. The samples were analyzed using CID (collision-induced dissociation) method.
242 The first scan (MS spectra) was recorded in the Orbitrap analyzer ($R = 60\,000$) with the mass
243 range m/z 400–1800. Then, the 20 most intense ions were selected for MS₂ experiments. Singly
244 charged species were excluded for MS₂ experiments. Dynamic exclusion of already fragmented
245 precursor ions was applied for 30 s, with a repeat count of 2, a repeat duration of 30 s, and an
246 exclusion mass width of ± 5 ppm. The precursor isolation width was 2 m/z . Fragmentation
247 occurred in the linear ion trap analyzer with normalized collision energy of 35. All
248 measurements in the Orbitrap analyzer were performed with on-the-fly internal recalibration
249 (lock mass) at m/z 445.12002 (polydimethylcyclsiloxane).

250 Raw data files were processed using Proteome Discoverer 1.4 software (Thermo Scientific).
251 Peak lists were searched using the MASCOT search software (Matrix Science) against the *S.*
252 *lugdunensis* database. Database searches were performed with the following parameters: two
253 missed trypsin cleavage sites allowed; variable modifications: carbamidomethylation on
254 cysteine, and oxidation on methionine. The parent-ion and daughter-ion tolerances were 5 ppm

255 and 0.35 Da, respectively. False discovery rate (FDR) threshold for identifications was
256 specified at 1% (for proteins and peptides).

257 For protein quantification, a label-free experiment was performed as previously described by³⁷.
258 Briefly, after MS analysis, raw data were imported in Progenesis LC-MS software (Nonlinear
259 Dynamics, version 4.0.4441.29989, Newcastle, UK). For comparison, one sample was set as a
260 reference and the retention times of all other samples within the experiment were aligned. After
261 alignment and normalization, statistical analysis was performed for one-way analysis of
262 variance (ANOVA) calculations. For quantitation, peptide features presenting a p value and a
263 q-value less than 0.05, and a power greater than 0.8 were retained. MS/MS spectra from selected
264 peptides were exported for peptide identification with Mascot (Matrix Science, version 2.2.04).
265 Database searches were performed with the following parameters: 1 missed trypsin cleavage
266 site allowed; variable modifications: carbamidomethylation of cysteine and oxidation of
267 methionine. Mass tolerances for precursor and fragment ions were set at 5 ppm and 0.35 Da,
268 respectively. False discovery rates (FDR) were calculated using a decoy-fusion approach in
269 Mascot (version 2.2.04). Identified peptide-spectrum-matches with $-10\log P$ value of 13 or
270 higher were kept, at a FDR threshold of 5%. Mascot search results were imported into
271 Progenesis. For each growth condition, the total cumulative abundance of the protein was
272 calculated by summing the abundances of peptides. Proteins identified with less than 2 peptides
273 were discarded. Only the proteins that varied by 2-fold in their average normalized abundances
274 between growth conditions were retained.

275

276

277 **Results**

278 **Confirmation of the genetic diversity of the *ess* locus of *S. lugdunensis***

279 To assess the CC-dependent variability of the *ess* locus observed in a recent genomic
280 comparison analysis, we have explored the distribution of the major polymorphisms described
281 within the modules 3 and 5 on a collection of 94 clinical isolates previously characterized by
282 MLST (Table 1). The cluster of seven genes previously described in the module 3 of the CC3
283 and CC7 strains was found in 48 clinical isolates of *S. lugdunensis* belonging to the CC1
284 (11/48), CC2 (17/48) and CC3 (20/48). The transposon identified in the module 5 of CC1 strains
285 was observed only in 19 isolates belonging to the CC1 (17/19) and CC2 (2/19). The transposon
286 identified in strains of CC1 was especially associated to the isolates belonging to the ST1
287 (16/19).

288

289 **Transcriptional analysis of the *ess* locus**

290 To probe the co-transcription of *ess* genes in *S. lugdunensis*, RNA was isolated from seven
291 *S. lugdunensis* strains (DSM WT, SL13, SL29, SL55, SL117, SL118 and SL122) during
292 exponential phase. For all strains, an amplification at the expected size was obtained indicating
293 that all *ess* genes are on the same transcriptional unit and thereby that they are co-transcribed
294 (Figure 1).

295 Of note, 5'-Rapid Amplification cDNA-Ends PCR was used to identify promoter region and
296 transcription start sites within the intergenic region between *esxA* and *esaA* or upstream of *esxA*,
297 but experiment was not conclusive.

298

299 **Kinetic expression analysis of the *esxA*, *lxB* and *essC* genes**

300 In order to determine whether the T7SS is expressed in the growth conditions studied and to
301 assess the bacterial density corresponding to the highest expression of the genes *esxA*, *lxB* and
302 *essC* to carry out a secretome analysis a kinetic expression analysis was performed. To this end,
303 RNA was isolated from DSM WT at different growth phases (exponential [OD₆₀₀ of 1.0 and
304 2.0] and stationary phases [OD₆₀₀ of 3.0 and 5.0]). A reverse transcription of the RNA followed
305 by qPCR was performed to determine the normalized RNA levels by using the 16S rRNA gene
306 as the endogenous control in these experiments.

307 No significant difference in *essC* RNA levels was found at OD₆₀₀ of 1.0 compared to RNA
308 levels obtained at OD₆₀₀ of 2.0 and 3.0 (Figure 2). However, at OD₆₀₀ 5.0, the transcript levels
309 were significantly decreased ($P < 0.05$) compared to those at early exponential phase but not

310 with that at OD₆₀₀ of 2.0 and 3.0. The transcript levels of *esxA* and *lxg* were not significantly
311 modified at the different experimental conditions.

312 These results show that the expression of the *essC* gene decreased during the bacterial growth
313 whereas the expression of *esxA* and *lxg* genes was stable during bacterial culture.

314 Of note, we also analyzed the impact of *essC* in-frame deletion on the genes localized
315 downstream of the *essC* genes by RT-qPCR and found no difference between DSM WT and
316 the DSM Δ *essC* (data not shown).

317

318 **Hemolytic activity assay and camp-test**

319 We examined the capacity of the DSM WT and DSM Δ *essC* strains to induce hemolysis. Large
320 areas of β -hemolysis were observed with DSM WT (7.9 cm \pm 0.75 cm) compared to DSM Δ *essC*
321 which was no more hemolytic (Figure 3A and 3B). Similarly, synergic hemolysis assessed by
322 CAMP-test was completely abolished by the in-frame deletion of the gene *essC* (Figure 3C).
323 Otherwise, no difference was observed for the bacterial growth between the two *S. lugdunensis*
324 strains.

325 Of note, no difference in hemolytic activity was observed between the two strains in liquid
326 medium with human or horse blood (data not shown).

327

328 **Cytotoxicity towards HaCaT cells *in vitro***

329 Using the human keratinocyte cell line HaCaT, we evaluated the cytotoxicity of both *S.*
330 *lugdunensis* DSM WT strain or with Δ *essC* mutant strain. We observed that after 24h of contact
331 between bacteria and HaCaT cells, the release of LDH by HaCaT cells that reflects necrosis
332 was identical for the two strains (Figure 4).

333

334 **Virulence towards *Caenorhabditis elegans***

335 In order to evaluate virulence traits of the DSM Δ *essC* mutant strain in host organism, we used
336 the nematode *C. elegans* as host model organism³⁸. We first used the slow killing test allowing
337 to see worm's death due to an establishment of bacteria into the *C. elegans* intestine. When
338 worms were in contact with DSM WT, it took ten days to kill 50 % of
339 *C. elegans* (Figure 5A). In contrast, 50 % of worms dead after 13 days when they were placed
340 on Δ *essC* mutant strain lawn (Figure 5A). However, using the statistical Log rank test, we
341 observed that the survival kinetic of worms in contact with *S. lugdunensis* DSM WT or with
342 DSM Δ *essC* were identical (P = 0.7233).

343 In order to check a putative direct killing effect of *S. lungdunensis* against *C. elegans*, we used
344 the *C. elegans* liquid killing test adapted from³⁶. In this condition, we observed that *C. elegans*
345 in contact with *S. lungdunensis* DSM WT or *C. elegans* in contact with DSM Δ essC died rapidly
346 and with the same kinetics (Figure 5B) (P = 0.0864).

347

348

349

350 **References**

- 351 1. Schneewind, O. & Missiakas, D. M. Protein secretion and surface display in Gram-positive
352 bacteria. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1123–1139 (2012).
- 353 2. Stoop, E. J. M., Bitter, W. & van der Sar, A. M. Tubercle bacilli rely on a type VII army
354 for pathogenicity. *Trends Microbiol.* **20**, 477–484 (2012).
- 355 3. Unnikrishnan, M., Constantinidou, C., Palmer, T. & Pallen, M. J. The Enigmatic Esx
356 Proteins: Looking Beyond Mycobacteria. *Trends Microbiol.* **25**, 192–204 (2017).
- 357 4. Gröschel, M. I., Sayes, F., Simeone, R., Majlessi, L. & Brosch, R. ESX secretion systems:
358 mycobacterial evolution to counter host immunity. *Nat. Rev. Microbiol.* **14**, 677–691
359 (2016).
- 360 5. Stanley, S. A., Raghavan, S., Hwang, W. W. & Cox, J. S. Acute infection and macrophage
361 subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc.*
362 *Natl. Acad. Sci.* **100**, 13001–13006 (2003).
- 363 6. Baptista, C., Barreto, H. C. & São-José, C. High Levels of DegU-P Activate an Esat-6-Like
364 Secretion System in *Bacillus subtilis*. *PLoS ONE* **8**, e67840 (2013).
- 365 7. Burts, M. L., Williams, W. A., DeBord, K. & Missiakas, D. M. EsxA and EsxB are secreted
366 by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus*
367 infections. *Proc. Natl. Acad. Sci.* **102**, 1169–1174 (2005).
- 368 8. Kneuper, H. *et al.* Heterogeneity in *ess* transcriptional organization and variable
369 contribution of the Ess/Type VII protein secretion system to virulence across closely related
370 *Staphylococcus aureus* strains: Type VII secretion in *S. aureus*. *Mol. Microbiol.* **93**, 928–
371 943 (2014).
- 372 9. Anderson, M., Aly, K. A., Chen, Y.-H. & Missiakas, D. Secretion of atypical protein
373 substrates by the ESAT-6 Secretion System of *Staphylococcus aureus*: ESAT-6 secretion
374 in *S. aureus*. *Mol. Microbiol.* **90**, 734–743 (2013).
- 375 10. Cao, Z., Casabona, M. G., Kneuper, H., Chalmers, J. D. & Palmer, T. The type VII secretion
376 system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor bacteria.
377 *Nat. Microbiol.* **2**, 16183 (2016).
- 378 11. Burts, M. L., DeDent, A. C. & Missiakas, D. M. EsaC substrate for the ESAT-6 secretion
379 pathway and its role in persistent infections of *Staphylococcus aureus*. *Mol. Microbiol.* **69**,
380 736–746 (2008).
- 381 12. Anderson, M. *et al.* EssE Promotes *Staphylococcus aureus* ESS-Dependent Protein
382 Secretion To Modify Host Immune Responses during Infection. *J. Bacteriol.* **199**, (2017).
- 383 13. Casabona, M. G. *et al.* Functional analysis of the EsaB component of the *Staphylococcus*
384 *aureus* Type VII secretion system. *Microbiology* **163**, 1851–1863 (2017).
- 385 14. Warne, B. *et al.* The Ess/Type VII secretion system of *Staphylococcus aureus* shows
386 unexpected genetic diversity. *BMC Genomics* **17**, (2016).
- 387 15. Argemi, X., Hansmann, Y., Riegel, P. & Prévost, G. Is *Staphylococcus lugdunensis*
388 Significant in Clinical Samples? *J. Clin. Microbiol.* **55**, 3167–3174 (2017).
- 389 16. Frank, K. L., del Pozo, J. L. & Patel, R. From Clinical Microbiology to Infection
390 Pathogenesis: How Daring To Be Different Works for *Staphylococcus lugdunensis*. *Clin.*
391 *Microbiol. Rev.* **21**, 111–133 (2008).

- 392 17. Yeh, C.-F. *et al.* Clinical Features, Outcomes, and Molecular Characteristics of
393 Community- and Health Care-Associated *Staphylococcus lugdunensis* Infections. *J. Clin.*
394 *Microbiol.* **54**, 2051–2057 (2016).
- 395 18. Arias, M. *et al.* Skin and soft tissue infections caused by *Staphylococcus lugdunensis*:
396 Report of 20 cases. *Scand. J. Infect. Dis.* **42**, 879–884 (2010).
- 397 19. Bellamy, R. & Barkham, T. *Staphylococcus lugdunensis* Infection Sites: Predominance of
398 Abscesses in the Pelvic Girdle Region. *Clin. Infect. Dis.* **35**, e32–e34 (2002).
- 399 20. Heldt Manica, L. A. & Cohen, P. R. *Staphylococcus lugdunensis* Infections of the Skin and
400 Soft Tissue: A Case Series and Review. *Dermatol. Ther.* **7**, 555–562 (2017).
- 401 21. Liu, P.-Y. *et al.* *Staphylococcus lugdunensis* Infective Endocarditis: A Literature Review
402 and Analysis of Risk Factors. *J. Microbiol. Immunol. Infect.* **43**, 478–484 (2010).
- 403 22. Kyaw, H., Shaikh, A. Z., Yaratha, G. & Deepika, M. Septic Embolic Stroke Resulting From
404 *Staphylococcus lugdunensis* Endocarditis. *Ochsner J.* **17**, 184–188 (2017).
- 405 23. Argemi, X. *et al.* Kinetics of biofilm formation by *Staphylococcus lugdunensis* strains in
406 bone and joint infections. *Diagn. Microbiol. Infect. Dis.* **88**, 298–304 (2017).
- 407 24. Douiri, N. *et al.* *Staphylococcus lugdunensis* : a virulent pathogen causing bone and joint
408 infections. *Clin. Microbiol. Infect.* **22**, 747–748 (2016).
- 409 25. Bocher, S., Tonning, B., Skov, R. L. & Prag, J. *Staphylococcus lugdunensis*, a Common
410 Cause of Skin and Soft Tissue Infections in the Community. *J. Clin. Microbiol.* **47**, 946–
411 950 (2009).
- 412 26. Lambe, D. W., Ferguson, K. P., Keplinger, J. L., Gemmell, C. G. & Kalbfleisch, J. H.
413 Pathogenicity of *Staphylococcus lugdunensis*, *Staphylococcus schleiferi*, and three other
414 coagulase-negative staphylococci in a mouse model and possible virulence factors. *Can. J.*
415 *Microbiol.* **36**, 455–463 (1990).
- 416 27. Sandoe, J. A. T. & Longshaw, C. M. Ventriculoperitoneal shunt infection caused by
417 *Staphylococcus lugdunensis*. *Clin. Microbiol. Infect.* **7**, 385–387 (2001).
- 418 28. Tan, T. Y., Ng, S. Y. & Ng, W. X. Clinical Significance of Coagulase-Negative
419 Staphylococci Recovered from Nonsterile Sites. *J. Clin. Microbiol.* **44**, 3413–3414 (2006).
- 420 29. Chassain, B. *et al.* Multilocus Sequence Typing Analysis of *Staphylococcus lugdunensis*
421 Implies a Clonal Population Structure. *J. Clin. Microbiol.* **50**, 3003–3009 (2012).
- 422 30. Didi, J., Lemee, L., Gibert, L., Pons, J.-L. & Pestel-Caron, M. Multi-Virulence-Locus
423 Sequence Typing of *Staphylococcus lugdunensis* Generates Results Consistent with a
424 Clonal Population Structure and Is Reliable for Epidemiological Typing. *J. Clin. Microbiol.*
425 **52**, 3624–3632 (2014).
- 426 31. Arnaud, M., Chastanet, A. & Débarbouillé, M. New Vector for Efficient Allelic
427 Replacement in Naturally Nontransformable, Low-GC-Content, Gram-Positive Bacteria.
428 *Appl. Environ. Microbiol.* **70**, 6887–6891 (2004).
- 429 32. Szabados, F. *et al.* Occurrence of genes of putative fibrinogen binding proteins and
430 hemolysins, as well as of their phenotypic correlates in isolates of *S. lugdunensis* of
431 different origins. *BMC Res. Notes* **4**, 113 (2011).
- 432 33. Marlinghaus, L. *et al.* Construction and characterization of three knockout mutants of the
433 *fbl* gene in *Staphylococcus lugdunensis*: characterization of isogenic mutants of *fbl*. *APMIS*
434 **120**, 108–116 (2012).

- 435 34. N'Diaye, A. *et al.* Effect of Substance P in *Staphylococcus aureus* and *Staphylococcus*
436 *epidermidis* Virulence: Implication for Skin Homeostasis. *Front. Microbiol.* **7**, (2016).
437 35. Blier, A.-S. *et al.* C-type natriuretic peptide modulates quorum sensing molecule and toxin
438 production in *Pseudomonas aeruginosa*. *Microbiol. Read. Engl.* **157**, 1929–1944 (2011).
439 36. Chua, S. L. *et al.* Dispersed cells represent a distinct stage in the transition from bacterial
440 biofilm to planktonic lifestyles. *Nat. Commun.* **5**, 4462 (2014).
441 37. Kentache, T., Ben Abdelkrim, A., Jouenne, T., Dé, E. & Hardouin, J. Global Dynamic
442 Proteome Study of a Pellicle-forming *Acinetobacter baumannii* Strain. *Mol. Cell.*
443 *Proteomics MCP* **16**, 100–112 (2017).
444 38. Sifri, C. D., Begun, J., Ausubel, F. M. & Calderwood, S. B. *Caenorhabditis elegans* as a
445 model host for *Staphylococcus aureus* pathogenesis. *Infect. Immun.* **71**, 2208–2217 (2003).
446

447 **Acknowledgements**

448 This study was supported by an educational grant for Jérémie Lebeurre from the Charles Nicolle
449 Foundation.

450

451 **Author contributions statement**

452 J.L., S.D. and M.P.C. designed the study. J.L and C.D. performed the genetic and transcriptional
453 analyses. J.L., C.D. and S.L. performed the mutagenesis experiments. O.L and T.C. performed
454 cytotoxicity assay and *C. elegans* virulence assay. P.C., J.H., and C.G. performed protein
455 extractions and secretome analysis. J.L., S.D. and M.P.C. wrote the manuscript.

456

457 **Additional information**

458 Competing financial interests: the authors declare that they have no competing interests.

459

460

461 **Figure legends**

462 **Figure 1.**

463 Transcriptional analysis of the *ess* locus of *S. lugdunensis*.

464 A. Schematic representation of the *S. lugdunensis* *ess* locus. Genes encoding putative secreted
465 proteins are coloured in red, putative transmembrane proteins in green, putative cytoplasmic
466 protein in blue and hypothetical proteins in white.

467 B. Co-transcription analysis by RT-PCR on RNA from the seven strains of *S. lugdunensis*. The
468 expected sizes for PCR products 1 to 10: 268, 282, 465, 560, 378, 558, 356, 433, 311 and
469 315 bp respectively.

470

471 **Figure 2.**

472 Kinetic analysis of *essC*, *esxA* and *lvg* genes expression by qRT-PCR. RNA was isolated from
473 DSM WT at OD₆₀₀ of 1.0, 2.0, 3.0 and 5.0. A reverse transcription of the RNA followed by
474 qPCR was performed to determine the normalized RNA levels ($2^{-\Delta Ct}$ method) by using the 16S
475 rRNA gene as the endogenous control. Histograms are expressed as mean \pm SEM. * $P < 0.05$
476 was determined by using the non-parametric Kruskal Wallis test.

477

478 **Figure 3.**

479 A. Analysis of the hemolytic activity of the strains DSM WT and DSM Δ *essC* on Colombia agar
480 plates with 5% horse blood. B. Comparison of the β -hemolysis diameter measured between the
481 DSM WT and DSM Δ *essC*. Histograms are expressed as mean \pm SEM. *** $P < 0.001$ was
482 determined by using the unpaired *t*-test. C. Synergic hemolysis assessed by CAMP-test on
483 tryptic soy agar with 5% sheep blood. White arrows indicate the synergic hemolysis areas
484 between *S. aureus* ATCC29123 and *S. lugdunensis* DSM WT strains.

485

486 **Figure 4.**

487 Cytotoxicity towards HaCaT cells. Comparative effect on the cytotoxicity of *S. lugdunensis*
488 DSM WT and DSM Δ *essC* strains toward HaCaT keratinocytes. Cytotoxicity levels were
489 measured by the release of lactate dehydrogenase (LDH). Histograms are expressed as mean \pm
490 SEM. The results are representative of three independent experiments.

491

492 **Figure 5.**

493 *S. lugdunensis* virulence toward *Caenorhabditis elegans* worms. Kinetics of survival of *C.*
494 *elegans* exposed to DSM WT and DSM Δ *essC* (A) in solid killing assay (slow killing) and (B)

495 in liquid killing assay (fast killing). Nematode survival was calculated by the Kaplan-Meier
496 method, and survival differences were tested for significance by using the log rank test.
497
498

Figure 1.

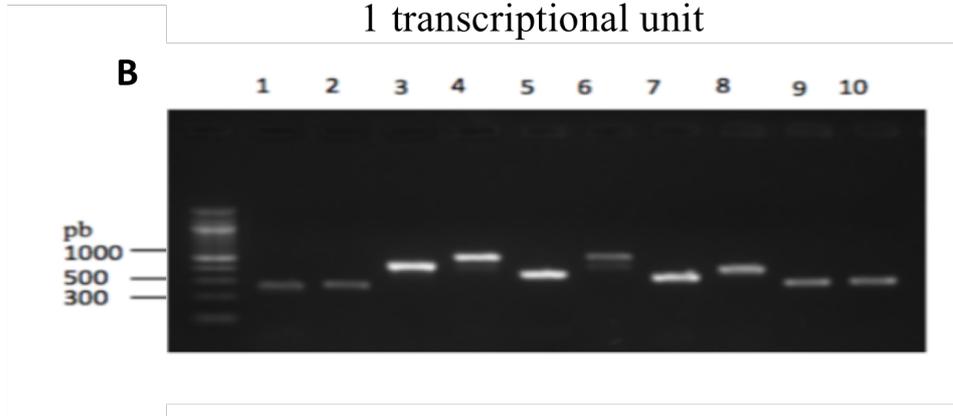
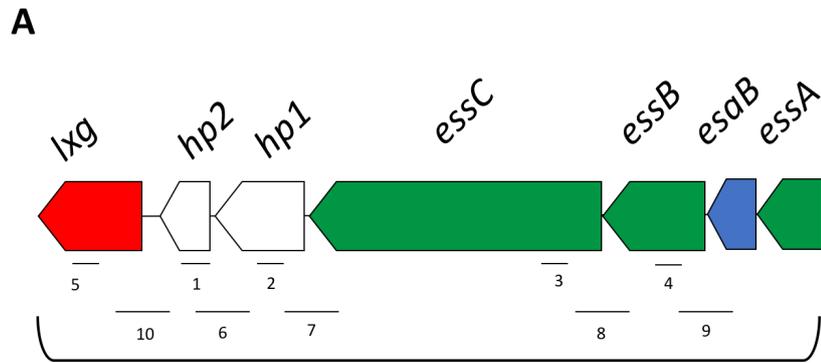


Figure 2.

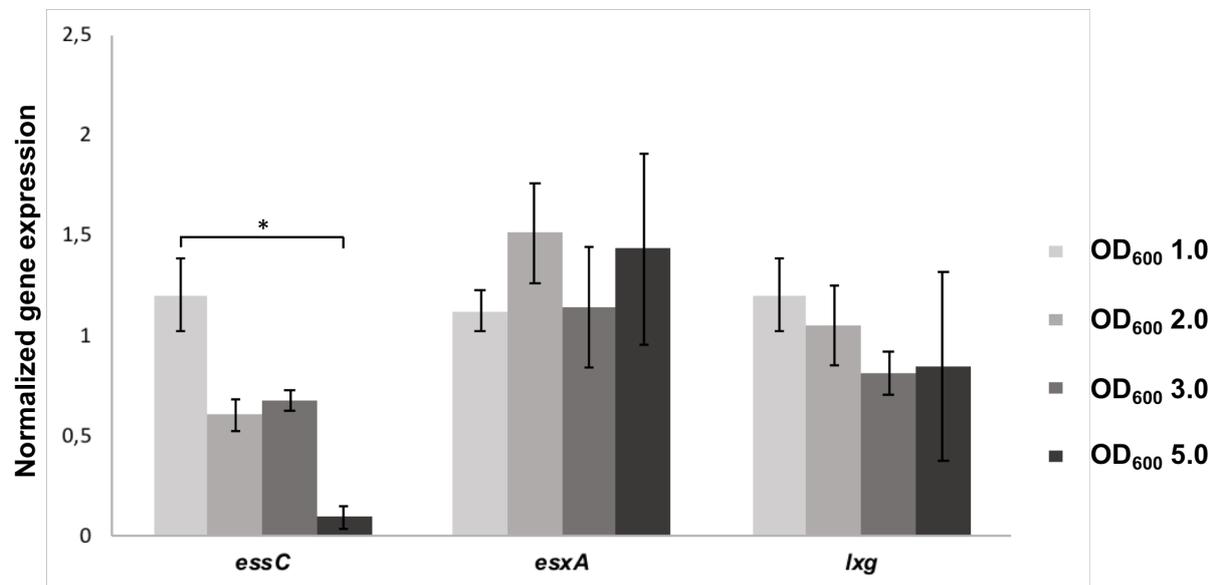


Figure 3.

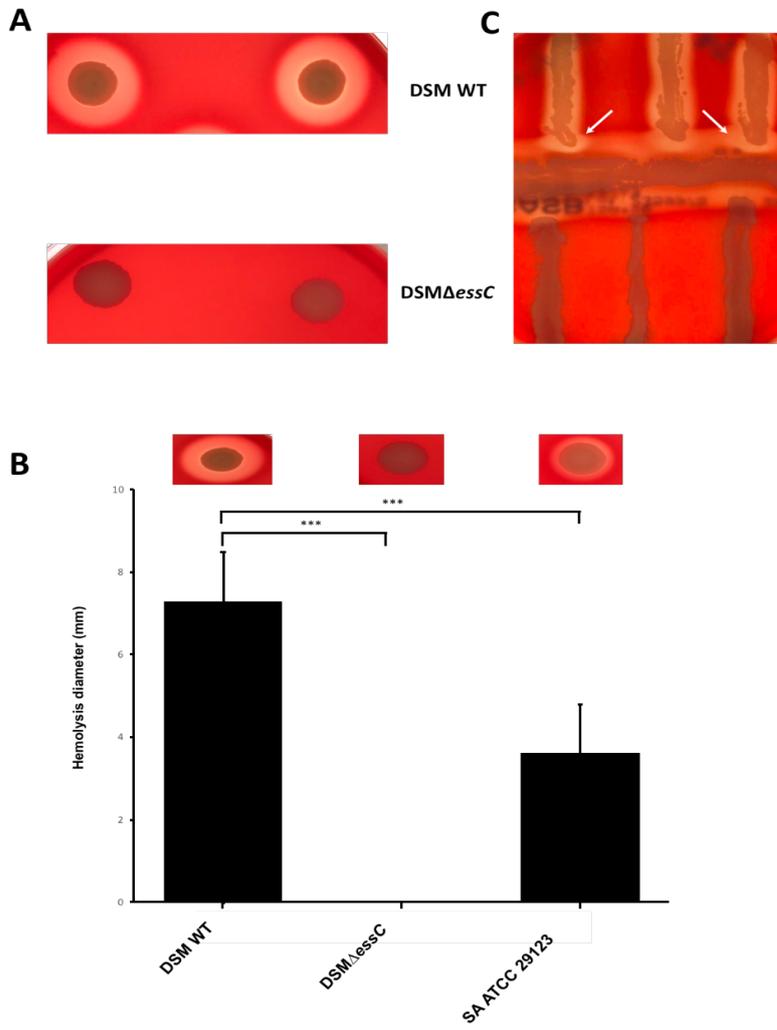


Figure 4.

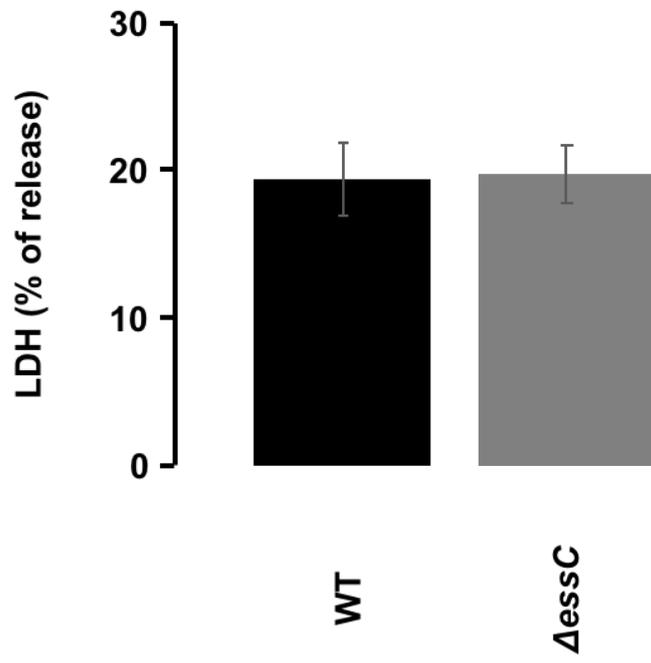
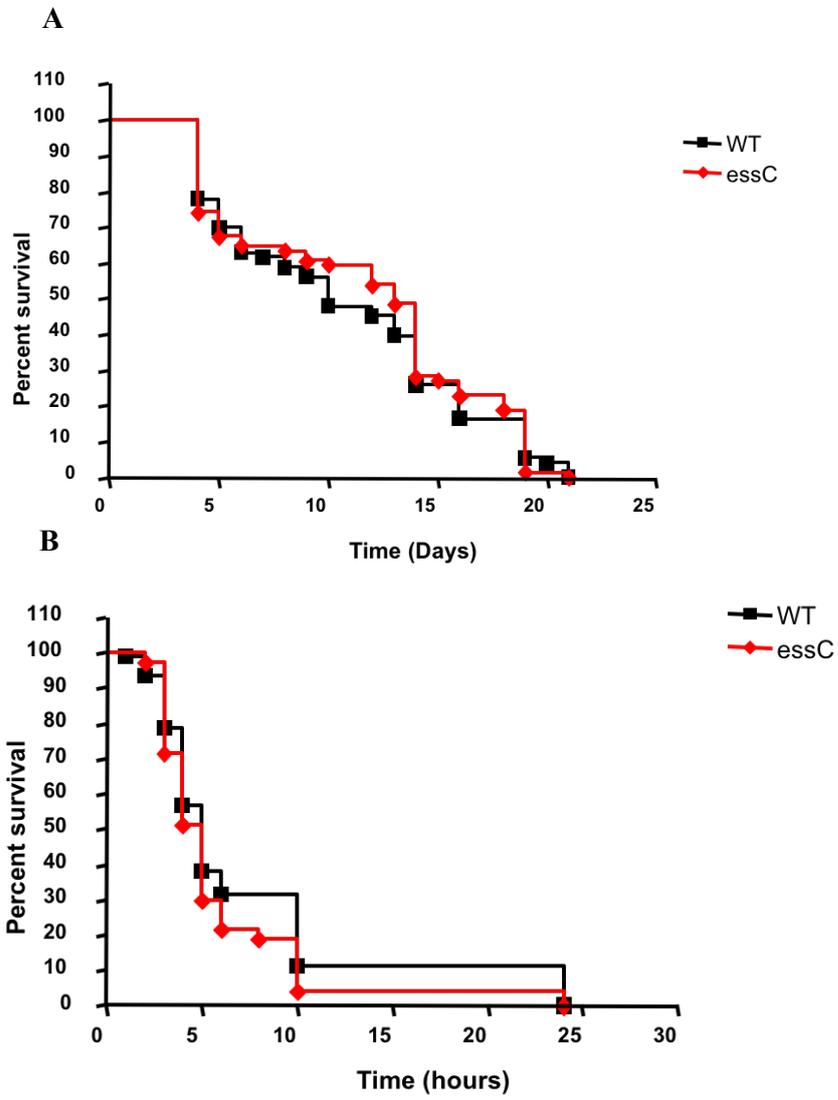


Figure 5.



Tables

Table 1. Distribution of the major polymorphisms of the *ess* locus in the collection of 94 clinical isolates of *S. lugdunensis*.

CC ^a	Number of isolates per CC	ST ^b	Number of isolates per ST	Number of isolates with cluster of 7 genes	Number of isolates with Tn ^c
1	34	1	16	9	16
		6	8	1	0
		7	1	1	1
		12	5	0	0
		15	4	0	0
2	17	2	15	15	2
		14	1	1	0
		19	1	1	0
3	20	3	18	18	0
		16	1	1	0
		20	1	1	0
4	5	4	3	0	0
		9	1	0	0
		17	1	0	0
5	5	5	4	0	0
		18	1	0	0
6	9	10	9	0	0
Singletons			4	0	0
Total			94	48	19

^aComplex Clonal, ^bSequence Type, ^cTransposon

Supplementary informations

Table S1. Primer pairs used in this study for identification of the transposon and the cluster of seven genes.

Primers	Sequence 5'-3'
TrEss_ExtF	ATTGGGATGCGGTAGAT
TrEss_ExtR	ACTGTACCGGATAGTGATGAA
TrEss_IntF	CACCTATTATCAACGCAACTA
TrEss_IntR	GGAGTAATCGCACCAAATA
TrEss_ItrR	CTTTCGCCTCGTTCAA
CDS_ExtF	CCTGGAATCATATCACTCA
CDS_ExtR	ATCAAGCATGAAAAATCTAAC
CDS_IntF	CAAATATGAGTAATGCCGTCA
CDS_intR	ATATTCAAACGTCACGGGTTC
CDS_ItrR	GGGTATTCATGGTCCTATC

Table S2. RT-PCR and RT-qPCR primer pairs used in this study

Primers	Sequence 5'-3'
HP1_1-F	TAATGATGCAGTTTTCGT
HP1_1-R	CAAATGAAGGCTATATAAGTA
HP2_2-F	CGATATTTTATCTGCATT
HP2_2-R	GATTTATTTAGATCCAATAGA
essC_3-F	CGCCAACATAGTGTTTC
essC_3-R	GTAAACGCTTATTAAGTCAAC
essB_4-F	ATGAGAAGTTACGACGATAAT
essB_4-R	TATGCGAGAGATTCCTAAA
lxg_5-F	ATAATCCTCTGAACCCAATCT
lxg_5-R	CCTTGGTATAAGAGCGAAGTA
HP1_HP2_6-F	TTTCATGGCATTAGATAACTC
HP1_HP2_6-R	TTCAATTACAATTACGCA
HP2_essC_7-F	ATATATTGTGCGAATACGATT
HP2_essC_7-R	GGGGGTATAAGCTCAC
essC_essB_8-F	AATGCTTTAGCGTGAT
essC_essB_8-R	ACATATTTAGATGATAACGAT
essB_esaB_9-F	TTTAGGAATCTCTCGCATATC
essB_esaB_9-R	AAACTATCAACACGGAAGC
lxg_HP2_10-F	TTATAGCTCTCGAAATTGTAC
lxg_HP2_10-R	GCCCAAATGATAAGTA

3. Discussion et perspectives

La diversité génétique de *S. lugdunensis*, espèce particulièrement virulente de SCN, a jusqu'à ce jour été principalement explorée par des analyses *multilocus* telles que la MLST^{24,208,249} ou la MVLST⁹⁸. La MLST, considérée comme une méthode de référence en phylogénie, a permis de montrer le caractère clonal marqué de cette espèce et l'existence de sept CC sans qu'aucun lien ne puisse être identifié entre lignées phylogénétiques et virulence.

Les objectifs des travaux présentés ici étaient (i) de séquencer le génome de trois souches pathogènes et de trois souches de portage, (ii) de comparer ces génomes avec ceux disponibles sur la base de données NCBI et (iii) de rechercher des déterminants spécifiques de la virulence de *S. lugdunensis*.

Le choix des souches utilisées pour le séquençage *Pacbio* a été contraint par la disponibilité des souches de portage⁹⁸. En effet, le portage de *S. lugdunensis*, naturellement sensible aux antibiotiques, n'étant pas recherché spécifiquement en routine, le souchier du laboratoire ne comportait que six souches de portage de deux ST éloignés : cinq de ST1 (CC1) et une de ST10 (CC6). Nous avons alors pris le parti de retenir deux souches de portage de ST1 (SL117 et SL122) d'une part et une souche de ST10 (SL118) d'autre part afin d'augmenter la diversité des génomes complets disponibles. Ces souches ont été recueillies à la faveur d'une étude française de portage inguinal réalisée chez des patients prélevés à leur arrivée aux urgences (SL122¹³) et d'une étude de portage suédoise réalisée chez des sujets non hospitalisés (SL117 et SL118¹⁴). Dans le but d'effectuer des comparaisons entre les génomes, les souches pathogènes ont été choisies pour représenter l'épidémiologie clinique des infections à *S. lugdunensis*^{2,122} : infection de la peau et des tissus mous (SL55), infection sur matériel (SL29) et endocardite infectieuse (SL13). Les souches pathogènes ont également été sélectionnées selon leur proximité phylogénétique avec les souches de portage de ST1 et de ST10 de façon à ce que les souches aient un fond génétique le plus proche possible. Ainsi, les souches SL29 et SL55 appartiennent respectivement au ST1 et au ST10. Toutefois, aucune souche isolée d'un contexte d'endocardite infectieuse de notre collection n'appartenait au ST1. La souche SL13 (ST6) a alors été sélectionnée car ce ST est phylogénétiquement proche du ST1. En effet, ces deux ST ne diffèrent l'un par rapport à l'autre par le polymorphisme d'un seul *locus* de MLST²⁰⁸.

Au moment du séquençage des six souches, il n'existait que deux génomes complets de *S. lugdunensis* publiés, celui de la souche HKU09-01¹³³ et celui de la souche N920143⁷³, qui sont toutes les deux des souches pathogènes, isolées respectivement à partir de pus d'une plaie cutanée et d'un abcès de sein. Ainsi, notre travail correspondait initialement au premier séquençage complet du génome de souches isolées en simple condition de portage et de souches de différents contextes cliniques sévères.

Depuis, deux études comportant un séquençage de génomes complets de cette espèce ont été publiées. L'équipe de Argemi *et al.*⁹⁶ a notamment séquencé le génome de six souches responsables d'infections et d'une souche isolée dans un contexte de colonisation (souche isolée d'un prélèvement réalisé à visée diagnostique mais dont le rôle dans l'infection n'a pas été retenu). Même si ces importants travaux ont permis d'augmenter significativement le nombre de génomes disponibles, leur impact était limité par la diversité génétique explorée puisque ces sept souches ont été isolées dans le cadre d'un essai clinique français monocentrique¹¹. Les souches que nous avons sélectionnées étaient d'origines géographiques très différentes puisqu'elles ont été collectées dans trois villes françaises et une suédoise⁹⁸. A l'heure actuelle, plusieurs autres génomes de souches provenant des États-Unis (quatre) et de Hong Kong (deux) ont été séquencés mais non publiés dans la littérature.

La diversité génétique de l'espèce n'est pas encore totalement explorée puisque plusieurs lignées phylogénétiques ne sont pas représentées dans NCBI, comme le CC2, alors qu'il s'agit d'une lignée prédominante dans plusieurs études épidémiologiques en France et en Asie^{23,24,208}. Le CC4, quant à lui, n'est pas une lignée très représentée dans ces études, mais il serait tout de même intéressant d'explorer le génome de souches de ce CC afin d'évaluer si elles présentent des événements génétiques particuliers et des variations spécifiques.

Par rapport à d'autres espèces bactériennes pathogènes ou à d'autres espèces de staphylocoques, le nombre de génomes disponibles de *S. lugdunensis* reste très faible puisqu'il s'élève aujourd'hui à 25 dont seulement 21 génomes complets. Pour *S. aureus* et *S. epidermidis*, il existe à ce jour 10011 et 482 séquences génomiques dont 335 et 16 génomes complets respectivement. De ce fait, *S. lugdunensis* reste une espèce particulièrement sous-étudiée par rapport à ces deux espèces de staphylocoques. De manière intéressante, pour *S. aureus* et *S. epidermidis*, il existe un très grand nombre de

draft genomes, mais le niveau de finition d'un génome dépend du type d'analyse. En effet, plusieurs études ont déjà permis la caractérisation du virulome, du résistome et du mobilome de ces espèces. Ainsi, l'obtention de nouveaux génomes complets, souvent onéreuse en termes de moyens et de temps, ne se justifie pas toujours pour des comparaisons génomiques. Pour *S. lugdunensis*, peu de facteurs de virulence ont été décrits et seules trois études à ce jour se sont attachées à comparer les génomes de cette espèce^{11,73,141}. Ainsi, l'obtention de génomes complets était nécessaire afin de réaliser une approche complète d'exploration génomique.

Le niveau de finition d'un génome dépend également de la technique de séquençage haut débit utilisée. Avant d'opter pour le séquençage *Pacbio*, il avait été entrepris au laboratoire, en collaboration avec l'unité Inserm U1245 de l'Université de Rouen, un séquençage du génome de ces six souches avec la technique *Illumina* Gallx générant des *reads* d'une taille de seulement 75 pb. Les travaux d'assemblage *de novo* avaient alors permis d'obtenir entre 15 et 33 contigs et malgré une tentative de circularisation des génomes, de nombreux *gaps* étaient encore présents, constituant une perte d'informations importante pour des analyses génomiques comparatives.

C'est pour cette raison que nous avons choisi le séquençage de troisième génération *Pacbio* (en collaboration avec le *Genomic Research Laboratory* des Hôpitaux Universitaires de Genève) qui présente plusieurs avantages. Il permet de générer de longues séquences de l'ordre de plusieurs dizaines de milliers de nucléotides permettant un assemblage *de novo* facilité. Le séquençage *Pacbio* permet alors l'obtention d'un unique *contig* correspondant au chromosome bactérien circularisé²⁵⁰. Ainsi, le séquençage *Pacbio* a généré un *contig* de l'ordre de 2,6 Mb pour nos six souches par rapport au séquençage *Illumina* qui avait généré de nombreux *contigs* d'une taille moyenne de seulement 100 kb. Le séquençage *Pacbio* est aussi une technique innovante pour détecter la présence de plasmides au sein des souches bactériennes ; ce qui n'avait pas été possible avec les données de séquençage *Illumina*.

Malgré ces avantages, le séquençage *Pacbio* souffre d'un important taux d'erreur, de l'ordre de 10 à 15%²⁵¹. En effet, les *reads* de séquençage peuvent contenir des insertions, des délétions ou des substitutions²⁵². Toutefois, ce taux d'erreur peut être compensé par le nombre de lectures d'une même base, appelé profondeur de séquençage. En moyenne, chaque position du génome, autrement dit chaque nucléotide, est séquencé plusieurs fois.

Ainsi, une profondeur suffisamment importante permet d'augmenter la précision du séquençage et la fiabilité de ses résultats²⁵¹. Par ailleurs, le séquençage *Pacbio* reste encore aujourd'hui à un prix élevé (\$1000 par Gigabases) comparativement à d'autres techniques de séquençage comme le séquençage *Illumina* HiSeq (\$7 par Gigabases)²⁵¹, constituant alors un frein au nombre de génomes séquencés lors de notre étude.

Le séquençage *Pacbio* a également été choisi en raison de notre souhait d'analyser le *pan-genome* de *S. lugdunensis*. En effet, afin de caractériser précisément le *core-genome* et le génome accessoire de cette espèce, il était nécessaire de n'utiliser que des séquences génomiques complètes pour s'assurer de la présence d'un gène chez toutes les souches ou non. De cette manière, l'analyse du *pan-genome* nous a permis de confirmer le caractère "fermé" du génome de *S. lugdunensis* puisque l'apport de six génomes complets n'a que peu augmenté le nombre de gènes total de l'espèce, autour de 3000 gènes¹⁴¹. Argemi *et al.*¹⁴¹ ont montré qu'à l'inverse, *S. aureus* et *S. epidermidis* possèdent un *pan-genome* ouvert qui s'enrichit régulièrement des nouveaux génomes séquencés pour comporter actuellement 3828 et 3864 gènes respectivement. De la même façon, et comme suggéré par les données de MLST, le *core-genome* de *S. lugdunensis* est très conservé avec un pourcentage moyen d'identité nucléotidique qui varie entre les souches de 99,5% à 99,9% alors que celui-ci varie de 97,5 à 99,8% pour *S. aureus* et de 96,6 à 99,7% pour *S. epidermidis*¹⁴¹.

