



HAL
open science

Évolution génotypique et phénotypique d'une souche épidémique de *Pseudomonas aeruginosa* au cours des 11 ans de sa diffusion hospitalière

Marie Petitjean

► **To cite this version:**

Marie Petitjean. Évolution génotypique et phénotypique d'une souche épidémique de *Pseudomonas aeruginosa* au cours des 11 ans de sa diffusion hospitalière. Bactériologie. Université Bourgogne Franche-Comté, 2017. Français. NNT : 2017UBFCE019 . tel-01823039

HAL Id: tel-01823039

<https://theses.hal.science/tel-01823039v1>

Submitted on 25 Jun 2018

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.

THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTE

École doctorale n°554

Environnements-Santé

Doctorat de Sciences de la Vie et de la Santé

Par

Marie Petitjean

Évolution génotypique et phénotypique d'une souche épidémique de *Pseudomonas aeruginosa* au cours des 11 ans de sa diffusion hospitalière.

Thèse présentée et soutenue à Besançon, le 31 octobre 2017

Composition du Jury :

Pr. Xavier Bertrand	PU-PH, UMR 6249 (Besançon)	Président du jury
Dr. Dominique Blanc	Privat docent & MER, CHUV (Lausanne)	Rapporteur
Pr. Anne-Marie Rogues	PU-PH, INSERM U657 (Bordeaux)	Rapporteur
Dr. Thilo Köhler	PhD, Privat docent, CMU (Genève)	Examineur
Pr. Didier Hocquet	PU-PH, UMR 6249 (Besançon)	Directeur de thèse
Dr. Benoît Valot	Ingénieur de recherche, UMR 6249 (Besançon)	Co-encadrant de thèse

Remerciements

Ce travail de thèse a pu être effectué grâce aux financements conjoints du Centre Hospitalier Régional Universitaire de Besançon et de la région Franche-Comté que je remercie vivement.

Il a eu lieu à l'UMR 6249 Chrono-Environnement dans le laboratoire d'Hygiène Hospitalière du CHRU de Besançon.

Je tiens à remercier,

Didier Hocquet, Directeur de cette thèse, qui m'a encadré durant ces trois ans. Merci pour tous les conseils avisés et tout le temps accordé qui m'a permis d'arriver au bout de cette thèse.

Benoît Valot, co-encadrant de cette thèse. Merci d'avoir été là pour m'aider à avancer, et merci pour tout ce temps consacré à réaliser des analyses bioinformatiques parfois infernales.

Xavier Bertrand, responsable du laboratoire d'hygiène hospitalière du CHU de Besançon, qui m'a accueillie au sein du labo.

Anne-Marie Rogues, responsable du laboratoire d'hygiène hospitalière au CHU de Bordeaux, et **Dominique Blanc**, Privat docent et Maître en Enseignement et Recherche au CHUV de Lausanne, les deux rapporteurs, pour leur temps consacré à la lecture et correction de ce manuscrit.

Thilo Köhler, PhD et Privat docent à l'université de Genève, pour avoir accepté d'évaluer mon travail de thèse.

Marie-Cécile Ploy et **Jean-Philippe Lavigne** pour leurs précieux conseils lors des Comités de Suivi de Thèse.

Un grand merci également à toutes les personnes que j'ai croisé durant ces trois ans et qui m'ont aidé à avancer. Merci à Alicia Silvant, Daniel Martak et Alexandre Meunier pour le temps qu'ils ont passé sur les manips. Merci à toute l'équipe d'hygiène qui m'a accueillie, vous êtes tous géniaux. Et pour finir merci à la fine équipe des doctorants des Hauts-du-Chazal avec qui j'ai passé de très bon moment.

Table des matières

Abréviations	5
1 Introduction	7
1.1 <i>Pseudomonas aeruginosa</i>	7
1.1.1 Généralités	7
1.1.2 Niches écologiques	7
1.1.3 Aspects cliniques	8
1.1.4 Aspects génomiques	9
1.2 Épidémiologie	10
1.2.1 Outils épidémiologiques modernes	10
1.2.2 Épidémiologie de <i>P. aeruginosa</i>	13
1.2.3 Clones à haut risque ayant une diffusion internationale	13
1.2.4 Caractéristiques des clones à haut risque épidémiques	14
1.2.5 Génomique d'épidémies hospitalières massives	15
1.2.6 Notion de vagues épidémiques : autres pathogènes	17
1.3 Évolution génomique	18
1.3.1 Mécanismes d'évolution	18
1.3.2 Plasticité particulière du génome de <i>P. aeruginosa</i>	24
1.4 Épidémiogénicité : résistance et virulence	24
1.4.1 Généralités sur les antibiotiques	25
1.4.2 Résistance naturelle	26
1.4.3 Résistance par mutation de gènes chromosomiques	28
1.4.4 Résistance aux antibiotiques par acquisition de matériel étranger	33
1.4.5 Facteurs de virulence	33
1.4.6 Évolution intrahôte	35
1.5 Outils d'analyse bioinformatique	36
1.5.1 Notion de séquençage haut débit	36
1.5.2 Techniques d'assemblage et d'alignement	43
1.5.3 Analyse de données	47
1.6 Axe de recherche	50
1.6.1 Sujet principal	50
1.6.2 Travaux annexes	51

2	Contexte de l'étude et objectifs	53
2.1	Description de l'épidémie	53
2.2	Hypothèse	54
2.3	Matériel et Méthodes	54
2.3.1	Collection d'isolats	54
2.3.2	Objectifs et méthodologie générale	56
3	DHS01	57
3.1	Objectifs et méthodologie	57
3.2	Résultats	58
3.2.1	Abstract	58
3.2.2	Impact Statement	58
3.2.3	Introduction	59
3.2.4	Methods	60
3.2.5	Results and Discussion	63
4	Évolution d'un clone de <i>P. aeruginosa</i>	79
4.1	Objectifs et méthodologie	79
4.2	Méthodes	79
4.2.1	Choix de l'outil d'assemblage	79
4.2.2	Construction d'une base de données	82
4.3	Résultats : Article en rédaction (Août 2017)	83
4.3.1	Abstract	84
4.3.2	Introduction	84
4.3.3	Material and Methods	84
4.3.4	Results & discussion	86
5	Conclusion et perspectives	97
5.1	Méthodologie	97
5.1.1	Choix de l'outil WGS	97
5.1.2	Limites de l'approche WGS	97
5.1.3	Limite de l'analyse	98
5.2	Compréhension de la biologie du pathogène	99
	Bibliographie	103
A	Article DHS01	118
B	Article <i>Enterobacter cloacae</i>	129
C	Article <i>P. aeruginosa</i> bruns	134
D	Article <i>Escherichia coli</i> ST131	147
E	Bande dessinée de vulgarisation scientifique	160

Abréviations

AAC(3')	<i>Aminoglycoside N(3')-acetyltransferase</i>
AAC(6')	<i>Aminoglycoside N(6')-acetyltransferase</i>
ADN/DNA	Acide désoxyribonucléique
ANT(2'')	<i>Aminoglycoside-2''-nucleotidyltransferase</i>
ARN/RNA	Acide ribonucléique
BES	<i>Besançon Epidemic Strain</i>
bp	<i>base pair</i>
C12-HSL	<i>N-(3-oxododecanoyl)homosérine-L-lactone</i>
C13-HSL	<i>N-butanoyl-L-homosérine lactone</i>
CC	Complexe clonal
CD-HIT	<i>Cluster Database at High Identity with Tolerance</i>
CRISPR-Cas	<i>Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated</i>
EUCAST	<i>European Committee on Antimicrobial Susceptibility Testing</i>
GI	<i>Genomic Island</i>
GTR	<i>General Time Reversible</i>
HSL	Homosérine lactone
ICE	<i>Integrative and Conjugative Element</i>
indel	Insertion-délétion
IS	<i>Insertion Sequence</i>
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
kb/kbp	kilo paire de base
LES	<i>Liverpool Epidemic Strain</i>
LPS	lipopolysaccharide
Mbp	Mega paire de base (1.000.000)
MCMC	<i>Markov Chain Monte Carlo</i>
MDCK	<i>Madin-Darby canine kidney</i>
MGE	<i>Mobile Genetic Element</i>
MH	Mueller Hinton
MLST	<i>Multi-Locus Sequence Typing</i>
NCBI	<i>National Center for Biotechnology Information</i>
NGS	<i>Next Generation Sequencing</i>
ORF	<i>Open Reading Frame</i>
PacBio	Pacific Biosciences
PCR	<i>Polymerase Chain Reaction</i>
PFGE	<i>Pulsed Field Gel Electrophoresis</i>