Le caractère fermé du *pan-genome* de *S. lugdunensis* peut être expliqué par la présence de plusieurs systèmes pouvant limiter les transferts horizontaux de gènes tels que le système CRISPR/Cas, les SRM de type I et II¹⁴¹, mais aussi par la présence d'un peptidoglycane épais¹⁴⁰. La comparaison génomique effectuée par Argemi *et al.*¹⁴¹ et nous-mêmes a mis en évidence que la moitié (9/21 soit 43%) des souches de *S. lugdunensis* avaient acquis un système CRISPR/Cas. Ce pourcentage est très supérieur aux 15% que Rossi *et al.*²⁵³ ont déterminé à partir de 122 génomes de 15 espèces de SCN différentes. En outre, ils ont montré chez des espèces de SCN telles que *S. epidermidis* ou *S. haemolyticus*, que seuls 9% et 3% respectivement possédaient un système CRISPR/Cas. Cependant, ce fort pourcentage observé chez *S. lugdunensis* doit être pondéré puisque toutes les souches possédant ce système CRISPR/Cas appartenaient uniquement au CC1, surreprésenté (9/21, soit 43%) par rapport aux autres CC (CC3 [5/21, soit 24%], CC5 [1/21, soit 4%], CC6 [4/21, soit 19%] et CC7 [2/21, soit 10%]).

Ce résultat est toutefois inattendu et paradoxal puisque les souches appartenant à ce CC1 sont aussi les souches qui présentent le plus d'EGM. De plus, les souches appartenant au CC3, qui ne possèdent ni de système CRIPR/Cas, ni de SRM, sont à l'inverse des souches qui présentent peu d'EGM (une seule souche sur cinq possède un plasmide). Il est alors possible d'émettre plusieurs hypothèses : (i) un système CRISPR/Cas non fonctionnel chez les souches de CC1, (ii) le sérotype capsulaire particulier des souches de CC3 (cf *infra*) pourrait être une barrière physique limitant la transformation de la cellule bactérienne ou (iii) l'acquisition des prophages et des plasmides pourraient être antérieure à celle du système CRISPR/Cas chez les souches de CC1. Afin de valider ou d'invalidier ces hypothèses, il pourrait être intéressant d'étudier le nombre de gènes total pour chaque lignée phylogénétique de *S. lugdunensis* et d'étudier si certains CC sont plus réceptifs que d'autres aux transferts de gènes horizontaux.

Malgré la présence du système CRISPR/Cas et de plusieurs SRM de type I et II, la majorité des souches (15/21, soit 71%) ont acquis plusieurs EGM. En effet, dans cette étude, cinq régions phagiques, deux îlots génomiques et un plasmide, jamais décrits chez *S. lugdunensis*, ont été identifiés ; ce qui est surprenant pour une espèce réfractaire à la transformation.

Compte tenu des homologues observées, les prophages identifiés ont probablement pour origine *S. aureus*, *S. epidermidis* et *S. hominis*, suggérant donc un transfert de gènes horizontal entre ces espèces de staphylocoques. Aucun gène de virulence n'a été identifié dans ces régions, mais plusieurs gènes ont été identifiés comme pouvant conférer un avantage aux souches possédant ces régions phagiques puisqu'ils codent pour une autolysine bifonctionnelle et une hydrolase de paroi cellulaire LytN. Chez *S. lugdunensis*, l'autolysine bifonctionnelle AtIL est impliquée dans le remodelage du peptidoglycane et la croissance cellulaire mais aussi dans la virulence puisqu'elle contribue à la formation du biofilm⁸¹. D'autre part la protéine LytN de *S. aureus* favorise également le remodelage du peptidoglycane et la croissance cellulaire²⁵⁴. Ainsi, les souches de *S. lugdunensis* ayant acquis ces régions phagiques pourraient y gagner un avantage en termes de croissance et de capacité à coloniser l'hôte.

Dans cette étude, aucun déterminant de la virulence n'a été identifié chez les souches de *S. lugdunensis* responsables d'infection par rapport aux souches de portage. Néanmoins, une région de 13 kb partiellement homologue aux îlots de pathogénicité de

S. aureus (SaPI) SaPITokyo11212 et SaPlivm60 a été identifiée chez les souches SL13 et VISLISI_33. Chez *S. aureus*, les SaPI jouent un rôle important dans la virulence puisqu'ils permettent l'acquisition de facteurs de virulence, tels que des toxines et des superantigènes²⁵⁵. Chez les deux souches de *S. lugdunensis*, la région identifiée portait des gènes codant une protéine de liaison à la pénicilline Ear et un transporteur du fer FhuD. Ces gènes ont déjà été décrits par Novick *et al.*²⁵⁵ dans plusieurs SaPI de *S. aureus* tels que SaPlmw2, SaPlm4 et SaPI1, où ils sont généralement associés à des gènes codant pour les toxines Seq, Sek et TSST-1. Cependant, aucun de ces gènes n'a été identifié chez *S. lugdunensis*. Le rôle de la protéine Ear n'a encore jamais été démontré à ce jour²⁵⁵ tandis que la protéine FhuD est impliquée dans le transport du fer et facilite l'acquisition de l'hème dans le sang chez *S. aureus*²⁵⁶.

A la faveur de notre analyse comparative génomique, plusieurs gènes de résistance ont également été identifiés tels que les gènes de résistance à la fosfomycine *fosB* (21/21, soit 100%), à la pénicilline *blaZ* (7/21, soit 33%) et une résistance plasmidique encore jamais décrite chez *S. lugdunensis* à la tétracycline *tetK* (1/21, soit < 1%). L'identification *in silico* de tels gènes est un élément en faveur d'une résistance à ces antibiotiques, mais ne témoigne en rien du phénotype. Il serait alors pertinent de déterminer le phénotype de résistance de ces souches.

Concernant la résistance à la pénicilline, ces résultats sont conformes à ceux de McHardy *et al.*²⁰ qui ont montré que le gène *blaZ* avait été détecté par PCR pour 39,1% (36/92) des isolats sensibles à l'oxacilline. Chez *S. lugdunensis*, le taux de résistance à la pénicilline varie considérablement selon les études allant de 48% à 69% probablement du fait de problèmes méthodologiques de détection *in vitro*²⁰.

En revanche, parmi nos 21 souches, aucune n'a présenté de gène de résistance *mecA* porté par une cassette SCC*mec*, malgré l'émergence récente de clones résistants à la méticilline²¹. Cette augmentation de la résistance (20²¹ à 32%²⁴) à cet antibiotique a surtout été décrite en Asie, notamment à Hong Kong, liée à la dissémination de clones porteurs de la cassette SCC*mec* dans un contexte d'infections associées aux soins²⁴⁶. Toutefois, ce niveau de résistance ne semble concerner que l'Asie puisqu'une étude internationale a montré que la résistance à la méticilline reste peu élevée chez *S. lugdunensis*, soit 8%²⁰. Pour ce qui est de la résistance à la tétracycline, elle est rare chez *S. lugdunensis*. En effet, il n'a été fait état que de quelques résistances isolées^{2,257},

mais qui pourraient s'avérer problématiques en cas d'émergence compte-tenu de l'intérêt de ces molécules dans le traitement des infections ostéoarticulaires.

Un gène de résistance à l'acide fusidique *fusB* a également été identifié chez une souche de *S. lugdunensis*. Ce gène était porté par un îlot génomique de 16 kb portant également un gène de virulence putatif *vapE*. Cet îlot génomique est homologue à celui de 16,6 kb de *S. aureus* nommé SaRIfusB46, décrit dans le clone épidémique CS6 responsable d'impétigo²⁵⁸. Chez *S. epidermidis*, des îlots génomiques porteurs du gène *vapE* et du gène *fusB* ont été identifiés dans des isolats cliniques résistants à l'acide fusidique²⁵⁹. La protéine VapE a initialement été identifiée chez des souches de *Dichelobacter nodosus* responsables d'infections sévères chez les ovins²⁶⁰. Le gène codant cette protéine fait partie d'une région associée à la virulence nommée *vap* chez cette bactérie. La présence d'un gène d'intégrase en amont du gène *vapE* suggère que les bactériophages pourraient jouer un rôle dans l'évolution et le transfert de ce facteur de virulence²⁶¹. De plus, chez *Streptococcus suis*, il a été montré dans un modèle murin de colonisation le rôle de la protéine VapE dans la virulence par la formation d'un mutant de délétion²⁶². En effet, dans ce modèle, les souris infectées par voie intra-péritonéale avec la souche sauvage manifestaient des symptômes cliniques graves, notamment une dépression, une apathie, une fièvre, une anorexie, une émaciation, des yeux gonflés et des troubles nerveux. Toutes les souris infectées avec la souche sauvage sont mortes dans les 2 jours suivant l'infection. En revanche, les souris inoculées avec la souche mutante n'ont présenté que de légers symptômes et ont survécu. Le mécanisme d'action de cette protéine dans la virulence n'a pas encore été déterminé et son rôle dans la virulence des staphylocoques reste hypothétique. Il serait alors intéressant de confirmer si cette protéine constitue un réel facteur de virulence chez *S. lugdunensis*.

Ainsi, l'identification de ces gènes de résistance et de virulence dans les génomes de *S. lugdunensis* suggère des transferts de gènes horizontaux d'origine staphylococcique par l'intermédiaire d'îlots génomiques, de prophages et de plasmides, même si ce phénomène est probablement rare du fait de la difficulté à transformer cette espèce¹⁴¹. La faible efficacité de transformation de cette espèce semble être l'hypothèse la plus probable pour expliquer pourquoi *S. lugdunensis* reste, à la différence de *S. aureus* et des autres SCN, particulièrement sensible à la plupart des molécules anti-staphylococciques. Les barrières aux transferts de gènes horizontaux telles que le système CRISPR/Cas, les SRM

de type I et II, la paroi et la capsule sont autant d'obstacles à l'intégration d'ADN étranger portant des déterminants de la résistance, mais également de la virulence.

Aucun polymorphisme de séquence de gènes régulateurs putatifs qui pourrait se traduire par une expression différentielle de facteurs de virulence n'a été détecté, en dehors du *locus agr*. Aucun lien entre le *agr-type* de *S. lugdunensis* et le type d'infection ou le caractère pathogène des souches n'a été démontré dans notre étude. Cependant, nous avons mis en évidence que toutes les souches de CC3 appartenaient au *agr-type* II alors que toutes les autres souches étaient associées au *agr-type* I. Tseng *et al.*¹²¹ ont montré que sept sur 11 souches isolées de bactériémie qui appartenaient au *agr-type* II, présentaient une activité hémolytique et protéasique supérieure à celle des quatre souches appartenant au *agr-type* I, mais une capacité comparable à former du biofilm.

Chez *S. lugdunensis*, il a été identifié un nombre plus faible d'allèles du *locus agr* comparativement à *S. aureus* (*agr-type* I à IV) et à *S. epidermidis* (*agr-type* I à III)²⁶³. Plusieurs études ont recherché un lien entre *agr-type* et différentes caractéristiques moléculaires, phénotypiques ou cliniques^{120,264–266}. Jarraud *et al.*¹²⁰ ont ainsi montré à partir de 198 souches de *S. aureus* qu'il existait une forte corrélation entre l'*agr-type* et le contexte clinique. En effet, les souches responsables d'endocardite de *S. aureus* étaient principalement associées aux *agr-type* I et II tandis que les souches d'*agr-type* IV étaient principalement isolées de syndrome d'exfoliation généralisé ou d'impétigo bulleux. Il a également été observé une corrélation entre l'*agr-type* et la capacité à former du biofilm. Effectivement, les souches de *S. aureus* associées aux *agr-types* II et III ont été montrées comme formant plus de biofilm que les autres souches²⁶⁷.

Par ailleurs, il a aussi été montré que les *agr-types* pouvaient varier en fonction du CC, du profil de résistance aux antibiotiques, de la présence de gènes codant des toxines, de l'expression des facteurs de virulence, de la structure du peptide auto-inducteur et de la présence d'EGM²⁶⁶. Compte-tenu de la parenté du *locus agr* de *S. lugdunensis* avec celui de *S. aureus* et de *S. epidermidis*, il est probable que ce système intervienne également dans la régulation de l'expression d'exoprotéines et de protéines de surface chez cette espèce. Cependant, aucune étude à ce jour ne s'est attachée à caractériser le rôle de ce régulateur global de la virulence chez *S. lugdunensis*. Cette possible corrélation entre *agr-types* et CC de *S. lugdunensis* devra être confirmée par l'étude d'une plus large collection de souches bien documentées sur le plan clinique et investiguée parallèlement

avec des approches phénotypiques, évaluant notamment la capacité comparée de souches de différents CC et/ou *agr*-types à former du biofilm.

De façon intéressante, les souches appartenant au CC3 se sont aussi distinguées des autres souches par la présence de l'allèle *cap8* du *locus* codant la capsule polysaccharidique. Hormis l'étude de Lambe *et al.*²⁶⁸ au cours de laquelle ils ont observé la capsule de *S. lugdunensis* par microscopie électronique en transmission, aucune étude n'a décrit à ce jour la composition, le rôle de la capsule dans la virulence ou même la prévalence des sérotypes capsulaires chez cette espèce. Chez *S. aureus*, les gènes codant les capsules de sérotypes 5 et 8 sont situés sur une région du chromosome de 17 kb ; le *locus* comporte 16 gènes, de *cap5A* (ou 8A) à *cap5P* (ou 8P). Parmi ces 16 gènes, 12 présentent une forte homologie entre les deux sérotypes capsulaires²⁶⁹. Les gènes régissant la spécificité du sérotype sont situés dans la région centrale du *locus*, comprenant quatre gènes *cap5H* (ou 8H) à *cap5K* (ou 8K), qui présentent une faible homologie. Par la formation d'un mutant de délétion du *locus agr*, il a été montré chez cette espèce la régulation positive de la transcription des gènes codant la capsule par *agr*^{269,270}. Le rôle de la capsule *in vivo* a été démontré dans des modèles murins de bactériémie dans lesquels les souches capsulées étaient plus virulentes que les souches délétées du *locus cap*²⁷¹. Il a aussi été montré que la capsule pouvait interférer dans la capacité d'adhésion des cellules bactériennes en masquant directement les adhésines de la paroi bactérienne²⁷². De telles approches mériteraient d'être développées chez *S. lugdunensis*.

En plus des variations génétiques spécifiques des souches de CC3 pour les *loci agr* et *cap*, d'importantes variations CC-dépendantes ont également été identifiées dans les gènes codant des protéines SIs. En effet, nous avons identifié que plusieurs protéines SIs étaient absentes ou potentiellement non fonctionnelles chez certaines souches de *S. lugdunensis*. Par exemple, les protéines SIsD et SIsE seraient non-fonctionnelles dans la plupart des souches appartenant au CC1, en raison d'une mutation non-sens et d'un décalage du cadre de lecture. À l'inverse, toutes les souches de CC1 présentaient le gène codant pour les protéines SIsA alors qu'il était absent de toutes les souches de CC3. Ces variations génétiques CC-dépendantes pourraient entraîner des différences phénotypiques, en particulier dans la capacité de formation de biofilm et d'adhésion entre les isolats de différentes lignées phylogénétiques. En effet, la protéine SIsA contient des domaines protéiques et une organisation structurale similaire à celle de la protéine Bhp, qui chez

S. epidermidis, promeut l'adhésion aux surfaces abiotiques et l'adhésion intercellulaire lors de la formation de biofilm^{83,273}. En outre, la protéine SIsD présente une organisation structurale similaire à celle des protéines appartenant à la famille des adhésines Sdr telles que SdrF et SdrG²⁷⁴. Chez *S. epidermidis*, ces protéines sont impliquées dans l'adhésion au fibrinogène humain ou au collagène de type I²⁷⁵.

Il a été suggéré par Otto *et al.*²⁷⁶ que la redondance fonctionnelle des nombreuses MSCRAMM de *S. epidermidis* souligne l'importance de l'adhésion au tissu de l'hôte dans la physiologie de cette espèce. Ainsi, chez *S. lugdunensis*, la présence de nombreuses MSCRAMM avérées comme Fbl⁶⁶ et vWbl⁶⁸, ou putatives comme les SIs⁷³, suggère le rôle central de l'adhésion dans les mécanismes de la pathogénèse de *S. lugdunensis*. Ces protéines sont particulièrement intéressantes car elles pourraient jouer un rôle dans l'adhésion aux tissus épithéliaux de l'hôte et dans la formation de biofilm⁷³. A ce jour, la fonction de ces protéines n'a toujours pas été caractérisée, du fait de leur taille importante et de la difficulté de transformation de *S. lugdunensis*.

Enfin, parmi les *loci* analysés codant des protéines potentiellement associées à la virulence lors de notre comparaison génomique, le *locus* composé de 21 à 29 gènes homologues du *locus ess* de *S. aureus* a été identifié comme un marqueur potentiellement spécifique des CC. Nous avons en effet identifié au sein de ce *locus* dit *ess*, une région très conservée entre toutes les souches (module 1), composée de six gènes homologues à ceux codant les protéines essentielles au fonctionnement du SST7 de *S. aureus*. Chez *S. aureus*, le SST7 est composé de quatre protéines membranaires (EsaA, EssA, EssB et EssC), une protéine cytoplasmique EsaB et une protéine sécrétée EsxA^{163,166}. Il a été montré dans plusieurs études que la délétion des gènes codant les protéines transmembranaires entraînait une abolition de la translocation des protéines sécrétées^{165,167,169}. De façon intéressante, *S. lugdunensis* est le seul SCN à ce jour à disposer des six protéines "core" du SST7¹⁶⁶.

Nous avons également mis en évidence de nombreuses variations génétiques et plusieurs organisations du *locus ess* en aval des gènes codant les protéines principales du SST7. Deux organisations principales ont été observées, chacune associée à un variant du gène codant la protéine EssC, *essC1* ou *essC2* (module 2). Le variant *essC1* était le plus prévalent, retrouvé chez 15 souches appartenant au CC1, CC5, CC6 et au ST26 alors que le variant *essC2* n'a été identifié que chez six souches dont celles du CC3 et du ST27. Dans

notre étude, seuls deux variants du gène *essC* ont été observés alors que chez *S. aureus*, il en existe jusqu'à quatre, chacun associé à un *cluster* de quatre à six gènes spécifiques¹⁶⁶. Chez *S. aureus*, chaque variant du gène *essC* est aussi associé spécifiquement à des CC particuliers sans qu'un lien avec la virulence des souches n'ait été établi. Par exemple, le variant *essC2* n'a été décrit que chez des souches appartenant au CC15 et au ST398 alors que le variant *essC4* a été uniquement identifié chez des souches du CC22¹⁶⁶.

Par ailleurs, la comparaison des *loci ess* de *S. lugdunensis* et de *S. aureus* a montré que l'organisation associée au variant *essC1* de *S. lugdunensis* était similaire à celle du variant *essC4* de *S. aureus* tandis que l'organisation associée au variant *essC2* de *S. lugdunensis* était similaire à celle du variant *essC2* de *S. aureus*. La similarité des organisations entre les souches de *S. lugdunensis* et de *S. aureus* a été confirmée par la comparaison de leurs séquences et de leurs domaines protéiques. Cependant, pour plusieurs protéines, il n'a pas été possible de prédire leur fonction car il s'agit de protéines qui n'ont pas encore été caractérisées et qui présentent des domaines de fonction inconnue (DUF).

La majorité des études consacrées à la caractérisation du SST7 chez *S. aureus* concerne des souches qui présentent le variant *essC1*^{165,167,169,180}. Ainsi, ce dernier est le mieux caractérisé à ce jour. Il code notamment les protéines sécrétées EsxB¹⁶⁵, EsxC¹⁶⁸ et EsxD¹⁶⁹ impliquées dans la virulence de *S. aureus*, mais aussi les protéines EsaD, EsaE et EsaG formant un complexe T/AT et jouant un rôle dans la compétition bactérienne¹⁸⁰. Aucune étude sur le variant *essC4* n'a été réalisée et seules deux études ont porté sur le variant *essC2* à ce jour^{182,277}. La caractérisation de ce dernier a permis de mettre en évidence la présence d'un gène codant une protéine EsxX possédant un domaine LXG N-terminal, et impliquée dans la virulence de *S. aureus* dans un modèle murin de formation d'abcès cutanés¹⁸². Ce domaine LXG a été rapporté comme présent chez des protéines utilisant le SST7 pour leur translocation dans le milieu extérieur²⁴⁸. Il a également été identifié au sein de la protéine EsaD, une toxine anti-bactérienne, qui présente par ailleurs un domaine à activité nucléasique¹⁸⁰. De manière intéressante, nous avons identifié dans les deux types d'organisation du *locus ess* de *S. lugdunensis* un gène codant des protéines avec un tel domaine LXG mais ne comportant aucun domaine à activité nucléasique.

Des variations CC-dépendantes dans l'organisation du *locus ess* ont été identifiées chez *S. lugdunensis* en aval des modules précédemment décrits. Outre le nombre variable

de gènes codant des protéines non caractérisées (avec un domaine DUF), il a aussi été observé un nombre variable de gènes codant des protéines avec des domaines DUF600 selon le CC des souches. De la même façon, deux à six gènes codant des protéines DUF600 selon le variant du gène *essC*¹⁶⁶ ont été mis en évidence chez *S. aureus*. Par la suite, Cao *et al*¹⁸⁰ ont montré que les souches dépourvues du gène codant la toxine EsaD possédaient au moins deux gènes codant des protéines DUF600, qui sont homologues de l'antitoxine EsaG et sont capables d'abolir l'activité de la toxine EsaD en se fixant à son domaine nucléasique.

Enfin, parmi les polymorphismes observés figuraient également la présence d'un *cluster* de sept gènes partagés entre les souches de CC1 et de CC3 et celle d'un transposon uniquement chez quatre souches de ST1 (CC1). La recherche de ces importantes variations sur une collection complémentaire de 94 isolats a confirmé les résultats de la comparaison génomique en montrant (i) que le *cluster* de sept gènes n'était retrouvé que chez des souches appartenant au CC1, CC2 et CC3 et (ii) que le transposon était majoritairement présent chez les souches de ST1 appartenant au CC1. L'origine et le mode d'acquisition de transposon au sein du *locus ess* selon les souches de *S. lugdunensis* n'est pas claire. Cependant, deux gènes codant des protéines sans lien avec le fonctionnement du SST7 ont été décrits dans le *locus ess* comportant le variant *essC4* de la souche HO 5096 0412 de *S. aureus*¹⁶⁶. Ces gènes codent pour une ATPase et une transposase (IS21). Cela suggère la présence d'une séquence d'insertion dans le *locus ess* et confirme l'organisation similaire entre le *locus ess* comportant le variant *essC4* chez *S. aureus* et celui comportant le variant *essC1* chez *S. lugdunensis*.

Au total, huit organisations génétiques du *locus ess* ont été identifiées à la faveur de l'analyse de seulement 21 génomes de *S. lugdunensis* ; leur impact sur la virulence devra être exploré. A ce stade, il est difficile de formuler des hypothèses sur le retentissement que pourraient avoir ces différentes organisations du *locus ess*, dépendantes du variant *essC*, sur le rôle du SST7 chez *S. lugdunensis*. Pour cela, il faudrait disposer d'une caractérisation phénotypique de ces protéines constitutives chez l'une des deux espèces de staphylocoques qui possède le *locus ess* entier, *S. aureus* ou *S. lugdunensis*.

Les principaux résultats de ce premier volet de notre travail consistent en (i) la confirmation d'un *pan-genome* fermé pour l'espèce *S. lugdunensis* malgré l'identification d'EGM (un plasmide, deux îlots génomiques et cinq prophages nouvellement décrits à la

faveur du séquençage de seulement six nouveaux génomes), (ii) l'absence de déterminants spécifiques de souches pathogènes par rapport aux souches de portage, et (iii) l'identification de variations génétiques CC-dépendantes et notamment l'identification de considérables variations génétiques au sein du *locus ess*, impliqué dans le développement d'abcès chroniques et la modulation de la réponse immune lors des infections à *S. aureus*^{165,168,169,182}. Ce travail met particulièrement en évidence des caractéristiques génomiques propres aux souches de CC3 (une absence de barrières connues au transfert horizontal d'ADN étranger, une quasi-absence d'EGM, la présence d'allèles spécifiques pour plusieurs gènes putativement associés à la virulence de l'espèce), qu'il conviendra de confirmer par l'étude d'une plus large collection d'isolats et dont il faudra étudier la signification par des approches phénotypiques.

Parmi les pistes qui s'offraient à nous pour donner suite aux résultats de cette comparaison génomique, nous avons choisi de contribuer à la caractérisation du rôle du SST7 de *S. lugdunensis*. Pour cela, nous avons entrepris la construction d'un mutant de délétion du gène *essC* homologue du gène codant la protéine transmembranaire EssC chez *S. aureus*. Chez cette espèce, la protéine EssC a été montrée comme indispensable au fonctionnement du SST7 et jouant un rôle clé dans la sécrétion de nombreuses protéines impliquées dans la virulence^{165,167,169}. Ainsi, notre objectif était de déterminer l'impact de la délétion du gène *essC* (i) dans la translocation des protéines EsxA et Lxg prédites comme sécrétées par ce type de système et (ii) dans la virulence de *S. lugdunensis in vitro* et *in vivo* dans un modèle d'infection du nématode *C. elegans*.

Au laboratoire, plusieurs essais de transformation de souches cliniques de *S. lugdunensis* se sont révélés infructueux. Cette espèce bactérienne, réfractaire à la transformation, présente un peptidoglycane épais qui agit comme une barrière physique¹⁴⁰. De plus, Heilbronner *et al.*¹⁴⁰ ont montré que les gènes *hsdR*, *hsdM* et *hsdS* codant un SRM de type I (*SluI*) était un frein à la transformation de *S. lugdunensis*. Ainsi, un nouveau protocole d'électroporation a été mis au point utilisant une souche d'*Escherichia coli* exprimant les gènes *hsdM* et *hsdS* permettant la méthylation du plasmide et donc sa protection contre l'endonucléase HsdR, une fois transformé chez *S. lugdunensis*. Par ailleurs, il a été mis en évidence dans cette même étude que des souches appartenant au CC3 et CC4 étaient plus difficiles à transformer que les souches de CC1, CC2 et CC5¹⁴⁰.

La première souche mutante de *S. lugdunensis* n'a été construite que récemment par la méthode de formation de protoplastes chez la souche DSM 4804⁷². Cette souche ayant déjà permis au laboratoire GRAM la construction d'un mutant de délétion du gène codant l'autolysine AtIL grâce à un protocole rapporté de l'équipe allemande dirigée par le Pr S. Gatermann (Département de microbiologie médicale, université de la Ruhr, Bochum, Allemagne)⁸, nous l'avons choisie pour construire ce mutant Δ essC.

Le typage MLST de la souche DSM 4804 a montré que cette souche appartenait au ST24 (CC6). D'après notre analyse génomique, les souches de CC6 ne présentent pas de SRM de type I ; cela pourrait donc expliquer pourquoi cette souche est plus facilement transformable. Toutefois, comme nous avons montré que les quatre souches appartenant au CC6 possédaient toutes un SRM de type II, il serait intéressant de connaître le fond génétique de cette souche en séquençant son génome complet pour identifier de potentielles variations dans les gènes codant ce SRM de type II.

Une fois la souche délétée du gène *essC* obtenue, nous avons entrepris des travaux de complémentation. Cependant, de nombreuses difficultés techniques ont été rencontrées. En premier lieu, nous avons rencontré des difficultés après clonage du gène *essC* chez *E. coli*. En effet, de nombreuses mutations sont apparues au sein du gène inséré sur le plasmide, malgré l'utilisation d'une ADN polymérase de haute-fidélité et une vérification préalable de la séquence du gène par un séquençage Sanger. Nous avons alors synthétisé le gène *essC* par PCR en plusieurs fois par des étapes de restriction-ligation pour obtenir une séquence du gène *essC* sans mutation. Toutefois, malgré de nombreuses tentatives, les transformations de protoplastes sont restées infructueuses. Ceci pourrait être dû à la trop grande taille du gène (4,5 kb) puisque nous avons réussi à transformer la souche sauvage avec le plasmide vide.

Afin d'évaluer le rôle du SST7 chez *S. lugdunensis*, et en particulier le rôle de la protéine EssC, nous avons entrepris une analyse comparée du sécrétome de la souche délétée et de la souche parentale par nano-chromatographie liquide couplée à une spectrométrie de masse en tandem. Le but de cette analyse était triple : (i) caractériser pour la première fois le sécrétome d'une souche de *S. lugdunensis*, (ii) vérifier la diminution voire l'abolition de la sécrétion des protéines EsxA et Lxg dans le milieu extérieur et (iii) identifier d'autres protéines dont la translocation pourrait être impactée par la délétion du gène *essC*.

Afin de définir les conditions optimales pour réaliser l'analyse du sécrétome, nous avons réalisé une cinétique d'expression des gènes *essC*, *esxA* et *lxxg*. Ces travaux ont permis de déterminer le maximum d'expression de ces gènes au cours de la croissance bactérienne. Aucune différence d'expression n'a été observée pour les gènes codant les protéines sécrétées. Cependant, une différence d'expression significative a été mesurée pour le gène *essC* entre le début de la phase de croissance exponentielle (DO 1) et la phase stationnaire (DO 5). Ainsi, l'expression du gène *essC* diminue au cours de la croissance bactérienne, mais semble avoir un pic d'expression à DO 1. Pour cette raison, nous avons choisi de préparer les surnageants de culture pour l'analyse du sécrétome à cette DO.

Lors de la mise au point de la méthode de préparation des échantillons protéiques avant analyse du sécrétome, nous avons rencontré de très nombreuses difficultés techniques, malgré des conditions de cultures bien standardisées et reproductibles. Dans un premier temps, la présence résiduelle de protéines provenant du milieu de culture liquide, bouillon trypticase soja (Bio-Rad), interférait avec le sécrétome bactérien lors de l'analyse des protéines extraites du surnageant. Nous avons alors opté pour un milieu de culture minimum supplémenté avec 5% de glucose pour s'affranchir des peptides présents dans le milieu de culture précédent. Toutefois, de nouvelles difficultés techniques ont été rencontrées : des concentrations protéiques trop faibles, des problèmes de solubilisation des protéines entraînant un faible taux de reproductibilité entre les différentes séries d'échantillons et les réplicats biologiques. Ces travaux devront donc être poursuivis par l'optimisation des conditions de préparation des surnageants de culture et d'extraction protéique.

Malgré ces aléas techniques, une comparaison de plusieurs surnageants de culture de la souche sauvage et de la souche mutée a pu être réalisée et conduire à quelques résultats préliminaires significatifs. Cependant, les résultats présentés ici devront être confirmés par l'analyse de nouveaux réplicats biologiques. Nous ne sommes pas en mesure à l'heure actuelle de décrire les protéines du sécrétome de la souche sauvage de *S. lugdunensis*. Toutefois, une première analyse quantitative des protéines présentes dans le sécrétome de la souche sauvage a montré qu'elle sécrèterait dix fois plus la protéine *EsxA* et sept fois plus la protéine *Lxxg* que la souche délétée du gène *essC*. Cela signifierait que l'absence de la protéine *EssC* abolirait la translocation de ces protéines dans le milieu extérieur. Notre analyse génomique du *locus ess* de *S. lugdunensis* a montré que le gène

essC code une protéine avec des domaines ATPase de la famille SpoIIIE-FtsK similaires à ceux de la protéine EssC de *S. aureus*. Zoltner *et al.*¹⁷³ ont mis en évidence chez *S. aureus* que la délétion de la protéine EssC ou de l'un des domaines SpoIIIE-FtsK entraînait une abolition complète de la sécrétion des protéines EsxA et EsxC. Ainsi, la protéine EssC de *S. lugdunensis* pourrait jouer le même rôle que chez *S. aureus*, à savoir celui d'un transporteur membranaire utilisé pour la translocation des protéines sécrétées du SST7.

Par ailleurs, chez la souche sauvage, deux protéines particulièrement intéressantes ont été montrées comme plus sécrétées dans nos premières expériences : la protéine de liaison au fibrinogène, Fbl, et la transglycolase IsaA.

La protéine Fbl est une des protéines majeures de la virulence de *S. lugdunensis*. Elle présente une identité nucléotidique de 58% avec la protéine ClfA de *S. aureus*. L'analyse du sécrétome de 15 souches de référence de *S. aureus* a montré que 27 protéines ancrées à la paroi *via* un motif LPXTG pouvaient être retrouvées dans le milieu extérieur, dont la protéine ClfA²⁷⁸. Ainsi, il n'est pas surprenant de retrouver la protéine Fbl dans le surnageant de *S. lugdunensis*.

La transglycolase IsaA a été retrouvée comme l'une des protéines majoritaires dans plusieurs études du sécrétome de *S. aureus*²⁷⁹⁻²⁸², jouant le rôle d'une autolysine chez cette espèce²⁸¹. Par ailleurs, Ziebandt *et al.*²⁸⁰ ont analysé le sécrétome de 25 isolats cliniques de *S. aureus* responsables d'IPTM, d'arthrites septiques et de sepsis et ont montré que 30 à 40 protéines étaient sécrétées par chaque isolat. Toutefois, seules 12 protéines étaient communes à tous, dont les autolysines IsaA, Aly et LytM, la lipase Lip, la nucléase Nuc, l'hémolysine Hla, la sérine protéase SspA et plusieurs autres protéines hypothétiques^{280,283}. De plus, l'autolysine IsaA a été montrée comme impliquée dans la virulence de *S. aureus* dans un modèle murin d'arthrite septique par comparaison du pouvoir pathogène de la souche délétée du gène *isaA* et de la souche parentale²⁸¹. Ainsi, cette protéine pourrait être impliquée dans le *turn over* et la modification du peptidoglycane mais aussi dans la colonisation des fosses nasales²⁸¹.

L'étude du sécrétome de la souche sauvage a montré qu'elle sécrèterait deux fois plus la protéine Fbl et cinq fois plus la transglycolase IsaA. Toutefois, les valeurs obtenues par le test Anova sont très proches du seuil de significativité (0,0479 et 0,0495 respectivement ; $P < 0,05$). Ainsi, au vu des problèmes de reproductibilité rencontrés et du

faible nombre de répliquats biologiques utilisables, cette différence entre la souche sauvage et la souche mutée pourrait être gommée par l'analyse de nouveaux échantillons. Quoiqu'il en soit, cette analyse suggère que la protéine IsaA pourrait constituer un nouveau facteur de virulence de *S. lugdunensis*.

Enfin, nous avons évalué si la délétion du gène *essC* affectait le pouvoir pathogène de *S. lugdunensis* dans un modèle *in vitro* de cytotoxicité sur une lignée de kératinocytes (HaCaT) et dans un modèle *in vivo* d'infection du nématode *C. elegans*. Aucune atténuation de la virulence du mutant n'a été observée malgré une potentielle diminution de sécrétion des protéines du SST7 et une perte de la capacité à lyser les érythrocytes. Chez *S. aureus*, comme pour la délétion des gènes *esxA*, *esxB* et *esxC* du SST7, la délétion du gène *essC* entraîne une diminution de la virulence dans des modèles murins de formation d'abcès^{165,168}.

Le modèle d'infection du nématode *C. elegans* avait déjà été utilisé par le GRAM chez *S. lugdunensis* pour caractériser le pouvoir pathogène du mutant de délétion du gène codant l'autolysine AtIL ; la souche sauvage DSM 4804, identique, présentait un niveau de virulence comparable à celui observé dans notre étude. Ainsi, ce modèle, standardisé et reproductible, nous a permis d'examiner rapidement et à faible coût la virulence de la souche sauvage et de la souche déléetée du gène *essC*, mais d'autres modèles pourraient être utilisés. Dai *et al.*¹⁸² ont montré que la protéine EsxX du SST7 de *S. aureus* était impliquée dans l'échappement à la réponse immunitaire innée et contribuait notamment à la lyse des polynucléaires neutrophiles. A côté des modèles murins très performants pour caractériser la virulence mais beaucoup plus coûteux^{140,165,168}, il pourrait alors être intéressant de recourir à un modèle d'infection des PNN afin d'étudier si le SST7 peut conférer à *S. lugdunensis* ce moyen d'échappement à la réponse de l'hôte.

Le séquençage et la comparaison de génomes complets de *S. lugdunensis* sont des approches qui produisent un très grand volume de données. Toutefois, cette méthode exclusivement basée sur des analyses *in silico* reste exploratoire et descriptive. En effet, la présence de variations génétiques dans certains gènes de virulence putatifs ne constitue en rien une preuve d'une différence de fonctionnement entre les souches. De plus, nombre de variations ont été retrouvées pour des gènes sans fonction établie ou de fonction putative mais elles offrent tout de même des pistes intéressantes à explorer. Ainsi, les approches et

comparaisons phénotypiques, très peu développées chez *S. lugdunensis*, restent cruciales pour la caractérisation des mécanismes qui sous-tendent sa virulence. D'un autre côté, il reste nécessaire de développer la base de données avec un plus grand nombre de génomes de *S. lugdunensis* et en particulier de souches appartenant aux CC2 et CC4 pour avoir une représentation complète de la diversité génétique de *S. lugdunensis*.

La caractérisation génétique apporte ici plusieurs pistes à explorer chez cette espèce bactérienne présentant une virulence particulière par rapport aux autres SCN. Notre étude génomique comparative nous a permis d'identifier de nombreuses variations et organisations génétiques dans le *locus ess* codant le SST7 chez une espèce présentant pourtant une structure de population clonale. Ainsi, la protéine EssC permettrait la sécrétion de protéines potentiellement impliquées dans la virulence, mais ces résultats doivent être confirmés par de nouvelles expérimentations. Les modèles utilisant *C. elegans* et, malgré quelques critiques dans la littérature²⁸⁴, les larves de *G. mellona*, se sont révélés utiles pour le screening de la virulence de souches de *S. lugdunensis*. Leur faible coût et leur facilité de mise en œuvre sont adaptés à l'évaluation de la virulence de larges collections d'isolats. Ils pourraient dans notre projet s'avérer utiles à l'investigation de la virulence selon le CC ou selon les variations du *locus ess* et faciliter ainsi le choix de la (les) souche(s) qu'il serait intéressant de muter pour caractériser le rôle des différentes protéines du système SST7.

TRAVAUX ASSOCIES

1. Apport phylogénétique de l'analyse de VNTR

1.1. Objectifs et principaux résultats de l'étude

Développement des méthodes de MLVA et de TRST chez *S. lugdunensis*

Dans le but d'explorer la structure de population et de rechercher un éventuel lien entre lignées phylogénétiques et virulence, plusieurs méthodes de typage telles que la MLST et la MVLST ont été développées sur une collection de souches de *S. lugdunensis* issues de différents contextes cliniques et de différentes origines géographiques^{98,208}. Les analyses phylogénétiques basées sur les résultats de génotypage avec l'une ou l'autre de ces méthodes ont révélé une structure de population clonale de l'espèce, mais aucune lignée hypervirulente associée à un contexte clinique sévère n'a été identifiée.

Dans ce contexte, nous avons choisi d'étudier l'apport du polymorphisme de *loci* de VNTR dans le typage et la phylogénie de *S. lugdunensis*. Leur rapidité d'évolution en font en effet une importante source de polymorphisme génétique et une méthode de typage extrêmement discriminante²⁸⁵. Nous avons ainsi développé deux nouvelles méthodes pour le typage de l'espèce : la MLVA et la TRST. La première consiste à déterminer le nombre de motifs répétés de plusieurs *loci* alors que la seconde évalue non seulement le nombre de répétitions mais aussi les SNP au sein de ces motifs répétés. Pour cela, nous avons identifié et séquencé des VNTR afin (i) d'évaluer le pouvoir discriminant de ces nouvelles méthodes et (ii) de le comparer à celui de la MLST (méthode de référence) et de la MVLST.

L'analyse des génomes des souches de référence HKU09-01¹³³ et N920143⁷³ a permis d'identifier 28 *loci* de VNTR dont le pouvoir discriminant a été évalué sur un premier panel d'isolats. Ce panel était composé de 30 isolats préalablement caractérisés par MLST et MVLST et représentatifs des différents CC^{98,208}. Ont été inclus à ce panel cinq paires d'isolats épidémiologiquement liés (isolés chez un même patient) pour évaluer la concordance épidémiologique de la méthode de MLVA.

Un *locus* de VNTR a été validé comme un bon candidat s'il (i) était présent chez tous les isolats, (ii) montrait une taille minimum de 18 pb et (iii) ne nécessitait l'utilisation que d'un seul couple d'amorces pour l'amplification par PCR et le séquençage Sanger. Sept VNTR, nommés SLU1 à SLU7, ont alors été sélectionnés pour le typage par MLVA et TRST.

Les schémas de ces deux nouvelles méthodes de typage ont été appliqués sur un ensemble de 98 isolats cliniques de *S. lugdunensis* collectés entre Novembre 2013 et Mars 2016 à l'hôpital universitaire de Strasbourg¹¹ (82 responsables d'infections sévères et 16 isolés en situation de colonisation). En parallèle, un typage par MLST et par MVLST avec le schéma *trilocus*⁹⁸ a été réalisé sur ces mêmes isolats.

Le génotypage par MLST et de MVLST sur les 128 isolats de *S. lugdunensis* a généré 25 ST et 21 *Trilocus Virulence Types* (VT^T) respectivement. Les techniques de MLVA et de TRST se sont montrées plus discriminantes puisqu'elles ont permis d'identifier 55 MT et 69 TRST-Types (TRT) respectivement. Les valeurs des DI ont confirmé que la MLVA ($DI_{MLVA} = 0,933$) et la TRST ($DI_{TRST} = 0,943$) étaient plus discriminantes que la MLST ($DI_{MLST} = 0,898$) et la MVLST ($DI_{MVLST} = 0,845$). Par ailleurs, les *clusters* (regroupements phylogénétiques) formés par les quatre méthodes étaient très congruents, mais la MLVA et la TRST présentaient l'avantage de permettre de subdiviser certains CC définis par MLST et en particulier les CC1, CC4, CC5 et CC6, en plusieurs nouveaux génotypes.

Même si aucun *cluster* spécifique à un contexte clinique n'a été identifié lors de cette étude, nous proposons ici deux nouvelles méthodes de typage dont la typabilité, la reproductibilité, la stabilité et la concordance épidémiologique sont excellentes. Ainsi, le séquençage des *loci* de VNTR constitue un outil très prometteur pour l'épidémiologie moléculaire des infections à *S. lugdunensis* car en plus de prédire le CC, il permet aussi de distinguer les isolats entre eux par de nouveaux génotypes (MT et TRT) au sein de plusieurs CC.

Développement de la méthode de *fbl*-typing et étude de l'impact du polymorphisme du gène *fbl* sur la capacité de liaison au fibrinogène

Lors de la mise au point des méthodes de génotypage de MLVA et de TRST, un *locus* de VNTR particulièrement intéressant a été identifié. Ce dernier est localisé dans la région R du gène *fbl* codant la protéine de liaison au fibrinogène, présente chez toutes les souches de *S. lugdunensis*¹⁰. Cette protéine possède une organisation identique à celle de la protéine de liaison au fibrinogène de *S. aureus* ClfA et partage avec celle-ci une homologie nucléotidique de 58%⁶⁶. Au vu de son polymorphisme important, ce *locus* de VNTR n'a pas été inclus au schéma de MLVA et de TRST afin de l'analyser de façon indépendante pour

développer le *fbl*-typing basé, sur celui de la méthode de *spa*-typing. Il consiste en l'analyse des séquences répétées de la région R du gène *fbl*.

Les deux premiers objectifs étaient donc (i) d'évaluer le pouvoir résolutif de cette nouvelle méthode et (ii) d'évaluer ses performances par rapport à celles de la MLST et de la TRST.

Le *fbl*-typing a été appliqué à une collection de 240 isolats incluant 106 isolats cliniques de *S. lugdunensis* collectés entre Janvier et Décembre 2016 à l'hôpital universitaire de Rouen, les 128 isolats de la précédente étude de MLVA/TRST et les six souches séquencées par *Pacbio*. Le *fbl*-typing a permis d'identifier 54 séquences répétées uniques de 18 pb et 92 *fbl*-types répartis en 10 *clusters* distincts. La comparaison des données de typage à partir des 128 isolats a révélé que le *fbl*-typing est une technique plus discriminante ($DI_{fbl\text{-typing}} = 0,964$) que la MLST ($DI_{MLST} = 0,899$) et aussi discriminante que la TRST ($DI_{TRST} = 0,943$). Par ailleurs, l'analyse phylogénétique basée sur ces résultats a montré une excellente concordance avec la méthode de typage de MLST ; le *fbl*-type permettant de prédire dans 100% des cas le CC d'un isolat. Dans le but de faire de cette nouvelle technique de typage une méthode portable, nous avons développé un outil en ligne appelé "*fbl*-typing server" (<http://fbl-typing.univ-rouen.fr>) comprenant la base de données des différents allèles identifiés lors de notre étude et permettant une identification rapide du *fbl*-type à partir d'une séquence fasta.

Grâce à l'analyse de la séquence du gène *fbl* à partir des 21 génomes complets de *S. lugdunensis*, une délétion de 4 pb a été identifiée dans la région 5' du gène entraînant l'apparition d'une mutation non-sens prématurée. Cette délétion a été recherchée chez 146 isolats et a été identifiée chez 27 souches dont 26 appartenant au CC1 et une au CC3.

Le dernier objectif de notre travail a alors été d'évaluer l'impact de la diversité génétique du gène *fbl* (la présence d'une délétion à l'extrémité 5' et/ou le nombre de répétitions du domaine R) sur la capacité des souches à lier le fibrinogène immobilisé *in vitro* chez 55 isolats représentatifs de la variabilité du gène *fbl*. La capacité de liaison au fibrinogène était variable selon les isolats. Toutefois, l'analyse statistique a révélé que les souches présentant plus de 43 motifs répétés dans la région R du gène *fbl* liaient le fibrinogène de façon significativement ($P < 0,05$) plus importante que les souches qui en présentaient un nombre inférieur. Par ailleurs, cette même analyse statistique a montré que

les souches présentant une délétion au niveau de la région 5' du gène *fbI* liaient significativement moins le fibrinogène ($P < 0,05$).

Devant ces résultats variables de liaison au fibrinogène selon les souches, nous avons évalué si la capacité à lier le fibrinogène était également liée à une différence d'expression du gène *fbI* entre les souches. Ainsi, nous avons étudié l'expression du gène *fbI* de cinq isolats qui liaient le fibrinogène de façon importante et cinq isolats qui ne le liaient pas, mais aucune différence significative d'expression n'a été observée entre les deux groupes d'isolats.

Ces travaux, réalisés à contribution égale avec le Dr Sandrine Dahyot, ont fait l'objet de deux articles scientifiques, l'un publié dans la revue "*Scientific Reports*" et le second est en cours de soumission.

1.2. Publication n°3

Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA) and Tandem Repeat Sequence Typing (TRST), helpful tools for subtyping Staphylococcus lugdunensis.

Sandrine Dahyot, Jérémie Lebeurre, Xavier Argemi, Patrice François, Ludovic Lemée, Gilles Prévost, Martine Pestel-Caron. *Sci Rep.* 2018 Aug 3;8(1):11669.

SCIENTIFIC REPORTS

OPEN

Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA) and Tandem Repeat Sequence Typing (TRST), helpful tools for subtyping *Staphylococcus lugdunensis*

Sandrine Dahyot¹, Jérémie Lebeurre², Xavier Argemi^{3,4}, Patrice François⁵, Ludovic Lemée¹, Gilles Prévost³ & Martine Pestel-Caron¹

Staphylococcus lugdunensis is an emergent virulent coagulase-negative *Staphylococcus* that is increasingly responsible for severe infections. In an attempt to generate informative sequence data for subtyping *S. lugdunensis*, we selected and sequenced seven polymorphic variable number of tandem repeats (VNTRs) to develop two new methods: a classic length-based multiple-locus VNTR analysis (MLVA) method and a tandem repeat sequence typing (TRST) method. We assessed their performances compared to two existing methods, multilocus sequence typing (MLST) and multivirulence-locus sequence typing (MVLST) for 128 isolates from diverse clinical settings and geographical origins. The clustering achieved by the four methods was highly congruent, with MLVA discriminating within clonal complexes as defined by MLST. Indeed, MLVA was highly discriminant compared to MLST and MVLST in terms of number of genotypes as well as diversity indexes. Sequencing of the seven VNTRs showed that they were stable, and analysis of sequence polymorphisms provided superior discriminatory power. The typeability, reproducibility, and epidemiological concordance of these new methods were excellent. Of note, no link between clustering and clinical settings was identified. This study demonstrates that MLVA and TRST provide valuable information for molecular epidemiological study of *S. lugdunensis*, and represent promising tools to distinguish between strains of homogenous lineages in this clonal species.

Staphylococcus lugdunensis is a coagulase-negative *Staphylococcus* (CoNS) belonging to the normal human skin microbiota^{1,2} that is increasingly recognized as a virulent pathogen in both community-acquired and healthcare associated infections^{3,4}. This commensal bacterium has mainly been associated with serious infections, such as skin and soft tissue infections^{5,6}, bone and joint infections^{7,8}, but also bacteremia and endocarditis^{9,10}. Thus, the pathogenicity of *S. lugdunensis* is closer to that of *S. aureus* than that of other CoNS; it may be related to virulence factors such as a fibrinogen binding protein¹¹, a von Willebrand factor-binding protein¹², synergistic hemolysins¹³ or a metalloprotease¹⁴.

Several molecular typing methods are currently available for *S. lugdunensis* strains such as macrorestriction analysis using pulsed-field gel electrophoresis¹⁵ and multilocus sequence typing (MLST)¹⁶. The latter is considered as the gold standard method for epidemiological typing and population genetic study of *S. lugdunensis*.

¹Normandie Univ, UNIROUEN, GRAM EA2656, Rouen University Hospital, F-76000, Rouen, France. ²Normandie Univ, UNIROUEN, GRAM EA2656, F-76000, Rouen, France. ³Université de Strasbourg, CHRU de Strasbourg, VBP EA7290, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Institut de Bactériologie, 3 rue Koeberlé, F-67000, Strasbourg, France. ⁴Hôpitaux Universitaires, Maladies Infectieuses et Tropicales, F-67000, Strasbourg, France. ⁵University of Geneva Hospitals, Genomic Research Laboratory, Service of Infectious Diseases, Geneva, Switzerland. Sandrine Dahyot and Jérémie Lebeurre contributed equally to this work. Correspondence and requests for materials should be addressed to S.D. (email: sandrine.dahyot@chu-rouen.fr)

Recently, we developed a multivirulence-locus sequence typing (MVLST) method, based on the sequence data of seven known or putative virulence-associated loci¹⁷. MVLST was significantly more discriminant than MLST, even allowing a trilocus sequence typing method for microepidemiological purposes. Phylogenetic analyses by MLST and MVLST revealed a clonal population structure, a mutational evolution of this pathogen, and a lack of hypervirulent lineages.

An interesting source of genetic polymorphism is provided by tandemly repeated sequences, known as variable number of tandem repeats (VNTRs), whose number of repetitions varies at different rates depending on the different loci and even alleles¹⁸. Molecular typing based on the analysis of repeat copy number at multiple VNTR loci, known as multiple-locus VNTR analysis (MLVA), is a genotyping method that is being used for strain comparison, and can also provide insights into population structure¹⁹. This method has already been successfully applied to many other bacterial species and showed a high level of discrimination^{19–23}. Especially, measuring repeat polymorphisms in clonal microorganisms can provide a solid basis for genetic type assignment²⁴. In addition, to increase the discriminatory power of MLVA, sequencing repeated sequences that display internal variability can be performed²⁵.

The aims of this study were to identify VNTR loci and to sequence them in order (i) to develop the first length-based MLVA method and a sequence-based MLVA (tandem repeat sequence typing) method for *S. lugdunensis*, and (ii) to evaluate and compare the added value of these new methods with the existing methods MLST and MVLST in terms of *S. lugdunensis* strain discrimination within clonal groups.

Results

MLST and MVLST genotyping. To determine their diversity and whether they belong to major clonal complexes, we analyzed 128 *S. lugdunensis* clinical isolates by MLST and MVLST. Data for 20 isolates were derived from our previous studies^{16,17}, while we performed MLST and MVLST analysis of 108 additional clinical isolates for the purposes of the present study.

MLST analysis on the 128 isolates identified 25 sequence types (STs). Two new STs, designated ST31 and ST32, were incorporated in the online database (<http://bigsdbs.web.pasteur.fr/staphlugdunensis>). ST3 was the most common ST (n = 29), followed by ST2 (n = 15) and ST24 (n = 15). The 25 STs were distributed over 7 clonal complexes (CC1 to 7) and 2 singletons, as shown in the minimum spanning tree (Fig. 1). The main clonal complex was CC1 (n = 38), consisting of 5 STs (1, 6, 7, 12 and 15, with ST6 as the primary founder) followed by CC3 (n = 31) with 3 STs (3, 16 and 20 with ST3 as the primary founder).

MVLST performed on the 128 isolates revealed that the number of individual alleles for each of the three genes ranged from 10 (for SLUG_16930 and *atlL_{R3}*) to 11 (for *isdJ*). Twenty-one trilocus virulence types (VT^Ts) were assigned; 3 major VT^Ts (VT^T1, VT^T15 and VT^T16) were shared by 81 isolates. Five new VT^T were identified (VT^T34 to 38). The clustering obtained by MVLST was similar to that observed by MLST. For instance, VT^T15 and VT^T16 corresponded to ST2 and ST3 respectively, VT^T1 clustered into 4 STs (ST1, ST6, ST12 and ST15) from CC1 and 3 linked VT^Ts (VT^T13, VT^T34 and VT^T37) belonged to ST24.

No correlation was found between genotype and clinical origin; for example, CC1 was shared by isolates from 8 different clinical contexts. Isolates recovered from hematogenic infections (blood or osteoarticular isolates) or from skin and soft tissue infections did not cluster in separate lineages. Nevertheless, an interesting trend was observed: 4 of 5 device-associated infection isolates belonged to CC3 (VT^T16). Pathogenic and carriage isolates overlapped in their genotype distribution, with 4 of 7 CCs containing isolates from both sources. However, although carriage isolates did not cluster in a unique lineage, 59% (10/17) belonged to CC3 and 53% (9/17) belonged to VT^T16.

MLVA genotyping and clustering. Based on the *S. lugdunensis* genomes available in GenBank database, 7 VNTR loci were selected and evaluated on the validation panel of 30 isolates. Four VNTR loci were located in open reading frames; SLU2, SLU3 and SLU6 were located in genes encoding hypothetical proteins while SLU7 was located in a gene encoding an AraC family transcriptional regulator. Three VNTRs (SLU1, SLU4 and SLU5) were located in non-coding regions. The tandem repeat (TR) unit size of the 7 VNTRs was between 24 and 58 bp (Table 1), while the number of TRs ranged from 1 to 11. MLVA discriminated 19 MLVA types (MTs) among the 30 isolates. The 5 pairs of isolates considered as clonal by MLST were also found to be clonal by MLVA, attesting the epidemiological concordance of the method.

The MLVA assay was then applied to the analysis of an expanded panel of 98 *S. lugdunensis* clinical isolates from highly diverse clinical situations. The 7 VNTR loci were amplified and sequenced for all 98 isolates. Diversity in the number of TRs varied among the various VNTR loci. The largest observed variation in VNTR size was found for SLU3 and spanned MTs 3 to 11 repeats (Table 1). MLVA discriminated 42 MTs among the 98 isolates. Four MTs were represented by more than 5 isolates: MT1 (n = 23), MT2 (n = 12), MT3 (n = 7) and MT4 (n = 6). Twenty-nine MTs were represented by only 1 isolate.

Finally, the 7 selected VNTRs enabled identification of 55 different MTs for the overall panel of 128 isolates tested in this study. A complete overview of these allelic profiles is presented in Table S4. Simpson's diversity index (DI) was calculated for each of the 7 loci to evaluate their discriminatory power. The SLU3 locus showed the highest DI value (0.679), with the highest number of TR; the lowest DI value was observed for the SLU7 locus (0.090) (Table 1). MLVA combined 2 markers (SLU7 and SLU5) with a low discriminatory power (DI < 0.5) with 5 highly discriminant markers.

The genetic relationships of the 128 isolates were deduced by the construction of an unweighted pair group method with arithmetic mean (UPGMA) tree (Fig. 2) and a minimum spanning tree (Fig. 3). For the minimum spanning tree analysis, MLVA complexes (MCs) were created if 6 of the 7 VNTR loci were identical. This analysis defined 6 different MCs (MC1 to 6) and 10 singletons (Fig. 3a). Two of the MCs, MC1 (30 isolates) and MC2 (54 isolates), were predominant. Of the 128 isolates analyzed in the study, 91% were part of an MC. With the UPGMA

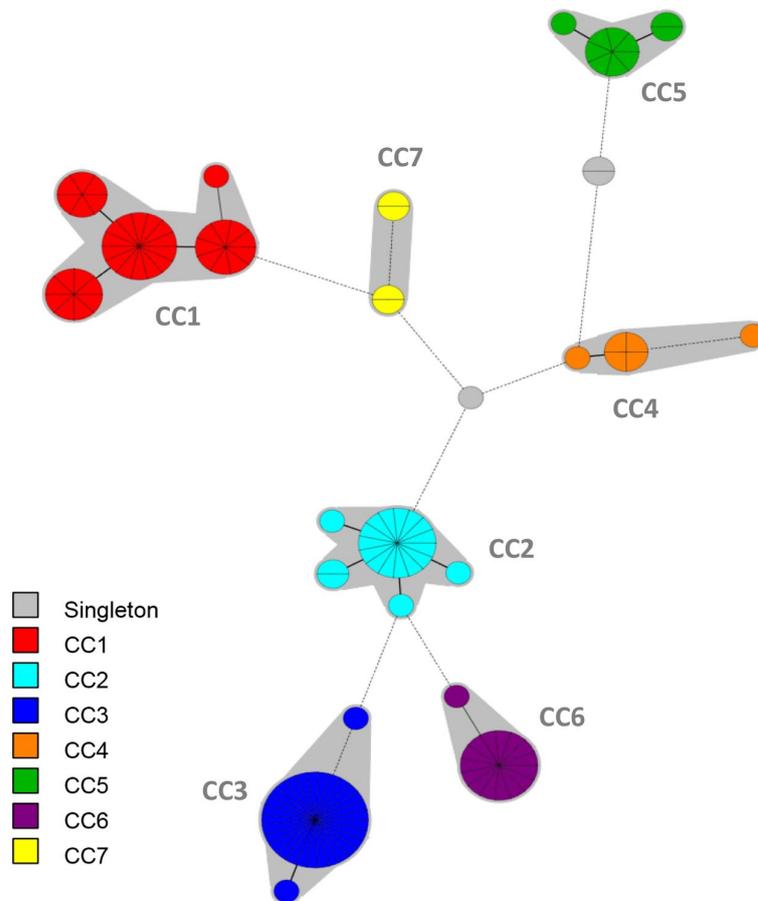


Figure 1. Minimum spanning tree of the 128 *S. lugdunensis* isolates typed by multilocus sequence typing (MLST). The MLST sequences were concatenated and analyzed in BioNumerics. Clustering of MLST profiles was done using a categorical coefficient. The colors used are based on clonal complexes (CCs). Each circle represents a sequence type and its size is proportional to the number of strains. Thick, short lines connecting two types denote types differing in a single locus; thin, longer lines connect double-locus variants; and dashed lines indicate the most likely connection between two types differing in more than two loci. Gray zones around circles delineate CCs. CCs are also indicated in characters e.g. CC1 denotes MLST complex 1.

algorithm, an MLVA cluster was defined by a cutoff value of 67% similarity (Fig. 2). Eight MLVA clusters (cluster I to VIII), each containing more than 2 isolates, and 7 singletons were identified. The predominant cluster, cluster I, included 51 isolates. Overall, clusters were highly congruent with the 2 algorithms, UPGMA and minimum spanning tree.

The distribution of the isolates did not allow us to identify any correlation between MCs and clinical settings. However, 12 of 17 (70%) carriage isolates and 4 of the 5 device-associated infection isolates clustered into MC2 (Fig. 3b), as observed with MLST and MVLST.

Compared to MLST, MLVA resulted in a higher differentiation of the *S. lugdunensis* isolates as 55 MTs were clustered into 6 MCs as opposed to 25 STs into 7 CCs. Interestingly, MLVA efficiently subdivided some CCs such as CC1 and CC5 (Table 2). On the contrary, isolates of CC2, represented by 5 STs, were less discriminated by MLVA with only 3 MTs. Despite obvious differences, there was considerable congruence between the clustering results obtained by the 2 typing methods, as shown by the MLVA-based minimum spanning tree (Fig. 3a). Especially, the CC1 isolates clustered into 2 linked MCs (MC1 and MC3), and the CC4 and CC5 isolates clustered into MC6 and MC5, respectively.

Tandem repeat sequence typing (TRST). Sequencing of the alleles of the 7 VNTRs was performed to determine their internal variability and size homoplasmy. Sequencing led to the identification of single nucleotide polymorphisms (SNPs) in the TRs of all loci, and showed that no insertion or deletion was present in flanking sequences. For each VNTR, alleles defined for the overall panel of 128 isolates ranged from 4 (SLU7) to 27 (SLU3) (Table 1). The Simpson's DI calculated for the 7 VNTRs based on TRST ranged from 0.120 for SLU7 to 0.889 for SLU3 (Table 1).

Analysis of the 7 VNTR sequences revealed a total of 21 TRST types (TRTs) among the 30 panel test isolates; epidemiologically-related isolates belonged to the same TRT, which is in favor of an *in vivo* stability of the VNTR

VNTR name	Primer sequences (5'–3')	Annealing temperature	Genome coordinates ^a	Putative function	Repeat size (bp)	Copy no. range	MLVA method		TRST method		
							DI ^b	CI ^c 95%	No. of alleles	DI ^b	CI ^c 95%
SLU1	F: ATCTCAGGTAAGGATATTTCGCATTG R: CACAATGTTTCGTATTGAATGGCTT	56 °C	2201528 2201881	Non-coding	57	1–3	0.531	0.500–0.562	6	0.656	0.609–0.702
SLU2	F: TTTATCCCACAATCATTTCCTTGC R: GTTCCAGTCTCTTGTCAATTAGTTT	55 °C	2247710 2248186	Hypothetical protein	58	1–4	0.595	0.523–0.667	21	0.801	0.745–0.856
SLU3	F: CACAAGACAATAGTAACCAGAAAAG R: CTTTATTATTGGTTGATTCTGTTGG	55 °C	1080477 1080983	Hypothetical protein	48	3–11	0.679	0.615–0.743	27	0.889	0.859–0.918
SLU4	F: ATATTTTCGTATTGTTGGCTCG R: GAAGCGCATAGTGTAGATGA	53 °C	569800 569359	Non-coding	57	1–6	0.666	0.617–0.715	14	0.757	0.706–0.808
SLU5	F: TTAAACATCATTGCAGGTCG R: CAGCAGAATACCATTTC AATTACA	53 °C	2497844 2497405	Non-coding	57	2–5	0.479	0.393–0.566	19	0.849	0.819–0.879
SLU6	F: TACATTAAGCTAGTTTGCAG R: AAATAAGTGAAGGACGTGATG	60 °C	2115216 2115837	Hypothetical protein	58	1–5	0.566	0.503–0.629	18	0.850	0.819–0.880
SLU7	F: CAATGATGATGCCTAACACCG R: AATTATGGATTATATCGAGCG	58 °C	2254827 2255435	AraC family transcriptional regulator	24	2–3	0.090	0.022–0.158	4	0.120	0.043–0.197

Table 1. Primers used for MLVA and TRST methods and characteristics of the 7 variable number of tandem repeat (VNTR) loci. ^aNucleotide coordinates in the reference strains *S. lugdunensis* N920143 chosen for primer design. ^bDI: Simpson's diversity index. ^cCI: confidence interval.

sequences. A total of 69 TRTs were identified for all the 128 isolates, suggesting that TRST is slightly more discriminant than MLVA (55 MTs). A complete overview of these allelic profiles is presented in Table S4. TRT1 was the most common ($n = 26$), followed by TRT2 ($n = 16$) and TRT3 ($n = 7$). Sixty-one TRTs were represented by only 1 isolate. A TRST cluster was defined by a cutoff value of 91% similarity with the UPGMA algorithm (Fig. S2). Nine TRST clusters (cluster I to IX), each containing more than 2 isolates were identified. The largest TRST cluster (cluster I) comprised 51 isolates represented by 11 TRTs. As illustrated by the UPGMA tree, these isolates belonged to the MC2 (Fig. S2).

Figure 4 illustrates the distribution of the distinct CCs onto the TRST diversity, similar to that obtained by MLST. Notwithstanding, as MLVA, TRST showed a variable level of discrimination according to CCs (Table 2). For example, TRST was highly discriminant for CC1 (28 TRTs versus 5 STs) and for CC6 (2 STs versus 10 TRTs).

As for MLVA, no correlation was observed between TRTs and clinical settings, but the majority of carriage ($n = 11$) and device-associated infection ($n = 4$) isolates remained clustered (TRST cluster I). Of note, 23 isolates collected from Strasbourg University Hospital between November 2013 and March 2016 could not be distinguished by both MLST (ST3) and MLVA/TRST (MT1/TRT1). As a fact, they could be considered as a clone. Interestingly, 14 isolates from 7 different clinical settings and 9 of the 16 carriage isolates belonged to this clone.

Discriminatory power and concordance between typing methods. To assess the respective discriminatory power of the MLST, MVLST, MLVA and TRST methods, we excluded 5 of the 10 isolates that were isolated from the same patient and that were shown to be identical by all typing methods²⁶. Simpson's DI calculated on the basis of the 123 remaining unrelated isolates showed that MLVA ($DI_{MLVA} = 0.933$) was more discriminatory than MLST ($DI_{MLST} = 0.898$) and MVLST ($DI_{MVLST} = 0.845$) (Table 3). Moreover, discriminatory power contributed by the sequencing of the VNTR loci ($DI_{TRST} = 0.943$) was higher than the analysis of the size of the VNTRs.

Table 2 shows Simpson's DI of the different methods according to the CCs defined by MLST. The higher level of discrimination of MLVA and TRST was particularly observed within the CC1 isolates, with DI of 0.949 and 0.962 respectively, compared to MLST ($DI = 0.737$). Conversely, MVLST had a limited discriminatory power ($DI = 0.206$) for isolates of CC1.

The adjusted Wallace coefficient (AW) of MLVA, compared with that of MLST and MVLST, showed that MLVA had high congruence with MLST ($AW = 0.826$) and MVLST ($AW = 0.815$) (Table 4). This indicates that isolates from a given MT will have an 83% probability of sharing the same ST. Conversely, the congruency of MLST with MLVA indicates that isolates from a given ST will have only a 52% probability of belonging to the same MT. This confirmed that MLVA was more discriminatory than MLST. Moreover, the predictive ability of TRST for *S. lugdunensis* clustering was highly congruent with MLVA typing, and the probability that a pair of isolates with the same MT also shared identical TRT was 84% (Table 4).

Stability of VNTR loci. Examination of VNTR loci stability revealed that, for each isolate, identical MLVA profiles (in terms of TR number and sequences) were generated for the original isolates and their 10th, 20th and 30th subcultures, indicating that the 7 VNTRs were stable.

Linkage disequilibrium. A quantitative analysis of the linkage between alleles from the 7 VNTR loci was performed by calculating the standardized index of association (I_A^S). The linkage disequilibrium was shown to be significant in the LIAN 3.7 analysis, indicating a clonal population structure for the *S. lugdunensis* isolates. I_A^S was calculated at 0.1819 for MLVA and 0.4639 for TRST at a significance of $P < 1.00 \times 10^{-5}$.

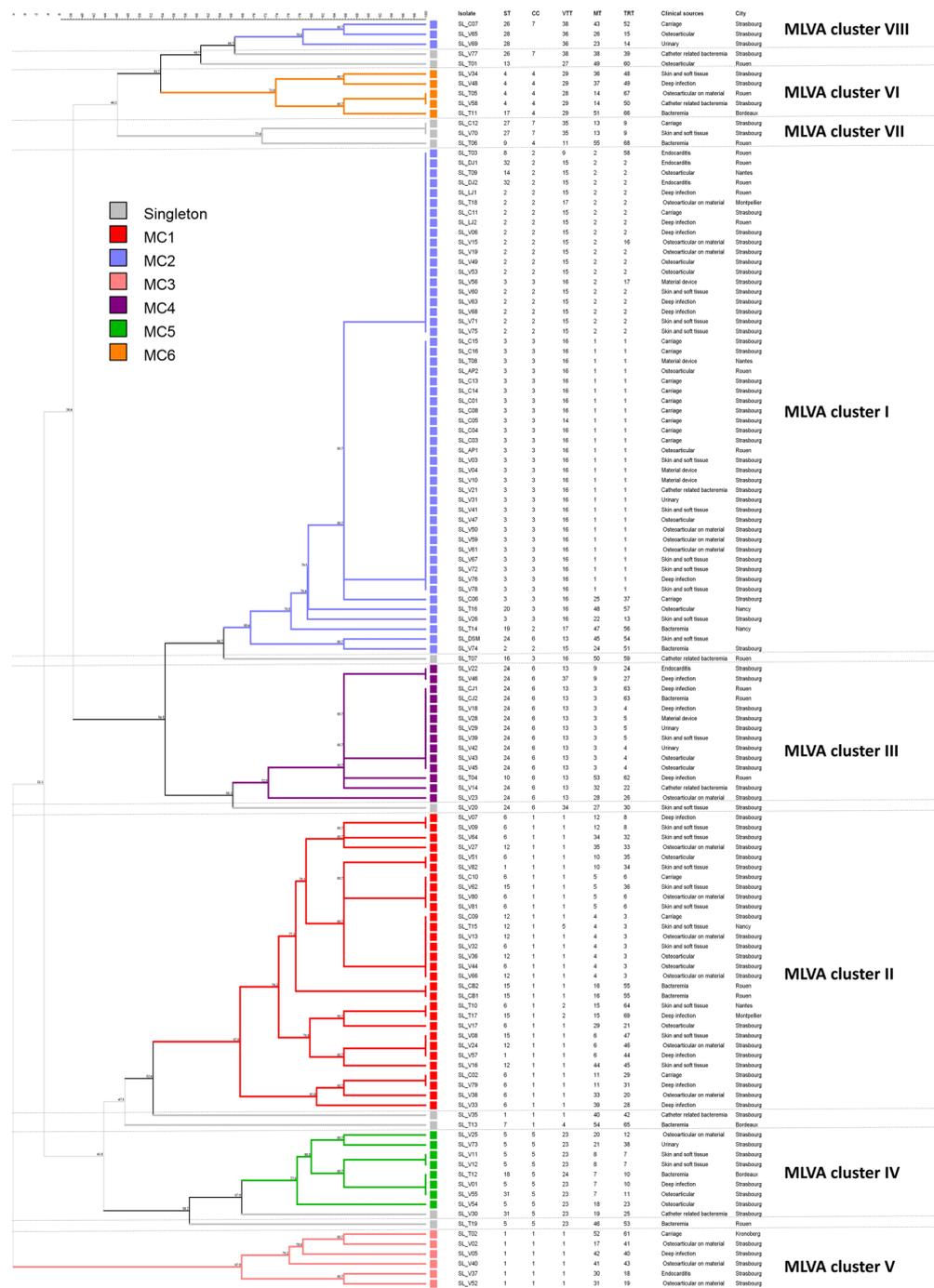


Figure 2. Multiple-locus VNTR analysis (MLVA) clustering of the 128 isolates of *S. lugdunensis* by the unweighted pair group method with arithmetic mean (UPGMA) method. The names of isolates, MLST sequence type (ST), MLST clonal complex (CC), MVLST trilocus virulence type (VT^T), MLVA type (MT), TRST type (TRT), the clinical sources and the city of isolates are shown on the right. A cutoff value of 67% similarity was applied to define MLVA clusters (named MLVA cluster I to VIII). The colors used are based on MLVA complexes (MCs).

Discussion

Because *S. lugdunensis* is more virulent than other CoNS²⁷, MLST and MVLST were developed to characterize the links between within-species genetic variations and characteristics, such as pathogenic potential, virulence, and epidemiology. To further delineate the lineages and clonal groups previously described by these typing methods,

CC (no. of isolates)	MLST		MVLST		MLVA		TRST	
	No. of STs	DI ^a	No. of VT [†] s	DI ^a	No. of MTs	DI ^a	No. of TRTs	DI ^a
1 (37)	5	0.737	4	0.206	22	0.949	28	0.962
2 (18)	5	0.405	3	0.307	3	0.216	5	0.405
3 (30)	3	0.131	2	0.067	6	0.310	6	0.310
4 (6)	3	0.600	3	0.600	5	0.933	6	1.000
5 (10)	3	0.511	2	0.200	7	0.911	8	0.956
6 (15)	2	0.133	3	0.257	7	0.724	10	0.914
7 (4)	2	0.667	2	0.667	3	0.833	3	0.833
Total	25	0.898	21	0.845	55	0.933	69	0.943

Table 2. Abilities of the four typing methods to discriminate isolates within major *S. lugdunensis* clonal complexes (CCs) defined by MLST. ^aDI: Simpson's diversity index.

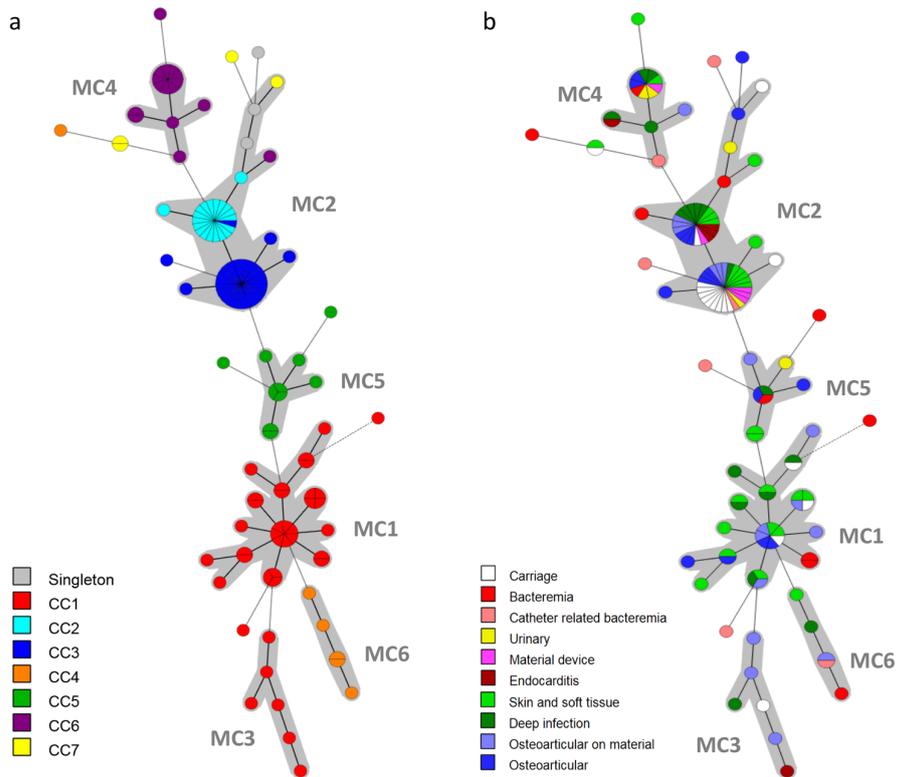


Figure 3. Minimum spanning tree of the 128 *S. lugdunensis* isolates typed by multiple-locus VNTR analysis (MLVA). The MLVA data were analyzed in BioNumerics. Clustering of MLVA profiles was done using a categorical coefficient. The colors used are based (a) on clonal complexes (CCs) defined by MLST and (b) on clinical contexts. Each circle represents an MLVA type. Thick, short lines connecting two types denote types differing in a single locus; thin, longer lines connect double-locus variants; and dashed lines indicate the most likely connection between two types differing in more than two loci. Gray zones around circles delineate MLVA complexes (MCs). MCs are also indicated in characters e.g. MC1 denotes MLVA complex 1.

MLVA has attracted intense interest. In this study, we developed a classic seven loci MLVA typing method and a sequence-based MLVA method (TRST), and compared them with MLST as a reference method and with the virulence-associated trilocus method (MVLST). The relevance of these new molecular tools was investigated against a genetically diverse set of 128 *S. lugdunensis* isolates collected from various clinical settings and geographical origins.

One hundred TR sequences contained in the *S. lugdunensis* genomes available in GenBank database were initially identified *in silico*. Finally, seven VNTRs were selected because they met the preconditions described in the methods section and for their discriminatory power. PCR amplification and sequencing of the seven VNTRs were achieved for all the 128 isolates tested representing a typeability of 100%. MLVA proposes here a set of

Typing method	Number of types	DI ^a	CI ^b 95%
MLST	25	0.898	0.871–0.925
MVLST	21	0.845	0.809–0.880
MLVA	55	0.933	0.905–0.961
TRST	69	0.943	0.915–0.971

Table 3. Simpson's diversity index of the four typing methods. ^aDI: Simpson's diversity index. ^bCI: confidence interval.

	MLST	MVLST	MLVA	TRST
MLST		0.858 (0.756–0.959)	0.524 (0.407–0.641)	0.482 (0.368–0.597)
MVLST	0.531 (0.445–0.618)		0.320 (0.228–0.412)	0.289 (0.199–0.380)
MLVA	0.826 (0.736–0.916)	0.815 (0.677–0.953)		0.838 (0.755–0.922)
TRST	0.906 (0.834–0.979)	0.878 (0.742–1.000)	1.000 (1.000–1.000)	

Table 4. Adjusted Wallace coefficients and 95% confidence interval of the four typing methods.

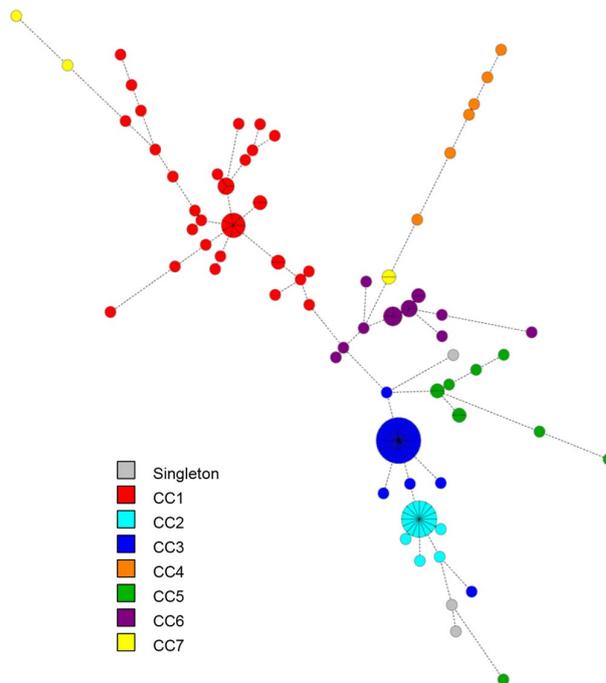


Figure 4. Minimum spanning tree of the 128 *S. lugdunensis* isolates typed by tandem repeat sequence typing (TRST). The VNTR sequences were combined into a composite dataset and analyzed in BioNumerics. Clustering of MLVA profiles was done using a categorical coefficient. Each circle represents a TRST type. The colors used are based on clonal complexes (CCs) defined by MLST. Thick, short lines connecting two types denote types differing in a single locus; thin, longer lines connect double-locus variants; and dashed lines indicate the most likely connection between two types differing in more than two loci.

markers with different DIs (0.090 to 0.679). Markers with a moderate DI (<0.5), presumably reflecting their slow rate of evolution, define clusters. Conversely, VNTRs displaying more diversity and more rapid evolution reflect variability within clusters. Our MLVA method is similar to those proposed for *S. aureus*²⁰ or *S. epidermidis*²¹, in terms of number of VNTRs and diversity indexes. Of note, 5 of the 7 VNTRs presented similarities (80 to 90% Blast identity) with other staphylococci, including *S. aureus* and *S. haemolyticus*, but the 7 primers are specific to the species *S. lugdunensis* excluding the possibility of false positive amplifications.

In the midst of all VNTRs identified with Tandem Repeats Finder software, some have been eliminated because fragments were too long to be amplified and sequenced in standard PCR conditions. For example, some were located in genes encoding putative surface-exposed proteins, like *sls* genes, as described by Heilbronner *et al.*²⁸. However, these VNTRs could be attractive because the number of repeats is possibly more subjected to

selective pressure. Indeed, previous MLVA studies of *Staphylococcus*^{21,29} and *Streptococcus agalactiae*²² found high variability in genes encoding Sdr adhesins and FbsA surface protein, respectively. Thus, these targets would be interesting to include in an MLVA method based on a whole genome sequencing approach.

We compared MLVA with the widely used MLST and the alternative trilocus MVLST. This latter method allowed us to define nearly the same number of profiles as MLST and was consistent with MLST, as unique STs or their single locus variants (SLVs) and double locus variants (DLVs) were found in each VT^T group. Thus, this comparison confirms the coevolution of virulence-associated loci studied with housekeeping genes as shown previously¹⁷, and validates the use of the trilocus method as a good epidemiological tool while significantly reducing costs.

Analysis of the seven VNTRs on the overall panel of isolates revealed that MLVA was highly discriminant compared to MLST in terms of number of profiles as well as discriminatory index. This could be explained by the fact that VNTRs evolve more rapidly than housekeeping genes, limiting the MLST resolution. However, although these two methods measure different evolutionary mechanisms at multiple genomic loci, similar clustering was obtained, probably due to the rarity of recombination among *S. lugdunensis* isolates¹⁶. Thus, our findings suggest that the loci targeted by MLVA provide a robust measure of genetic relationships. The greater discriminatory index of MLVA made it possible to distinguish between isolates within CCs defined by MLST, but this appeared highly CC-dependent. Indeed, some clones, especially CC1, showed a significant diversity with many genotypes identified by MLVA and TRST. On the contrary, CC2 was less discriminated by these methods, suggesting that CC2 has emerged more recently than CC1. The higher discriminatory power of MLVA in some lineages defined as clonal by MLST has previously been described for several other species such as *Listeria monocytogenes*²³ and *Streptococcus agalactiae*²².

On the other hand, MLVA is generally considered as an unreliable phylogenetic method because of the high frequency of the independent evolution of the same repeat array sizes in different isolates (*i.e.* size homoplasy)^{30,31}. Thus, sequencing was selected not only as a method for repeat number verification, but also to detect single nucleotide polymorphisms in addition to allele size, SNPs being less prone to homoplasy³². It allowed us to propose a TRST method, which proved more discriminating than MLVA. Indeed, 55 MTs were identified by analyzing repeat array sizes, whereas 69 TRTs were identified by sequencing, indicating that a quarter of these alleles were misclassified by conventional MLVA. Considering that the growing use of whole genome sequencing, with its continuously decreasing costs and increased rapidity³³, the present TRST method offers an additional tool for a worldwide traceable picture of *S. lugdunensis* clones. New alleles and TRST types will be certainly identified by *in silico*-TRST analysis as new genome sequences become available, contributing to a global evaluation of *S. lugdunensis* epidemiology.

MLVA and TRST have strong potential for inter-laboratory comparisons, providing accurate and reproducible data that can easily be exchanged. We have proved that the VNTRs studied were stable during laboratory subcultures, in terms of number of repeats as well as nucleotide sequences, and found complete reproducibility. Furthermore, the epidemiological concordance of the related strains has been validated. Thus, our new MLVA and TRST methods fulfill a number of criteria recognized as important for the successful implementation of a typing method and the interpretation of its results²⁶. Here, the use of seven VNTRs resulted in high discriminatory capacity. Interestingly, the combination of only five markers (SLU1, SLU2, SLU3, SLU4 and SLU6) resulted in a highly similar clustering of isolates (44 MTs) while slightly decreasing the discriminatory capacity (DI = 0.920). This simpler version of the MLVA method could allow potential future use of multiplex PCR assays with fluorescent primers and determination of allele size by capillary electrophoresis.

In this study, no strong correlation was observed between genotypes and clinical settings of the 128 isolates. This could suggest that host susceptibility plays an important role in the propensity of isolates to cause disease, and/or that the virulence potential of isolates is not reflected by MLST or MLVA data. This has been described for *S. aureus*, for which MLST data do not provide any information regarding the pathogenicity of isolates, because changes in virulence correspond primarily to changes in the accessory, rather than core genome³⁴. Like *S. aureus*, *S. lugdunensis* evolution is predominantly clonal^{16,17}, therefore the study of the accessory genome could contribute to a better understanding of the genetic basis of these different clinical behaviors. On the other hand, the majority of carriage isolates and those collected from device-associated infections of this work were distributed in one complex whatever the typing method. Thus, it would be interesting to analyze a larger number of worldwide invasive and carriage *S. lugdunensis* isolates to measure the ratio of pathogenic to carrier strains for the various lineages. This ratio, combined with the study of the accessory genome, would help to investigate the existence of successful commensal clones.

As previously reported in other studies^{4,16}, CC1 (n = 31) and CC3 (n = 26) were the most common clonal complexes identified by MLST for the 98 isolates collected at Strasbourg University Hospital. Within these CCs, the most prevalent STs were the primary founders of these complexes, ST6 (n = 14) and ST3 (n = 26) respectively. Interestingly, a particularly prevalent clone (n = 23), including virulent and carriage isolates, was identified by both MLST (ST3) and MLVA/TRST analysis (MT1/TRT1), even if the discriminatory power of these latter typing methods is high. Thus, whole genome (wg) sequencing, and particularly wgMLST, wgTRST and/or virulome determination, would be useful to identify markers to distinguish between such isolates. On the other hand, the fact that these isolates have been regularly identified from different patients and different hospital wards over a 24-month period suggests that this could be a clone capable of long-term persistence. While ST3 has already been identified^{4,16}, this present study is the first to describe it as a persistent clone. This has already been described by Cheng *et al.*³⁵, who identified a major endemic ST6 clone of oxacillin-resistant *S. lugdunensis* in a tertiary medical center in Taiwan. However, unlike Cheng *et al.*, ST3 isolates are not multiresistant to antibiotics, so their persistence cannot be directly related to antibiotic pressure; it could be linked to a higher fitness under certain environmental or clinical conditions. In the light of these results and that of Kassem *et al.*³⁶, who showed that clinical

surfaces are frequently contaminated by *S. lugdunensis*, further investigations are needed to explore the mode of transmission and environmental reservoirs of this species.

In conclusion, this work provides a description of the first two VNTR-based methods for *S. lugdunensis* typing. MLVA and TRST are the most discriminant typing methods available to date. Furthermore, the analysis of the polymorphism of seven VNTRs in terms of variations in copy number of repeats, as well as internal sequence, allowed us to predict CCs with high accuracy while distinguishing strains belonging to some lineages. Sequence-based MLVA may prove particularly useful because it yields unambiguous data that can easily be exchanged and used for inter-laboratory comparisons. Resolving phylogenetic diversity to a high level in this clonal species, MLVA and TRST represent promising tools which could help to elucidate the global epidemiology and evolution of *S. lugdunensis* as well as to identify outbreaks.

Methods

Bacterial isolates. A total of 128 *S. lugdunensis* human isolates were used in this study (see Table S1 in the supplemental material). A first set of 30 isolates collected from five French regions and Sweden was used for an initial validation of two new typing methods. Among them, 20 epidemiologically unrelated isolates were used, including the reference strain ATCC 43809, 18 clinical isolates (bacteremia [n = 8], bone and joint infections [n = 4], skin and soft tissue infections [n = 3], medical device [n = 2] and other [n = 1]) as well as a carriage isolate. All 20 isolates were representative of the main clusters previously defined by MLST¹⁶ and MVLST¹⁷. To evaluate the epidemiological concordance of MLVA, we also included 5 pairs of isolates from patients at time intervals ranging from 0 to 17 days. Second, to assess the performances of MLVA, 82 clinical isolates (bone and joint infections [n = 28], skin and soft tissue infections [n = 23], bacteremia [n = 9], urinary tract infections [n = 5], medical device [n = 4] and other [n = 13]), collected during the VISLISI trial (Virulence of *Staphylococcus lugdunensis* in Severe Infections)¹⁴, and 16 carriage isolates (isolated from various anatomical locations not associated with infection) were included. These 98 isolates were collected from November 2013 to March 2016 at Strasbourg University Hospital, France. All the 128 isolates were identified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometer.

DNA extraction and PCR amplification. Isolates (stored at -80°C before use) were grown overnight at 37°C on tryptic soy agar with 5% horse blood agar plate. DNA was extracted using the InstaGene Matrix kit (Bio-Rad, Marnes la Coquette, France) according to the manufacturer's recommendations. PCRs were performed on a Veriti Thermal Cycler (Applied Biosystems, Foster City, CA, USA) in a final volume of 25 μl containing 12.5 μl GoTaq G2 Green Master Mix (Promega, Charbonnières-Les-Bains, France), 0.50 μM each primer and 5 μl of extracted DNA.