QRDR	<i>Quinolone Resistance Determining Region</i>
RND	<i>Resistance Nodulation Division</i>
SMRT	<i>Single Molecule Real Time</i>
SNP	<i>Single Nucleotide Polymorphism</i>
ST	<i>Sequence Type</i>
VFDB	<i>Virulence Factor DataBase</i>
WGS	<i>Whole Genome Sequencing</i>

Chapitre 1

Introduction

1.1 *Pseudomonas aeruginosa*

1.1.1 Généralités

L'espèce *Pseudomonas aeruginosa* (figure 1.1) aussi appelée bacille pyocyanique a été isolée pour la première fois en 1882 par Charles Gessard sur des plaies purulentes. Son nom vient du grec *puon* et *kuanos* qui signifient pus bleu. Cette bactérie a une taille de 1 à 3 μm de long contre 0.5 à 1 μm de large. C'est un bacille à Gram négatif, il prend une couleur rose lors de la coloration de Gram.

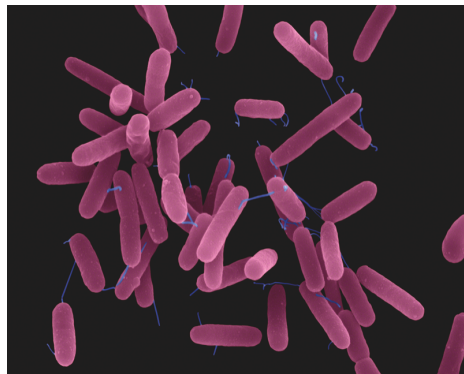


FIGURE 1.1 – *P. aeruginosa* vue au microscope électronique à balayage avec coloration [1]

1.1.2 Niches écologiques

La présence de *P. aeruginosa* dans des milieux très divers correspond au caractère ubiquiste de cette bactérie. Elle est principalement présente au niveau des zones humides (eaux douces, eaux de mer, eaux stagnantes ou courantes, sols, végétaux) et peut parfaitement vivre dans des environnements pollués (eaux ou sols) [2]. C'est une bactérie que l'on retrouve fréquemment dans l'environnement et qui est particulièrement adaptée au milieu hospitalier où elle se développe fortement.

A l'hôpital ou dans l'environnement, *P. aeruginosa* cohabite avec un certain nombre de bactéries, qu'elles soient du genre *Pseudomonas* ou non [3]. Une grande variété d'espèces du genre *Pseudomonas* peut survivre dans l'environnement hospitalier bien que les souches soient avant tout environnementales et majoritairement non pathogènes.

La capacité de survie de *P. aeruginosa* dans de nombreux environnements est liée d'une part à son large génome et d'autre part à sa capacité de modification. De nombreux gènes de régulation sont présents dans le génome de *P. aeruginosa*, lui permettant une adaptation rapide à un changement dans son environnement [4]. Cette capacité est d'ailleurs un des enjeux de la recherche autour de ce pathogène car l'acquisition de propriétés particulières dans l'environnement peut être un atout pour le développement dans l'environnement hospitalier [5]. En effet, l'acquisition d'éléments mobiles portant à la fois de la résistance aux métaux lourds ainsi que des gènes de résistance aux antibiotiques permettent à une souche environnementale d'être capable de se développer dans le milieu hospitalier où il y a une forte imprégnation en antibiotiques.

Ce gain de résistances par les souches d'origine environnementale leur permet de se propager dans les hôpitaux compliquant le traitement des patients. En effet, la difficulté à trouver un traitement contre ces souches augmente en lien avec le nombre de résistances acquises.

1.1.3 Aspects cliniques

P. aeruginosa est un des principaux pathogènes opportunistes. En termes de fréquence, elle est la troisième bactérie responsable d'infections nosocomiales (8,4% des infections) derrière *Escherichia coli* (26%) et *Staphylococcus aureus* (15,9%) lors de l'enquête nationale de prévalence menée en 2012 [6].

La diffusion de *P. aeruginosa* dans le milieu hospitalier entraîne des infections diverses chez les sujets les plus fragilisés que ce soient les patients immunodéprimés, ceux hospitalisés en réanimation, les grands brûlés ou les patients atteints de mucoviscidose. L'infection peut toucher plusieurs parties du corps (peau, oeil, plaies, etc.) avant d'être véhiculée par la circulation sanguine dans le cas de bactériémie [2].

Des infections pulmonaires surviennent principalement chez les patients atteints de mucoviscidose ainsi que chez les patients souffrant d'insuffisance respiratoire ou les patients sous ventilation assistée, surtout en réanimation. D'autres infections comme les infections urinaires sont également retrouvées, en particulier chez les patients sondés. La présence de la bactérie dans l'eau ou les solutions utilisées pour le nettoyage des lentilles de contact peut entraîner une contamination de l'oeil.