MultiLocus Sequence Typing (MLST). MLST genotyping was performed by sequencing seven gene loci (*aroE*, *dat*, *ddl*, *gmk*, *ldh*, *recA* and *yqjL*) as previously described¹⁶, except that the PCR products were purified and sequenced by GATC Biotech SARL (Konstanz, Germany). Allelic profiles and corresponding sequence types (STs) were assigned with BioNumerics software (Version 7.6, Applied Maths, Sint-Martens-Latem, Belgium). The sequences of all new alleles and the new STs identified were entered into the international MLST database (<http://bigsd.b.pasteur.fr/staphlugdunensis/>). STs were clustered into clonal complexes (CCs) when they shared at least five identical alleles (single [SLV] or double locus variants [DLV]) with another member of the group, by using the eBURST V3 program (<http://eburst.mlst.net>).

MultiVirulence-Locus Sequence Typing (MVLST). MVLST genotyping was based on the sequence determination of the intragenic regions of three virulence-associated loci, *atlLR3*, *isdJ*, and SLUG_16930 as previously described¹⁷, except that the PCR products were purified and sequenced by GATC Biotech SARL (Konstanz, Germany). Sequences of the three loci were analyzed using BioNumerics software (Version 7.6, Applied Maths, Sint-Martens-Latem, Belgium) to determine allelic profiles and trilocus virulence type (VT^T).

Multiple-Locus VNTR Analysis (MLVA) assay development. The genome sequences of *S. lugdunensis* isolates N920143²⁸ and HKU09-01³⁷ were screened *in silico* for the presence of tandem repeats (TRs) by using Tandem Repeats Finder software, version 4.07b³⁸. Primers flanking 28 potential VNTR loci were designed using the OLIGO Primer Design software (Molecular Biology Insights, USA); their specificity was evaluated by Blastn. Primers were tested on the panel of 30 genetically diverse *S. lugdunensis* isolates to assess amplification ability, repeatability and the level of polymorphism of these loci. A VNTR was validated as a good candidate for MLVA typing if the locus (i) was present in all isolates, (ii) exhibited a minimum size of 18 bp for an individual TR, and (iii) required a single primer pair for amplification and sequencing of the internal fragment. Of the 28 VNTRs tested, seven fulfilled the above mentioned criteria, and were named SLU (for *S. lugdunensis*) followed by a number.

The primers used for VNTR amplification and sequencing are presented in Table 1. VNTR loci were amplified in simplex PCRs using the following steps: denaturation at 95°C for 5 min, followed by 35 cycles of amplification including denaturation at 95°C for 30 s, annealing at a temperature depending on the primer pair for 30 s, and extension at 72°C for 30 s. A final extension step was performed at 72°C for 5 min. All PCR products were visualized using gel-electrophoresis, purified, and sequenced by GATC Biotech SARL (Konstanz, Germany).

The number of TRs of each VNTR was determined with BioNumerics software (Version 7.6, Applied Maths, Saint-Martens-Latem, Belgium) by using the "polymorphic VNTR typing" plugin. To check that the entire repeat array had been sequenced, the flanking sequences (start and stop pattern) were spotted but were not included as typing information. For incomplete repeats, the TR copy number was rounded down to the nearest complete copy number, in accordance with the guidelines published by Nadon *et al.*³⁹. The numbers of TRs at each locus were combined and listed in the order SLU1-SLU2-SLU3-SLU4-SLU5-SLU6-SLU7 e.g. 1-2-10-4-3-2-2. Each unique combination was converted into a distinct MLVA type (MT).

Tandem Repeat Sequence Typing (TRST) of VNTR loci. Sequences of TR were determined for each VNTR, and each unique repeat sequence was identified by the name of the VNTR in lower case, followed by a number (Table S2); for example repeat sequence slu1_01 (Fig. S1). Each unique combination of TR sequences defined an allele (Table S3). Each allele was named by a number, corresponding to the number of TRs, followed by a lowercase letter, corresponding to the combination of the different TR sequences of the VNTR, e.g. SLU1_2a (slu1_00-slu1_01) (for ease of viewing, the allele does not include the lowercase “vntr” for each repeat; e.g. 00–01) (Fig. S1). The assignment of VNTR alleles was performed with BioNumerics software (Version 7.6, Applied Maths, Sint-Martens-Latem, Belgium), based on the successive occurrence of user-defined TR sequences. Each unique combination of alleles defined an allelic profile (with VNTRs listed in the same order as MLVA profiles), converted into a distinct TRST type (TRT).

Data analysis. All typing data were uploaded into the BioNumerics software (Version 7.6, Applied Maths, Sint-Martens-Latem, Belgium) and clustered using the appropriate settings. Two different techniques were used to represent the relationships between isolates. A dendrogram was generated based on the categorical coefficient provided by the BioNumerics software and the algorithm of unweighted pair group method with arithmetic mean (UPGMA). Cutoff values of 67% and 91% similarity were applied to define respectively MLVA and TRST clusters. Minimum spanning trees were constructed using the categorical coefficient and the priority rule to first link types that have the highest number of single locus variants. MTs were grouped into MLVA complexes (MCs) if they differed by no more than a single VNTR. To generate a minimum spanning tree based on TRST data, a composite dataset was constructed using BioNumerics software, based on a multiple alignment of each VNTR allele.

The discriminatory ability as well as the congruence of the results obtained by using the different typing methods and/or their combination were evaluated. To this aim, both the Simpson's diversity index (DI)⁴⁰ and adjusted Wallace coefficients (AW)⁴¹ (plus confidence intervals (CI) as described by Grundmann *et al.*⁴²) were calculated by resorting to the online tool at <http://www.comparingpartitions.info/?link=Tool>.

The standardized index of association (I_A^S) between alleles was calculated to test for linkage disequilibrium between alleles of MLVA and TRST using the LIAN Linkage Analysis 3.7 online tool (<http://guanine.evolbio.mpg.de>) as described by Haubold and Hudson⁴³. The null hypothesis of linkage equilibrium, $I_A^S = 0$, was tested with 100000 Monte Carlo simulations.

Test of VNTR stability. Since VNTR regions are considered to be one of the fastest evolving sequences of a genome, it was important to test the stability of the selected VNTR regions. To determine this stability, two isolates were subcultured on tryptic soy agar for 30 consecutive days by streaking a single colony from each strain on agar plates. The original culture, subculture numbers 10, 20 and 30 were used for DNA extraction and the total DNA was subjected to the MLVA and TRST assays.

References

- Bieber, L. & Kahlmeter, G. *Staphylococcus lugdunensis* in several niches of the normal skin flora. *Clin. Microbiol. Infect.* **16**, 385–388 (2010).
- van der Mee-Marquet, N. *et al.* *Staphylococcus lugdunensis* infections: high frequency of inguinal area carriage. *J. Clin. Microbiol.* **41**, 1404–1409 (2003).
- Frank, K. L., del Pozo, J. L. & Patel, R. From clinical microbiology to infection pathogenesis: how daring to be different works for *Staphylococcus lugdunensis*. *Clin. Microbiol. Rev.* **21**, 111–133 (2008).
- Yeh, C.-F. *et al.* Clinical features, outcomes, and molecular characteristics of community- and health care-associated *Staphylococcus lugdunensis* infections. *J. Clin. Microbiol.* **54**, 2051–2057 (2016).
- Bocher, S., Tønning, B., Skov, R. L. & Prag, J. *Staphylococcus lugdunensis*, a common cause of skin and soft tissue infections in the community. *J. Clin. Microbiol.* **47**, 946–950 (2009).
- Heldt Manica, L. A. & Cohen, P. R. *Staphylococcus lugdunensis* infections of the skin and soft tissue: a case series and review. *Dermatol. Ther.* **7**, 555–562 (2017).
- Shah, N. B. *et al.* Laboratory and clinical characteristics of *Staphylococcus lugdunensis* prosthetic joint infections. *J. Clin. Microbiol.* **48**, 1600–1603 (2010).
- Argemi, X. *et al.* *Staphylococcus lugdunensis* small colony variant conversion resulting in chronic prosthetic joint infection. *Médecine Mal. Infect.* **47**, 498–501 (2017).
- Liu, P.-Y. *et al.* *Staphylococcus lugdunensis* infective endocarditis: a literature review and analysis of risk factors. *J. Microbiol. Immunol. Infect.* **43**, 478–484 (2010).
- Kyaw, H., Shaikh, A. Z., Yaratha, G. & Deepika, M. Septic embolic stroke resulting from *Staphylococcus lugdunensis* endocarditis. *Ochsner J.* **17**, 184–188 (2017).
- Mitchell, J. Characterization of the fibrinogen-binding surface protein Fbl of *Staphylococcus lugdunensis*. *Microbiology* **150**, 3831–3841 (2004).
- Nilsson, M. *et al.* A von Willebrand factor-binding protein from *Staphylococcus lugdunensis*. *FEMS Microbiol. Lett.* **234**, 155–161 (2004).
- Donvito, B. *et al.* Synergistic hemolytic activity of *Staphylococcus lugdunensis* is mediated by three peptides encoded by a non-agr genetic locus. *Infect. Immun.* **65**, 95–100 (1997).
- Argemi, X. *et al.* VISLISI trial, a prospective clinical study allowing identification of a new metalloprotease and putative virulence factor from *Staphylococcus lugdunensis*. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **23**, 334.e1–334.e8 (2017).
- Yeh, C.-F. *et al.* Molecular characteristics of disease-causing and commensal *Staphylococcus lugdunensis* Isolates from 2003 to 2013 at a tertiary hospital in Taiwan. *PLOS ONE* **10**, e0134859 (2015).
- Chassain, B. *et al.* Multilocus sequence typing analysis of *Staphylococcus lugdunensis* implies a clonal population structure. *J. Clin. Microbiol.* **50**, 3003–3009 (2012).
- Didi, J., Lemeé, L., Gibert, L., Pons, J.-L. & Pestel-Caron, M. Multi-virulence-locus sequence typing of *Staphylococcus lugdunensis* generates results consistent with a clonal population structure and is reliable for epidemiological typing. *J. Clin. Microbiol.* **52**, 3624–3632 (2014).
- García-Yoldi, D. *et al.* Comparison of multiple-locus variable-number tandem-repeat analysis with other PCR-based methods for typing *Brucella suis* isolates. *J. Clin. Microbiol.* **45**, 4070–4072 (2007).
- Visca, P. *et al.* Investigation of the population structure of *Legionella pneumophila* by analysis of tandem repeat copy number and internal sequence variation. *Microbiology* **157**, 2582–2594 (2011).

20. Schouls, L. M. *et al.* Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. *PLoS ONE* **4**, e5082 (2009).
21. Johansson, A., Koskiniemi, S., Gottfridsson, P., Wistrom, J. & Monsen, T. Multiple-locus variable-number tandem repeat analysis for typing of *Staphylococcus epidermidis*. *J. Clin. Microbiol.* **44**, 260–265 (2006).
22. Haguenoer, E. *et al.* A multi locus variable number of tandem repeat analysis (MLVA) scheme for *Streptococcus agalactiae* genotyping. *BMC Microbiol.* **11**, 171 (2011).
23. Chenal-Francois, V. *et al.* Optimized multilocus variable-number tandem-repeat analysis assay and its complementarity with pulsed-field gel electrophoresis and multilocus sequence typing for *Listeria monocytogenes* clone identification and surveillance. *J. Clin. Microbiol.* **51**, 1868–1880 (2013).
24. Melles, D. C. *et al.* High-throughput typing of *Staphylococcus aureus* by amplified fragment length polymorphism (AFLP) or multi-locus variable number of tandem repeat analysis (MLVA) reveals consistent strain relatedness. *Eur. J. Clin. Microbiol. Infect. Dis.* **28**, 39–45 (2009).
25. Brevik, Ø. J., Ottem, K. F. & Nylund, A. Multiple-locus, variable number of tandem repeat analysis (MLVA) of the fish-pathogen *Francisella noatunensis*. *BMC Vet. Res.* **7**, 5 (2011).
26. van Belkum, A. *et al.* Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **13**(Suppl 3), 1–46 (2007).
27. Argemi, X., Hansmann, Y., Riegel, P. & Prévost, G. Is *Staphylococcus lugdunensis* significant in clinical samples? *J. Clin. Microbiol.* **55**, 3167–3174 (2017).
28. Heilbronner, S. *et al.* Genome sequence of *Staphylococcus lugdunensis* N920143 allows identification of putative colonization and virulence factors: *Staphylococcus lugdunensis* genome sequence. *FEMS Microbiol. Lett.* **322**, 60–67 (2011).
29. Sabat, A. *et al.* Distribution of the serine-aspartate repeat protein-encoding *sdr* genes among nasal-carriage and invasive *Staphylococcus aureus* strains. *J. Clin. Microbiol.* **44**, 1135–1138 (2006).
30. Estoup, A., Jarne, P. & Cornuet, J.-M. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol. Ecol.* **11**, 1591–1604 (2002).
31. Reyes, J. F., Chan, C. H. S. & Tanaka, M. M. Impact of homoplasy on variable numbers of tandem repeats and spoligotypes in *Mycobacterium tuberculosis*. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **12**, 811–818 (2012).
32. Pearson, T. *et al.* Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc. Natl. Acad. Sci. USA* **101**, 13536–13541 (2004).
33. Sabat, A. J. *et al.* Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance* **18**, 20380 (2013).
34. Turner, K. M. E. & Feil, E. J. The secret life of the multilocus sequence type. *Int. J. Antimicrob. Agents* **29**, 129–135 (2007).
35. Cheng, C. *et al.* Persistence of a major endemic clone of oxacillin-resistant *Staphylococcus lugdunensis* sequence type 6 at a tertiary medical centre in northern Taiwan. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **36**, 72–77 (2015).
36. Kassem, I. I., Esseili, M. A. & Sigler, V. Detection and differentiation of staphylococcal contamination of clinical surfaces using denaturing gradient gel electrophoresis. *J. Hosp. Infect.* **78**, 187–193 (2011).
37. Tse, H. *et al.* Complete genome sequence of *Staphylococcus lugdunensis* strain HKU09-01. *J. Bacteriol.* **192**, 1471–1472 (2010).
38. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
39. Nadon, C. A. *et al.* Development and application of MLVA methods as a tool for inter-laboratory surveillance. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **18**, 20565 (2013).
40. Hunter, P. R. & Gaston, M. A. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**, 2465–2466 (1988).
41. Severiano, A., Pinto, F. R., Ramirez, M. & Carriço, J. A. Adjusted Wallace coefficient as a measure of congruence between typing methods. *J. Clin. Microbiol.* **49**, 3997–4000 (2011).
42. Grundmann, H., Hori, S. & Tanner, G. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J. Clin. Microbiol.* **39**, 4190–4192 (2001).
43. Haubold, B. & Hudson, R. R. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *LinkageAnalysis. Bioinforma. Oxf. Engl.* **16**, 847–848 (2000).

Acknowledgements

This study was supported by an educational grant for Jérémie Lebeurre from the Charles Nicolle Foundation. The authors would like to thank Estelle Bilak-Jumas and Simon Le Hello for helpful discussion. The authors are grateful to Nikki Sabourin-Gibbs, Rouen University Hospital, for her help in editing the manuscript.

Author Contributions

S.D., J.L., X.A. and M.P.C. designed the study. X.A. and G.P. provided the VISLISI clinical isolates. S.D. and J.L. performed all manipulations and analyzed the data. S.D., J.L. and M.P.C. wrote the manuscript. All authors have read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30144-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

1.3. Publication n°4

Fbi typing of Staphylococcus lugdunensis: a frontline tool for epidemiological studies, but not predictive of fibrinogen binding ability

Sandrine Dahyot, Jérémie Lebeurre, Floriane Laumay, Xavier Argemi, Charline Dubos, Ludovic Lemée, Gilles Prévost, Patrice François, Martine Pestel-Caron. En cours de soumission.

1 **Fbi typing of *Staphylococcus lugdunensis* : a frontline tool for epidemiological studies,**
2 **but not predictive of fibrinogen binding ability**

3
4
5 Sandrine Dahyot^{1*}, ⁶, Jérémie Lebeurre^{2, 6}, Floriane Laumay³, Xavier Argemi^{4, 5}, Charline
6 Dubos¹, Ludovic Lemée¹, Gilles Prévost⁴, Patrice François³, Martine Pestel-Caron¹.

7
8 ¹Normandie Univ, UNIROUEN, GRAM EA2656, Rouen University Hospital, F-76000
9 Rouen, France

10 ²Normandie Univ, UNIROUEN, GRAM EA2656, F-76000 Rouen, France

11 ³University of Geneva Hospitals, Genomic Research Laboratory, Service of Infectious
12 Diseases, Geneva, Switzerland

13 ⁴Université de Strasbourg, CHRU de Strasbourg, VBP EA7290, Fédération de Médecine
14 Translationnelle de Strasbourg (FMTS), Institut de Bactériologie, 3 rue Koeberlé, F-67000
15 Strasbourg, France

16 ⁵Hôpitaux Universitaires, Maladies Infectieuses et Tropicales, F-67000 Strasbourg, France

17
18 ⁶S. Dahyot and J. Lebeurre contributed equally to this work

19
20
21 *Corresponding author:

22 E-mail: sandrine.dahyot@chu-rouen.fr

23 Phone number : +33235148299 ; fax number : +33235148203

24 Groupe de Recherche sur l'Adaptation Microbienne (EA2656), Normandie Univ,
25 UNIROUEN, Rouen University Hospital, F-76000 Rouen, France

28 **Abstract**

29 *Staphylococcus lugdunensis* has been increasingly recognized as a virulent pathogen,
30 responsible for severe infections with an outcome resembling that of *Staphylococcus aureus*.
31 Here, we developed a repeat-based typing scheme based on the sequence analysis of the R-
32 domain within the *fbl* gene encoding the fibrinogen-binding protein Fbl. We typed 240 *S.*
33 *lugdunensis* isolates of various clinical and geographic origins. The length of the R-domain
34 ranged from 9 to 52 repeats. *fbl* typing identified 54 unique 18-bp repeat sequences and 92
35 distinct *fbl* types. *fbl* typing revealed a discriminatory power higher than that of multilocus
36 sequence typing (MLST) and equivalent to that of tandem repeat sequence typing. *fbl* types
37 could assign isolates to MLST clonal complexes with a very good predictive power. The
38 ability to promote adherence to immobilized human fibrinogen was evaluated for 55 isolates
39 chosen to reflect the genetic diversity of the *fbl* gene. The lowest percentage of fibrinogen
40 binding was observed for isolates carrying a 5'-end frameshift mutation of the *fbl* gene and for
41 those harboring less than 43 repeats in the R-domain. qRT-PCR assays for a few strains
42 revealed no correlation between *fbl* gene expression and fibrinogen binding capacity. In
43 conclusion, this study shows that *fbl* typing is a useful tool in *S. lugdunensis* epidemiology,
44 especially because it is an easy to use, cost-effective, rapid and portable method ([http://fbl-](http://fbl-typing.univ-rouen.fr/)
45 [typing.univ-rouen.fr/](http://fbl-typing.univ-rouen.fr/)). The impact of *fbl* polymorphisms on the structure of the protein, its
46 expression on the cell surface and in virulence remains to be determined.

47

48

49 Introduction

50 *Staphylococcus lugdunensis*, first described by Freney in 1988¹, is a member of the coagulase-
51 negative staphylococci (CoNS). It belongs to the normal human skin flora, colonizing several
52 distinct niches such as perineal and inguinal areas² but has been increasingly recognized as a
53 virulent pathogen³. Described as a “wolf in sheep’s clothing”, *S. lugdunensis* behaves in many
54 ways more like the coagulase positive *Staphylococcus aureus*, exhibiting increased virulence
55 compared to that of other CoNS³. It can cause many types of infections, ranging from
56 localized to systemic^{4,5}. It has mainly be reported in skin and soft tissue infections⁶, but it is
57 also responsible for catheter-related bloodstream infections, bone and joint infections⁷ and
58 severe infective endocarditis⁸. Infective endocarditis can occur in native or prosthetic valves,
59 and are associated with an unusual rate of morbidity and mortality⁹.

60 In contrast to *S. aureus*, only very few pathogenicity factors have been characterized¹⁰. *S.*
61 *lugdunensis* can in particular express proteins that are covalently anchored to the cell wall
62 peptidoglycan by sortases, like the fibrinogen-binding protein Fbl¹¹ and the von Willebrand
63 factor-binding protein vWbl¹². Of note, *fbl* and *vwbl* genes were highly prevalent in *S.*
64 *lugdunensis* strains assayed previously (100% and 98.2% respectively)^{13,14}. In addition to
65 these two virulence factors, it is noteworthy that *S. lugdunensis* is the only CoNS to have a
66 locus encoding iron-regulated surface determinant proteins, involved in iron acquisition from
67 haemoglobin¹⁵.

68 A variety of molecular typing methods have been developed for *S. lugdunensis*
69 characterization, including pulsed-field gel electrophoresis (PFGE)¹⁶, multilocus sequence
70 typing (MLST)¹⁷ and multivirulence-locus sequence typing (MVLST)¹⁸. Phylogenetic
71 analyses by MLST and MVLST revealed a clonal population structure, a mutational evolution
72 of this pathogen, and the lack of hypervirulent lineages. MLST has become the tool of choice
73 for unambiguously defining clones, however portable tools that are even more discriminating
74 than MLST are needed.

75 Another method involves the typing of multiple, rapidly-evolving *loci*, called variable number
76 of tandem repeats (VNTRs). By using these markers, one can construct allelic profiles in
77 multiple locus VNTR analysis (MLVA) systems, making strain differentiation of clonal
78 bacteria populations possible¹⁹. Recently, we developed the first VNTRs-based schemes for *S.*
79 *lugdunensis* typing: a classic length-based MLVA method and a sequence-based MLVA

80 method called tandem repeat sequence typing (TRST) method²⁰. These typing methods have
81 been shown to be more discriminatory than MLST and MVLST, and represent very promising
82 tools for molecular epidemiological studies of *S. lugdunensis*.

83 A more practical alternative to the multilocus approach is the sequence-based typing of one
84 (or two) highly polymorphic *loci*, since – analyzing a clonal population – phylogenetic
85 inferences will rarely be confounded through homologous genetic recombination²¹. This
86 approach have previously been established for a number of pathogens, including
87 *Staphylococcus aureus* (*spa*)²², *Streptococcus pyogenes* (*emm*)²³, *Listeria monocytogenes*
88 (*LmMT1*, *LmMT2*)²⁴ and *Clostridium difficile* (TR6, TR10)²¹.

89 For instance, *spa* has been one of the most successful single genetic markers used for the
90 typing of *S. aureus* strains²⁵. *spa* typing is an established typing method based on the
91 assessment of the number and the sequence variation in the X-region of the *Staphylococcus*
92 protein A gene (*spa*), containing 2 to 15 repetitive sequences consisting of 21 to 27 bp
93 (mainly 24 bp)²⁶. The X-region of *spa* consists of direct repeats exhibiting an extensive
94 polymorphism based on point mutations, deletions, duplications, and insertions. *spa* typing
95 has been shown to be highly concordant with other typing methods like MLST or
96 PFGE^{22,27,28}. Due to the clonal population structure of *S. aureus*, *spa* typing is regarded a
97 highly discriminatory method that can be used for outbreak investigations as well as for
98 assigning strains to phylogenetic lineages in population studies^{29,30}. Moreover, *spa* sequencing
99 has significant advantages in terms of speed, ease of use, ease of interpretation, and
100 standardization among laboratories²².

101 Like *spa* typing, typing using *clf*-R-domains (18-pb repeats) of clumping factors ClfA (*clfA*)
102 and ClfB (*clfB*) has been shown as useful for *S. aureus* typing and grouping host-specific
103 lineages^{31,32}. ClfA is a cell-wall associated adhesin that mediates bacterial adherence to
104 fibrinogen (Fg). The Fg-binding domain of ClfA is located within region A. Between the
105 region A and the C-terminal cell wall-associated region is the dipeptide repeat region (region
106 R) acting as a flexible stalk to extend the ligand-binding A domain from the cell surface³³.
107 The *clfA*-R region varies in size among different *S. aureus* strains^{31,34}, and it has been shown
108 that copy number of repeats affect adherence and clumping titers of *S. aureus* strains^{33,35}.

109 Interestingly, the protein Fbl encoded by the *fbl* gene of *S. lugdunensis* is closely related to the
110 ClfA of *S. aureus*, showing 62% amino acid identity in the minimal Fg-binding region^{11,36}.

111 However, the R-region of Fbl (SDSDSA hexapeptide motif) differs slightly from that of ClfA
112 (SD repeats only). The size of the R-region of *fbl* varies according to the strains¹¹. In this
113 context, the analysis of the *fbl* R-domain could be a potential useful marker for *S. lugdunensis*
114 typing.

115 Therefore, the objectives of this study were (i) to develop and evaluate the potential use of a
116 *fbl* R-domain repeat-based system for typing diverse *S. lugdunensis* isolates (ii) to evaluate its
117 typing performances with MLST and TRST and (iii) to determine the potential impact of *fbl*
118 gene polymorphisms detected on the ability of isolates to bind Fg *in vitro*.

119

120 Results

121 *fbl* R-domain polymorphism

122 Two hundred and forty *S. lugdunensis* isolates from diverse clinical settings and geographical
123 origins were characterized by *fbl* typing. All repeats regardless of length were successfully
124 sequenced and assembled. Of note, the R-region of the *fbl* gene of one isolate (SL_V70) was
125 more difficult to amplify because of the presence of an insertion sequence (IS) located
126 downstream the repeated region. A Blast search revealed a significant similarity (similarity
127 score of 841) with the IS6 family. The number of repeats varied between 9 and 52,
128 characterized by PCR products ranging from 586 to 1360 bp in length. The average number of
129 repeats in each strain's R-region was 42. Analysis of the sequences resulted in the
130 identification of 54 unique 18-bp repeat sequences, each given a numeric code (Table S2).
131 Each repeat varied from another by at least one point mutation. The deduced amino acid
132 sequences of the repeats allowed to identify 13 types of 6-amino-acid units (Table S2). The
133 most frequent pattern was DSDSDA. Polymorphisms resulted for most repeats in
134 synonymous substitutions, with a ratio of nonsynonymous to synonymous substitutions of
135 0.045 (a ratio of > 1 indicates positive selection, a ratio of 1 indicates no selection pressure
136 and a ratio of < 1 indicates purifying selection), suggesting a potential role of environmental
137 selective pressure in their evolution (Table S2).

138

139 *fbl* typing results

140 The organization of the repeats of the *fbl* R-domain (composition, number and order of
141 repeats) from each of the isolates was represented as a *fbl* type repeat code. Ninety two
142 distinct *fbl* types were defined for the 240 isolates, designated fbl9a to fbl52a. A complete
143 overview of these profiles is presented in Table S3. The most common *fbl* types were fbl47b
144 ($n = 43$), fbl45f ($n = 24$) and fbl41a ($n = 13$). There were 68 *fbl* types that were found only
145 once among the 230 patients. The length of the variable region was not an accurate indicator
146 of *fbl* type, as many had the same number of repeats (for example, 15 unique *fbl* types had 45
147 repeat elements). To assess the discriminatory power of *fbl* typing, we excluded 10 of the 20
148 isolates that were isolated from the same patient and that had the same *fbl* type. Simpson's DI
149 calculated on the basis of the 230 remaining unrelated isolates was 0.946 (Table 2).

150 Analysis of amino acid profiles of the *fbl* repeats revealed interesting structural patterns; for
151 example, 80% of *fbl* types ended with the same three motifs of six amino acids (DSDSNS-
152 GSDSNV-DSDSDS), although major variations exist at the DNA level (Table S3). For 20%

153 of the other *fbl* types, the three the last amino acids motif differed slightly (DSDSDA-
154 DSDSDA-DSDSDS).

155

156 ***fbl* clustering**

157 A minimum spanning tree was constructed for the 240 isolates in order to visualize the
158 relationships among the *fbl* types (Fig. 2a). The 92 unique *fbl* types were distributed over 10
159 clusters (cluster 1 to 10), each containing more than 2 isolates, and 46 singletons. *fbl*
160 clustering allowed the grouping of isolates that had similar repeat organizations. The main
161 cluster was cluster 1 (57 isolates), consisting of 20 *fbl* types, followed by cluster 2 (52
162 isolates) with 6 *fbl* types. These two largest *fbl* clusters comprised 45% of the isolates. Of the
163 240 isolates analyzed in the study, 81% were part of a cluster. The majority of isolates
164 belonging to *fbl* cluster 2 shared as the first repeats of the R-region the repeat codes 0-6-20-
165 23, and the repeat codes 7-18-14-26-44 for the last repeats. Isolates of cluster 1 have in
166 common repeat 0 as the first repeat unit, which is followed by repeats 6-13-1 in a rather
167 conserved order (arrangement identical for 56 of 57 isolates) (Table S3).

168 The distribution of the isolates did not allow to identify any correlation between *fbl* types and
169 clinical settings (Fig. 2b). However, the *fbl* type 47b was shared by 20 of 79 (25%) carriage
170 isolates.

171

172 ***Discriminatory power and concordance between typing methods***

173 The 128 isolates representative of the breadth of the *S. lugdunensis* genetic diversity by using
174 MLST and TRST²⁰ were *fbl* typed. A total of 25 sequence types (STs) belonging to 7 clonal
175 complexes (CCs), 69 TRST types (TRTs) and 60 *fbl* types was identified for these isolates
176 (Table 2). Simpson's DI for the 123 of these 128 isolates (only one genotype included by
177 patient to avoid any bias) was calculated at 0.964 for *fbl* typing, at 0.899 for MLST and at
178 0.943 for TRST (Table 2). Thus, MLST was the least discriminatory technique, whereas *fbl*
179 typing and TRST showed both a high comparable discriminatory power. The five pairs of
180 epidemiologically-related isolates considered as clonal by MLST and TRST belonged to the
181 same *fbl* types. This is in favor of an *in vivo* stability of the *fbl* R-domain, and attests the
182 epidemiological concordance of the method.

183 The minimum spanning tree analysis of the 240 isolates showed that the grouping by *fbl*
184 typing was similar to the grouping found by MLST (Fig. 2a). For example, isolates of *fbl*
185 clusters 1 and 2 belonged to CC1 and CC3 respectively. One of most notable findings was

186 that each *fbl* type was associated with only a single CC. Thus, *fbl* typing was able to provide a
187 clonal complex assessment for the isolates (Table S3). This is confirmed by the calculation of
188 adjusted Wallace coefficient (AW) of *fbl* typing (Table 3). It was found that if any two
189 isolates shared the same *fbl* type, they had 100% of probability of belonging to the same CC.
190 On the basis of these results, a table is provided to be used as a preliminary guideline for the
191 correspondence between *fbl* types and MLST CCs (Table S3).

192 However, *fbl* typing was less able to predict the ST (AW = 0.680) (Table 3), as an *fbl* type
193 could correspond with several STs. For example, *fbl* type 45f was associated with ST1, ST6
194 and ST15; and *fbl* type 47b was associated with ST3, ST16 and ST20. Conversely, some STs
195 could contain up to 11 different *fbl* types, like the ST2 and ST3. On the other hand, the low
196 AW coefficients between *fbl* typing and TRST showed that *fbl* types could predict the TRST
197 types with 35% probability whereas the probability with which the TRST types predicted the
198 *fbl* types was only 21% (Table 3).

199 Thus, *fbl* typing further distinguished new genotypes within the MLST groups on many
200 occasions (for six of the seven CCs analyzed), more rarely (for only two out of the seven
201 CCs) for TRST groups (Table S4). For example, 12 *fbl* types were identified for the CC2
202 isolates whereas only five STs and five TRTs were described for these isolates by MLST and
203 TRST respectively. Strains of CC1 encompassed 20 unique *fbl* types while MLST only
204 distinguishes 5 STs but TRST allowed the identification of 28 TRTs.

205

206 ***Stability of fbl R-domain***

207 Examination of the *fbl* R-domain stability revealed that, for each isolate, identical *fbl* types
208 were generated for the original isolates and their 10th, 20th and 30th subcultures, indicating that
209 the R-region of *fbl* was stable.

210

211 ***Construction and validation of “fbl typing Server”***

212 Fbl typing server was constructed to perform *fbl* typing using sequencing data saved in a
213 FASTA format. It is a BLAST-based publicly available Web-based service hosted by the
214 Rouen University (<http://fbl-typing.univ-rouen.fr/>). The tool contained the 92 *fbl* types
215 available at the time of the study. The *fbl* typing Server Web tool was evaluated on
216 preassembled pairs of Sanger sequences from a data set consisting of 92 samples identified
217 with BioNumerics software, covering the 92 *fbl* types. The *fbl* typing Server identified *fbl*
218 types correctly at a 100% identity match.

219

220 ***fbl* 5'-end polymorphism**

221 While examining the complete sequence variation of *fbl* gene in published *S. lugdunensis*
222 genomes, we discovered a 4 bp deletion occurring at position 16 from the start codon for two
223 strains (Fig. 1b). *In silico* analysis suggested that this deletion caused a frameshift mutation
224 leading to a premature translational stop (L22STOP) (Fig. 1c). One hundred and forty six of
225 the 240 *S. lugdunensis* isolates were screened for this deletion by Sanger sequencing. The
226 deletion was found in 27 isolates, belonging to different *fbl* types (Table S1). Twenty six of
227 the 27 isolates belonged to CC1, with a distribution that varies according to the ST. All
228 isolates characterized as ST12 ($n = 8$) and ST15 ($n = 6$) had this deletion, whereas isolates
229 belonging to ST1 ($n = 12$) and ST7 ($n = 1$) did not. Within ST6 (CC1), 11 of the 15 isolates
230 (73%) had the deletion. The only isolate that had the deletion and was not of CC1 belonged to
231 ST3 (CC3).

232

233 **Fibrinogen binding**

234 The 240 isolates were first tested with rapid latex agglutination test. It yielded positive results
235 for 77 isolates (32%) (Table 4). Then, the capacity of binding to immobilized human
236 fibrinogen was tested for 55 isolates with different *fbl* gene characteristics in terms of *fbl*
237 types as well as the presence or absence of the 5'-end deletion. The level of bacterial binding
238 to immobilized fibrinogen varied considerably within the clinical isolates tested (0.00 to
239 132.83%). Six of the 55 isolates significantly bound to solid-phase fibrinogen (++++) and 11
240 bound with poorly strength (+) (Table 4, Fig. S1). Thirty eight strains did not bind to solid-
241 phase fibrinogen (-).

242 Statistical comparisons of the ability of isolates to bind to fibrinogen showed that the group of
243 19 isolates that had the 5'-end deletion of the *fbl* gene was significantly associated with a
244 lower ability to bind to fibrinogen (6.47%) than the group of 36 isolates that didn't have the
245 deletion (25.43%) ($P < 0.001$). Moreover, within the group of isolates without the *fbl* deletion,
246 the isolates with ≥ 43 repeats bound significantly more strongly to fibrinogen (37.64%) than
247 those with < 43 repeats (6.23%) ($P < 0.001$). All strains that significantly bound to solid-
248 phase fibrinogen were tested positive in the agglutination test. Among the 11 strains that
249 bound poorly to solid-phase fibrinogen, six tested positive in the agglutination test. Only three
250 isolates reacted in the agglutination test but did not bind to solid-phase fibrinogen (Table 4).

251 No difference was observed in our study with regard to the type of clinical specimen from
252 which the isolate was obtained and the fibrinogen binding phenotype.

253

254 ***fbl gene expression analysis***

255 The level of *fbl* gene transcript in mid-exponential phase cells was determined by real-time
256 qRT-PCR for ten isolates selected to represent five Fg-binding-negative and five Fg-binding-
257 positive isolates. Isolates displayed variability in the relative *fbl* expression, ranging from
258 0.002 to 0.382. However, there was no significant difference ($P > 0.05$) in *fbl* gene expression
259 level between the two groups (Fig. 3).

260

261

262 **Discussion**

263 *S. lugdunensis* has emerged as a significant human pathogen with notable clinical and
264 microbiological characteristics that stand out among other CoNS^{3,37}. In an attempt to better
265 understand the genetic background and population structure of *S. lugdunensis*, several
266 sequence-based methods have been developed, such as MLST, MVLST and more recently
267 TRST^{17,18,20}. Phylogenetic analysis revealed a clonal population structure and a mutational
268 evolution of this species. Therefore, single-locus DNA sequencing of repeat regions could be
269 used for reliable and accurate typing of *S. lugdunensis*.

270 The *fbl* R-domain appears to have several properties of a reasonable marker for typing. *fbl*
271 gene encodes the Fbl protein, a homolog of ClfA of *S. aureus*, and contains a SD-repeat
272 motif. SD-repeat (Sdr) containing proteins constitute a subfamily of the Microbial Surface
273 Components Recognizing Adhesive Matrix Molecules (MSCRAMM) family³⁸, including
274 members such as ClfA, ClfB, SdrC, SdrD and SdrE of *S. aureus*³⁹ and SdrF and SdrG of
275 *Staphylococcus epidermidis*⁴⁰. SD repeats have been explored for genotyping due to their
276 polymorphisms in both the copy numbers and the sequence of the repeats^{31,32,41}. In the present
277 study, we have developed and evaluated a new typing system for *S. lugdunensis* based on the
278 DNA sequencing of the *fbl* R-domain as an alternative to current techniques for use in
279 research and clinical applications.

280 Our strain collection fulfilled Van Belkum's criteria¹⁹ for validation and application of a
281 typing methods for use in bacterial epidemiology, because it included a large “test
282 population” of isolates representative of the breadth of the *S. lugdunensis* genetic diversity as
283 well as a set of clinical isolates with well-defined inclusion criteria. Unambiguous *fbl* types
284 were achieved for all the 240 isolates used in this study. They were organized to describe a
285 total of 92 different *fbl* types, grouped into 10 distinct clusters. To improve *fbl* cluster
286 definition, it would be interesting to develop and implement an automated and objective tool
287 to infer clonal relatedness from *fbl* repeat regions, like the Based Upon Repeat Pattern
288 (BURP) developed for *spa* typing of *S. aureus*⁴². The comparison of *fbl* typing with
289 previously obtained MLST and TRST typing data revealed that *fbl* typing was highly
290 discriminant compared to MLST, both in terms of genotypes and discriminatory index.
291 Especially, *fbl* typing revealed a considerable degree of *fbl* gene repeat polymorphism within
292 particular clonal groups defined by MLST, like CC1. Moreover, we have shown that there
293 was a fairly good correlation between clustering of isolates obtained by *fbl* typing and that

294 obtained by MLST. As *fbl* types were shown to predict MLST CCs at 100%, this allowed us
295 to provide a preliminary guideline for assigning CC from *fbl* type. Typing data resulting from
296 future studies may be used to enlarge our knowledge of *fbl*-MLST mappings, which will be
297 extremely useful for the daily routine typing of *S. lugdunensis*. Notably, this can be
298 particularly useful to detect putative clonal dissemination of methicillin resistant clones, like
299 described in a tertiary medical center in Taiwan⁴³.

300 The fact that *fbl* typing can group isolates in congruence with MLST indicates a non
301 recombinogenic nature of the *fbl* locus in a clonal background. This may account for the
302 observation that, while variable enough to provide adequate strain discrimination, this region
303 has the stability to group related strains for use as a typing tool. Analysis revealed that repeat
304 composition and organization, rather than the number of repeats, allowed *fbl* typing to
305 correlate with the MLST data. As described for *spa* typing of *S. aureus*, primary assignment
306 of lineages is *via* nucleotide mutation, followed by intralocus variation *via* addition and
307 deletion of repeats, suggesting that point mutations occur at a much lower rate than repeat
308 number variation²⁹. Therefore, *fbl* typing appears as a technique that is capable of
309 simultaneously indexing genetic variations that accumulate both rapidly and slowly by two
310 independent mechanisms, enabling it to be used in both long- and short-term epidemiologic
311 questions and population-based studies. On the other hand, *fbl* typing was less able to predict
312 the ST, as different STs could be included within a *fbl* type. Those ambiguities might be due
313 to the parallel evolution of strains originally belonging to the same clonal lineage. The lack of
314 discrimination following from this convergent evolution of different lineages can be
315 compensated by additional lineage-specific markers. *fbl* typing had the same discriminatory
316 power as TRST, while analyzing a single VNTR. This is probably because *fbl* repeats are
317 more prone to duplication and deletion via slipped-strand mispairing because they are smaller
318 (18 bp repeats) than those used for TRST (50 bp on average)²⁰.

319 The main advantage of *fbl* typing is that adequate typing information is obtained from a single
320 locus, as opposed to MLST and TRST, which require the combination of allelic information
321 from seven loci. The use of a single-locus marker is inherently less expensive, less time-
322 consuming, and less error prone than multilocus techniques. Moreover, another advantage of
323 *fbl* typing is the unambiguity and portability of the sequence data obtained. We have
324 developed an Internet Web site which allows to assign online *fbl* types from the sequence
325 data, and to download the *fbl* sequences available. This greatly simplifies the sharing of
326 information between laboratories and facilitates the creation of a large-scale database for the

327 study of global (such as the study of internationally circulating clones) as well as local (such
328 as transmission from patient to patient) epidemiology. It is also important to note that *fbl*
329 typing has *in vitro* stability. Furthermore, the finding of the same *fbl* types for isolates from
330 each clonal lineage collected in different years and from rather dispersed geographical areas
331 suggests a good *in vivo* stability of the *fbl* locus as a molecular marker in epidemiological
332 investigations.

333 Some limitations of *fbl* typing can however be evoked. First, *fbl* typing may technically be
334 more difficult than *spa* typing, even though sequencing technology is improving, because of
335 the large average of the amplified region size (1180 bp), compared to that of *spa* (556 bp)²².
336 Moreover, for some clonal lineages, like CC1 and CC3, we could demonstrate that although
337 they encompass a large variety of different *fbl* types, only a few of them seem to have the
338 ability to spread efficiently within hospitals . For example, *fbl47b* was the most predominant
339 *fbl* type at Strasbourg University Hospital ($n = 13/98 - 13\%$) as well as at Rouen University
340 Hospital ($n = 29/125 - 23\%$), over different study periods. This was also demonstrated in a
341 previous study by alternative typing methods as TRST²⁰. The same phenomenon was
342 described for *S. aureus*, for which endemic spread of highly successful *spa* types may finally
343 lead to a lack of discrimination in local hospital epidemiology³⁰. However, the rapidly
344 increasing number of clinical bacterial isolates being whole genome sequenced will certainly
345 overcome these limitations. A study has shown that *spa* typing by whole-genome sequencing
346 (WGS) can reliably replace Sanger sequencing. Indeed, despite series of repeats that can be
347 missed by the WGS assembly, a virtually total agreement was found between the two
348 methods⁴⁴. Thus, future *fbl* types of worldwide *S. lugdunensis* isolates will be certainly
349 identified by *in silico* analysis, with the advantage that other genes of interest can easily be
350 analyzed, including genes used for typing like MLST and TRST genes, alternative
351 polymorphic regions, virulence genes and resistance genes. These additional markers could be
352 considered to improve discrimination of predominant *fbl* types.

353 Since Fbl has been shown to be the major Fg-binding protein of *S. lugdunensis*⁴⁵, we wanted
354 to determine whether the genetic variability of *fbl* observed was related to their ability to bind
355 to fibrinogen. It has been previously described that some strains of *S. lugdunensis* could be
356 misidentified as *S. aureus* using agglutination-based kits that usually include antibodies
357 against fibrinogen binding proteins⁴⁶. Thus, all isolates were first tested by latex agglutination
358 assay. A positive clumping reaction was observed for 32% of the isolates, which was
359 comparable to that obtained by Szabados *et al.* (29.3%)¹⁴; but inferior to some previous

360 studies that reported a positivity rate of 64.7%⁴⁷ and 83.7%⁴⁸. As previously shown, these
361 differences can be explained by using different commercial assays⁴⁹. Furthermore, although
362 Fbl is probably the only fibrinogen-binding surface protein of *S. lugdunensis*⁴⁵, we cannot
363 judge if other fibrinogen binding proteins than Fbl could react in these assays. Indeed, other
364 genes encoding proteins potentially involved in Fg-binding have been described, like a
365 putative fibrinogen/fibronectin binding adhesin which is homologous to FbpA of *S.*
366 *aureus*^{10,14,18}, AtlL⁵⁰, and SlsD that has putative structural similarity to the Fg-binding domain
367 of SdrG of *S. epidermidis*¹⁰. The ability of 55 isolates to promote adherence to immobilized
368 human fibrinogen was then investigated. Isolates tested were chosen to be representative of
369 the different phylogenetic positions. In contrast to the high prevalence of *fbl* gene (100%), the
370 fibrinogen binding was only observed for 31% of our isolates. Similar results have been
371 described for *S. aureus* for which fibrinogen binding greatly varies within the strains⁵¹. These
372 differences in fibrinogen binding may be explained by several factors.

373 Our first hypothesis comes from the observation that some isolates revealed a key genetic
374 difference in the 5'-end of the *fbl* gene. A deletion of 4 bp was observed, resulting in a
375 frameshift mutation which potentially radically truncates the Fbl protein from 881 to 21
376 amino acids. Frameshift mutations tend to occur at repetitive DNA sequences when a
377 misalignment produces either addition or deletion of nucleotides⁵². Here, the deletion was
378 possibly linked to the repeat motif AAAG. We showed that isolates having this deletion
379 bound significantly less strongly to fibrinogen than the group of isolates that didn't have the
380 deletion ($P < 0.001$). To confirm this results, the expression of Fbl at the protein level should
381 be performed by Western-blot analysis from both bacterial cells and culture supernatants. We
382 actually assume that for isolates with the deletion no Fbl protein will be detected, neither in
383 the bacterial pellet nor in the supernatant. Furthermore, this deletion appeared interestingly
384 highly CC-dependent, since it was observed only within CC1 except for one isolate. Most of
385 the isolates belonging to ST6, which is the predicted founder of CC1, presented the deletion.
386 Of note, the deletion was not found for isolates belonging to ST1 and ST7, whereas ST12 and
387 ST15 clones were uniformly associated with the deletion, all four STs evolving from ST6.
388 Therefore, we can assume that this particular event might have arisen in some ST6 strains and
389 has been "transmitted" to ST12 and ST15, constituting a special feature of these STs. This
390 finding has to be confirmed on a larger collection of isolates.

391 The second explanation is that the copy number of repeats in *fbl* R-region would affect the
392 Fg-binding capacity. This region is thought to span the thick peptidoglycan layer of the cell

393 wall and to allow exposure of the ligand-binding domain at the cell surface by analogy with
394 other Sdr proteins³³. It is known that a certain number of SD repeats are required for
395 functional expression of the ligand-binding domain of ClfA on the cell surface³³. By analogy,
396 it is possible that the length of the repeat domain within *fbl* is critical for protein function. Our
397 findings show that isolates displaying fewer than 43 repeats were significantly weakly
398 adherent to fibrinogen. To confirm this trend and determine the minimum length that would
399 be required to span the entire cell wall and to display the Fg-binding domain in a form that
400 can participate fully in fibrinogen binding, construction of variants of Fbl with truncated R-
401 domains would be necessary.

402 However, some Fg-binding-negative isolates carried more than 43 repeats and didn't have the
403 deletion. Other possible reasons for these results might include that in some isolates, Fbl
404 protein might not be expressed sufficiently, due to loss or inadequate expression of *fbl* gene.
405 To test this hypothesis, we performed qRT-PCR for Fg-binding negative and Fg-binding
406 positive isolates. Differences in *fbl* expression were observed between individual isolates
407 tested but no significant difference in *fbl* transcript levels between the two groups of isolates
408 was observed. Thus, despite the expression of *fbl*, some isolates didn't bind fibrinogen.
409 Therefore, the last hypothesis is that the production of an extracellular matrix such as a
410 capsule could mask the accessibility of the surface protein³. This has been described for *S.*
411 *aureus*, for which capsular polysaccharide expression inhibits *S. aureus* ClfA-mediated
412 binding to fibrinogen and platelets, and a full length repeat region cannot overcome this
413 inhibition³⁵.

414 The genetic diversity of the *fbl* gene could make it an important target for adaptive evolution
415 by means of host specialization and other environmental factors. Thus, it could influence the
416 pathogenicity of *S. lugdunensis*, by increasing fitness to colonize the host surfaces or inducing
417 infective endocarditis. Our observations suggest that *fbl* type was not predictive of clinical
418 context. It was demonstrated previously that all lineages of *S. lugdunensis* were equally likely
419 to cause an infection^{17,18,20}. This suggested that most, if not all, natural isolates of *S.*
420 *lugdunensis* are capable of infection and that human host factors are probably crucial for *S.*
421 *lugdunensis* infection. More surprisingly, although a wide range of Fg adherence phenotypes
422 was observed for all types of *S. lugdunensis* isolates, no correlation with clinical context was
423 observed. Notably, an infectious endocarditis isolate did not display Fg adhesion while a
424 carrier isolate exhibiting very strong one. In *S. aureus*, experiments demonstrated that ClfA
425 was essential to promote valve colonization in rats with experimental endocarditis as well as

426 septic arthritis in mice^{53,54}. On the contrary, it has also been reported that the fibrinogen and
427 fibronectin binding ability of *S. aureus* strains does not predict the outcome in a rat model of
428 infective endocarditis⁵¹. Authors explained this apparent paradox by possible differential gene
429 expression *in vivo*, human host factors or adhesin redundancy. In a previous study, it was
430 demonstrated that a *S. lugdunensis* mutant defective in the surface protein Fbl was not
431 significantly less virulent than the wild-type in a rat endocarditis model, although a trend
432 toward reduced virulence was observed⁵⁵. However, only one clinical strain, that caused a
433 milder course of disease than a *S. aureus* strain, was tested. Thus, it would be interesting to
434 compare the virulence of several other clinical isolates of *S. lugdunensis* with their *fbl* mutants
435 in animal models of endocarditis, but also in murine nasal colonization assays to study the
436 implication of *fbl* in adhesion to host surfaces.

437

438 **Conclusion**

439 In conclusion, we demonstrated that the repeat region of *fbl* is polymorphic enough to provide
440 a useful strain resolution. *fbl* typing is highly discriminant compared to MLST and equivalent
441 to TRST even though a single locus is analyzed. Predicting CCs with high accuracy, *fbl*
442 typing can serve as a frontline tool for typing *S. lugdunensis* strains in both basic and applied
443 research. The development of a Web tool publicly available to identify *fbl* types yields
444 interchangeable information that may be used both for local epidemiology as well as
445 international multicenter surveillance of *S. lugdunensis* lineages. Moreover, we showed that *S.*
446 *lugdunensis* isolates displayed natural variability in adherence to immobilized human
447 fibrinogen, which could be partly explained by some genetic variations of the *fbl* gene. It still
448 remains to be determined whether these differences influence the expression of Fbl protein on
449 the cell surface and have an impact on the colonization of the host tissues and the subsequent
450 development of infection.

451

452

453 **Materials and methods**

454 ***Bacterial isolates***

455 A total of 240 *S. lugdunensis* human isolates were used in this study (see Table S1 in the
456 supplemental material). A first panel, called “TRST panel”, included 128 *S. lugdunensis*
457 isolates that have previously been characterized by MLST and TRST, collected from five
458 French regions and Sweden²⁰. Among them, 5 pairs of isolates were collected from patients at
459 time intervals ranging from 0 to 17 days. The precise description of this collection of strains
460 was published previously²⁰. A further set of 106 clinical isolates recovered from 101 patients
461 were collected at the University Hospital of Rouen in France from January 2016 to December
462 2016. In addition, six clinical strains whose complete genome sequence was available in our
463 laboratory, were included in this study. Overall, the 240 isolates were recovered from skin and
464 soft tissue infections ($n = 50$), bone and joint infections ($n = 41$), deep infections ($n = 22$),
465 bacteremia ($n = 12$), endocarditis ($n = 9$), medical device infections ($n = 8$), other infections
466 ($n = 12$), carriage ($n = 79$) and from undocumented origin ($n = 7$). All the 240 isolates were
467 identified by matrix-assisted laser desorption/ionization time-of-flight mass.

468

469 ***DNA extraction and PCR amplification***

470 Isolates (stored at -80°C before use) were grown overnight at 37°C on tryptic soy agar (TSA)
471 with 5% horse blood agar plate. DNA was extracted using the InstaGene Matrix kit (Bio-Rad,
472 Marnes la Coquette, France) according to the manufacturer’s recommendations. PCRs were
473 performed on a Veriti Thermal Cycler (Applied Biosystems, Foster City, CA, USA) in a final
474 volume of 25 µl containing 12.5 µl GoTaq G2 Green Master Mix (Promega, Charbonnières-
475 Les-Bains, France), 0.50 µM each primer and 5 µl of extracted DNA.

476

477 ***PCR amplification and DNA sequencing of the fbl gene***

478 *The fbl* gene of the 15 genomes of *S. lugdunensis* available in the GenBank database at the
479 time of the study and of six complete genomes sequenced by our laboratory was aligned by
480 using clustal multiple sequence alignment (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).
481 Primers fbl_R_F and fbl_R_R were designed for the *fbl* typing assay development to flank the
482 *fbl* R-domain (Table 1 and Fig. 1a). Primers fbl_N_F and fbl_N_R were designed in the 5’
483 part of the gene to amplify and sequence a 264 bp fragment corresponding to the region
484 upstream the *fbl* gene and its first 125 bp (Table 1). All primers were designed using the
485 OLIGO Primer Design software (Molecular Biology Insights, USA). Sequences were BLAST

486 checked for homologous regions). The conditions that allowed amplification of single PCR
487 products were as follows: denaturation at 95°C for 5min, followed by 35 cycles of
488 amplification including denaturation at 95°C for 30s, annealing at 55°C for 30s, and extension
489 at 72°C for 30s. A final extension step at 72°C for 5min was performed. All PCR products
490 were separated on 1% agarose gel by electrophoresis, purified, and sequenced in both forward
491 and reverse directions by GATC Biotech SARL (Konstanz, Germany).

492

493 ***Identification of *fbl* types and cluster analysis***

494 The number as well as the sequences of *fbl* repeats were determined with BioNumerics
495 software (Version 7.6, Applied Maths, Saint-Martens-Latem, Belgium) by using the
496 "polymorphic VNTR typing" plugin. To verify that the entire repeat array has been
497 sequenced, the flanking sequences were spotted but were not included as typing information.
498 Each unique 18-bp repeat sequence was identified by a numeric code, called "repeat code".
499 Each unique combination of repeat sequences defined an "*fbl* type", named by a number
500 followed by a lowercase letter, where the number corresponds to the number of repeats and
501 the letter to the combination of the different repeat sequences, e.g. fbl42a. In this way, each
502 *fbl* type denotes a collection of specific repeat units arranged in a precise pattern.

503 Cluster analysis of *fbl* sequences was performed using the "polymorphic VNTR typing"
504 plugin of BioNumerics software (Version 7.6, Applied Maths, Saint-Martens-Latem,
505 Belgium). With this plugin, sequences are compared and aligned using an algorithm based on
506 the DSI (duplication, substitution, and indels) model for pairwise alignment of repeats, which
507 considers that modification of sequences can occur through duplication of tandem repeats,
508 substitutions, insertions, and deletions (the latter two events are collectively termed indels)⁵⁶.
509 A minimum spanning tree was generated from the similarity matrix with the root node
510 assigned to the sequence type with the greatest number of related types. The default software
511 parameters were used for analysis with a bin distance of 1%, i.e., two entries having a
512 similarity of 99% to 100% will have a distance of 0 and two entries that have a similarity of
513 98% to 99% will have a distance of 1. For cluster analysis, only *fbl* types separated by a
514 minimum spanning tree (MST) distance of ≤ 2 (i.e., if they were $\geq 97\%$ similar) were
515 considered closely related and assigned to the same cluster.

516

517 ***Data analysis***

518 The discriminatory ability as well as the congruence of the results obtained by using the
519 different typing methods were evaluated. To this aim, both the Simpson's diversity index
520 (DI)⁵⁷, and adjusted Wallace coefficients (AW)⁵⁸, (plus confidence intervals [CI] as described
521 by Grundmann *et al.*⁵⁹) were calculated by resorting to the online tool at
522 <http://www.comparingpartitions.info/?link=Tool>. If the CI of any two methods overlaps, one
523 cannot exclude the hypothesis that they have similar discriminatory powers at a 95%
524 confidence level. The ratio of nonsynonymous to synonymous substitutions (d_N/d_S) was
525 calculated using the MEGA version 7 software by using the Maximum Likelihood analysis of
526 natural selection codon-by-codon.

527

528 ***Test of fbl stability***

529 To determine the stability of the *fbl* R-domain, two isolates were subcultured on TSA for 30
530 consecutive days by streaking a single colony from each strain on agar plates. The original
531 culture, subculture number 10, 20 and 30 were used for DNA extraction and the total DNA
532 was subjected to the *fbl* assay.

533

534 ***Construction and validation of “fbl typing Server”***

535 A *fbl* allele database was created to contain all identified *fbl* types. The database was
536 constructed as a single FASTA file. Perfect identity hits (100% of identity) report the
537 corresponding *fbl* type whereas imperfect hits report an “unknown” *fbl* type, and then the user
538 is encouraged to contact the curator of *fbl* typing Server for updating the database with this
539 new variant. The Web tool has been made publicly available and is hosted by the Rouen
540 University (<http://fbl-typing.univ-rouen.fr/>). Individual FASTA assemblies of the paired
541 Sanger sequences of 92 isolates identified with BioNumerics software (Version 7.6, Applied
542 Maths, Saint-Martens-Latem, Belgium) were analyzed with this newly developed Web tool,
543 and the output results were compared.

544

545 ***Rapid slide latex agglutination tests***

546 A rapid latex agglutination test was performed on all isolates using the kit Pastorex™ Staph
547 Plus (Bio-Rad, Marnes la Coquette, France) according to the manufacturers' instructions.

548

549 ***Adherence of bacterial cells to immobilized fibrinogen***

550 Bacterial adhesion to solid phase-adsorbed human fibrinogen was assessed in a 96-well plate
551 format. Flat-bottom microtiter plates were coated for 1h at 37°C with 50 µl per well of 10

552 $\mu\text{g/ml}$ human fibrinogen or phosphate-buffered saline (PBS). The plates were washed three
553 times with PBS and incubated at 37°C for 1h after addition of $250\ \mu\text{l}$ of 2% human serum
554 albumin (HSA). The plates were washed three times with PBS prior the inoculation bacterial
555 suspension prepared as follows. Bacterial isolates were inoculated onto Mueller Hinton broth
556 and incubated overnight at 37°C under constant agitation. A volume of $100\ \mu\text{l}$ of suspension
557 was inoculated in fresh broth (10 ml) and incubated at 37°C for 3h. Bacteria were washed
558 twice with PBS and adjusted to an $\text{OD}_{600\text{nm}} = 0.9$ UA. Fifty microliters of adjusted bacterial
559 suspension were added per well and after 1h30 of incubation at 37°C , the wells were rinsed
560 three times with PBS. Adherent bacteria were fixed for 30min at 60°C , stained with $95\ \mu\text{l}$ of
561 0.5% crystal violet in 0.5% ethanol for 15 min and air-dried after washings. Crystal violet
562 staining was solubilized into $100\ \mu\text{l}$ of dimethyl sulfoxide (DMSO) and the absorbance
563 measured at $\text{OD}_{595\text{nm}}$ using a microplate reader. Adherence of strains was evaluated in at least
564 five wells, and the experiment was performed at least two independent times. *S. aureus*
565 DU5925 was used as a negative control (non-adherent); *S. aureus* 8325-4 and 8325-4 pCF4
566 were used as positive ones (adherent and hyper-adherent, respectively). Adherence of strains
567 was expressed relative to that of 8325-4 (100%). All results were expressed as mean
568 percentages \pm standard error of mean (SEM). After suppression of significant outliers
569 ($P < 0.05$, online tool: <https://www.graphpad.com/quickcalcs/grubbs2/>), adherence of strains
570 was compared to that of DU5925 using Dunnett's multiple comparisons test, and the global
571 difference was analyzed using ordinary one-way ANOVA (GraphPad Prism statistical
572 analysis). The impact of *fbl* genetic diversity (presence of the 5'-end frameshift mutation or
573 number of repeats in the R-region) on the adherence was evaluated by using a nonparametric
574 analysis of variance test (Kruskal-Wallis) on R software (version 3.5.1). *P* value of < 0.05 was
575 regarded as significant.

576

577 **Total RNA extraction**

578 Fresh trypticase soy broth (TSB) was inoculated with an overnight culture to an initial
579 $\text{OD}_{600\text{nm}}$ of 0.01 and grown at 37°C in a shaking incubator set at 150 rpm. Cells were grown
580 to mid-exponential phase ($\text{OD}_{600\text{nm}} = 1$), and 1 ml was centrifuged (5min at $8,000\times g$, 4°C).
581 Supernatants were carefully removed and the pellets were resuspended in $200\ \mu\text{l}$ TE Buffer
582 (10 mM Tris-HCl pH 8 and 1 mM EDTA) containing 2,1 mg lysozyme (Sigma-Aldrich, St.
583 Louis, MO, USA) and $10\ \mu\text{g}$ lysostaphin (Sigma-Aldrich, St. Louis, MO, USA). Samples
584 were vortexed for 30s, incubated at 37°C for 10min, and lysed using the RA1 lysis buffer
585 provided in the NucleoSpin® RNA kit (Macherey Nagel, Hoerdt, France) supplemented with

586 β -mercaptoethanol (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's
587 instructions. Next, a bead-beating procedure alternating 1-min cycles of beating with
588 incubation on ice was performed. RNA were isolated using the NucleoSpin® RNA kit
589 (Macherey Nagel, Hoerd, France) according to the manufacturer's instructions. The amount
590 of RNA yielded was assessed with a NanoDrop spectrophotometer (Thermofisher Scientific,
591 Montigny le Bretonne, France). RNA samples were then treated by TURBO DNA-*free*TM kit
592 (Ambion, Austin, TX, USA) according to the manufacturer's instructions.

593

594 ***Quantitative real-time PCR***

595 The cDNA synthesis was performed using the Omniscript® RT kit (Qiagen, Hilden,
596 Germany) according to the manufacturer's instructions. A total of 2 μ g of RNA from
597 respective samples was reverse transcribed in a 20 μ L volume reaction using random primers
598 (Promega, Madison, USA) and RNase inhibitor (40 U/ μ l) (Invitrogen, Carlsbad, USA),
599 according to the manufacturer's instructions. Prior to use on Real-time PCR assays, the
600 synthesized cDNA samples were diluted with nuclease-free water (1:10 and 1:100 dilutions).
601 The quantitative RT-PCR (qRT-PCR) primers listed in Table 1 were used in 20 μ l PCR
602 mixtures which included SYBR® Green PCR Master Mix (Bio-Rad, Marnes la Coquette,
603 France) and 1 μ l of diluted cDNA template (1:10 for *fbl* and 1:100 for 16S rRNA). A CFX96
604 real-time PCR detection system (Bio-Rad, Marnes la Coquette, France) was used for qRT-
605 PCR with the following PCR amplification program: initial denaturation step at 95°C for
606 10min, followed by 40 cycles of amplification including denaturation at 95°C for 30s,
607 annealing at 55°C for 30s, and extension at 72°C for 30s. Melting curve analysis was used to
608 identify specific products increasing in 0.5°C increment every 5s from 55 to 95°C. Relative
609 *fbl* expression levels for each isolate were quantified using the gene expression analysis
610 module of the CFX ManagerTM software, with 16S rRNA gene as reference to normalize the
611 results. All *S. lugdunensis* isolates were assayed as triplicates in each experimental run, and
612 three independent biological assays were performed. Box plots were drawn using R software
613 (version 3.5.1) to compare the gene expression ratios of the groups of Fg-binding-negative (-)
614 and Fg-binding-positive (+++) *S. lugdunensis* strains. The statistical significance of *fbl* gene
615 expression differences between the two groups was determined based on Mann-Whitney-
616 Wilcoxon measurements. *P* value of <0.05 was regarded as significant.

617 **Acknowledgements**

618 This study was supported by an educational grant for Jérémie Lebeurre from the Charles
619 Nicolle Foundation. The authors would like to thank the “Service Informatique de l’UFR
620 Santé” of Rouen University (Florence Charles and Raphaël Chauvin) and Amandine Paulay
621 for development of the Web tool; and Agnès Cottalorda for statistical analyses.

622

623

624 **Conflict of interests**

625 Competing financial interests: The authors declare that they have no competing interests.

626

627

628 **Author contributions statement**

629 SD, JL, PF and MPC designed the study. XA and GP provided the VISLISI clinical isolates.
630 SD, JL, CD, FL performed the experiments. SD and JL analyzed the data. SD, JL and MPC
631 wrote the manuscript. All authors have read and approved the final version of the manuscript.

632

633

634 Supplementary information accompanies the manuscript on the Emerging Microbes &
635 Infections website <http://www.nature.com/emi>

636

637 **References**

- 638 1. Freney, J. *et al.* *Staphylococcus lugdunensis* sp. nov. and *Staphylococcus schleiferi* sp. nov., two
639 species from human clinical specimens. *Int. J. Syst. Evol. Microbiol.* **38**, 168–172 (1988).
- 640 2. Bieber, L. & Kahlmeter, G. *Staphylococcus lugdunensis* in several niches of the normal skin
641 flora. *Clin. Microbiol. Infect.* **16**, 385–388 (2010).
- 642 3. Frank, K. L., del Pozo, J. L. & Patel, R. From Clinical Microbiology to Infection Pathogenesis:
643 How Daring To Be Different Works for *Staphylococcus lugdunensis*. *Clin. Microbiol. Rev.* **21**,
644 111–133 (2008).
- 645 4. Zinkernagel, A. S. *et al.* Significance of *Staphylococcus lugdunensis* Bacteremia: Report of 28
646 Cases and Review of the Literature. *Infection* **36**, 314–321 (2008).
- 647 5. Heldt Manica, L. A. & Cohen, P. R. *Staphylococcus lugdunensis* Infections of the Skin and Soft
648 Tissue: A Case Series and Review. *Dermatol. Ther.* **7**, 555–562 (2017).
- 649 6. Papapetropoulos, N., Papapetropoulou, M. & Vantarakis, A. Abscesses and wound infections
650 due to *Staphylococcus lugdunensis*: report of 16 cases. *Infection* **41**, 525–528 (2013).
- 651 7. Argemi, X. *et al.* Kinetics of biofilm formation by *Staphylococcus lugdunensis* strains in bone
652 and joint infections. *Diagn. Microbiol. Infect. Dis.* **88**, 298–304 (2017).
- 653 8. Non, L. R. & Santos, C. A. Q. The occurrence of infective endocarditis with *Staphylococcus*
654 *lugdunensis* bacteremia: A retrospective cohort study and systematic review. *J. Infect.* **74**, 179–
655 186 (2017).
- 656 9. Anguera, I. *Staphylococcus lugdunensis* infective endocarditis: description of 10 cases and
657 analysis of native valve, prosthetic valve, and pacemaker lead endocarditis clinical profiles.
658 *Heart* **91**, e10–e10 (2005).
- 659 10. Heilbronner, S. *et al.* Genome sequence of *Staphylococcus lugdunensis* N920143 allows
660 identification of putative colonization and virulence factors: *Staphylococcus lugdunensis*
661 genome sequence. *FEMS Microbiol. Lett.* **322**, 60–67 (2011).
- 662 11. Mitchell, J. Characterization of the fibrinogen-binding surface protein Fbl of *Staphylococcus*
663 *lugdunensis*. *Microbiology* **150**, 3831–3841 (2004).
- 664 12. Nilsson, M. A von Willebrand factor-binding protein from *Staphylococcus lugdunensis*. *FEMS*
665 *Microbiol. Lett.* **234**, 155–161 (2004).
- 666 13. Pereira, E. M., Oliveira, F. L. F., Schuenck, R. P., Zoletti, G. O. & dos Santos, K. R. N.
667 Detection of *Staphylococcus lugdunensis* by a new species-specific PCR based on the *fbl* gene.
668 *FEMS Immunol. Med. Microbiol.* **58**, 295–298 (2010).
- 669 14. Szabados, F. *et al.* Occurrence of genes of putative fibrinogen binding proteins and hemolysins,
670 as well as of their phenotypic correlates in isolates of *S. lugdunensis* of different origins. *BMC*
671 *Res. Notes* **4**, 113 (2011).
- 672 15. Heilbronner, S. *et al.* Competing for Iron: Duplication and Amplification of the *isd* Locus in
673 *Staphylococcus lugdunensis* HKU09-01 Provides a Competitive Advantage to Overcome
674 Nutritional Limitation. *PLOS Genet.* **12**, e1006246 (2016).
- 675 16. Yeh, C.-F. *et al.* Molecular Characteristics of Disease-Causing and Commensal *Staphylococcus*
676 *lugdunensis* Isolates from 2003 to 2013 at a Tertiary Hospital in Taiwan. *PLOS ONE* **10**,
677 e0134859 (2015).

- 678 17. Chassain, B. *et al.* Multilocus Sequence Typing Analysis of *Staphylococcus lugdunensis*
679 Implies a Clonal Population Structure. *J. Clin. Microbiol.* **50**, 3003–3009 (2012).
- 680 18. Didi, J., Lemee, L., Gibert, L., Pons, J.-L. & Pestel-Caron, M. Multi-Virulence-Locus Sequence
681 Typing of *Staphylococcus lugdunensis* Generates Results Consistent with a Clonal Population
682 Structure and Is Reliable for Epidemiological Typing. *J. Clin. Microbiol.* **52**, 3624–3632
683 (2014).
- 684 19. Van Belkum, A. *et al.* Guidelines for the validation and application of typing methods for use in
685 bacterial epidemiology. *Clin. Microbiol. Infect.* **13**, 1–46 (2007).
- 686 20. Dahyot, S. *et al.* Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA) and
687 Tandem Repeat Sequence Typing (TRST), helpful tools for subtyping *Staphylococcus*
688 *lugdunensis*. *Sci. Rep.* **8**, (2018).
- 689 21. ZaiSS, N. H. *et al.* Typing *Clostridium difficile* strains based on tandem repeat sequences. *BMC*
690 *Microbiol.* **9**, 6 (2009).
- 691 22. Shopsin, B. *et al.* Evaluation of protein A gene polymorphic region DNA sequencing for typing
692 of *Staphylococcus aureus* strains. *J. Clin. Microbiol.* **37**, 3556–3563 (1999).
- 693 23. Beall, B., Facklam, R. & Thompson, T. Sequencing *emm*-specific PCR products for routine and
694 accurate typing of group A streptococci. *J. Clin. Microbiol.* **34**, 953–958 (1996).
- 695 24. Edlind, T. & Liu, Y. Development and Evaluation of a Commercial Sequence-Based Strain
696 Typing Service for *Listeria monocytogenes*. *J. Microb. Biochem. Technol.* **07**, (2015).
- 697 25. Asadollahi, P. *et al.* Distribution of the Most Prevalent *Spa* Types among Clinical Isolates of
698 Methicillin-Resistant and -Susceptible *Staphylococcus aureus* around the World: A Review.
699 *Front. Microbiol.* **9**, (2018).
- 700 26. Frenay, H. M. E. *et al.* Molecular typing of methicillin-resistant *Staphylococcus aureus* on the
701 basis of protein A gene polymorphism. *Eur. J. Clin. Microbiol. Infect. Dis.* **15**, 60–64 (1996).
- 702 27. Harmsen, D. *et al.* Typing of Methicillin-Resistant *Staphylococcus aureus* in a University
703 Hospital Setting by Using Novel Software for *spa* Repeat Determination and Database
704 Management. *J. Clin. Microbiol.* **41**, 5442–5448 (2003).
- 705 28. Golding, G. R. *et al.* A preliminary guideline for the assignment of methicillin-resistant
706 *Staphylococcus aureus* to a Canadian pulsed-field gel electrophoresis epidemic type using *spa*
707 typing. *Can. J. Infect. Dis. Med. Microbiol.* **19**, 273–281 (2008).
- 708 29. Koreen, L. *et al.* *spa* Typing Method for Discriminating among *Staphylococcus aureus* Isolates:
709 Implications for Use of a Single Marker To Detect Genetic Micro- and Macrovariation. *J. Clin.*
710 *Microbiol.* **42**, 792–799 (2004).
- 711 30. Strommenger, B. *et al.* *spa* Typing of *Staphylococcus aureus* as a Frontline Tool in
712 Epidemiological Typing. *J. Clin. Microbiol.* **46**, 574–581 (2008).
- 713 31. Said, K. B., Ramotar, K., Zhu, G. & Zhao, X. Repeat-based subtyping and grouping of
714 *Staphylococcus aureus* from human infections and bovine mastitis using the R-domain of the
715 clumping factor A gene. *Diagn. Microbiol. Infect. Dis.* **63**, 24–37 (2009).
- 716 32. Koreen, L. *et al.* Comparative Sequencing of the Serine-Aspartate Repeat-Encoding Region of
717 the Clumping Factor B Gene (*clfB*) for Resolution within Clonal Groups of *Staphylococcus*
718 *aureus*. *J. Clin. Microbiol.* **43**, 3985–3994 (2005).

- 719 33. Hartford, O., Francois, P., Vaudaux, P. & Foster, T. J. The dipeptide repeat region of the
720 fibrinogen-binding protein (clumping factor) is required for functional expression of the
721 fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. *Mol. Microbiol.* **25**,
722 1065–1076 (1997).
- 723 34. McDevitt, D. & Foster, T. J. Variation in the size of the repeat region of the fibrinogen receptor
724 (clumping factor) of *Staphylococcus aureus* strains. *Microbiology* **141**, 937–943 (1995).
- 725 35. Risley, A. L., Loughman, A., Cywes-Bentley, C., Foster, T. J. & Lee, J. C. Capsular
726 Polysaccharide Masks Clumping Factor A–Mediated Adherence of *Staphylococcus aureus* to
727 Fibrinogen and Platelets. *J. Infect. Dis.* **196**, 919–927 (2007).
- 728 36. Nilsson, M., Bjerketorp, J., Guss, B. & Frykberg, L. A fibrinogen-binding protein of
729 *Staphylococcus lugdunensis*. *FEMS Microbiol. Lett.* **241**, 87–93 (2004).
- 730 37. Becker, K., Heilmann, C. & Peters, G. Coagulase-Negative Staphylococci. *Clin. Microbiol. Rev.*
731 **27**, 870–926 (2014).
- 732 38. McCrea, K. W. *et al.* The serine-aspartate repeat (Sdr) protein family in *Staphylococcus*
733 *epidermidis*. *Microbiology* **146**, 1535–1546 (2000).
- 734 39. Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: the
735 many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* **12**, 49–
736 62 (2014).
- 737 40. Bowden, M. G. Identification and preliminary characterization of cell-wall-anchored proteins of
738 *Staphylococcus epidermidis*. *Microbiology* **151**, 1453–1464 (2005).
- 739 41. Said, K. B. *et al.* Regional profiling for determination of genotype diversity of mastitis-specific
740 *Staphylococcus aureus* lineage in Canada by use of clumping factor A, pulsed-field gel
741 electrophoresis, and spa typing. *J. Clin. Microbiol.* **48**, 375–386 (2010).
- 742 42. Mellmann, A. *et al.* Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-
743 term evolution of *Staphylococcus aureus* populations based on spa polymorphisms. *BMC*
744 *Microbiol.* **7**, 98 (2007).
- 745 43. Cheng, C. *et al.* Persistence of a major endemic clone of oxacillin-resistant *Staphylococcus*
746 *lugdunensis* sequence type 6 at a tertiary medical centre in northern Taiwan. *Int. J. Infect. Dis.*
747 **36**, 72–77 (2015).
- 748 44. Bartels, M. D. *et al.* Comparing Whole-Genome Sequencing with Sanger Sequencing for spa
749 Typing of Methicillin-Resistant *Staphylococcus aureus*. *J. Clin. Microbiol.* **52**, 4305–4308
750 (2014).
- 751 45. Marlinghaus, L. *et al.* Construction and characterization of three knockout mutants of the *fbl*
752 gene in *Staphylococcus lugdunensis*: CHARACTERIZATION OF ISOGENIC MUTANTS OF
753 FBL. *APMIS* **120**, 108–116 (2012).
- 754 46. Shin, J. H. *et al.* Prevalence, identification, and antimicrobial susceptibility of *Staphylococcus*
755 *lugdunensis* from various clinical specimens in Korea. *Jpn. J. Infect. Dis.* **60**, 312 (2007).
- 756 47. Mateo, M. *et al.* Genotypic versus phenotypic characterization, with respect to susceptibility and
757 identification, of 17 clinical isolates of *Staphylococcus lugdunensis*. *J. Antimicrob. Chemother.*
758 **56**, 287–291 (2005).
- 759 48. Iorio, N. L. P. *et al.* Simplified and Reliable Scheme for Species-Level Identification of
760 *Staphylococcus* Clinical Isolates. *J. Clin. Microbiol.* **45**, 2564–2569 (2007).

- 761 49. Personne, P. *et al.* Comparative performances of six agglutination kits assessed by using typical
762 and atypical strains of *Staphylococcus aureus*. *J. Clin. Microbiol.* **35**, 1138–1140 (1997).
- 763 50. Hussain, M., Steinbacher, T., Peters, G., Heilmann, C. & Becker, K. The adhesive properties of
764 the *Staphylococcus lugdunensis* multifunctional autolysin AtlL and its role in biofilm formation
765 and internalization. *Int. J. Med. Microbiol.* **305**, 129–139 (2015).
- 766 51. Ythier, M. *et al.* Natural variability of in vitro adherence to fibrinogen and fibronectin does not
767 correlate with in vivo infectivity of *Staphylococcus aureus*. *Infect. Immun.* **78**, 1711–1716
768 (2010).
- 769 52. Lovett, S. T. Encoded errors: mutations and rearrangements mediated by misalignment at
770 repetitive DNA sequences: Replication slippage. *Mol. Microbiol.* **52**, 1243–1253 (2004).
- 771 53. Que, Y.-A. *et al.* Fibrinogen and fibronectin binding cooperate for valve infection and invasion
772 in *Staphylococcus aureus* experimental endocarditis. *J. Exp. Med.* **201**, 1627–1635 (2005).
- 773 54. Josefsson, E., Hartford, O., O'Brien, L., Patti, J. M. & Foster, T. Protection against
774 experimental *Staphylococcus aureus* arthritis by vaccination with clumping factor A, a novel
775 virulence determinant. *J. Infect. Dis.* **184**, 1572–1580 (2001).
- 776 55. Heilbronner, S., Hanses, F., Monk, I. R., Speziale, P. & Foster, T. J. Sortase A Promotes
777 Virulence in Experimental *Staphylococcus lugdunensis* Endocarditis. *Microbiology* (2013).
778 doi:10.1099/mic.0.070292-0
- 779 56. Benson, G. Sequence alignment with tandem duplication. *J. Comput. Biol. J. Comput. Mol. Cell*
780 *Biol.* **4**, 351–367 (1997).
- 781 57. Hunter, P. R. & Gaston, M. A. Numerical index of the discriminatory ability of typing systems:
782 an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**, 2465–2466 (1988).
- 783 58. Severiano, A., Pinto, F. R., Ramirez, M. & Carriço, J. A. Adjusted Wallace Coefficient as a
784 Measure of Congruence between Typing Methods. *J. Clin. Microbiol.* **49**, 3997–4000 (2011).
- 785 59. Grundmann, H., Hori, S. & Tanner, G. Determining Confidence Intervals When Measuring
786 Genetic Diversity and the Discriminatory Abilities of Typing Methods for Microorganisms. *J.*
787 *Clin. Microbiol.* **39**, 4190–4192 (2001).
- 788 60. Tseng, C.-P. *et al.* Broad-range ribosomal RNA real-time PCR after removal of DNA from
789 reagents: melting profiles for clinically important bacteria. *Clin. Chem.* **49**, 306–309 (2003).
- 790
791
792

791 **Figure legends**

792

793 **Figure 1. Schematic representation of the Fbl protein and alignments of the 5'-end**

794 **region. a** The different regions indicated are S (signal peptide), N1-, N2-, and N3-domains,
795 R-repeat region and WMC (W, wall spanning; M, membrane spanning; C, cytoplasmic
796 positively charge tail). Conserved motifs are indicated at the top. The locations of the forward
797 and reverse primers used to amplify and sequence the R-region are shown. **b** Alignment of the
798 nucleotide sequence from the 5'-end of the *fbl* gene of isolates exhibiting the wild type
799 sequence (*fbl*_WT) and isolates with a 4 bp deletion (*fbl*_D). Numbering begins at the start
800 codon. Identical sequences are denoted by asterisks. The position of the 4 bp deletion is
801 shown with a red box. Alignment was obtained using the Clustal Omega tool
802 <https://www.ebi.ac.uk/Tools/msa/clustalo/>. **c** Wild type (*fbl*_WT) and truncated Fbl (*fbl*_D)
803 translation products of the signal peptide region of Fbl (amino acids 1-39). Beneath the
804 alignment, preserved amino acid are indicated with asterisks. The deletion causing the
805 frameshift mutation (E7fs) is indicated with a red box. The asterisk denotes the stop codon
806 (L22STOP).

807

808 **Figure 2. Minimum spanning tree analysis of 240 *S. lugdunensis* isolates based on *fbl***

809 **types.** Cluster analysis was performed using the polymorphic VNTR typing plugin of
810 BioNumerics. *fbl* types separated by a minimum spanning tree (MST) distance of ≤ 2 (i.e., if
811 they were $\geq 97\%$ similar) were considered closely related and assigned to the same cluster.
812 Each circle represents a *fbl* type and its size is proportional to the number of isolates. The
813 length of the branches expressed the MST distance between two *fbl* types. Gray zones around
814 circles delineate *fbl* clusters. The colors used are based **a** on clonal complexes (CCs) defined
815 by MLST ; undetermined = unknown CC (isolates non characterized by MLST) and **b** on
816 clinical contexts.

817

818 **Figure 3. Box plots showing the *fbl* relative gene expression levels in Fg-binding-negative**

819 **(-) ($n = 5$) and Fg-binding-positive (+++) ($n = 5$) *S. lugdunensis* isolates.** No significant
820 differences was detected in *fbl* relative gene expression levels between the two groups of *S.*
821 *lugdunensis* isolates.

822

823

Tables

Table 1. Primers used in this study

Target	Name of primer	Primer sequence (5'-3')	Reference
<i>fbl</i> R-domain amplification and sequencing by conventional PCR	fbl_R_F	TGGCATGGGATAATGAAGTAG	This study
	fbl_R_R	CCACTCCCGTATAGTAACT	
<i>fbl</i> Nter-domain amplification and sequencing by conventional PCR	fbl_N_F	TTGCCTGTATAATGCTATCTT	This study
	fbl_N_R	TTTGACAAATTACGCTCC	
<i>fbl</i> gene expression by qRT-PCR	fbl_qRT_F	AATAGCGAGGCACAAGCAATA	This study
	fbl_qRT_R	GCAGGTGATACATCGGTGG	
16S rRNA gene expression by qRT-PCR	Q3	GAGGAAGGIGIGGAIGACGT	60
	Q4	AGICCCGIGAACGTATTCAC	

Table 2. Discriminatory power of the three typing methods for the unrelated isolates.

Typing method	All isolates (<i>n</i> = 230)			TRST panel (<i>n</i> = 123)		
	No. of genotypes	DI ^a	CI ^b 95%	No. of genotypes	DI ^a	CI ^b 95%
<i>fbl</i> typing	92	0.946	0.929-0.964	60	0.964	0.949-0.979
MLST	ND ^c	ND ^c	ND ^c	25	0.899	0.872-0.926
TRST	ND ^c	ND ^c	ND ^c	69	0.943	0.915-0.971

^aDI: Simpson's diversity index. ^bCI: confidence interval. ^cND: not determined.

Table 3. Adjusted Wallace coefficients and 95% confidence interval for the 123 unrelated isolates typed by MLST, TRST and *fbl* typing.
Clonal complex (CC). Sequence type (ST).

	CC (MLST)	ST (MLST)	TRST type	<i>fbl</i> type
CC (MLST)		0.477 (0.398-0.555)	0.257 (0.176-0.339)	0.158 (0.089-0.226)
ST (MLST)	1.000 (1.000- 1.000)		0.487 (0.371-0.602)	0.225 (0.126-0.324)
TRST type	1.000 (1.000- 1.000)	0.901 (0.828-0.974)		0.215 (0.076-0.353)
<i>fbl</i> type	1.000 (1.000- 1.000)	0.680 (0.521-0.840)	0.351 (0.202-0.500)	

Table 4. Bacterial adherence to immobilized fibrinogen. The fibrinogen binding was performed in two independent experiments measured in at least five wells. Statistical differences compared to *S. aureus* DU5925 used Dunnett's multiple comparisons test ($P < 0.05$). Ordinary one-way ANOVA: $P < 0.0001$ ****. Clonal complex (CC). Not determined (ND). Skin and soft tissue infection (SSTI). Bone and joint infection (BJI).