Une infection peut survenir à la suite d'otite externe, en particulier chez les nageurs. Des infections cutanées, surtout chez les grands brûlés entraînent des risques de septicémie dont le taux de mortalité est élevé. Les plaies opératoires peuvent aussi être des portes d'entrées pour *P. aeruginosa*.

D'autre part, cette bactérie peut tout à fait être retrouvée chez un patient asymptomatique.

En effet, *P. aeruginosa* étant un pathogène opportuniste, il n'est pathogène que pour les sujets affaiblis ou dont le système immunitaire est déficient. Il est donc possible d'être porteur sain de la bactérie, avec un risque d'infection si les défenses immunitaires sont altérées. De plus, un porteur asymptomatique peut transmettre la bactérie à un patient fragilisé.

1.1.4 Aspects génomiques

Le génome bactérien est composé d'un chromosome circulaire et parfois d'un ou plusieurs plasmides.

Le premier génome complet de l'espèce est celui de la souche de référence PAO1 qui a été séquencé en 2000 par *Stover et al.* [4]. Il contient 6,3 millions de paires de bases (6,3 Mbp), 5570 cadres ouverts de lecture (ou ORFs pour *Open Reading Frame*) et un taux de G+C de 66,6%. Depuis de nombreux génomes ont été séquencés, 2427 sont disponibles sur le site du NCBI (juillet 2017) parmi lesquels 94 sont des génomes complets (chromosome circulaire, avec le ou les plasmides éventuels, qui contient la séquence complète au nucléotide près) dont seulement 5 d'entre eux possèdent au moins un plasmide.

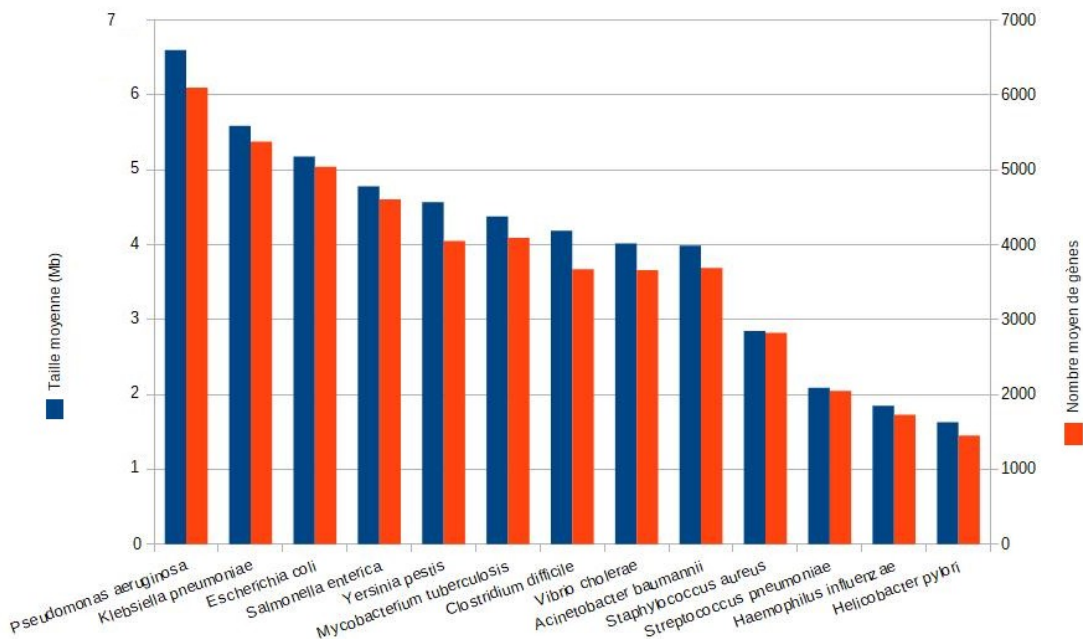


FIGURE 1.2 – Tailles des génomes des principales espèces bactériennes pathogènes déposés sur le NCBI

Le génome de *P. aeruginosa* est un des génomes bactériens contenant le plus d'information génétique (cf. figure 1.2). En effet, le génome moyen de cette bactérie est de 6,60 Mbp ce qui correspond à une moyenne de 6610 gènes alors que le génome moyen d'*Escherichia coli* ne contient que 5043 gènes soit environ 5,18 Mbp. Le génome moyen de *Staphylococcus aureus* ne mesure que 2,85 Mbp codant pour 2824 gènes alors que *Haemophilus influenzae*, bactérie responsable de méningites ou d'otites, contient 1729 gènes répartis sur 1,85 Mbp. La plupart

des gènes présents chez *P. aeruginosa* ne sont donc pas impliqués dans les capacités génétiques de base telles que la croissance cellulaire et le métabolisme de la bactérie.