Clinical isolate	<i>fbl</i> type	Premature stop codon in <i>fbl</i> gene	MLST CC	Binding to solid-phase fibrinogen	Agglutination test	Clinical source	City
SL_T29	43g	–	1	+++ ****	Positive	Material device infection	Rouen
SL_R30	45l	–	ND	+++ **	Positive	SSTI	Rouen
SL_V03	47b	–	3	+++ **	Positive	SSTI	Strasbourg
SL_V74	45b	–	2	+++ *	Positive	Bacteremia	Strasbourg
SL_T117	45f	–	1	+++ *	Positive	Carriage	Kronoberg
SL_V51	45h	–	1	+++ *	Positive	BJI	Strasbourg
SL_T118	40b	–	6	+	Negative	Carriage	Kronoberg
SL_T122	43g	–	1	+	Positive	Carriage	Tours
SL_V06	45c	–	2	+	Negative	Deep infection	Strasbourg
SL_R34	45f	–	ND	+	Positive	SSTI	Rouen
SL_V05	45g	–	1	+	Positive	Deep infection	Strasbourg
SL_V35	45i	–	1	+	Positive	Catheter related bacteremia	Strasbourg
SL_R44	47b	–	ND	+	Negative	Carriage	Rouen
SL_V34	52a	–	4	+	Positive	SSTI	Strasbourg
SL_V07	21a	L22STOP	1	+	Negative	Deep infection	Strasbourg
SL_V24	44a	L22STOP	1	+	Negative	BJI on material	Strasbourg
SL_C47	13a	–	3	-	Negative	Carriage	Strasbourg
SL_R59	13a	–	ND	-	Negative	BJI	Rouen
SL_R87	18a	–	ND	-	Negative	Carriage	Rouen
SL_V50	24a	–	3	-	Negative	BJI on material	Strasbourg
SL_R74	37a	–	ND	-	Negative	Carriage	Rouen
SL_DSM	41b	–	6	-	Positive	SSTI	Unknown
SL_T55	41b	–	6	-	Negative	SSTI	Nantes
SL_V70	41c	–	7	-	Negative	SSTI	Strasbourg
SL_C35	42a	–	3	-	Negative	Carriage	Strasbourg
SL_C72	42d	–	3	-	Negative	Carriage	Strasbourg
SL_R72	42f	–	ND	-	Negative	Deep infection	Rouen
SL_V78	43a	–	3	-	Negative	SSTI	Strasbourg
SL_R41	45b	–	ND	-	Negative	SSTI	Rouen
SL_V77	45d	–	7	-	Positive	Catheter related bacteremia	Strasbourg
SL_R29	45d	–	ND	-	Positive	Carriage	Rouen
SL_C42	45e	–	7	-	Positive	Carriage	Strasbourg
SL_V31	47a	–	3	-	Negative	Urinary infection	Strasbourg
SL_V10	47c	–	3	-	Negative	Material device infection	Strasbourg
SL_V21	47d	–	3	-	Negative	Catheter related bacteremia	Strasbourg

SL_V48	52a	_	4	-	Negative	Deep infection	Strasbourg
SL_R99	9a	_	ND	-	Negative	Carriage	Rouen
SL_T10	9a	_	Singleton	-	Negative	BJI	Rouen
SL_V81	30a	L22STOP	1	-	Negative	SSTI	Strasbourg
SL_V08	40c	L22STOP	1	-	Negative	SSTI	Strasbourg
SL_C60	43c	L22STOP	1	-	Negative	Carriage	Strasbourg
SL_V13	43c	L22STOP	1	-	Negative	BJI on material	Strasbourg
SL_V27	43c	L22STOP	1	-	Negative	BJI on material	Strasbourg
SL_V36	43c	L22STOP	1	-	Negative	BJI	Strasbourg
SL_V66	43c	L22STOP	1	-	Negative	BJI on material	Strasbourg
SL_T13	45f	L22STOP	1	-	Negative	Endocarditis	Rouen
SL_V09	45f	L22STOP	1	-	Negative	SSTI	Strasbourg
SL_V32	45f	L22STOP	1	-	Negative	SSTI	Strasbourg
SL_V64	45f	L22STOP	1	-	Negative	SSTI	Strasbourg
SL_V79	45f	L22STOP	1	-	Negative	Deep infection	Strasbourg
SL_R15	45k	L22STOP	ND	-	Negative	Carriage	Rouen
SL_V38	46a	L22STOP	1	-	Negative	BJI on material	Strasbourg
SL_C62	46b	L22STOP	1	-	Negative	Carriage	Strasbourg
SL_V62	46c	L22STOP	1	-	Negative	SSTI	Strasbourg
SL_V33	45a	L22STOP	1	-	Negative	Deep infection	Strasbourg

Figures

Figure 1



Figure 2

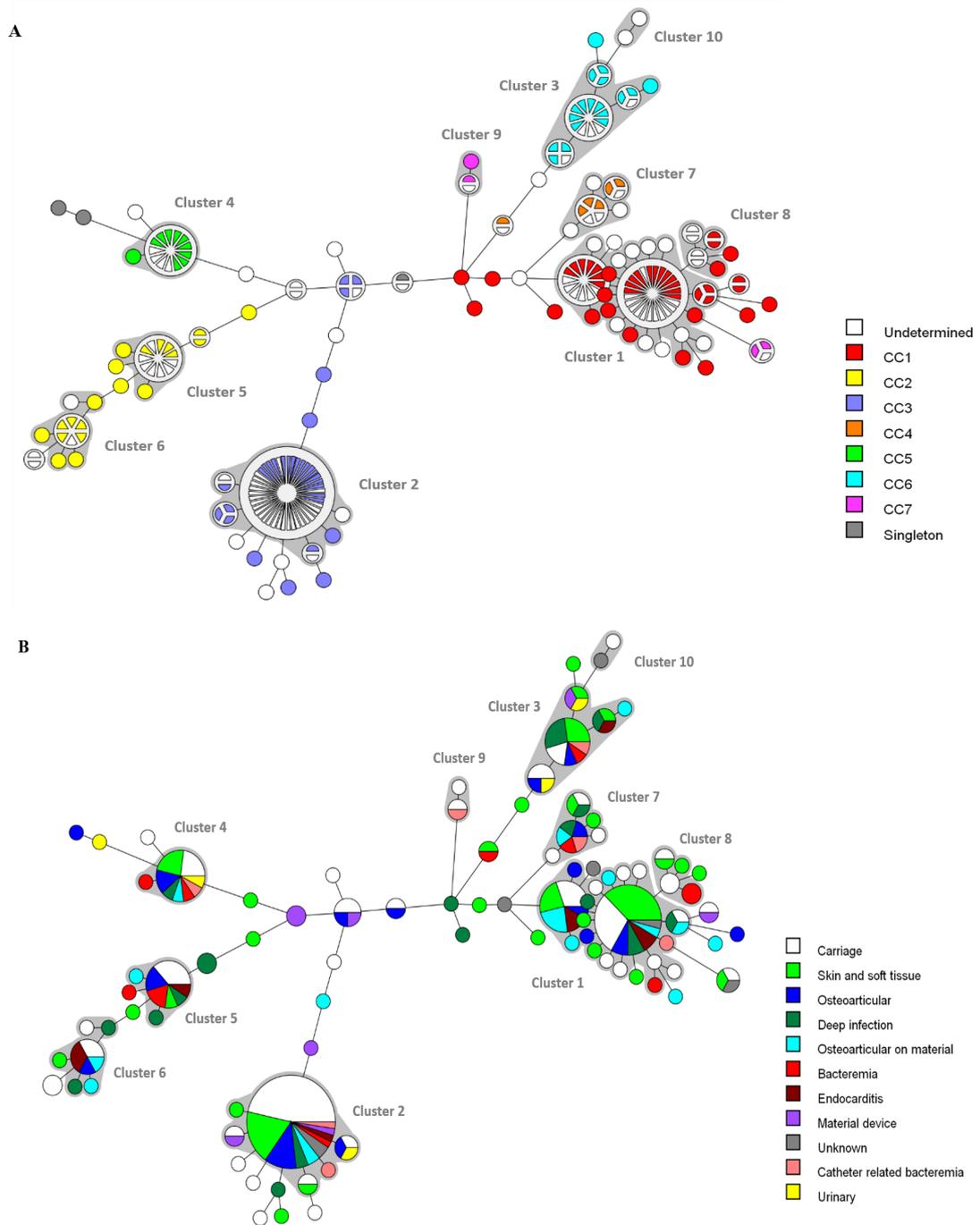


Figure 3

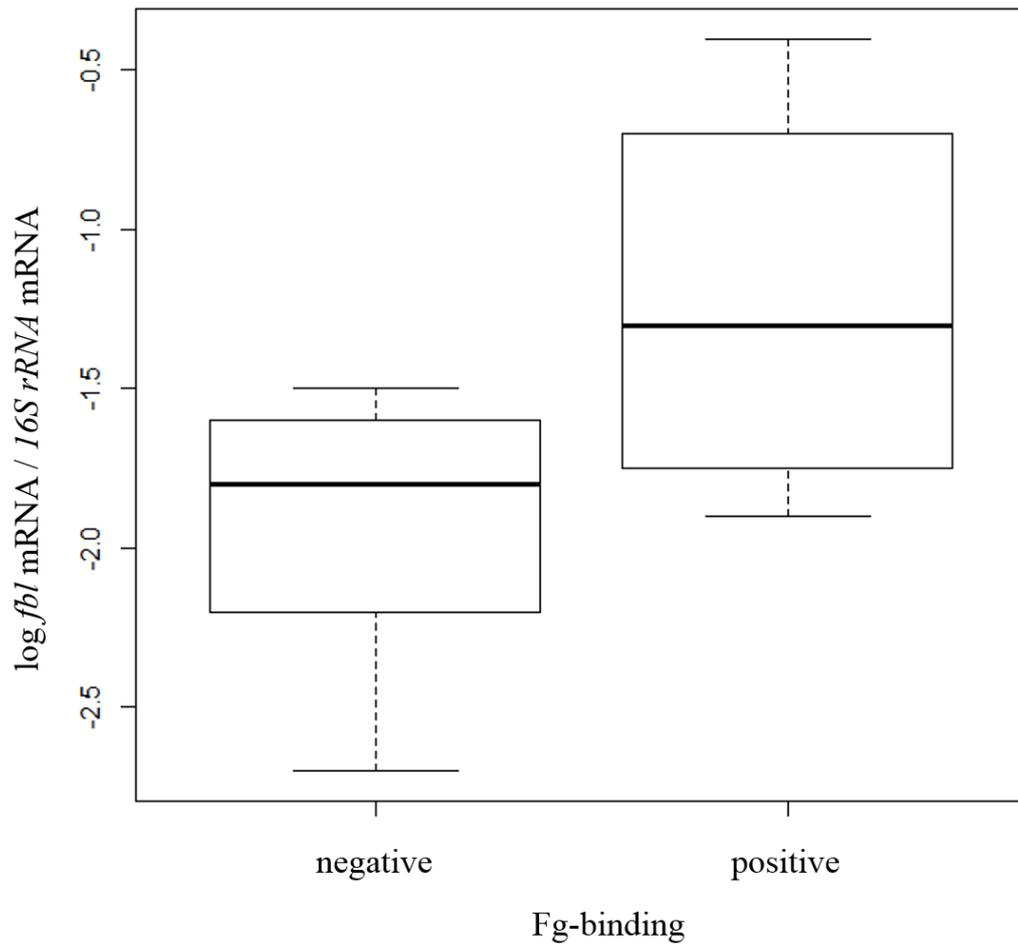


Table S1 Characteristics of the 240 *S. lugdunensis* isolates included in this study.

ND : not determined. ^a N: negative, P: positive

Isolate ID	City	Clinical source	MLST		TRST	<i>fbl</i> type	Premature stop codon in <i>fbl</i> gene	Agglutination test ^a
			Sequence Type (ST)	Clonal Complex (CC)	TRST type (TRT)			
"TRST panel" (n = 128)								
SL_V01	Strasbourg	Deep infection	5	5	10	41a	—	N
SL_V02	Strasbourg	Osteoarticular on material	1	1	41	45g	—	P
SL_V03	Strasbourg	Skin and soft tissue	3	3	1	47b	—	P
SL_V04	Strasbourg	Material device	3	3	1	47b	—	N
SL_V05	Strasbourg	Deep infection	1	1	40	45g	—	P
SL_V06	Strasbourg	Deep infection	2	2	2	45c	—	N
SL_V07	Strasbourg	Deep infection	6	1	8	21a	L22STOP	N
SL_V08	Strasbourg	Skin and soft tissue	15	1	47	40c	L22STOP	N
SL_V09	Strasbourg	Skin and soft tissue	6	1	8	45f	L22STOP	N
SL_V10	Strasbourg	Material device	3	3	1	47c	—	N
SL_V11	Strasbourg	Skin and soft tissue	5	5	7	41a	—	N
SL_V12	Strasbourg	Skin and soft tissue	5	5	7	41a	—	N
SL_V13	Strasbourg	Osteoarticular on material	12	1	3	43c	L22STOP	N
SL_V14	Strasbourg	Catheter related bacteremia	24	6	22	41b	—	P
SL_V15	Strasbourg	Osteoarticular on material	2	2	16	38a	—	N
SL_V16	Strasbourg	Skin and soft tissue	12	1	45	43c	L22STOP	N
SL_V17	Strasbourg	Osteoarticular	6	1	21	45f	—	P
SL_V18	Strasbourg	Deep infection	24	6	4	41b	—	P
SL_V19	Strasbourg	Osteoarticular on material	2	2	2	38b	—	N
SL_V20	Strasbourg	Skin and soft tissue	24	6	30	39a	—	P
SL_V21	Strasbourg	Catheter related bacteremia	3	3	1	47d	—	N
SL_V22	Strasbourg	Endocarditis	24	6	24	42b	—	N
SL_V23	Strasbourg	Osteoarticular on material	24	6	26	42c	—	N
SL_V24	Strasbourg	Osteoarticular on material	12	1	46	44a	L22STOP	N
SL_V25	Strasbourg	Osteoarticular on material	5	5	12	41a	—	N
SL_V26	Strasbourg	Skin and soft tissue	3	3	13	47b	—	N
SL_V27	Strasbourg	Osteoarticular on material	12	1	33	43c	L22STOP	N
SL_V28	Strasbourg	Material device	24	6	5	43b	—	P
SL_V29	Strasbourg	Urinary	24	6	5	43b	—	P
SL_V30	Strasbourg	Catheter related bacteremia	31	5	25	41a	—	N
SL_V31	Strasbourg	Urinary	3	3	1	47a	—	N
SL_V32	Strasbourg	Skin and soft tissue	6	1	3	45f	L22STOP	N
SL_V33	Strasbourg	Deep infection	6	1	28	45a	L22STOP	N
SL_V34	Strasbourg	Skin and soft tissue	4	4	48	52a	—	P
SL_V35	Strasbourg	Catheter related bacteremia	1	1	42	45i	—	P
SL_V36	Strasbourg	Osteoarticular	12	1	3	43c	L22STOP	N
SL_V37	Strasbourg	Endocarditis	1	1	18	45f	—	P
SL_V38	Strasbourg	Osteoarticular on material	6	1	20	46a	L22STOP	N
SL_V39	Strasbourg	Skin and soft tissue	24	6	5	43b	—	P
SL_V40	Strasbourg	Osteoarticular on material	1	1	43	39b	—	P
SL_V41	Strasbourg	Skin and soft tissue	3	3	1	47b	—	N
SL_V42	Strasbourg	Urinary	24	6	4	40b	—	P
SL_V43	Strasbourg	Osteoarticular	24	6	4	40b	—	N
SL_V44	Strasbourg	Osteoarticular	6	1	3	39c	—	P
SL_V45	Strasbourg	Osteoarticular	24	6	4	41b	—	P
SL_V46	Strasbourg	Deep infection	24	6	27	41b	—	N
SL_V47	Strasbourg	Osteoarticular	3	3	1	47a	—	N
SL_V48	Strasbourg	Deep infection	4	4	49	52a	—	N
SL_V49	Strasbourg	Osteoarticular	2	2	2	38a	—	N
SL_V50	Strasbourg	Osteoarticular on material	3	3	1	24a	—	N
SL_V51	Strasbourg	Osteoarticular	6	1	35	45h	—	P

SL_V52	Strasbourg	Osteoarticular on material	1	1	19	35a	_	P
SL_V53	Strasbourg	Osteoarticular	2	2	2	45b	_	N
SL_V54	Strasbourg	Osteoarticular	5	5	23	41a	_	N
SL_V55	Strasbourg	Osteoarticular	31	5	11	41a	_	N
SL_V56	Strasbourg	Material device	3	3	17	13a	_	N
SL_V57	Strasbourg	Deep infection	1	1	44	21b	_	P
SL_V58	Strasbourg	Catheter related bacteremia	4	4	50	49a	_	N
SL_V59	Strasbourg	Osteoarticular on material	3	3	1	47b	_	N
SL_V60	Strasbourg	Skin and soft tissue	2	2	2	36a	_	N
SL_V61	Strasbourg	Osteoarticular on material	3	3	1	47b	_	N
SL_V62	Strasbourg	Skin and soft tissue	15	1	36	46c	L22STOP	N
SL_V63	Strasbourg	Deep infection	2	2	2	34a	_	N
SL_V64	Strasbourg	Skin and soft tissue	6	1	32	45f	L22STOP	N
SL_V65	Strasbourg	Osteoarticular	28	Singleton	15	48a	_	P
SL_V66	Strasbourg	Osteoarticular on material	12	1	3	43c	L22STOP	N
SL_V67	Strasbourg	Skin and soft tissue	3	3	1	45j	_	N
SL_V68	Strasbourg	Deep infection	2	2	2	38c	_	N
SL_V69	Strasbourg	Urinary	28	Singleton	14	48b	_	P
SL_V70	Strasbourg	Skin and soft tissue	27	7	9	41c	_	N
SL_V71	Strasbourg	Skin and soft tissue	2	2	2	23a	_	N
SL_V72	Strasbourg	Skin and soft tissue	3	3	1	47b	_	N
SL_V73	Strasbourg	Urinary	5	5	38	41a	_	P
SL_V74	Strasbourg	Bacteremia	2	2	51	45b	_	P
SL_V75	Strasbourg	Skin and soft tissue	2	2	2	40a	_	N
SL_V76	Strasbourg	Deep infection	3	3	1	47b	_	N
SL_V77	Strasbourg	Catheter related bacteremia	26	7	39	45d	_	P
SL_V78	Strasbourg	Skin and soft tissue	3	3	1	43a	_	N
SL_V79	Strasbourg	Deep infection	6	1	31	45f	L22STOP	N
SL_V80	Strasbourg	Osteoarticular on material	6	1	6	45f	_	P
SL_V81	Strasbourg	Skin and soft tissue	6	1	6	30a	L22STOP	N
SL_V82	Strasbourg	Skin and soft tissue	1	1	34	25a	_	P
SL_C08	Strasbourg	Carriage	3	3	1	47b	_	N
SL_C18	Strasbourg	Carriage	6	1	29	45f	L22STOP	N
SL_C21	Strasbourg	Carriage	3	3	1	47a	_	N
SL_C22	Strasbourg	Carriage	3	3	1	47b	_	N
SL_C27	Strasbourg	Carriage	3	3	1	47b	L22STOP	N
SL_C35	Strasbourg	Carriage	3	3	37	42a	_	N
SL_C42	Strasbourg	Carriage	26	7	52	45e	_	P
SL_C47	Strasbourg	Carriage	3	3	1	13a	_	N
SL_C60	Strasbourg	Carriage	12	1	3	43c	L22STOP	N
SL_C62	Strasbourg	Carriage	15	1	6	46b	L22STOP	N
SL_C68	Strasbourg	Carriage	2	2	2	38a	_	N
SL_C71	Strasbourg	Carriage	27	7	9	41c	_	N
SL_C72	Strasbourg	Carriage	3	3	1	42d	_	N
SL_C75	Strasbourg	Carriage	3	3	1	47b	_	P
SL_C77	Strasbourg	Carriage	3	3	1	47b	_	P
SL_C85	Strasbourg	Carriage	3	3	1	13a	_	N
SL_T02	Rouen	Endocarditis	8	2	58	45b	_	N
SL_T03	Rouen	Bacteremia	9	4	68	23b	_	P
SL_T09	Rouen	Bacteremia	5	5	53	42e	_	N
SL_T10	Rouen	Osteoarticular	13	Singleton	60	9a	_	N
SL_T119	Kronoberg	Carriage	1	1	61	45g	_	P
SL_T21	Rouen	Deep infection	10	6	62	42b	_	N
SL_T27	Rouen	Osteoarticular on material	4	4	67	49a	_	P
SL_T37	Rouen	Catheter related bacteremia	16	3	59	47b	_	N
SL_T53	Nantes	Material device	3	3	1	39d	_	N
SL_T56	Nantes	Osteoarticular	14	2	2	45b	_	P
SL_T59	Nantes	Skin and soft tissue	6	1	64	45f	L22STOP	P
SL_T62	Bordeaux	Bacteremia	17	4	66	49a	_	P

SL_T66	Bordeaux	Bacteremia	18	5	10	41a	_	N
SL_T68	Bordeaux	Bacteremia	7	1	65	44c	_	P
SL_T73	Nancy	Bacteremia	19	2	56	42h	_	N
SL_T74	Nancy	Skin and soft tissue	12	1	3	43c	L22STOP	P
SL_T76	Nancy	Osteoarticular	20	3	57	47b	_	P
SL_T84	Montpellier	Deep infection	15	1	69	45f	L22STOP	N
SL_T85	Montpellier	Osteoarticular on material	2	2	2	45o	_	N
SL_AP1	Rouen	Osteoarticular	3	3	1	47b	_	N
SL_AP2	Rouen	Osteoarticular	3	3	1	47b	_	P
SL_CB1	Rouen	Bacteremia	15	1	55	46d	L22STOP	N
SL_CB2	Rouen	Bacteremia	15	1	55	46d	L22STOP	N
SL_CJ1	Rouen	Deep infection	24	6	63	41b	_	P
SL_CJ2	Rouen	Bacteremia	24	6	63	41b	_	P
SL_DJ1	Rouen	Endocarditis	32	2	2	38a	_	N
SL_DJ2	Rouen	Endocarditis	32	2	2	38a	_	N
SL_LJ1	Rouen	Deep infection	2	2	2	41d	_	N
SL_LJ2	Rouen	Deep infection	2	2	2	41d	_	N
SL_DSM		Skin and soft tissue	24	6	54	41b	_	P
Other isolates (n = 112)								
SL_T117	Kronoberg	Carriage	1	1	ND	45f	_	P
SL_T118	Kronoberg	Carriage	10	6	ND	40b	_	N
SL_T122	Tours	Carriage	1	1	ND	43g	_	P
SL_T13	Rouen	Endocarditis	6	1	ND	45f	L22STOP	N
SL_T29	Rouen	Material device	1	1	ND	43g	_	P
SL_T55	Nantes	Skin and soft tissue	10	6	ND	41b	_	N
SL_R01	Rouen	Carriage	ND	ND	ND	43c	ND	N
SL_R02	Rouen	Carriage	ND	ND	ND	43c	ND	P
SL_R03	Rouen	Endocarditis	ND	ND	ND	43c	ND	N
SL_R04	Rouen	Endocarditis	ND	ND	ND	43c	ND	P
SL_R05	Rouen	Carriage	ND	ND	ND	45b	ND	N
SL_R06	Rouen	Carriage	ND	ND	ND	45b	ND	P
SL_R07	Rouen	Material device	ND	ND	ND	14a	ND	N
SL_R08	Rouen	Material device	ND	ND	ND	14a	ND	N
SL_R09	Rouen	Carriage	ND	ND	ND	30b	ND	N
SL_R10	Rouen	Carriage	ND	ND	ND	30b	ND	N
SL_R11	Rouen	Bacteremia	ND	ND	ND	47b	_	P
SL_R12	Rouen	Carriage	ND	ND	ND	44f	ND	P
SL_R13	Rouen	Carriage	ND	ND	ND	44d	ND	P
SL_R14	Rouen	Skin and soft tissue	ND	ND	ND	45f	ND	N
SL_R15	Rouen	Carriage	ND	ND	ND	45k	L22STOP	N
SL_R16	Rouen	Carriage	ND	ND	ND	44e	ND	N
SL_R17	Rouen	Unknown	ND	ND	ND	41e	ND	P
SL_R18	Rouen	Carriage	ND	ND	ND	45b	ND	N
SL_R19	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R20	Rouen	Skin and soft tissue	ND	ND	ND	47b	ND	N
SL_R21	Rouen	Carriage	ND	ND	ND	45f	ND	N
SL_R22	Rouen	Carriage	ND	ND	ND	24b	ND	N
SL_R23	Rouen	Osteoarticular	ND	ND	ND	49a	ND	P
SL_R24	Rouen	Carriage	ND	ND	ND	43c	ND	N
SL_R25	Rouen	Osteoarticular	ND	ND	ND	47b	ND	N
SL_R26	Rouen	Skin and soft tissue	ND	ND	ND	49b	ND	P
SL_R27	Rouen	Carriage	ND	ND	ND	41a	ND	P
SL_R28	Rouen	Carriage	ND	ND	ND	42g	ND	N
SL_R29	Rouen	Carriage	ND	ND	ND	45d	_	P
SL_R30	Rouen	Skin and soft tissue	ND	ND	ND	45l	_	P
SL_R31	Rouen	Carriage	ND	ND	ND	38a	ND	N
SL_R32	Rouen	Unknown	ND	ND	ND	41c	ND	N
SL_R33	Rouen	Carriage	ND	ND	ND	45f	ND	P
SL_R34	Rouen	Skin and soft tissue	ND	ND	ND	45f	_	P

SL_R35	Rouen	Carriage	ND	ND	ND	49c	ND	P
SL_R36	Rouen	Unknown	ND	ND	ND	43d	ND	P
SL_R37	Rouen	Carriage	ND	ND	ND	41a	ND	N
SL_R38	Rouen	Skin and soft tissue	ND	ND	ND	48c	ND	N
SL_R39	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R40	Rouen	Skin and soft tissue	ND	ND	ND	41b	ND	P
SL_R41	Rouen	Skin and soft tissue	ND	ND	ND	45b	–	N
SL_R42	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R43	Rouen	Unknown	ND	ND	ND	29a	ND	N
SL_R44	Rouen	Carriage	ND	ND	ND	47b	–	N
SL_R45	Rouen	Carriage	ND	ND	ND	47b	ND	P
SL_R46	Rouen	Carriage	ND	ND	ND	47c	ND	N
SL_R47	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R48	Rouen	Carriage	ND	ND	ND	46e	ND	N
SL_R49	Rouen	Carriage	ND	ND	ND	41a	ND	N
SL_R50	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R51	Rouen	Carriage	ND	ND	ND	45m	ND	P
SL_R52	Rouen	Carriage	ND	ND	ND	46e	ND	N
SL_R53	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R54	Rouen	Unknown	ND	ND	ND	45f	ND	P
SL_R55	Rouen	Unknown	ND	ND	ND	47b	ND	N
SL_R56	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R57	Rouen	Skin and soft tissue	ND	ND	ND	47b	ND	N
SL_R58	Rouen	Skin and soft tissue	ND	ND	ND	47b	ND	N
SL_R59	Rouen	Osteoarticular	ND	ND	ND	13a	–	N
SL_R60	Rouen	Skin and soft tissue	ND	ND	ND	45f	ND	P
SL_R61	Rouen	Skin and soft tissue	ND	ND	ND	46f	ND	N
SL_R62	Rouen	Skin and soft tissue	ND	ND	ND	42b	ND	N
SL_R63	Rouen	Skin and soft tissue	ND	ND	ND	43c	ND	N
SL_R64	Rouen	Skin and soft tissue	ND	ND	ND	41a	ND	N
SL_R65	Rouen	Skin and soft tissue	ND	ND	ND	36b	ND	P
SL_R66	Rouen	Deep infection	ND	ND	ND	49a	ND	N
SL_R67	Rouen	Carriage	ND	ND	ND	49d	ND	P
SL_R68	Rouen	Carriage	ND	ND	ND	44b	ND	P
SL_R69	Rouen	Carriage	ND	ND	ND	40b	ND	P
SL_R70	Rouen	Carriage	ND	ND	ND	45f	ND	P
SL_R71	Rouen	Carriage	ND	ND	ND	45b	ND	N
SL_R72	Rouen	Deep infection	ND	ND	ND	42f	–	N
SL_R73	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R74	Rouen	Carriage	ND	ND	ND	37a	–	N
SL_R75	Rouen	Deep infection	ND	ND	ND	45b	ND	P
SL_R76	Rouen	Skin and soft tissue	ND	ND	ND	47b	ND	N
SL_R77	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R78	Rouen	Carriage	ND	ND	ND	46f	ND	N
SL_R79	Rouen	Carriage	ND	ND	ND	45j	ND	N
SL_R80	Rouen	Skin and soft tissue	ND	ND	ND	45f	ND	N
SL_R81	Rouen	Osteoarticular	ND	ND	ND	43e	ND	N
SL_R82	Rouen	Endocarditis	ND	ND	ND	47b	ND	N
SL_R83	Rouen	Carriage	ND	ND	ND	41b	ND	P
SL_R84	Rouen	Carriage	ND	ND	ND	45f	ND	N
SL_R85	Rouen	Carriage	ND	ND	ND	33a	ND	N
SL_R86	Rouen	Carriage	ND	ND	ND	45f	ND	N
SL_R87	Rouen	Carriage	ND	ND	ND	18a	–	N
SL_R88	Rouen	Skin and soft tissue	ND	ND	ND	20a	ND	N
SL_R89	Rouen	Skin and soft tissue	ND	ND	ND	45f	ND	P
SL_R90	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R91	Rouen	Carriage	ND	ND	ND	47b	ND	N
SL_R92	Rouen	Deep infection	ND	ND	ND	47b	ND	N
SL_R93	Rouen	Carriage	ND	ND	ND	52a	ND	N

SL_R94	Rouen	Carriage	<i>ND</i>	<i>ND</i>	<i>ND</i>	47b	<i>ND</i>	N
SL_R95	Rouen	Carriage	<i>ND</i>	<i>ND</i>	<i>ND</i>	43f	<i>ND</i>	N
SL_R96	Rouen	Unknown	<i>ND</i>	<i>ND</i>	<i>ND</i>	47b	<i>ND</i>	N
SL_R97	Rouen	Carriage	<i>ND</i>	<i>ND</i>	<i>ND</i>	47b	<i>ND</i>	N
SL_R98	Rouen	Carriage	<i>ND</i>	<i>ND</i>	<i>ND</i>	41b	<i>ND</i>	P
SL_R99	Rouen	Carriage	<i>ND</i>	<i>ND</i>	<i>ND</i>	9a	-	N
SL_R100	Rouen	Carriage	<i>ND</i>	<i>ND</i>	<i>ND</i>	41f	<i>ND</i>	N
SL_R101	Rouen	Skin and soft tissue	<i>ND</i>	<i>ND</i>	<i>ND</i>	45n	<i>ND</i>	N
SL_R102	Rouen	Skin and soft tissue	<i>ND</i>	<i>ND</i>	<i>ND</i>	23b	<i>ND</i>	N
SL_R103	Rouen	Osteoarticular	<i>ND</i>	<i>ND</i>	<i>ND</i>	47b	<i>ND</i>	N
SL_R104	Rouen	Skin and soft tissue	<i>ND</i>	<i>ND</i>	<i>ND</i>	41g	<i>ND</i>	N
SL_R105	Rouen	Bacteremia	<i>ND</i>	<i>ND</i>	<i>ND</i>	45b	<i>ND</i>	P
SL_R106	Rouen	Osteoarticular	<i>ND</i>	<i>ND</i>	<i>ND</i>	45f	<i>ND</i>	N

Table S2 DNA and amino acid sequences of the 54 individual *fbI* repeats and their numeric codes identified from the 240 *S. lugdunensis* isolates studied.

Repeat code	DNA sequence	Amino Acid Sequence	Repeat code	DNA sequence	Amino Acid Sequence
0	GATTCCGATAGTGATGCA	DSDDSA	27	GACTCCGACAGTGACGCG	DSDDSA
1	GACTCCGATAGTGATGCA	DSDDSA	28	TATTCCGATAGCGACGCA	YSDDSA
2	GATTCCGACAGTGATGCA	DSDDSA	29	GATTCCGACAGTGACGCA	DSDDSA
3	GATTCCGATAGTGACGCA	DSDDSA	30	GAATCAGATAGCGATGCA	ESDDSA
4	GATTCCGATAGCGATGCA	DSDDSA	31	GACTCCGATAGCGATGCA	DSDDSA
5	GACTCCGATAGTGACGCA	DSDDSA	32	GACTCCGACAGTGATGCA	DSDDSA
6	GATTCCGATAGCGACGCA	DSDDSA	33	GACTTAGATAGTGACGCA	DLDDSA
7	GATTCCGACAGCGATGCA	DSDDSA	34	GATTCCGACAGTGATGCG	DSDDSA
8	GATTCCGACAGCGACGCA	DSDDSA	35	GATTCCGATAGTGATTCA	DSDDSD
9	GACTCCGACAGCGATGCA	DSDDSA	36	GACTCTGACAGTGACGCA	DSDDSA
10	GACTCCGACAGCGACGCA	DSDDSA	37	GACTCCGATAGTAATGCG	DSDDNA
11	GATTCCGATAGTGATGCG	DSDDSA	38	GACTCTGACAGCGATGCG	DSDDSA
12	GACTCCGATAGTGATGCG	DSDDSA	39	GACTCCGATAGTGACGCG	DSDDSA
13	GACTCCGACAGTGATGCG	DSDDSA	40	GATTCAGACAGCGATGCA	DSDDSA
14	GATTCCGACAGCGATGCG	DSDDSA	41	GATTCTGACAGCGACACA	DSDDST
15	GATTCAGACAGCGATGCG	DSDDSA	42	GACTCCGACAGCGATGTG	DSDDSDV
16	GATTCTGACAGTGATGCG	DSDDSA	43	GACTCTGACAGCGACGCA	DSDDSA
17	GACTCTGACAGTGATGCG	DSDDSA	44	GACTCAGATAGTGATTCA	DSDDSD
18	GACTCAGATAGTGACGCA	DSDDSA	45	GATTCTGACAGCAATTCC	DSDDNS
19	GACTCAGATAGCGATGCA	DSDDSA	46	GGTCCGACAGCAATGTG	GSDDNV
20	GACTCAGATAGCGACGCA	DSDDSA	47	GACTCAGACAGCGACGCA	DSDDSA
21	GATTCTGACAGCGACGCA	DSDDSA	48	GACTCAGATAGCGATGCG	DSDDSA
22	GATTCAGACAGCGACGCA	DSDDSA	49	GACACAGATAGTGACGCA	DTDDSA
23	GACTCCGACAGCGATGCG	DSDDSA	50	AATTCCGATAGTGATGCA	NSDDSA
24	GACTCCGACAGTGACGCA	DSDDSA	51	GAGTCAGATAGCGATGCA	ESDDSA
25	GACTCAGATAGTGATGCA	DSDDSA	52	GATTCCGACAGTGACGCG	DSDDSA
26	GACTCCGATAGCGACGCA	DSDDSA	53	GCATGGGATAGTGATGCA	AWDDSA

Table S3 List of the 92 *fbI* types identified from the 240 *S. lugdunensis* isolates studied; and preliminary guideline for the assignment of MLST clonal complexes (CCs) using *fbI* typing. CC*: CC predicted from the repeat alignment of the *fbI* types.

<i>fbI</i> type	<i>fbI</i> cluster	Repeat succession	CC	ST
fbI21a	-	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-30-6-6-45-46-44	CC1	ST6
fbI21b	-	0-6-13-1-7-18-2-11-17-12-13-5-4-18-16-20-9-6-45-46-44	CC1	ST1
fbI25a	-	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-45-46-44	CC1	ST1
fbI29a	-	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-4-18-16-20-9-6-45-46-44	CC1*	-
fbI30a	-	0-6-13-1-7-18-2-11-38-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-9-6-45-46-44	CC1*	-
fbI35a	-	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-21-1-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-44	CC1	ST1
fbI39b	-	0-6-13-1-7-1-7-18-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-8-2-18-16-20-9-6-45-46-44	CC1	ST1
fbI39c	-	0-7-18-2-11-17-12-13-5-14-10-19-18-16-20-16-20-0-9-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST6
fbI40c	-	0-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST15
fbI41e	1	0-6-13-1-7-18-34-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI41g	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI43c	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST12
fbI43e	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-19-15-19-22-2-3-10-18-16-20-9-6-45-46-44	CC1*	-
fbI43g	-	0-6-13-5-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST1
fbI44a	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST12
fbI44b	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-10-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI44c	1	0-6-13-1-7-18-2-0-11-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST7
fbI44d	1	0-6-13-1-7-18-2-11-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI44e	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI44f	1	6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI45a	-	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST6
fbI45f	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST1, ST6, ST15
fbI45g	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST1
fbI45h	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST6
fbI45i	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST1
fbI45k	1	0-6-13-1-7-49-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI45l	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-5-4-18-16-20-9-6-45-46-44	CC1	-
fbI45m	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-8-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI45n	-	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-31-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI46a	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST6
fbI46b	1	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST15
fbI46c	8	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-13-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1	ST15
fbI46d	8	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-8-2-3-4-18-16-20-9-6-45-46-44	CC1	ST15
fbI46e	8	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-8-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI46f	8	0-6-13-1-7-18-2-11-17-12-13-5-14-10-0-19-6-11-10-0-9-6-21-0-10-19-18-16-20-16-19-15-19-22-2-3-4-18-16-20-9-6-45-46-44	CC1*	-
fbI23a	-	0-28-20-23-1-7-5-7-11-12-6-12-0-6-20-23-9-11-13-13-26-45-46-44	CC2	ST2
fbI30b	-	0-28-20-23-1-7-26-23-12-1-7-18-7-11-13-18-23-10-0-19-8-19-9-11-13-13-26-45-46-44	CC2*	-
fbI33a	6	0-28-20-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-26-45-46-44	CC2*	-
fbI34a	6	0-28-20-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI36a	6	0-28-20-23-1-7-26-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-8-19-9-11-13-26-45-46-44	CC2	ST2
fbI38a	6	0-28-20-23-1-7-26-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2, ST32
fbI38b	6	0-28-20-23-1-7-26-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-33-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI38c	6	0-7-26-23-1-7-26-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI40a	5	0-28-20-23-12-6-20-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI41d	-	0-28-20-23-1-7-5-7-11-26-12-0-6-20-23-12-1-7-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI42h	-	0-28-20-23-1-7-5-7-11-26-12-0-6-20-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-8-19-9-11-13-26-45-46-44	CC2	ST19
fbI45b	5	0-28-20-23-1-7-5-7-11-26-12-0-6-20-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2, ST8, ST14
fbI45c	5	0-28-20-23-1-7-5-7-11-26-12-0-6-20-9-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI45o	-	0-28-20-23-1-7-18-7-11-26-12-0-6-20-23-12-1-7-18-7-11-13-18-23-10-0-19-3-0-27-29-0-18-7-7-8-19-9-11-13-13-26-45-46-44	CC2	ST2
fbI13a	-	0-6-20-23-12-1-7-20-7-18-14-26-44	CC3	ST3
fbI18a	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-26-44	CC3*	-
fbI24a	-	0-6-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI24b	-	50-6-20-23-12-1-7-20-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3*	-
fbI27a	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-7-18-14-26-44	CC3*	-
fbI39d	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-52-21-1-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI42a	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-0-27-21-1-18-7-8-20-23-25-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI42d	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-26-44	CC3	ST3
fbI42f	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3*	-
fbI42g	-	0-6-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3*	-
fbI43a	-	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-47-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI44j	2	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-25-7-7-26-44	CC3	ST3
fbI47a	2	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-14-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI47b	2	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3, ST16, ST20
fbI47c	2	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-20-23-1-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI47d	2	0-6-20-10-11-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3	ST3
fbI48c	2	0-6-20-23-12-1-7-20-7-11-13-18-14-10-0-11-9-11-8-0-23-1-27-21-1-18-7-8-20-23-25-7-8-20-23-25-7-8-19-23-25-7-7-18-14-26-44	CC3*	-
fbI23b	-	53-23-1-9-11-13-12-13-5-14-10-0-31-6-14-1-27-2-11-20-45-46-44	CC4	ST9
fbI49a	7	0-6-23-1-7-18-7-11-13-13-12-5-14-10-0-19-6-11-8-0-23-1-27-2-11-20-21-1-8-19-34-19-7-21-11-19-9-21-2-3-4-18-16-20-9-6-45-46-44	CC4	ST4, ST17
fbI49b	7	0-6-23-1-7-18-7-11-13-13-12-5-14-10-0-19-6-11-8-0-23-1-27-2-20-21-1-8-19-34-19-7-21-11-19-9-21-2-3-4-18-16-20-9-6-45-46-44	CC4*	-
fbI49c	7	0-6-23-1-7-18-7-11-13-13-12-5-14-10-0-51-6-11-8-0-23-1-27-2-11-20-21-1-8-19-34-19-7-21-11-19-9-21-2-3-4-18-16-20-9-6-45-46-44	CC4*	-
fbI49d	7	0-6-23-1-7-18-7-11-13-13-12-5-14-10-0-19-6-11-8-0-23-1-27-2-11-20-21-1-8-19-34-19-7-21-11-19-9-21-2-3-4-18-16-20-9-6-45-46-44	CC4	-
fbI52a	7	0-6-23-1-6-23-1-7-18-7-11-13-13-12-5-14-10-0-19-6-11-8-0-23-1-27-2-11-20-21-1-8-19-34-19-7-21-11-19-9-21-2-3-4-18-16-20-9-6-45-46-44	CC4	ST4
fbI14a	-	0-6-20-23-1-14-20-9-11-13-26-45-46-44	CC5*	-
fbI20a	-	0-11-20-24-0-18-8-20-23-25-7-8-19-9-11-13-26-45-46-44	CC5*	-
fbI41f	4	0-6-20-23-1-14-20-9-11-13-5-14-10-0-19-6-7-8-14-13-5-7-11-20-24-0-18-8-20-23-25-7-8-19-9-11-13-26-45-46-44	CC5	ST5, ST18, ST31
fbI41f	-	0-6-20-23-1-14-20-9-11-13-5-14-9-0-25-6-4-7-4-13-5-7-11-20-24-0-18-8-20-23-25-7-8-19-9-11-13-26-45-46-44	CC5*	-
fbI42c	4	0-6-20-23-1-14-20-9-11-13-5-14-10-0-19-6-7-8-14-13-5-7-11-20-24-0-18-8-20-23-25-7-8-19-9-11-13-26-45-46-44	CC5	ST5
fbI36b	-	0-6-23-1-6-11-8-0-23-1-9-11-13-12-13-5-14-10-0-31-6-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6*	-
fbI39a	-	0-6-23-1-6-11-8-0-42-19-11-13-13-5-14-10-0-31-6-2-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6	ST24
fbI40b	3	0-6-23-1-6-11-8-0-23-1-9-11-13-12-13-5-14-10-0-31-6-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6	ST10, ST24
fbI41b	3	0-6-23-1-6-11-8-0-23-1-9-11-13-12-13-5-14-10-0-31-6-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6	ST10, ST24
fbI42b	3	0-6-23-1-6-11-8-0-23-1-9-11-13-12-13-5-14-10-0-31-6-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6	ST10, ST24
fbI42c	3	0-6-23-1-6-11-8-0-23-1-9-11-32-12-13-5-14-10-0-31-6-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6	ST24
fbI43b	3	0-6-23-1-6-11-8-6-11-8-0-23-1-9-11-13-13-5-14-10-0-31-6-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6	ST24
fbI43d	10	0-6-23-1-6-11-8-6-11-8-0-23-1-23-8-0-23-1-9-11-13-13-5-14-32-27-2-11-20-21-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6*	-
fbI43f	10	0-6-23-1-6-11-8-6-11-8-0-23-1-23-7-0-23-1-9-11-13-13-5-14-32-27-2-11-20-8-1-8-20-31-4-18-16-20-9-6-45-46-44	CC6*	-

Table S4 Abilities of the three typing methods to discriminate the 123 isolates within major *S. lugdunensis* clonal complexes (CCs) defined by MLST. Sequence type (ST). TRST type (TRT).

CC (No. of isolates)	MLST		TRST		<i>fbl</i> typing	
	No. of STs	DI ^a	No. of TRTs	DI ^a	No. of <i>fbl</i> types	DI ^a
1 (37)	5	0.752	28	0.962	20	0.896
2 (18)	5	0.405	5	0.405	12	0.922
3 (30)	3	0.131	6	0.310	11	0.710
4 (6)	3	0.600	6	1	3	0.733
5 (10)	3	0.511	8	0.956	2	0.200
6 (15)	2	0.133	10	0.914	6	0.810
7 (4)	2	0.667	3	0.833	3	0.833
singleton (3)	2	0,667	3	1	3	1
Total	25	0.899	69	0.943	60	0,964

^aDI: Simpson's diversity index.

2. Discussion et perspectives

L'objectif des travaux présentés ici était d'évaluer l'apport phylogénétique de l'analyse de *loci* de VNTR, à travers le développement de deux nouvelles méthodes de génotypage, pour le suivi de l'épidémiologie globale et l'évolution de l'espèce *S. lugdunensis*.

Dans un premier temps, nous avons développé un schéma de *Multiple Locus VNTR Analysis* (MLVA) et de *Tandem Repeat Sequence Typing* (TRST). Nous avons ensuite comparé leur pouvoir discriminant et leur apport phylogénétique par rapport aux méthodes de typage par MLST²⁰⁸ et MVLST *trilocus*⁹⁸.

Afin de déterminer le pouvoir discriminant de ces nouvelles méthodes de typage, deux collections différentes de souches ont été utilisées. Selon les recommandations de Nadon *et al.*²¹⁴, il est tout d'abord conseillé d'évaluer la diversité des *loci* à partir de 10 à 20 isolats non liés épidémiologiquement et présentant un fond génétique différent déterminé à l'aide d'autres méthodes de typage. Ainsi, nous avons sélectionné un panel "test" de 20 isolats de *S. lugdunensis* provenant d'hôpitaux de cinq villes de France (Bordeaux, Montpellier, Nancy, Nantes et Rouen) et de Suède (Kronoberg). Ces isolats, préalablement caractérisés par MLST et MVLST^{98,208}, étaient représentatifs des différents ST et CC de l'espèce. L'intérêt de ce premier panel était double : (i) éliminer les VNTR présentant peu ou pas de variabilité entre les isolats, et (ii) écarter les VNTR pour lesquels des amplifications aspécifiques par PCR étaient obtenues, même après l'utilisation de plusieurs couples d'amorces.

Le principal intérêt des méthodes de typage basées sur l'analyse des VNTR est leur apport dans les études micro-épidémiologiques dans lesquelles il s'agit de savoir si des isolats collectés dans une même aire géographique sont une seule et même souche. Par exemple, des isolats épidémiologiquement apparentés issus d'épidémies présumées sont normalement attribués à des génotypes identiques ou proches²⁸⁵. Ainsi, selon les recommandations de van Belkum *et al.*²⁸⁵, lors de la validation d'une méthode de typage, il est souhaitable que cinq à dix isolats épidémiologiquement liés soient inclus à la population "test". Nous avons ainsi inclus cinq paires d'isolats, isolés chez des mêmes patients à différents intervalles de temps, de zéro à 17 jours.

Une autre étape primordiale pour le développement d'un schéma de MLVA et de TRST est la sélection des VNTR. Pour identifier de tels *loci*, il est nécessaire d'utiliser des logiciels spécialisés comme *Tandem Repeat Finder* (TRF), *TredD* ou *GeneQuest*, à partir de génomes complets. L'analyse *in silico* des génomes des souches de référence HKU09-01¹³³ et N920143⁷³ a été effectuée à partir du logiciel TRF et a permis d'identifier une centaine de VNTR.

Selon les recommandations de Nadon *et al.*²¹⁴, des VNTR dont les motifs répétés ont une taille inférieure ou égale à 5 pb sont à écarter. En effet, dans le cas d'une éventuelle utilisation de l'électrophorèse capillaire, la résolution de l'automate pourrait être une limite à la détermination fiable et reproductible du nombre de motifs répétés. De plus, les VNTR de type microsatellite sont moins stables au cours du temps que ceux de type minisatellite. Par ailleurs, il est conseillé de ne pas utiliser des VNTR présentant des insertions ou des délétions afin de faciliter le compte des motifs répétés. Nadon *et al.*²¹⁴ recommandent également de n'utiliser que des VNTR avec des régions flanquantes conservées à 100% pour ne pas limiter la typabilité, c'est-à-dire la capacité de la méthode à assigner un génotype aux isolats testés.

Dans cette étude, nous avons utilisé le séquençage Sanger pour déterminer le nombre de motifs répétés. Une des contraintes de notre étude a alors été d'écarter les VNTR qui ne permettraient pas une amplification par PCR et un séquençage Sanger à l'aide d'un seul couple d'amorces. Ainsi, parmi la centaine de VNTR identifiés dans notre étude, 28 ont été conservés selon ces recommandations. Nous avons évalué le pouvoir discriminant des 28 VNTR sur le panel "test". Sept VNTR, dont quatre intra-géniques et trois dans des régions non-codantes, présentant des TR d'une taille allant de 24 à 48 pb, ont été sélectionnés pour leur pouvoir discriminant. Par ailleurs, les paires d'isolats d'un même patient présentaient le même génotype, permettant de valider la concordance épidémiologique de la méthode.

La seconde collection d'isolats consistait en un panel de 98 isolats cliniques de *S. lugdunensis* collectés entre 2013 et 2016 à l'hôpital universitaire de Strasbourg ; 82 isolats responsables d'infections sévères collectés lors de l'étude prospective VISLISI (*Virulence of Staphylococcus lugdunensis In Severe Infections*) et 16 isolés en situation de colonisation¹¹. La méthode de MLVA appliquée aux 98 isolats a permis de générer 42 MT avec des DI par VNTR compris entre 0,090 et 0,679. En outre, notre schéma de MLVA est similaire en termes de nombre de VNTR et de DI à ceux proposés pour les espèces de

*S. aureus*²²⁷ et *S. epidermidis*²²⁰. En effet, le schéma de MLVA développé par Schouls *et al.*²²⁷ pour le typage de *S. aureus* utilise huit VNTR avec des DI compris entre 0,340 et 0,800 alors que le schéma développé pour *S. epidermidis* se base sur le polymorphisme de cinq VNTR avec des DI compris entre 0,510 et 0,800²²⁰.

La méthode de MLVA appliquée aux 128 isolats a montré une typabilité de 100% et un pouvoir résolutif plus important que la méthode de MLST en termes de nombre de génotypes et de DI. Cette différence peut s'expliquer par le fait que les VNTR évoluent plus rapidement que les gènes domestiques, ce qui limite le pouvoir discriminant de la MLST. La MLST mesure l'héritabilité des mutations présentes au sein des gènes domestiques alors que la MLVA mesure la variabilité des VNTR, principalement liée au glissement de l'ADN polymérase²¹³. Bien que ces deux méthodes se basent sur des mécanismes d'évolution différents, un regroupement similaire des isolats a été observé, confirmant que les phénomènes de recombinaison sont rares chez *S. lugdunensis*^{98,208}. Ainsi, la MLVA présentait l'avantage de prédire le CC voire le ST définis par MLST. En effet, le calcul du coefficient de Wallace a montré que deux isolats d'un même MT avait 83% de probabilité de partager le même ST. A l'inverse, deux isolats d'un même ST n'avaient que 52% de probabilité d'appartenir au même MT, confirmant par ailleurs que la technique de typage de MLVA était bien plus discriminante que la MLST.

Un autre avantage de cette technique est qu'elle permettait de discriminer les isolats de *S. lugdunensis* au sein de certains CC. Par exemple, la MLVA subdivisait particulièrement le CC1 dont les 37 isolats appartenaient par MLST à 5 ST différents alors que par MLVA ces mêmes isolats appartenaient à 22 MT distincts. De la même façon, il a été montré chez d'autres espèces clonales comme *S. agalactiae*²¹⁸ ou *L. monocytogenes*²¹⁷ que la MLVA permettait la discrimination de nouveaux génotypes au sein de certains CC. Cependant, certains CC comme le CC2 étaient moins discriminés par MLVA ; les 18 isolats de ce CC appartenaient à cinq ST différents alors que par MLVA ces mêmes isolats appartenaient seulement à trois MT. Pris ensemble, ces résultats suggèrent des évolutions différentes entre ces lignées phylogénétiques ; celles pour lesquelles la MLVA ne permet pas de définir plus de génotypes suggèrent une émergence récente alors que celles pour lesquelles les VNTR ont eu le temps d'évoluer seraient plus anciennes.

La MLVA est généralement considérée comme une méthode pouvant conduire à une classification erronée de certains isolats du fait de phénomènes de recombinaison et/ou

d'homoplasie, phénomènes définis par l'observation d'un même nombre de répétitions entre deux isolats, indépendant d'un ancêtre commun^{241,286}. Le séquençage des VNTR n'a donc pas été utilisé tant pour la vérification du nombre de motifs répétés que pour l'apport des SNP moins assujettis à l'homoplasie. L'identification des mutations au sein des VNTR a permis d'augmenter sensiblement leur DI. La méthode de TRST s'est ainsi montrée plus discriminante que la MLVA tant en termes de génotypes (55 MT vs 69 TRT) que de DI (0,933 vs 0,943).

Dans notre étude, aucune corrélation entre tableau clinique et génotype n'a été mise en évidence, et donc aucune lignée hyper-virulente ou susceptible d'être associée à des formes d'infection plus sévères n'a été identifiée. Cela suggère que les méthodes de typage telles que la MLST ou la MLVA ne reflètent probablement pas le pouvoir pathogène des souches de *S. lugdunensis* et/ou l'importance du rôle des facteurs de l'hôte dans la propension des isolats de *S. lugdunensis* à causer des infections. Il a été montré chez *S. aureus* que les données de typage de MLST ne permettent pas de prédire le potentiel infectieux des isolats, car celui-ci serait lié à l'évolution du génome accessoire et non à celui du *core-genome*¹⁸⁹.

Comme observé précédemment dans plusieurs études épidémiologiques chez *S. lugdunensis*^{24,208}, les isolats appartenant au CC1 et au CC3 étaient les plus prévalents dans notre étude, représentant respectivement 32% (31/98) et 27% (26/98) des isolats collectés lors de l'étude VISLISI¹¹. Au sein de ces CC, les ST les plus prévalents correspondaient aux génotypes ancestraux (ancêtre commun) de ces complexes. Fait intéressant, 23 isolats étaient regroupés en un même génotype quelle que soit la méthode de typage utilisée : ST3 (CC3) par MLST et MT1/TRT1 par MLVA et TRST. Ce clone particulièrement prévalent comprenait des isolats issus d'infection ou d'un contexte de colonisation. Ces isolats ont été collectés à partir de plusieurs patients de différents services hospitaliers sur une période de 24 mois, suggérant que ce clone était capable de persistance sur le long terme. Cheng *et al.*²³ ont déjà décrit des faits similaires dans leur étude à Taïwan dans laquelle ils avaient identifié à l'hôpital un clone de ST6 endémique résistant à la méticilline. A la différence de Cheng *et al.*²³, les isolats de ST3 de notre étude n'étaient pas résistants aux antibiotiques, ainsi leur persistance n'est pas due à une pression de sélection liée aux antibiotiques. Cette persistance pourrait alors être liée à une meilleure capacité d'adaptation de ce clone aux conditions de l'environnement, comme une capacité

élevée à former du biofilm. En effet, Kassem *et al.*²⁸⁷ ont d'ailleurs montré que *S. lugdunensis* pouvait être un contaminant fréquent de différentes surfaces hospitalières. Au total, ces données suggèrent la possible persistance de clones de *S. lugdunensis* à l'hôpital, mais des études supplémentaires sont nécessaires afin de déterminer son mode de transmission, son réservoir hospitalier et les mécanismes de cette persistance.

L'étude du polymorphisme des VNTR s'impose comme la méthode de génotypage la plus discriminante et la plus adaptée existant à ce jour pour les études à court et à long terme chez *S. lugdunensis*. La MLVA et la TRST sont à la fois des outils performants pour les études micro-épidémiologiques mais aussi pour caractériser l'évolution de la population de cette espèce bactérienne. En outre, ces méthodes, fiables et reproductibles, génèrent des données non-ambiguës facilement échangeables pour des comparaisons entre laboratoires. De plus, la démocratisation dans de nombreuses structures de recherche et la diminution constante des coûts du séquençage haut débit permettront d'enrichir la base de données de ces méthodes, contribuant à améliorer le suivi de l'épidémiologie globale de *S. lugdunensis*.

Lors de la mise au point des méthodes de génotypage de MLVA et de TRST précédemment décrites, un *locus* de VNTR particulièrement intéressant a été identifié. Ce dernier, localisé dans la région R du gène *fbl*, présentait un tel pouvoir discriminant, qu'il nous a semblé pertinent de l'utiliser pour développer un schéma de typage basé sur ce seul *locus*, appelé *fbl-typing*.

Un second argument en faveur d'une analyse de ce VNTR à part entière est le fait que le gène *fbl* code une protéine présentant une identité de 58% avec la protéine ClfA de *S. aureus*⁶⁶. Comme cette dernière, la protéine Fbl contient des motifs de répétition SD en nombre variable. Les protéines composées de ces motifs appartiennent à la famille des adhésines Sdr telles que ClfB, SdrC, SrdD et SdrE de *S. aureus*²⁸⁸ et SdrF et SdrG de *S. epidermidis*²⁷³. Chez *S. aureus*, des méthodes de typage basées sur le nombre de motifs répétés²¹⁹ et sur les SNP présents au sein de ces motifs ont déjà prouvé leur efficacité²⁸⁹. Ainsi, nous avons évalué le pouvoir discriminant d'un nouveau schéma de typage pour l'espèce de *S. lugdunensis* basé sur le séquençage de la région R du gène *fbl*. Ce pouvoir discriminant a ensuite été comparé à ceux des méthodes de MLST et de TRST.

La collection de souches utilisées dans cette étude correspondait aux recommandations formulées par van Belkum *et al.*²⁸⁵ puisque nous avons utilisé comme panel "test" les 128 souches de la précédente étude pour le développement des méthodes de typage de MLVA et de TRST. De plus, pour évaluer le pouvoir discriminant, une collection de 106 isolats cliniques, récoltés sur l'année 2016 à l'hôpital universitaire de Rouen, a également été testée.

Le schéma de *fbl-typing* appliqué aux 240 isolats de *S. lugdunensis* a montré une typabilité de 100% et a permis de décrire 92 *fbl-types* distincts regroupés en dix *clusters*. Ces *clusters* ont été définis sur la base de la distance génétique entre les *fbl-types* en utilisant les paramètres par défaut du logiciel BioNumerics. Il pourrait alors être intéressant de développer un logiciel basé sur un algorithme plus adapté afin de mieux définir les *clusters* comme celui déjà existant, *Based Upon Repeat Pattern* (BURP), pour le *spa-typing* chez *S. aureus*²⁹⁰. Au vu de la similitude de ces deux méthodes de typage, la mise au point de cet outil informatique pourrait être facilement mise en place.

La comparaison du pouvoir discriminant des différentes méthodes de typage a montré que le *fbl-typing* était plus discriminant en termes de génotypes et de DI que la MLST et aussi discriminant que la TRST. Par ailleurs, nous avons observé une excellente congruence entre le regroupement des souches par MLST et *fbl-typing*. En effet, le *fbl-type* permettait de prédire le CC défini par MLST avec 100% de probabilité. Ainsi, le *fbl-typing* pourrait être utilisé comme outil de typage de *S. lugdunensis* en première intention en routine et notamment pour suivre la dissémination de clones endémiques résistants à la pénicilline comme décrits en Asie^{23,24}. Cette méthode permet tout à la fois de discriminer les isolats entre eux, en générant de nombreux génotypes, et de prédire leur CC. Cela suggère alors que la région R du gène *fbl* n'évolue pas par recombinaison, mais qu'elle est assujettie à des mécanismes d'évolution indépendants. Le premier est un mécanisme d'évolution lent d'apparition de SNP au sein des motifs répétés du gène *fbl*. Le second est un mécanisme d'évolution rapide de la région R liée à la modification du nombre de motifs répétés par *strand-slippage mispairing*²¹³. De plus, ce phénomène peut expliquer pourquoi le *fbl-typing* est aussi discriminant que la TRST qui se base sur le polymorphisme de sept *loci*. En effet, les motifs de 18 pb de la région R du gène *fbl* sont probablement plus sujets à des duplications et des délétions par *strand-slippage mispairing* du fait de leur petite taille par rapport aux motifs répétés (de 50 pb en moyenne) des *loci* utilisés dans le schéma de

TRST. Ainsi, cette méthode peut être utilisée pour des études micro-épidémiologiques, grâce à l'évolution rapide du nombre de motifs répétés, mais également macro-épidémiologique par l'étude des SNP au sein des motifs répétés, stables dans le temps.

Le principal avantage de cette méthode de *fbl-typing* est qu'elle permet de déterminer plusieurs génotypes pour un isolat (*fbl-type* et CC) en ne séquençant qu'un seul *locus* par rapport à la MLST et la TRST qui en utilisent sept. L'utilisation d'un unique *locus* est par nature moins coûteuse et moins chronophage que les techniques de MLST et TRST. En outre, cette méthode génère des données non-ambiguës et échangeables entre les laboratoires, et ce d'autant plus que nous avons développé un outil en ligne utilisable par tous pour une identification rapide du *fbl-type* d'un isolat à partir d'une séquence fasta, permettant la création d'une base de données pour des études à grande échelle.

La protéine Fbl a été décrite comme la protéine majeure de liaison au fibrinogène chez *S. lugdunensis*⁶⁶. Nous avons alors évalué l'impact de la diversité génétique du gène *fbl* (la présence d'une délétion à l'extrémité 5' et/ou le nombre de répétitions du domaine R) sur la capacité des souches à lier le fibrinogène immobilisé *in vitro* chez 55 isolats représentatifs de la variabilité du gène. Malgré la présence du gène *fbl* chez tous les isolats, seuls 31% (17/55) des isolats avaient la capacité de lier le fibrinogène. Cette variation dans la capacité de liaison avait aussi été observée chez 133 souches de portage de *S. aureus*²⁹¹. Plusieurs raisons peuvent expliquer ces différences.

En effet, il existe chez *S. lugdunensis* plusieurs gènes codant des protéines potentiellement impliquées dans la liaison au fibrinogène telles que la protéine putative de liaison au fibrinogène homologue à la protéine FbpA de *S. aureus*^{8,73}, l'autolysine AtlL^{79,81} et la protéine SlsD⁷³. Ces protéines pourraient contribuer à des phénomènes compensatoires pour la liaison au fibrinogène.

Cette variabilité pourrait être expliquée par la présence d'une délétion de 4 pb dans la région 5' du gène *fbl* entraînant un décalage du cadre de lecture. Nous avons observé que les isolats qui présentaient cette délétion liaient significativement moins le fibrinogène. Pour confirmer ces résultats, il serait intéressant d'étudier par *western blot* l'expression de la protéine Fbl et de vérifier sa localisation cellulaire par l'étude comparée de la fraction totale et du surnageant de souches présentant ou non la délétion. Par ailleurs, l'étude de la distribution de cette délétion au sein de 146 isolats a montré qu'elle était CC-dépendante,

puisque retrouvée presque exclusivement chez des souches de CC1. Parmi ces isolats de CC1, la délétion a été identifiée chez la majorité des isolats appartenant au génotype ancestral, le ST6. De plus, la délétion a été retrouvée chez tous les isolats de ST12 et ST15, mais aucun de ST1 et de ST7. Comme les isolats de ces quatre ST sont issus du génotype ancestral (ST6), ces résultats suggèrent que la délétion a pu initialement apparaître chez des isolats de ST6 et être transmise aux isolats appartenant aux ST12 et ST15, mais pas à ceux de ST1 et ST7. Cette hypothèse devra être validée sur une plus large collection d'isolats appartenant au CC1.

Une autre explication de la variabilité dans la capacité de liaison au fibrinogène pourrait être imputée au nombre de motifs répétés de la protéine Fbl. Par analogie avec les adhésines Sdr, la région R de la protéine Fbl s'étendrait à travers la couche épaisse du peptidoglycane et permettrait l'exposition du domaine de liaison au fibrinogène à la surface bactérienne. En effet, chez *S. aureus*, il a été montré pour la protéine ClfA qu'un nombre de motifs SD minimum était requis pour l'exposition optimale de son domaine de liaison⁷¹. Par ailleurs, il a été montré chez *S. agalactiae* la présence d'un VNTR au sein de la séquence du gène codant la protéine de liaison au fibrinogène FbsA²⁹². Pour cette protéine un nombre de 4 à 7 motifs répétés était corrélé à sa capacité maximale de liaison au fibrinogène. Nos résultats ont montré que les isolats présentant plus de 43 motifs répétés liaient le fibrinogène de façon significativement plus importante que les souches qui en présentaient un nombre inférieur. Ainsi, le nombre de motifs répétés pourrait être déterminant dans le rôle de la protéine Fbl. Cette tendance pourrait être confirmée par le clonage de variants du gène *fbl* présentant un nombre de motifs répétés différents chez *Lactococcus lactis* et l'étude de leur capacité de liaison au fibrinogène.

Toutefois, plusieurs isolats qui ne liaient pas le fibrinogène comportaient également plus de 43 motifs répétés dans la région R du gène *fbl* et ne présentaient pas la délétion. Nous avons alors évalué si la capacité à lier le fibrinogène pouvait être corrélée à une différence d'expression du gène *fbl* entre les souches. Nous avons étudié l'expression du gène *fbl* de 5 isolats qui liaient le fibrinogène de façon importante et cinq isolats qui ne le liaient pas, mais aucune différence significative d'expression n'a été observée entre les deux groupes d'isolats. Par ailleurs, aucune variation n'a été identifiée dans la région promotrice du gène *fbl*. L'expression du gène n'a été évaluée que sur un faible nombre d'isolats et devra donc être poursuivie sur un plus grand nombre de souches.

Une dernière hypothèse sur la variabilité de la capacité de liaison au fibrinogène entre les souches pourrait être imputée à la capsule polysaccharidique de *S. lugdunensis*. En effet, il a été montré que la capsule de *S. aureus* pouvait contribuer au masquage de la protéine ClfA et interférer avec ses propriétés d'adhésion²⁹³. Comme évoqué précédemment, la capsule de *S. lugdunensis* n'a que très peu été étudiée ; il serait donc intéressant de déterminer la présence ou non d'une capsule chez les souches qui ne lient pas le fibrinogène.

Dans notre étude, aucune corrélation entre le *fbl*-type et le contexte clinique n'a été observée, suggérant que la variabilité de la région R du gène *fbl* n'est pas un déterminant de la virulence de *S. lugdunensis*. De manière plus surprenante, aucun lien entre la capacité de liaison au fibrinogène et le contexte clinique n'a été observé. Par exemple, un isolat responsable d'une endocardite infectieuse s'est avéré incapable de lier le fibrinogène alors qu'à l'inverse un isolat de colonisation liait fortement le fibrinogène. Il a été montré expérimentalement chez le rat que la protéine ClfA de *S. aureus* était essentielle à la colonisation et à l'invasion des valves cardiaques²⁹⁴. *A contrario*, il a aussi été montré que la capacité de liaison au fibrinogène de *S. aureus* n'était pas corrélée avec la virulence des souches dans un modèle d'endocardite chez le rat²⁹¹. Ythier *et al.*²⁹¹ ont alors suggéré que ce paradoxe pouvait être dû à une différence d'expression des gènes *in vivo*, aux facteurs de l'hôte et à une redondance fonctionnelle des adhésines. Heilbronner *et al.*¹⁴⁰ ont montré précédemment chez *S. lugdunensis* qu'un mutant de délétion du gène *fbl* présentait une virulence diminuée mais non significative par rapport à la souche parentale dans un modèle d'endocardite chez le rat. Il pourrait être intéressant de comparer la virulence de plusieurs isolats cliniques de *S. lugdunensis* et de leurs mutants de délétion du gène *fbl* dans des modèles d'endocardite chez le rat ou le lapin, mais également dans des modèles murins de colonisation afin d'étudier l'implication de la protéine Fbl dans l'adhésion aux tissus de l'hôte.

En conclusion, nous avons démontré que la région R du gène *fbl* est très variable au sein de l'espèce de *S. lugdunensis*. Sur la base du polymorphisme d'un seul *locus*, le *fbl*-typing a montré un pouvoir très discriminant par rapport à celui de la méthode de référence de MLST et équivalent à celui de la TRST. En plus de son haut pouvoir discriminant, le *fbl*-typing permet de prédire le CC des isolats avec une très grande fiabilité et peut donc servir en première intention pour le typage des isolats de *S. lugdunensis* en

routine. Le développement d'un outil accessible en ligne pour l'identification du *fbl*-type permettra d'enrichir la base de données avec de nouvelles séquences et permettra un suivi de l'épidémiologie mondiale pour cette espèce de SCN à la virulence particulière. Par ailleurs, nous avons montré une grande diversité phénotypique dans la capacité de liaison au fibrinogène humain immobilisé *in vitro* pouvant être expliquée par la diversité génétique du gène. Il reste à déterminer si ces différences ont un impact sur la production, l'export, l'ancrage et l'expression de la protéine Fbl à la surface de la paroi bactérienne, mais également un retentissement sur la colonisation des tissus de l'hôte.

CONCLUSION GENERALE

Le travail présenté dans ce mémoire s'inscrit dans la thématique de recherche du "Groupe de Recherche sur l'Adaptation Microbienne" EA2656 ; il a consisté en une analyse comparative de génomes complets de souches pathogènes et de portage de *Staphylococcus lugdunensis* et en une caractérisation du système de sécrétion Ess/type VII chez cette espèce.

Ces travaux ont consisté en un séquençage du génome complet de trois souches pathogènes et trois souches de portage pour les comparer aux 15 génomes de *S. lugdunensis* déjà disponibles sur la base de données du NCBI. L'analyse génomique comparative a confirmé le caractère fermé du *pan-genome* de *S. lugdunensis*, mais la présence de plusieurs éléments génétiques mobiles portant des gènes de virulence putatifs et des gènes de résistance suggère des transferts horizontaux avec d'autres espèces de staphylocoques. Aucun déterminant génétique associé spécifiquement au contexte de virulence ou de portage de *S. lugdunensis* n'a été identifié au cours de notre étude. Cependant, la présence d'îlots génomiques et de variations CC-dépendantes au sein de *loci* potentiellement associés à la virulence, notamment chez les souches appartenant au CC3, offre plusieurs pistes à explorer.

La plus notable de ces variations CC-dépendantes est celle identifiée au sein du *locus* codant le système de sécrétion *ess/type VII*. Nous avons mis en évidence jusqu'à huit organisations génétiques différentes chez une espèce présentant pourtant une structure de population clonale. Nos travaux de caractérisation phénotypique suggèrent que ce système de sécrétion pourrait contribuer à la translocation de protéines prédites *in silico* comme sécrétées et impliquées dans la virulence. Néanmoins, dans les modèles d'infection *in vitro* de cytotoxicité cellulaire et d'infection *in vivo* du nématode *Caenorhabditis elegans*, aucune atténuation de la virulence n'a été observée chez la souche délétée du gène *essC* malgré une perte de sa capacité à lyser les érythrocytes, comparativement à la souche sauvage. Ces travaux seront poursuivis par l'étude comparée des sécrétomes afin de confirmer le rôle du système de sécrétion Ess/type VII et plus particulièrement de la protéine EssC, chez *S. lugdunensis*. Il serait également pertinent d'évaluer la virulence de ces souches dans des modèles murins de colonisation et de formation d'abcès cutanés.

La mise en perspective de nos travaux avec la revue de la littérature souligne le faible nombre de génomes disponibles et d'approches phénotypiques chez *S. lugdunensis* par rapport à *S. aureus*. Il conviendrait de développer de telles études sur de larges collections de souches pour appréhender les mécanismes qui sous-tendent la virulence de cette espèce de staphylocoque à coagulase négative.

Nos travaux ont également permis de développer et d'évaluer le pouvoir discriminant de trois nouvelles méthodes de typage (MLVA, TRST et *fbl-typing*) constituant des outils très prometteurs pour l'épidémiologie moléculaire des infections à *S. lugdunensis*. Nous avons de plus créé pour le *fbl-typing* un logiciel accessible en ligne qui permettra d'enrichir la base de données avec de nouvelles séquences et sera de ce fait un outil de premier plan pour le suivi de l'épidémiologie mondiale de ces infections.

REFERENCES BIBLIOGRAPHIQUES

1. Freney, J. *et al.* *Staphylococcus lugdunensis* sp. nov. and *Staphylococcus schleiferi* sp. nov., two species from human clinical specimens. *International Journal of Systematic Bacteriology* **38**, 168–172 (1988).
2. Frank, K. L., del Pozo, J. L. & Patel, R. From Clinical Microbiology to Infection Pathogenesis: How Daring To Be Different Works for *Staphylococcus lugdunensis*. *Clinical Microbiology Reviews* **21**, 111–133 (2008).
3. Fleurette, J. *et al.* Clinical isolates of *Staphylococcus lugdunensis* and *S. schleiferi*: bacteriological characteristics and susceptibility to antimicrobial agents. *Res. Microbiol.* **140**, 107–118 (1989).
4. Seifert, H., Oltmanns, D., Becker, K., Wisplinghoff, H. & von Eiff, C. *Staphylococcus lugdunensis* pacemaker-related infection. *Emerging Infect. Dis.* **11**, 1283–1286 (2005).
5. Leung, M. J., Nuttall, N., Pryce, T. M., Coombs, G. W. & Pearman, J. W. Colony Variation in *Staphylococcus lugdunensis*. *J Clin Microbiol.* 1998 Oct;**36**(10):3096-8.
6. Argemi, X. *et al.* *Staphylococcus lugdunensis* small colony variant conversion resulting in chronic prosthetic joint infection. *Médecine et Maladies Infectieuses* **47**, 498–501 (2017).
7. Tan, T. Y., Ng, S. Y. & He, J. Microbiological Characteristics, Presumptive Identification, and Antibiotic Susceptibilities of *Staphylococcus lugdunensis*. *J Clin Microbiol* **46**, 2393–2395 (2008).
8. Szabados, F. *et al.* Occurrence of genes of putative fibrinogen binding proteins and hemolysins, as well as of their phenotypic correlates in isolates of *S. lugdunensis* of different origins. *BMC research notes* **4**, 113 (2011).
9. Vandenesch, F. *et al.* Delta-like haemolysin produced by *Staphylococcus lugdunensis*. *FEMS Microbiol Lett* **62**, 65–68 (1991).
10. Chatzigeorgiou, K.-S., Siafakas, N., Petinaki, E. & Zerva, L. *fbl* gene as a species-specific target for *Staphylococcus lugdunensis* identification. *Journal of Clinical Laboratory Analysis* **24**, 119–122 (2010).
11. Argemi, X. *et al.* VISLISI trial, a prospective clinical study allowing identification of a new metalloprotease and putative virulence factor from *Staphylococcus lugdunensis*. *Clin. Microbiol. Infect.* **23**, 334.e1-334.e8 (2017).
12. Argemi, X. *et al.* Implementation of Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry in Routine Clinical Laboratories Improves Identification of Coagulase-Negative Staphylococci and Reveals the Pathogenic Role of *Staphylococcus lugdunensis*. *Journal of Clinical Microbiology* **53**, 2030–2036 (2015).
13. van der Mee-Marquet, N. *et al.* *Staphylococcus lugdunensis* Infections: High Frequency of Inguinal Area Carriage. *Journal of Clinical Microbiology* **41**, 1404–1409 (2003).
14. Bieber, L. & Kahlmeter, G. *Staphylococcus lugdunensis* in several niches of the normal skin flora.

- Clinical Microbiology and Infection* **16**, 385–388 (2010).
15. Bocher, S., Tønning, B., Skov, R. L. & Prag, J. *Staphylococcus lugdunensis*, a Common Cause of Skin and Soft Tissue Infections in the Community. *Journal of Clinical Microbiology* **47**, 946–950 (2009).
 16. Davis, M. F., Cain, C. L., Brazil, A. M. & Rankin, S. C. Two coagulase-negative staphylococci emerging as potential zoonotic pathogens: wolves in sheep's clothing? *Front Microbiol* **4**, (2013).
 17. Kasprówicz, A. *et al.* The Occurrence and Comparative Phenotypic Characteristics of *Staphylococcus spp.* from Healthy and Diseased, Household and Shelter Dogs, Based on Routine Biochemical Diagnostic Methods. **8**
 18. Deinhofer, M. *Staphylococcus spp.* as mastitis-related pathogens in goat milk. *Veterinary Microbiology* **43**, 161–166 (1995).
 19. Higaki, S., Kitagawa, T., Morohashi, M. & Yamagishi, T. Distribution and antimicrobial susceptibility of coagulase-negative staphylococci from skin lesions. *J. Int. Med. Res.* **27**, 191–195 (1999).
 20. McHardy, I. H. *et al.* Clinical and Microbiological Aspects of β -Lactam Resistance in *Staphylococcus lugdunensis*. *Journal of Clinical Microbiology* **55**, 585–595 (2017).
 21. Lin, J.-F. *et al.* Clinical experience and microbiologic characteristics of invasive *Staphylococcus lugdunensis* infection in a tertiary center in northern Taiwan. *Journal of Microbiology, Immunology and Infection* **48**, 406–412 (2015).
 22. Ferreira, A. M., Martins, K. B., Silva, V. R. da, Mondelli, A. L. & Cunha, M. de L. R. de S. da. Correlation of phenotypic tests with the presence of the blaZ gene for detection of beta-lactamase. *Brazilian Journal of Microbiology* **48**, 159–166 (2017).
 23. Cheng, C. *et al.* Persistence of a major endemic clone of oxacillin-resistant *Staphylococcus lugdunensis* sequence type 6 at a tertiary medical centre in northern Taiwan. *Int. J. Infect. Dis.* **36**, 72–77 (2015).
 24. Yeh, C.-F. *et al.* Clinical Features, Outcomes, and Molecular Characteristics of Community- and Health Care-Associated *Staphylococcus lugdunensis* Infections. *J. Clin. Microbiol.* **54**, 2051–2057 (2016).
 25. Sader, H. S. & Jones, R. N. Antimicrobial activity of daptomycin in comparison to glycopeptides and other antimicrobials when tested against numerous species of coagulase-negative *Staphylococcus*. *Diagnostic Microbiology and Infectious Disease* **73**, 212–214 (2012).
 26. Yeh, C.-F. *et al.* Molecular Characteristics of Disease-Causing and Commensal *Staphylococcus lugdunensis* Isolates from 2003 to 2013 at a Tertiary Hospital in Taiwan. *PLOS ONE* **10**, e0134859 (2015).
 27. Bourgeois, I., Pestel-Caron, M., Lemeland, J.-F., Pons, J.-L. & Caron, F. Tolerance to the Glycopeptides Vancomycin and Teicoplanin in Coagulase-Negative Staphylococci. *Antimicrobial Agents and Chemotherapy* **51**, 740–743 (2007).
 28. Poitevin-Later, F., Vandenesch, F., Dyke, K., Fleurette, J. & Etienne, J. Cadmium-resistance plasmid in *Staphylococcus lugdunensis*. *FEMS Microbiol. Lett.* **78**, 59–63 (1992).

29. Chaouni, L. B., Etienne, J., Greenland, T. & Vandenesch, F. Nucleic acid sequence and affiliation of pLUG10, a novel cadmium resistance plasmid from *Staphylococcus lugdunensis*. *Plasmid* **36**, 1–8 (1996).
30. Becker, K., Heilmann, C. & Peters, G. Coagulase-Negative Staphylococci. *Clinical Microbiology Reviews* **27**, 870–926 (2014).
31. Tan, T. Y., Ng, S. Y. & Ng, W. X. Clinical Significance of Coagulase-Negative Staphylococci Recovered from Nonsterile Sites. *Journal of Clinical Microbiology* **44**, 3413–3414 (2006).
32. Wu, A.-B. *et al.* Clinical and microbiological characteristics of community-acquired *Staphylococcus lugdunensis* infections in Southern Taiwan. *J. Clin. Microbiol.* **49**, 3015–3018 (2011).
33. Herchline, T. E. Occurrence of *Staphylococcus lugdunensis* in Consecutive Clinical Cultures and Relationship of Isolation to Infection. 3
34. Heldt Manica, L. A. & Cohen, P. R. *Staphylococcus lugdunensis* Infections of the Skin and Soft Tissue: A Case Series and Review. *Dermatology and Therapy* **7**, 555–562 (2017).
35. Arias, M. *et al.* Skin and soft tissue infections caused by *Staphylococcus lugdunensis*: Report of 20 cases. *Scandinavian Journal of Infectious Diseases* **42**, 879–884 (2010).
36. Delaunay, F., Pegot, A., Coquerel-Beghin, D., Aktouf, A. & Auquit-Auckbur, I. Fasciites nécrosantes à *Staphylococcus lugdunensis* après dermolipectomie abdominale : à propos de deux cas et revue de la littérature. *Annales de Chirurgie Plastique Esthétique* **59**, 136–139 (2014).
37. Zaaroura, H., Geffen, Y., Bergman, R. & Avitan-Hersh, E. Clinical and microbiological properties of *Staphylococcus lugdunensis* skin infections. *The Journal of Dermatology* **45**, 994–999 (2018).
38. Bellamy, R. & Barkham, T. *Staphylococcus lugdunensis* Infection Sites: Predominance of Abscesses in the Pelvic Girdle Region. *Clinical Infectious Diseases* **35**, e32–e34 (2002).
39. Guet-Revillet, H. *et al.* Bacterial Pathogens Associated with Hidradenitis Suppurativa, France. *Emerging Infectious Diseases* **20**, 1990–1998 (2014).
40. Douiri, N. *et al.* *Staphylococcus lugdunensis*: a virulent pathogen causing bone and joint infections. *Clinical Microbiology and Infection* **22**, 747–748 (2016).
41. Seng, P. *et al.* *Staphylococcus lugdunensis*: An Underreported Pathogen in Osteomyelitis. *The Journal of Foot and Ankle Surgery* **56**, 412–413 (2017).
42. Shah, N. B. *et al.* Laboratory and clinical characteristics of *Staphylococcus lugdunensis* prosthetic joint infections. *J. Clin. Microbiol.* **48**, 1600–1603 (2010).
43. Li, A. *et al.* Metalware-associated orthopaedic infections caused by *Staphylococcus lugdunensis* : An emerging pathogen. *Journal of Infection* **75**, 368–370 (2017).
44. Lourtet-Hascoët, J., Bicart-See, A., Félicé, M. P., Giordano, G. & Bonnet, E. *Staphylococcus lugdunensis*, a serious pathogen in periprosthetic joint infections: comparison to *Staphylococcus aureus* and *Staphylococcus epidermidis*. *International Journal of Infectious Diseases* **51**, 56–61 (2016).
45. Lin, J.-F. *et al.* Clinical experience and microbiologic characteristics of invasive *Staphylococcus lugdunensis* infection in a tertiary center in northern Taiwan. *J Microbiol Immunol Infect* (2014). doi:10.1016/j.jmii.2013.12.010

46. Pfaller, M. A. *et al.* Survey of blood stream infections attributable to gram-positive cocci: frequency of occurrence and antimicrobial susceptibility of isolates collected in 1997 in the United States, Canada, and Latin America from the SENTRY Antimicrobial Surveillance Program. *Diagnostic microbiology and infectious disease* **33**, 283–297 (1999).
47. Choi, S.-H. *et al.* Incidence, characteristics, and outcomes of *Staphylococcus lugdunensis* bacteremia. *J. Clin. Microbiol.* **48**, 3346–3349 (2010).
48. Castro, J. G. & Dowdy, L. Septic shock caused by *Staphylococcus lugdunensis*. *Clinical infectious diseases* **28**, 681–682 (1999).
49. Tee, W. S. N., Soh, S. Y., Lin, R. & Loo, L. H. *Staphylococcus lugdunensis* Carrying the mecA Gene Causes Catheter-Associated Bloodstream Infection in Premature Neonate. *Journal of Clinical Microbiology* **41**, 519–520 (2003).
50. Non, L. R. & Santos, C. A. Q. The occurrence of infective endocarditis with *Staphylococcus lugdunensis* bacteremia: A retrospective cohort study and systematic review. *Journal of Infection* **74**, 179–186 (2017).
51. Liu, P.-Y. *et al.* *Staphylococcus lugdunensis* Infective Endocarditis: A Literature Review and Analysis of Risk Factors. *Journal of Microbiology, Immunology and Infection* **43**, 478–484 (2010).
52. Anguera, I. *Staphylococcus lugdunensis* infective endocarditis: description of 10 cases and analysis of native valve, prosthetic valve, and pacemaker lead endocarditis clinical profiles. *Heart* **91**, e10–e10 (2005).
53. Patel, R. *et al.* Frequency of Isolation of *Staphylococcus lugdunensis* among Staphylococcal Isolates Causing Endocarditis: a 20-Year Experience. *Journal of clinical microbiology* **38**, 4262–4263 (2000).
54. Sandoe, J. A. T. & Longshaw, C. M. Ventriculoperitoneal shunt infection caused by *Staphylococcus lugdunensis*. *Clinical Microbiology and Infection* **7**, 385–387 (2001).
55. Kaabia, N., Scauarda, D., Lena, G. & Drancourt, M. Molecular Identification of *Staphylococcus lugdunensis* in a Patient with Meningitis. *Journal of Clinical Microbiology* **40**, 1824–1825 (2002).
56. Chiquet, C. *et al.* Acute Postoperative Endophthalmitis Caused by *Staphylococcus lugdunensis*. *Journal of Clinical Microbiology* **45**, 1673–1678 (2007).
57. Pinna, A. *et al.* Identification and antibiotic susceptibility of coagulase negative staphylococci isolated in corneal/external infections. *British Journal of Ophthalmology* **83**, 771–773 (1999).
58. Schnitzler, N., Meilicke, R., Conrads, G., Frank, D. & Haase, G. *Staphylococcus lugdunensis*: report of a case of peritonitis and an easy-to-perform screening strategy. *Journal of clinical microbiology* **36**, 812–813 (1998).
59. Haile, D. T. *et al.* Frequency of Isolation of *Staphylococcus lugdunensis* in Consecutive Urine Cultures and Relationship to Urinary Tract Infection. *Journal of Clinical Microbiology* **40**, 654–656 (2002).
60. Mouw, J. K., Ou, G. & Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nature Reviews Molecular Cell Biology* **15**, 771–785 (2014).
61. Josse, J., Laurent, F. & Diot, A. Staphylococcal Adhesion and Host Cell Invasion: Fibronectin-

- Binding and Other Mechanisms. *Frontiers in Microbiology* **8**, (2017).
62. Paulsson, M., Petersson, A.-C. & Ljungh, A. Serum and tissue protein binding and cell surface properties of *Staphylococcus lugdunensis*. *Journal of Medical Microbiology* **38**, 96–102 (1993).
63. Speziale, P., Pietrocola, G., Foster, T. J. & Geoghegan, J. A. Protein-based biofilm matrices in Staphylococci. *Frontiers in Cellular and Infection Microbiology* **4**, (2014).
64. Ton-That, H., Marraffini, L. A. & Schneewind, O. Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1694**, 269–278 (2004).
65. Dramsi, S., Magnet, S., Davison, S. & Arthur, M. Covalent attachment of proteins to peptidoglycan. *FEMS Microbiology Reviews* **32**, 307–320 (2008).
66. Mitchell, J. Characterization of the fibrinogen-binding surface protein Fbl of *Staphylococcus lugdunensis*. *Microbiology* **150**, 3831–3841 (2004).
67. Nilsson, M., Bjerketorp, J., Guss, B. & Frykberg, L. A fibrinogen-binding protein of *Staphylococcus lugdunensis*. *FEMS Microbiology Letters* **241**, 87–93 (2004).
68. Nilsson, M. *et al.* A von Willebrand factor-binding protein from *Staphylococcus lugdunensis*. *FEMS Microbiol. Lett.* **234**, 155–161 (2004).
69. Liesenborghs, L. *et al.* Shear-Resistant Binding to von Willebrand Factor Allows *Staphylococcus lugdunensis* to Adhere to the Cardiac Valves and Initiate Endocarditis. *Journal of Infectious Diseases* **213**, 1148–1156 (2016).
70. Mcdevitt, D. *et al.* Characterization of the Interaction Between the *Staphylococcus aureus* Clumping Factor (ClfA) and Fibrinogen. *European Journal of Biochemistry* **247**, 416–424 (1997).
71. Hartford, O., Francois, P., Vaudaux, P. & Foster, T. J. The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. *Molecular Microbiology* **25**, 1065–1076 (1997).
72. Marlinghaus, L. *et al.* Construction and characterization of three knockout mutants of the *fbl* gene in *Staphylococcus lugdunensis*: characterization of isogenic mutants of *fbl*. *APMIS* **120**, 108–116 (2012).
73. Heilbronner, S. *et al.* Genome sequence of *Staphylococcus lugdunensis* N920143 allows identification of putative colonization and virulence factors: *Staphylococcus lugdunensis* genome sequence. *FEMS Microbiology Letters* **322**, 60–67 (2011).
74. Ponnuraj, K. *et al.* A “dock, lock, and latch” Structural Model for a Staphylococcal Adhesin Binding to Fibrinogen. *Cell* **115**, 217–228 (2003).
75. Johansson, M. U. *et al.* Communication: Solution Structure of the Albumin-binding GA Module: A Versatile Bacterial Protein Domain. **7**
76. Wästfelt, M., Stålhammar-Carlemalm, M., Delisse, A.-M., Cabezon, T. & Lindahl, G. Identification of a Family of Streptococcal Surface Proteins with Extremely Repetitive Structure. *Journal of Biological Chemistry* **271**, 18892–18897 (1996).
77. Vollmer, W., Joris, B., Charlier, P. & Foster, S. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiology Reviews* **32**, 259–286 (2008).
78. Albrecht, T., Raue, S., Rosenstein, R., Nieselt, K. & Gotz, F. Phylogeny of the Staphylococcal

- Major Autolysin and Its Use in Genus and Species Typing. *Journal of Bacteriology* **194**, 2630–2636 (2012).
79. Hussain, M., Steinbacher, T., Peters, G., Heilmann, C. & Becker, K. The adhesive properties of the *Staphylococcus lugdunensis* multifunctional autolysin AtlL and its role in biofilm formation and internalization. *International Journal of Medical Microbiology* **305**, 129–139 (2015).
80. Bourgeois, I. *et al.* Characterization of AtlL, a bifunctional autolysin of *Staphylococcus lugdunensis* with N -acetylglucosaminidase and N -acetylmuramoyl-l-alanine amidase activities. *FEMS Microbiology Letters* **290**, 105–113 (2009).
81. Gibert, L. *et al.* The major autolysin of *Staphylococcus lugdunensis*, AtlL, is involved in cell separation, stress-induced autolysis and contributes to bacterial pathogenesis. *FEMS Microbiology Letters* **352**, 78–86 (2014).
82. Gotz, F. *Staphylococcus* and biofilms. *Molecular Microbiology* **43**, 1367–1378 (2002).
83. Otto, M. Staphylococcal Infections: Mechanisms of Biofilm Maturation and Detachment as Critical Determinants of Pathogenicity. *Annual Review of Medicine* **64**, 175–188 (2013).
84. Fitzpatrick, F., Humphreys, H. & O’Gara, J. P. Evidence for icaADBC-Independent Biofilm Development Mechanism in Methicillin-Resistant *Staphylococcus aureus* Clinical Isolates. *Journal of Clinical Microbiology* **43**, 1973–1976 (2005).
85. Kogan, G., Sadovskaya, I., Chaignon, P., Chokr, A. & Jabbouri, S. Biofilms of clinical strains of *Staphylococcus* that do not contain polysaccharide intercellular adhesin. *FEMS Microbiology Letters* **255**, 11–16 (2006).
86. Otto, M. Staphylococcal biofilms. *Curr Top Microbiol Immunol* **322**, 207–228 (2008).
87. Arciola, C. R., Campoccia, D., Speziale, P., Montanaro, L. & Costerton, J. W. Biofilm formation in *Staphylococcus* implant infections. A review of molecular mechanisms and implications for biofilm-resistant materials. *Biomaterials* **33**, 5967–5982 (2012).
88. Otto, M. *Staphylococcus epidermidis* — the ‘accidental’ pathogen. *Nature Reviews Microbiology* **7**, 555–567 (2009).
89. O’Neill, E. *et al.* A Novel *Staphylococcus aureus* Biofilm Phenotype Mediated by the Fibronectin-Binding Proteins, FnBPA and FnBPB. *Journal of Bacteriology* **190**, 3835–3850 (2008).
90. Montanaro, L. *et al.* Extracellular DNA in Biofilms. *The International Journal of Artificial Organs* **34**, 824–831 (2011).
91. Frank, K. L., Reichert, E. J., Piper, K. E. & Patel, R. In Vitro Effects of Antimicrobial Agents on Planktonic and Biofilm Forms of *Staphylococcus lugdunensis* Clinical Isolates. *Antimicrobial Agents and Chemotherapy* **51**, 888–895 (2007).
92. Rajendran, N. B. *et al.* Important Contribution of the Novel Locus *comEB* to Extracellular DNA-Dependent *Staphylococcus lugdunensis* Biofilm Formation. *Infection and Immunity* **83**, 4682–4692 (2015).
93. Giormezis, N. *et al.* Virulence factors among *Staphylococcus lugdunensis* are associated with infection sites and clonal spread. *European Journal of Clinical Microbiology & Infectious Diseases* **34**,

773–778 (2015).

94. Argemi, X. *et al.* Kinetics of biofilm formation by *Staphylococcus lugdunensis* strains in bone and joint infections. *Diagnostic Microbiology and Infectious Disease* **88**, 298–304 (2017).
95. Grumann, D., Nübel, U. & Bröker, B. M. *Staphylococcus aureus* toxins – Their functions and genetics. *Infection, Genetics and Evolution* **21**, 583–592 (2014).
96. Argemi, X. *et al.* Whole genome sequencing of 7 strains of *Staphylococcus lugdunensis* allows identification of mobile genetic elements. *Genome Biology and Evolution* (2017). doi:10.1093/gbe/evx077
97. Hébert, G. A. Hemolysins and Other Characteristics That Help Differentiate and Biotype *Staphylococcus lugdunensis* and *Staphylococcus schleiferi*. *J. Clin. Microbiol.* **28**, 7 (1990).
98. Didi, J., Lemee, L., Gibert, L., Pons, J.-L. & Pestel-Caron, M. Multi-Virulence-Locus Sequence Typing of *Staphylococcus lugdunensis* Generates Results Consistent with a Clonal Population Structure and Is Reliable for Epidemiological Typing. *Journal of Clinical Microbiology* **52**, 3624–3632 (2014).
99. Donvito, B. *et al.* Synergistic hemolytic activity of *Staphylococcus lugdunensis* is mediated by three peptides encoded by a non-agr genetic locus. *Infection and immunity* **65**, 95–100 (1997).
100. Rautenberg, M., Joo, H.-S., Otto, M. & Peschel, A. Neutrophil responses to staphylococcal pathogens and commensals *via* the formyl peptide receptor 2 relates to phenol-soluble modulins release and virulence. *The FASEB Journal* **25**, 1254–1263 (2011).
101. Vandenesch, F., Lina, G. & Henry, T. *Staphylococcus aureus* Hemolysins, bi-component Leukocidins, and Cytolytic Peptides: A Redundant Arsenal of Membrane-Damaging Virulence Factors? *Frontiers in Cellular and Infection Microbiology* **2**, (2012).
102. Contreras, H., Chim, N., Credali, A. & Goulding, C. W. Heme uptake in bacterial pathogens. *Current Opinion in Chemical Biology* **19**, 34–41 (2014).
103. Grigg, J. C., Ukpabi, G., Gaudin, C. F. M. & Murphy, M. E. P. Structural biology of heme binding in the *Staphylococcus aureus* Isd system. *Journal of Inorganic Biochemistry* **104**, 341–348 (2010).
104. Zapotoczna, M., Heilbronner, S., Speziale, P. & Foster, T. J. Iron-regulated surface determinant (Isd) proteins of *Staphylococcus lugdunensis*. *J. Bacteriol.* **194**, 6453–6467 (2012).
105. Farrand, A. J. *et al.* An Iron-Regulated Autolysin Remodels the Cell Wall To Facilitate Heme Acquisition in *Staphylococcus lugdunensis*. *Infection and Immunity* **83**, 3578–3589 (2015).
106. Mazmanian, S. K. Passage of Heme-Iron Across the Envelope of *Staphylococcus aureus*. *Science* **299**, 906–909 (2003).
107. Reniere, M. L. & Skaar, E. P. *Staphylococcus aureus* haem oxygenases are differentially regulated by iron and haem. *Molecular Microbiology* **69**, 1304–1315 (2008).
108. Missineo, A. *et al.* IsdC from *Staphylococcus lugdunensis* Induces Biofilm Formation under Low-Iron Growth Conditions. *Infection and Immunity* **82**, 2448–2459 (2014).
109. Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516 (2016).
110. Bera, A., Biswas, R., Herbert, S. & Gotz, F. The Presence of Peptidoglycan O-Acetyltransferase in Various Staphylococcal Species Correlates with Lysozyme Resistance and Pathogenicity. *Infection*

- and *Immunity* **74**, 4598–4604 (2006).
111. Flannagan, R. S., Cosío, G. & Grinstein, S. Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nature Reviews Microbiology* **7**, 355–366 (2009).
112. Flannagan, R. S., Heit, B. & Heinrichs, D. E. Intracellular replication of *Staphylococcus aureus* in mature phagolysosomes in macrophages precedes host cell death, and bacterial escape and dissemination: *S. aureus* replicates in mature phagolysosomes in macrophages. *Cellular Microbiology* **18**, 514–535 (2016).
113. Flannagan, R. S., Watson, D. W., Surewaard, B. G. J., Kubes, P. & Heinrichs, D. E. The surreptitious survival of the emerging pathogen *Staphylococcus lugdunensis* within macrophages as an immune evasion strategy. *Cellular Microbiology* e12869 (2018). doi:10.1111/cmi.12869
114. Cassat, J. E. *et al.* A Secreted Bacterial Protease Tailors the *Staphylococcus aureus* Virulence Repertoire to Modulate Bone Remodeling during Osteomyelitis. *Cell Host & Microbe* **13**, 759–772 (2013).
115. Singh, R. & Ray, P. Quorum sensing-mediated regulation of staphylococcal virulence and antibiotic resistance. *Future Microbiology* **9**, 669–681 (2014).
116. Le, K. Y. & Otto, M. Quorum-sensing regulation in staphylococci—an overview. *Frontiers in Microbiology* **6**, (2015).
117. Vandenesch, F., Projan, S. J., Kreiswirth, B., Etienne, J. & Novick, R. P. Agr-related sequences in *Staphylococcus lugdunensis*. *FEMS Microbiol Lett* **111**, 115–122 (1993).
118. Thoendel, M., Kavanaugh, J. S., Flack, C. E. & Horswill, A. R. Peptide Signaling in the Staphylococci. *Chemical Reviews* **111**, 117–151 (2011).
119. Painter, K. L., Krishna, A., Wigneshweraraj, S. & Edwards, A. M. What role does the quorum-sensing accessory gene regulator system play during *Staphylococcus aureus* bacteremia? *Trends in Microbiology* **22**, 676–685 (2014).
120. Jarraud, S. *et al.* Exfoliatin-Producing Strains Define a Fourth agr Specificity Group in *Staphylococcus aureus*. *J. Bacteriol.* **182**, 6 (2000).
121. Tseng, S.-P. *et al.* Genotypes and phenotypes of *Staphylococcus lugdunensis* isolates recovered from bacteremia. *Journal of Microbiology, Immunology and Infection* **48**, 397–405 (2015).
122. Argemi, X., Hansmann, Y., Riegel, P. & Prévost, G. Is *Staphylococcus lugdunensis* Significant in Clinical Samples? *Journal of Clinical Microbiology* **55**, 3167–3174 (2017).
123. Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31–46 (2010).
124. Levy, S. E. & Myers, R. M. Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics* **17**, 95–115 (2016).
125. Desai, A. & Jere, A. Next-generation sequencing: ready for the clinics? *Clinical Genetics* **81**, 503–510 (2012).
126. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
127. Trapnell, C. & Salzberg, S. L. How to map billions of short reads onto genomes. *Nature*

- Biotechnology* **27**, 455–457 (2009).
128. Faino, L. & Thomma, B. P. H. J. Get your high-quality low-cost genome sequence. *Trends in Plant Science* **19**, 288–291 (2014).
129. Sohn, J. & Nam, J.-W. The present and future of *de novo* whole-genome assembly. *Briefings in Bioinformatics* bbw096 (2016). doi:10.1093/bib/bbw096
130. Liao, Y.-C., Lin, H.-H., Sabharwal, A., Haase, E. M. & Scannapieco, F. A. MyPro: A seamless pipeline for automated prokaryotic genome assembly and annotation. *Journal of Microbiological Methods* **113**, 72–74 (2015).
131. Koren, S., Treangen, T. J., Hill, C. M., Pop, M. & Phillippy, A. M. Automated ensemble assembly and validation of microbial genomes. *BMC Bioinformatics* **15**, 126 (2014).
132. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**, 563–569 (2013).
133. Tse, H. *et al.* Complete Genome Sequence of *Staphylococcus lugdunensis* Strain HKU09-01. *Journal of Bacteriology* **192**, 1471–1472 (2010).
134. Malachowa, N. & DeLeo, F. R. Mobile genetic elements of *Staphylococcus aureus*. *Cellular and Molecular Life Sciences* **67**, 3057–3071 (2010).
135. McCarthy, A. J. *et al.* Extensive Horizontal Gene Transfer during *Staphylococcus aureus* Co-colonization In Vivo. *Genome Biology and Evolution* **6**, 2697–2708 (2014).
136. Deghorain, M. & Van Melderen, L. The Staphylococci Phages Family: An Overview. *Viruses* **4**, 3316–3335 (2012).
137. Monk, I. R., Shah, I. M., Xu, M., Tan, M.-W. & Foster, T. J. Transforming the Untransformable: Application of Direct Transformation To Manipulate Genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *mBio* **3**, (2012).
138. Sadykov, M. R. Restriction–Modification Systems as a Barrier for Genetic Manipulation of *Staphylococcus aureus*. in *The Genetic Manipulation of Staphylococci* (ed. Bose, J. L.) **1373**, 9–23 (Springer New York, 2014).
139. Waldron, D. E. & Lindsay, J. A. Sau1: a Novel Lineage-Specific Type I Restriction-Modification System That Blocks Horizontal Gene Transfer into *Staphylococcus aureus* and between *S. aureus* Isolates of Different Lineages. *Journal of Bacteriology* **188**, 5578–5585 (2006).
140. Heilbronner, S., Hanses, F., Monk, I. R., Speziale, P. & Foster, T. J. Sortase A Promotes Virulence in Experimental *Staphylococcus lugdunensis* Endocarditis. *Microbiology* (2013). doi:10.1099/mic.0.070292-0
141. Argemi, X. *et al.* Comparative genomic analysis of *Staphylococcus lugdunensis* shows a closed pan-genome and multiple barriers to horizontal gene transfer. *BMC Genomics* **19**, (2018).
142. Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P. & van der Oost, J. The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiology and Molecular Biology Reviews* **78**, 74–88 (2014).
143. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas System and Its Role in Phage-Bacteria

- Interactions. *Annual Review of Microbiology* **64**, 475–493 (2010).
144. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. The basic building blocks and evolution of CRISPR–Cas systems. *Biochemical Society Transactions* **41**, 1392–1400 (2013).
145. Barrangou, R. CRISPR-Cas systems and RNA-guided interference: CRISPR-Cas systems and RNA-guided interference. *Wiley Interdisciplinary Reviews: RNA* **4**, 267–278 (2013).
146. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology* **13**, 722–736 (2015).
147. Schuster, C. F. *et al.* Characterization of a mazEF Toxin-Antitoxin Homologue from *Staphylococcus equorum*. *Journal of Bacteriology* **195**, 115–125 (2013).
148. Schuster, C. & Bertram, R. Toxin-Antitoxin Systems of *Staphylococcus aureus*. *Toxins* **8**, 140 (2016).
149. Fineran, P. C. *et al.* The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. *Proceedings of the National Academy of Sciences* **106**, 894–899 (2009).
150. Bélanger, M. & Moineau, S. Mutational Analysis of the Antitoxin in the Lactococcal Type III Toxin-Antitoxin System AbiQ. *Applied and Environmental Microbiology* **81**, 3848–3855 (2015).
151. Carlos Guimaraes, L. *et al.* Inside the Pan-genome - Methods and Software Overview. *Current Genomics* **16**, 245–252 (2015).
152. Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. The microbial pan-genome. *Current Opinion in Genetics & Development* **15**, 589–594 (2005).
153. Costa, T. R. D. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* **13**, 343–359 (2015).
154. Hsu, T. *et al.* The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proceedings of the National Academy of Sciences* **100**, 12420–12425 (2003).
155. Bitter, W. *et al.* Systematic Genetic Nomenclature for Type VII Secretion Systems. *PLoS Pathogens* **5**, e1000507 (2009).
156. Stoop, E. J. M., Bitter, W. & van der Sar, A. M. Tubercle bacilli rely on a type VII army for pathogenicity. *Trends in Microbiology* **20**, 477–484 (2012).
157. Bottai, D., Gröschel, M. I. & Brosch, R. Type VII Secretion Systems in Gram-Positive Bacteria. in *Protein and Sugar Export and Assembly in Gram-positive Bacteria* (eds. Bagnoli, F. & Rappuoli, R.) **404**, 235–265 (Springer International Publishing, 2015).
158. Sørensen, A. L., Nagai, S., Houen, G., Andersen, P. & Andersen, Å. B. Purification and Characterization of a Low-Molecular-Mass T-Cell Antigen Secreted by *Mycobacterium tuberculosis*. *Infect. Immun.* **63**, 8 (1995).
159. Guinn, K. M. *et al.* Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*: RD1 genes affect secretion and virulence of *M. tuberculosis*. *Molecular Microbiology* **51**, 359–370 (2004).
160. Houben, E. N. G., Korotkov, K. V. & Bitter, W. Take five — Type VII secretion systems of

- Mycobacteria. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 1707–1716 (2014).
161. Simeone, R., Bottai, D. & Brosch, R. ESX/type VII secretion systems and their role in host-pathogen interaction. *Current Opinion in Microbiology* **12**, 4–10 (2009).
162. Smith, J. *et al.* Evidence for Pore Formation in Host Cell Membranes by ESX-1-Secreted ESAT-6 and Its Role in *Mycobacterium marinum* Escape from the Vacuole. *Infection and Immunity* **76**, 5478–5487 (2008).
163. Unnikrishnan, M., Constantinidou, C., Palmer, T. & Pallen, M. J. The Enigmatic Esx Proteins: Looking Beyond *Mycobacteria*. *Trends in Microbiology* **25**, 192–204 (2017).
164. Baptista, C., Barreto, H. C. & São-José, C. High Levels of DegU-P Activate an ESAT-6-Like Secretion System in *Bacillus subtilis*. *PLoS ONE* **8**, e67840 (2013).
165. Burts, M. L., Williams, W. A., DeBord, K. & Missiakas, D. M. EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proceedings of the National Academy of Sciences* **102**, 1169–1174 (2005).
166. Warne, B. *et al.* The Ess/Type VII secretion system of *Staphylococcus aureus* shows unexpected genetic diversity. *BMC Genomics* **17**, (2016).
167. Kneuper, H. *et al.* Heterogeneity in *ess* transcriptional organization and variable contribution of the Ess/Type VII protein secretion system to virulence across closely related *Staphylococcus aureus* strains: Type VII secretion in *S. aureus*. *Molecular Microbiology* **93**, 928–943 (2014).
168. Burts, M. L., DeDent, A. C. & Missiakas, D. M. EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of *Staphylococcus aureus*. *Molecular Microbiology* **69**, 736–746 (2008).
169. Anderson, M., Aly, K. A., Chen, Y.-H. & Missiakas, D. Secretion of atypical protein substrates by the ESAT-6 Secretion System of *Staphylococcus aureus*: ESAT-6 secretion in *S. aureus*. *Molecular Microbiology* **90**, 734–743 (2013).
170. Anderson, M. *et al.* EssE Promotes *Staphylococcus aureus* ESS-Dependent Protein Secretion To Modify Host Immune Responses during Infection. *Journal of Bacteriology* **199**, (2017).
171. Aly, K. A., Anderson, M., Ohr, R. J. & Missiakas, D. Isolation of a Membrane Protein Complex for Type VII Secretion in *Staphylococcus aureus*. *Journal of Bacteriology* **199**, (2017).
172. Rosenberg, O. S. *et al.* Substrates Control Multimerization and Activation of the Multi-Domain ATPase Motor of Type VII Secretion. *Cell* **161**, 501–512 (2015).
173. Zoltner, M. *et al.* EssC: domain structures inform on the elusive translocation channel in the Type VII secretion system. *Biochemical Journal* **473**, 1941–1952 (2016).
174. Casabona, M. G. *et al.* Functional analysis of the EsaB component of the *Staphylococcus aureus* Type VII secretion system. *Microbiology* **163**, 1851–1863 (2017).
175. Poulsen, C., Panjikar, S., Holton, S. J., Wilmanns, M. & Song, Y.-H. WXG100 Protein Superfamily Consists of Three Subfamilies and Exhibits an α -Helical C-Terminal Conserved Residue Pattern. *PLoS ONE* **9**, 12 (2014).
176. Renshaw, P. S. *et al.* Structure and function of the complex formed by the tuberculosis virulence

- factors CFP-10 and ESAT-6. *The EMBO Journal* **24**, 2491–2498 (2005).
177. De Leon, J. *et al.* *Mycobacterium tuberculosis* ESAT-6 Exhibits a Unique Membrane-interacting Activity That Is Not Found in Its Ortholog from Non-pathogenic *Mycobacterium smegmatis*. *Journal of Biological Chemistry* **287**, 44184–44191 (2012).
178. Korea, C. G. *et al.* Staphylococcal Esx Proteins Modulate Apoptosis and Release of Intracellular *Staphylococcus aureus* during Infection in Epithelial Cells. *Infection and Immunity* **82**, 4144–4153 (2014).
179. Ohr, R. J., Anderson, M., Shi, M., Schneewind, O. & Missiakas, D. EssD, a Nuclease Effector of the *Staphylococcus aureus* ESS Pathway. *Journal of Bacteriology* **199**, e00528-16 (2017).
180. Cao, Z., Casabona, M. G., Kneuper, H., Chalmers, J. D. & Palmer, T. The type VII secretion system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor bacteria. *Nature Microbiology* **2**, 16183 (2016).
181. Fluit, A. C. Livestock-associated *Staphylococcus aureus*. *Clin Microbiol Infect* **18**, 735–744 (2012).
182. Dai, Y. *et al.* A Novel ESAT-6 Secretion System-Secreted Protein EsxX of Community-Associated *Staphylococcus aureus* Lineage ST398 Contributes to Immune Evasion and Virulence. *Frontiers in Microbiology* **8**, (2017).
183. Kahl, B. C. *et al.* Population Dynamics of Persistent *Staphylococcus aureus* Isolated from the Airways of Cystic Fibrosis Patients during a 6-Year Prospective Study. *Journal of Clinical Microbiology* **41**, 4424–4427 (2003).
184. Herbert, S. *et al.* Repair of Global Regulators in *Staphylococcus aureus* 8325 and Comparative Analysis with Other Clinical Isolates. *Infection and Immunity* **78**, 2877–2889 (2010).
185. Ishii, K. *et al.* Induction of Virulence Gene Expression in *Staphylococcus aureus* by Pulmonary Surfactant. *Infection and Immunity* **82**, 1500–1510 (2014).
186. Linhardt, F., Ziebuhr, W., Meyer, P., Witte, W. & Hacker, J. Pulsed-field gel electrophoresis of genomic restriction fragments as a tool for the epidemiological analysis of *Staphylococcus aureus* and coagulase-negative staphylococci. *FEMS Microbiol Lett* **74**, 181–185 (1992).
187. Maiden, M. C. J. *et al.* Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 3140–3145 (1998).
188. Maiden, M. C. J. Multilocus Sequence Typing of Bacteria. *Annual Review of Microbiology* **60**, 561–588 (2006).
189. Turner, K. M. E. & Feil, E. J. The secret life of the multilocus sequence type. *International Journal of Antimicrobial Agents* **29**, 129–135 (2007).
190. Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *Journal of Bacteriology* **186**, 1518–1530 (2004).
191. Falush, D., Stephens, M. & Pritchard, J. K. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. 22

192. Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. How clonal are bacteria? *Proceedings of the National Academy of Sciences* **90**, 4384–4388 (1993).
193. Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J. & Spratt, B. G. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* **38**, 1008–1015 (2000).
194. Williamson, D. A., Heffernan, H. & Nimmo, G. Contemporary genomic approaches in the diagnosis and typing of *Staphylococcus aureus*. *Pathology* **47**, 270–275 (2015).
195. O'Hara, F. P. *et al.* *spa* Typing and Multilocus Sequence Typing Show Comparable Performance in a Macroeconomic Study of *Staphylococcus aureus* in the United States. *Microbial Drug Resistance* **22**, 88–96 (2016).
196. Deurenberg, R. H. & Stobberingh, E. E. The evolution of *Staphylococcus aureus*. *Infection, Genetics and Evolution* **8**, 747–763 (2008).
197. Uhlemann, A.-C. *et al.* Identification of a Highly Transmissible Animal-Independent *Staphylococcus aureus* ST398 Clone with Distinct Genomic and Cell Adhesion Properties. *mBio* **3**, e00027-12-e00027-12 (2012).
198. Waring, R. Methicillin-Resistant *S. aureus* Infections among Patients in the Emergency Department. *n engl j med* **9** (2006).
199. Lakhundi, S. & Zhang, K. Methicillin-Resistant *Staphylococcus aureus*: Molecular Characterization, Evolution, and Epidemiology. *Clinical Microbiology Reviews* **31**, (2018).
200. on behalf of the Regional Infection Control Group of the Centre Region, Diene, S. M., Corvaglia, A. R., François, P. & van der Mee-Marquet, N. Prophages and adaptation of *Staphylococcus aureus* ST398 to the human clinic. *BMC Genomics* **18**, (2017).
201. Wang, X.-M. Evaluation of a multilocus sequence typing system for *Staphylococcus epidermidis*. *Journal of Medical Microbiology* **52**, 989–998 (2003).
202. Wisplinghoff, H. *et al.* Related Clones Containing SCCmec Type IV Predominate among Clinically Significant *Staphylococcus epidermidis* Isolates. *Antimicrobial Agents and Chemotherapy* **47**, 3574–3579 (2003).
203. Thomas, J. C. *et al.* Improved Multilocus Sequence Typing Scheme for *Staphylococcus epidermidis*. *Journal of Clinical Microbiology* **45**, 616–619 (2007).
204. Kozitskaya, S. *et al.* Clonal Analysis of *Staphylococcus epidermidis* Isolates Carrying or Lacking Biofilm-Mediating Genes by Multilocus Sequence Typing. *Journal of Clinical Microbiology* **43**, 4751–4757 (2005).
205. Cavanagh, J. P. *et al.* Core genome conservation of *Staphylococcus haemolyticus* limits sequence based population structure analysis. *Journal of Microbiological Methods* **89**, 159–166 (2012).
206. Zhang, L. *et al.* Multilocus Sequence Typing and Further Genetic Characterization of the Enigmatic Pathogen, *Staphylococcus hominis*. *PLoS ONE* **8**, e66496 (2013).
207. Solyman, S. M. *et al.* Multilocus Sequence Typing for Characterization of *Staphylococcus pseudintermedius*. *Journal of Clinical Microbiology* **51**, 306–310 (2013).

208. Chassain, B. *et al.* Multilocus Sequence Typing Analysis of *Staphylococcus lugdunensis* Implies a Clonal Population Structure. *Journal of Clinical Microbiology* **50**, 3003–3009 (2012).
209. Saber, H. *et al.* A Review of Staphylococcal Cassette Chromosome mec (SCCmec) Types in Coagulase-Negative Staphylococci (CoNS) Species. *Malaysian Journal of Medical Sciences* **24**, 7–18 (2017).
210. Lemee, L. Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*. *Microbiology* **151**, 3171–3180 (2005).
211. Zhang, W., Jayarao, B. M. & Knabel, S. J. Multi-Virulence-Locus Sequence Typing of *Listeria monocytogenes*. *Applied and Environmental Microbiology* **70**, 913–920 (2004).
212. Verghese, B., Schwalm, N. D., Dudley, E. G. & Knabel, S. J. A combined multi-virulence-locus sequence typing and Staphylococcal Cassette Chromosome mec typing scheme possesses enhanced discriminatory power for genotyping MRSA. *Infection, Genetics and Evolution* **12**, 1816–1821 (2012).
213. Zhou, K., Aertsen, A. & Michiels, C. W. The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiology Reviews* **38**, 119–141 (2014).
214. Nadon, C. A. *et al.* Development and application of MLVA methods as a tool for inter-laboratory surveillance. *Euro Surveill.* **18**, 20565 (2013).
215. Kjeldsen, M. K., Torpdahl, M., Pedersen, K. & Nielsen, E. M. Development and comparison of a generic multiple-locus variable-number tandem repeat analysis with pulsed-field gel electrophoresis for typing of *Salmonella enterica* subsp. *enterica*. *Journal of Applied Microbiology* **119**, 1707–1717 (2015).
216. Vignaud, M.-L. *et al.* MLVA for *Salmonella enterica* subsp. *enterica* Serovar Dublin: Development of a Method Suitable for Inter-Laboratory Surveillance and Application in the Context of a Raw Milk Cheese Outbreak in France in 2012. *Frontiers in Microbiology* **08**, (2017).
217. Chenal-Francisque, V. *et al.* Optimized Multilocus Variable-Number Tandem-Repeat Analysis Assay and Its Complementarity with Pulsed-Field Gel Electrophoresis and Multilocus Sequence Typing for *Listeria monocytogenes* Clone Identification and Surveillance. *Journal of Clinical Microbiology* **51**, 1868–1880 (2013).
218. Haguenoer, E. *et al.* A multi locus variable number of tandem repeat analysis (MLVA) scheme for *Streptococcus agalactiae* genotyping. *BMC microbiology* **11**, 171 (2011).
219. Sabat, A. *et al.* New Method for Typing *Staphylococcus aureus* Strains: Multiple-Locus Variable-Number Tandem Repeat Analysis of Polymorphism and Genetic Relationships of Clinical Isolates. *Journal of Clinical Microbiology* **41**, 1801–1804 (2003).
220. Johansson, A., Koskiniemi, S., Gottfridsson, P., Wistrom, J. & Monsen, T. Multiple-Locus Variable-Number Tandem Repeat Analysis for Typing of *Staphylococcus epidermidis*. *Journal of Clinical Microbiology* **44**, 260–265 (2006).
221. Wichelhaus, T. A. *et al.* Rapid Molecular Typing of Methicillin-Resistant *Staphylococcus aureus* by PCR-RFLP. *Infection Control & Hospital Epidemiology* **22**, 294–298 (2001).
222. Ryffel, C., Bucher, R., Kayser, F. H. & Berger-Bächi, B. The *Staphylococcus aureus* mec determinant comprises an unusual cluster of direct repeats and codes for a gene product similar to the

- Escherichia coli* sn-glycerophosphoryl diester phosphodiesterase. *Journal of Bacteriology* **173**, 7416–7422 (1991).
223. Sabat, A. *et al.* Distribution of the Serine-Aspartate Repeat Protein-Encoding *sdr* Genes among Nasal-Carriage and Invasive *Staphylococcus aureus* Strains. *Journal of Clinical Microbiology* **44**, 1135–1138 (2006).
224. Hardy, K. J. Distribution and characterization of staphylococcal interspersed repeat units (SIRUs) and potential use for strain differentiation. *Microbiology* **150**, 4045–4052 (2004).
225. Ikawaty, R., Willems, R. J. L., Box, A. T. A., Verhoef, J. & Fluit, A. C. Novel Multiple-Locus Variable-Number Tandem-Repeat Analysis Method for Rapid Molecular Typing of Human *Staphylococcus aureus*. *Journal of Clinical Microbiology* **46**, 3147–3151 (2008).
226. Pourcel, C. *et al.* Improved Multiple-Locus Variable-Number Tandem-Repeat Assay for *Staphylococcus aureus* Genotyping, Providing a Highly Informative Technique Together with Strong Phylogenetic Value. *Journal of Clinical Microbiology* **47**, 3121–3128 (2009).
227. Schouls, L. M. *et al.* Multiple-Locus Variable Number Tandem Repeat Analysis of *Staphylococcus aureus*: Comparison with Pulsed-Field Gel Electrophoresis and *spa*-Typing. *PLoS ONE* **4**, e5082 (2009).
228. Szczuka, E., Trawczynski, K. & Kaznowski, A. Clonal analysis of *Staphylococcus hominis* strains isolated from hospitalized patients. *Pol J Microbiol* **63**, 349–354 (2014).
229. Rakov, A. V., Ubukata, K. & Ashley Robinson, D. Population structure of hyperinvasive serotype 12F, clonal complex 218 *Streptococcus pneumoniae* revealed by multilocus *boxB* sequence typing. *Infection, Genetics and Evolution* **11**, 1929–1939 (2011).
230. ZaiSS, N. H. *et al.* Typing *Clostridium difficile* strains based on tandem repeat sequences. *BMC Microbiology* **9**, 6 (2009).
231. Visca, P. *et al.* Investigation of the population structure of *Legionella pneumophila* by analysis of tandem repeat copy number and internal sequence variation. *Microbiology* **157**, 2582–2594 (2011).
232. Brevik, Ø. J., Ottem, K. F. & Nylund, A. Multiple-locus, variable number of tandem repeat analysis (MLVA) of the fish-pathogen *Francisella noatunensis*. *BMC veterinary research* **7**, 5 (2011).
233. Frénay, H. M. E. *et al.* Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *European Journal of Clinical Microbiology & Infectious Diseases* **15**, 60–64 (1996).
234. Shopsin, B. *et al.* Evaluation of Protein A Gene Polymorphic Region DNA Sequencing for Typing of *Staphylococcus aureus* Strains. *J. Clin. Microbiol.* **37**, 8 (1999).
235. Mazi, W., Sangal, V., Sandstrom, G., Saeed, A. & Yu, J. Evaluation of *spa*-typing of methicillin-resistant *Staphylococcus aureus* using high-resolution melting analysis. *International Journal of Infectious Diseases* **38**, 125–128 (2015).
236. Strommenger, B. *et al.* *spa* Typing of *Staphylococcus aureus* as a Frontline Tool in Epidemiological Typing. *Journal of Clinical Microbiology* **46**, 574–581 (2008).
237. Malachowa, N. *et al.* Comparison of Multiple-Locus Variable-Number Tandem-Repeat Analysis with Pulsed-Field Gel Electrophoresis, *spa* Typing, and Multilocus Sequence Typing for Clonal

- Characterization of *Staphylococcus aureus* Isolates. *Journal of Clinical Microbiology* **43**, 3095–3100 (2005).
238. Asadollahi, P. *et al.* Distribution of the Most Prevalent Spa Types among Clinical Isolates of Methicillin-Resistant and -Susceptible *Staphylococcus aureus* around the World: A Review. *Frontiers in Microbiology* **9**, (2018).
239. Cookson, B. D. *et al.* Evaluation of Molecular Typing Methods in Characterizing a European Collection of Epidemic Methicillin-Resistant *Staphylococcus aureus* Strains: the HARMONY Collection. *Journal of Clinical Microbiology* **45**, 1830–1837 (2007).
240. Koreen, L. *et al.* spa Typing Method for Discriminating among *Staphylococcus aureus* Isolates: Implications for Use of a Single Marker To Detect Genetic Micro- and Macrovariation. *Journal of Clinical Microbiology* **42**, 792–799 (2004).
241. Estoup, A., Jarne, P. & Cornuet, J.-M. Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol. Ecol.* **11**, 1591–1604 (2002).
242. Tagini, F. & Greub, G. Bacterial genome sequencing in clinical microbiology: a pathogen-oriented review. *European Journal of Clinical Microbiology & Infectious Diseases* **36**, 2007–2020 (2017).
243. Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L. & Goering, R. V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clinical Microbiology and Infection* **24**, 350–354 (2018).
244. Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D. & Mellmann, A. Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes. *Journal of Clinical Microbiology* **52**, 2365–2370 (2014).
245. Köser, C. U. *et al.* Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. *The New England Journal of Medicine* **9** (2012).
246. Chang, S.-C. *et al.* Characterization of two novel variants of staphylococcal cassette chromosome mec elements in oxacillin-resistant *Staphylococcus lugdunensis*. *Journal of Antimicrobial Chemotherapy* **72**, 3258–3262 (2017).
247. Schneewind, O. & Missiakas, D. M. Protein secretion and surface display in Gram-positive bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1123–1139 (2012).
248. Zhang, D., Iyer, L. M. & Aravind, L. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic Acids Research* **39**, 4532–4552 (2011).
249. Cheng, C. *et al.* Persistence of a major endemic clone of oxacillin-resistant *Staphylococcus lugdunensis* sequence type 6 at a tertiary medical centre in northern Taiwan. *Int. J. Infect. Dis.* **36**, 72–77 (2015).
250. Koren, S. & Phillippy, A. M. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology* **23**, 110–120 (2015).
251. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation

- sequencing technologies. *Nature Reviews Genetics* **17**, 333–351 (2016).
252. Carneiro, M. O. *et al.* Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* **13**, 375 (2012).
253. Rossi, C. C., Souza-Silva, T., Araújo-Alves, A. V. & Giambiagi-deMarval, M. CRISPR-Cas Systems Features and the Gene-Reservoir Role of Coagulase-Negative Staphylococci. *Frontiers in Microbiology* **8**, (2017).
254. Frankel, M. B., Hendrickx, A. P. A., Missiakas, D. M. & Schneewind, O. LytN, a murein hydrolase in the cross-wall compartment of *Staphylococcus aureus*, is involved in proper bacterial growth and envelope assembly. *J. Biol. Chem.* **286**, 32593–32605 (2011).
255. Novick, R. P. & Ram, G. Staphylococcal pathogenicity islands — movers and shakers in the genomic firmament. *Current Opinion in Microbiology* **38**, 197–204 (2017).
256. Beasley, F. C. *et al.* Characterization of staphyloferrin A biosynthetic and transport mutants in *Staphylococcus aureus*. *Mol. Microbiol.* **72**, 947–963 (2009).
257. Vandenesch, F., Etienne, J., Reverdy, M. E. & Eykyn, S. J. Endocarditis due to *Staphylococcus lugdunensis*: report of 11 cases and review. *Clinical infectious diseases* **17**, 871–876 (1993).
258. O'Neill, A. J., Larsen, A. R., Skov, R., Henriksen, A. S. & Chopra, I. Characterization of the epidemic European fusidic acid-resistant impetigo clone of *Staphylococcus aureus*. *J. Clin. Microbiol.* **45**, 1505–1510 (2007).
259. Chen, H.-J. *et al.* New Structure of Phage-Related Islands Carrying *fusB* and a Virulence Gene in Fusidic Acid-Resistant *Staphylococcus epidermidis*. *Antimicrobial Agents and Chemotherapy* **57**, 5737–5739 (2013).
260. Bloomfield, G. A., Whittle, G., McDonagh, M. B., Katz, M. E. & Cheetham, B. F. Analysis of sequences flanking the *vap* regions of *Dichelobacter nodosus*: evidence for multiple integration events, a killer system, and a new genetic element. *Microbiology (Reading, Engl.)* **143 (Pt 2)**, 553–562 (1997).
261. Ma, Z. *et al.* Insight into the specific virulence related genes and toxin-antitoxin virulent pathogenicity islands in swine streptococcosis pathogen *Streptococcus equi* ssp. *zooepidemicus* strain ATCC35246. *BMC Genomics* **14**, 377 (2013).
262. Ji, X. *et al.* A novel virulence-associated protein, *vapE*, in *Streptococcus suis* serotype 2. *Mol Med Rep* **13**, 2871–2877 (2016).
263. Dufour, P. *et al.* High Genetic Variability of the *agr* Locus in *Staphylococcus* Species. *Journal of Bacteriology* **184**, 1180–1186 (2002).
264. Hellmark, B., Söderquist, B., Unemo, M. & Nilsson-Augustinsson, Å. Comparison of *Staphylococcus epidermidis* isolated from prosthetic joint infections and commensal isolates in regard to antibiotic susceptibility, *agr* type, biofilm production, and epidemiology. *Int. J. Med. Microbiol.* **303**, 32–39 (2013).
265. Harris, L. G. *et al.* Limitations in the use of PSMγ, *agr*, RNAIII, and biofilm formation as biomarkers to define invasive *Staphylococcus epidermidis* from chronic biomedical device-associated infections. *Int. J. Med. Microbiol.* **307**, 382–387 (2017).

266. Tan, L., Li, S. R., Jiang, B., Hu, X. M. & Li, S. Therapeutic Targeting of the *Staphylococcus aureus* Accessory Gene Regulator (*agr*) System. *Front. Microbiol.* **9**, (2018).
267. Ikonomidis, A. *et al.* Association of biofilm formation and methicillin-resistance with accessory gene regulator (*agr*) loci in Greek *Staphylococcus aureus* clones. *Microb. Pathog.* **47**, 341–344 (2009).
268. Lambe, D. W., Jeffery, C., Ferguson, K. P. & Cooper, M. D. Examination of the glycocalyx of four species of *Staphylococcus* by transmission electron microscopy and image analysis. *Microbios* **78**, 133–143 (1994).
269. O’Riordan, K. & Lee, J. C. *Staphylococcus aureus* capsular polysaccharides. *Clin. Microbiol. Rev.* **17**, 218–234 (2004).
270. Luong, T., Sau, S., Gomez, M., Lee, J. C. & Lee, C. Y. Regulation of *Staphylococcus aureus* capsular polysaccharide expression by *agr* and *sarA*. *Infect. Immun.* **70**, 444–450 (2002).
271. Thakker, M., Park, J. S., Carey, V. & Lee, J. C. *Staphylococcus aureus* serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. *Infect. Immun.* **66**, 5183–5189 (1998).
272. Nemeth, J. & Lee, J. C. Antibodies to capsular polysaccharides are not protective against experimental *Staphylococcus aureus* endocarditis. *Infect. Immun.* **63**, 375–380 (1995).
273. Bowden, M. G. *et al.* Identification and preliminary characterization of cell-wall-anchored proteins of *Staphylococcus epidermidis*. *Microbiology (Reading, Engl.)* **151**, 1453–1464 (2005).
274. Arrecubieta, C., Lee, M.-H., Macey, A., Foster, T. J. & Lowy, F. D. SdrF, a *Staphylococcus epidermidis* surface protein, binds type I collagen. *J. Biol. Chem.* **282**, 18767–18776 (2007).
275. Arrecubieta, C. *et al.* SdrF, a *Staphylococcus epidermidis* surface protein, contributes to the initiation of ventricular assist device driveline-related infections. *PLoS Pathog.* **5**, e1000411 (2009).
276. Otto, M. Molecular basis of *Staphylococcus epidermidis* infections. *Semin Immunopathol* **34**, 201–214 (2012).
277. Wang, Y. *et al.* Role of the ESAT-6 secretion system in virulence of the emerging community-associated *Staphylococcus aureus* lineage ST398. *Sci Rep* **6**, (2016).
278. Kusch, H. & Engelmann, S. Secrets of the secretome in *Staphylococcus aureus*. *International Journal of Medical Microbiology* **304**, 133–141 (2014).
279. Bushe. Comparative secretome analyses of human and zoonotic *Staphylococcus aureus* isolates of CC8, CC22 and CC398. 56 (2018).
280. Ziebandt, A.-K. *et al.* Proteomics uncovers extreme heterogeneity in the *Staphylococcus aureus* exoproteome due to genomic plasticity and variant gene regulation. *Proteomics.* **10**, 1634–1644 (2010).
281. Stapleton, M. R. *et al.* Characterization of IsaA and SceD, Two Putative Lytic Transglycosylases of *Staphylococcus aureus*. *Journal of Bacteriology* **189**, 7316–7325 (2007).
282. Quiblier, C. *et al.* Secretome Analysis Defines the Major Role of SecDF in *Staphylococcus aureus* Virulence. *PLoS ONE* **8**, e63513 (2013).
283. Hecker, M., Mäder, U. & Völker, U. From the genome sequence via the proteome to cell physiology – Pathoproteomics and pathophysiology of *Staphylococcus aureus*. *International Journal of*

Medical Microbiology **308**, 545–557 (2018).

284. Tsai, C. J.-Y., Loh, J. M. S. & Proft, T. *Galleria mellonella* infection models for the study of bacterial diseases and for antimicrobial drug testing. *Virulence* **7**, 214–229 (2016).
285. van Belkum, A. *et al.* Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect.* **13 Suppl 3**, 1–46 (2007).
286. Reyes, J. F., Chan, C. H. S. & Tanaka, M. M. Impact of homoplasmy on variable numbers of tandem repeats and spoligotypes in *Mycobacterium tuberculosis*. *Infect. Genet. Evol.* **12**, 811–818 (2012).
287. Kassem, I. I., Esseili, M. A. & Sigler, V. Detection and differentiation of staphylococcal contamination of clinical surfaces using denaturing gradient gel electrophoresis. *Journal of Hospital Infection* **78**, 187–193 (2011).
288. Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* **12**, 49–62 (2014).
289. Koreen, L. *et al.* Comparative sequencing of the serine-aspartate repeat-encoding region of the clumping factor B gene (clfB) for resolution within clonal groups of *Staphylococcus aureus*. *J. Clin. Microbiol.* **43**, 3985–3994 (2005).
290. Mellmann, A. *et al.* Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-term evolution of *Staphylococcus aureus* populations based on spa polymorphisms. *BMC Microbiol.* **7**, 98 (2007).
291. Ythier, M. *et al.* Natural variability of in vitro adherence to fibrinogen and fibronectin does not correlate with in vivo infectivity of *Staphylococcus aureus*. *Infect. Immun.* **78**, 1711–1716 (2010).
292. Rosenau, A. *et al.* Evaluation of the ability of *Streptococcus agalactiae* strains isolated from genital and neonatal specimens to bind to human fibrinogen and correlation with characteristics of the *fbxA* and *fbxB* genes. *Infect. Immun.* **75**, 1310–1317 (2007).
293. Riskey, A. L., Loughman, A., Cywes-Bentley, C., Foster, T. J. & Lee, J. C. Capsular polysaccharide masks clumping factor A-mediated adherence of *Staphylococcus aureus* to fibrinogen and platelets. *J. Infect. Dis.* **196**, 919–927 (2007).
294. Que, Y.-A. *et al.* Fibrinogen and fibronectin binding cooperate for valve infection and invasion in *Staphylococcus aureus* experimental endocarditis. *J. Exp. Med.* **201**, 1627–1635 (2005).

RESUME

La première partie de nos travaux a consisté au séquençage de génomes complets de trois souches pathogènes et de trois souches de portage de *Staphylococcus lugdunensis* pour les comparer aux 15 génomes complets disponibles sur NCBI. Aucun déterminant génétique associé au contexte de virulence ou de portage de *S. lugdunensis* n'a été identifié. Cependant, nous avons mis en évidence la présence d'éléments génétiques mobiles et des variations dépendantes des complexes clonaux, définis par *MultiLocus Sequence Typing*, au sein de *loci* potentiellement associés à la virulence. Des variations ont été observées dans un *locus* homologue à celui de *Staphylococcus aureus* codant le système de sécrétion Ess/type VII (SST7). Nous avons mis en évidence huit organisations génétiques chez cette espèce présentant pourtant une structure de population clonale.

La seconde partie de nos travaux a consisté à la caractérisation phénotypique et moléculaire du SST7 chez *S. lugdunensis* par la formation d'un mutant de délétion du gène *essC* codant une protéine essentielle à la sécrétion. Nos résultats suggèrent que le SST7 serait impliqué dans la translocation de protéines prédites *in silico* comme impliquées dans la virulence. Néanmoins, dans des modèles de cytotoxicité cellulaire et d'infection du nématode *Caenorhabditis elegans*, aucune atténuation de la virulence n'a été observée chez la souche mutante malgré une perte de sa capacité à lyser les érythrocytes, comparativement à la souche sauvage.

Nos travaux ont également permis de développer et d'évaluer le pouvoir discriminant de trois nouvelles méthodes de typage constituant des outils très prometteurs pour l'épidémiologie moléculaire des infections à *S. lugdunensis*.

Mots clés : *Staphylococcus lugdunensis*, analyse génomique comparative, système de sécrétion Ess/type VII, virulence, génotypage, *Variable Number of Tandem Repeat*.

ABSTRACT

The first part of this study consisted in whole genome sequencing of three pathogenic and three carriage strains of *S. lugdunensis* and comparison with the 15 genomes available in the NCBI. No genetic determinant was associated to the pathogenic or carriage context. However, we have highlighted the presence of mobile genetic elements and MultiLocus Sequence Typing clonal complex dependent variations within *loci* potentially associated with virulence. Variations were observed in the *ess* locus homologous to that of *Staphylococcus aureus* encoding the type VII secretion system (T7SS). We showed eight genetic organizations in this species with a clonal population structure.

The second part of our work consisted in a phenotypic and molecular characterization of T7SS in *S. lugdunensis* by construction of a deletion *essC* gene mutant. This gene encodes a protein required for protein secretion. Our results suggest that T7SS could be involved in translocation of proteins predicted as implicated in virulence *in silico*. Nevertheless, no virulence attenuation was observed in cells cytotoxicity assay and *Caenorhabditis elegans* virulence assays between wild-type and mutant strains which yet has lost the ability to lyse erythrocytes.

We also developed and evaluated discriminating power of three new typing methods, which are very promising tools for the molecular epidemiology of *S. lugdunensis* infections.

Keywords: *Staphylococcus lugdunensis*, genomic comparative analysis, type VII secretion system, virulence, genotyping, *Variable Number of Tandem Repeat*.