Le *core genome* correspond à l'ensemble des gènes qui sont conservés au sein d'un groupe de souches. Cela concerne un grand nombre de gènes, entre 4455 et 5316, pour l'espèce *P. aeruginosa* [7]. C'est d'ailleurs cette diversité qui permet à ce pathogène de s'adapter à des environnements divers et lui confère son caractère ubiquiste. Cette souche possède en particulier la capacité de développer de la résistance à divers antibiotiques surtout dans les milieux où la pression de sélection est forte, comme c'est le cas pour les hôpitaux où les antibiotiques sont largement utilisés [8].

1.2 Épidémiologie

1.2.1 Outils épidémiologiques modernes

L'épidémiologie moléculaire des pathogènes bactériens s'attache à étudier les populations microbiennes, les caractériser et comprendre leur évolution. Ici, on ne s'intéressera qu'à la caractérisation bactérienne des maladies afin d'observer leur dangerosité (facteurs de virulence et résistance aux antibiotiques) et être en mesure de les classer. A ce jour, de nombreux outils existent pour caractériser des souches bactériennes afin de comprendre leurs liens épidémiologiques et étudier leur diffusion. Nous verrons ci-après les trois techniques les plus utilisées.

PFGE

Pulsed Field Gel Electrophoresis est une technique créée en 1983 par *Schwartz et Cantor* [9] pour pouvoir analyser le génome complet.

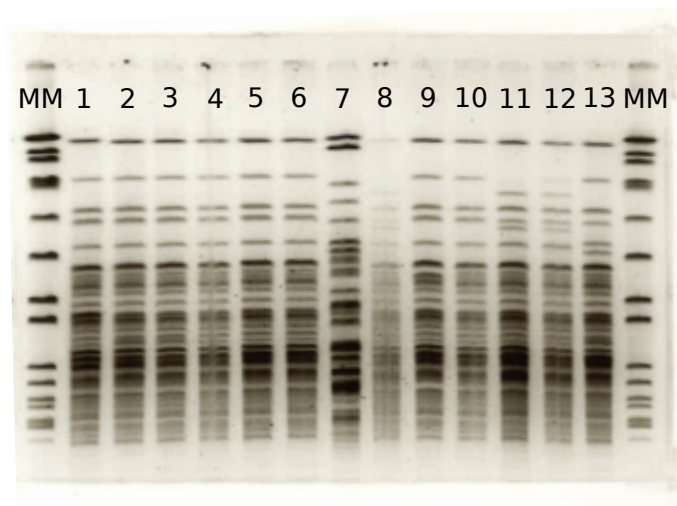


FIGURE 1.3 – Exemple d'un gel de PFGE typant des souches de *P. aeruginosa*. Les colonnes 1 à 13 correspondent aux souches testées tandis que les colonnes MM se réfèrent aux bandes connues de la digestion de *S. aureus* NCTC 8325 par l'enzyme *SmaI* [10].

Cette technique consiste à fragmenter le génome des bactéries, en utilisant des enzymes de

restriction, et à les faire migrer dans un gel d'électrophorèse afin de séparer les fragments d'ADN en fonction de leur taille. Les bandes d'ADN sont ensuite révélées. Plus le profil de bande (ou pulsotype) entre deux souches est identique, plus les souches sont proches. Les profils ainsi obtenus permettent de classer les souches bactériennes en clusters ou groupes semblables à 75% [10].

On peut voir sur la figure 1.3 un exemple de typage par PFGE de treize souches de *P. aeruginosa*. La colonne correspondant à la souche numéro 7 se démarque des autres par des bandes de profils différents, elle ne sera donc pas classée dans le même groupe que les autres.

MLST

Multi-Locus Sequence Typing permet de séparer les souches en fonction de leur *Sequence Type* (ST) d'appartenance correspondant à un regroupement d'allèles de 5 à 10 gènes de ménage (gènes toujours présents dans le génome car essentiels et très conservés) en fonction des espèces.

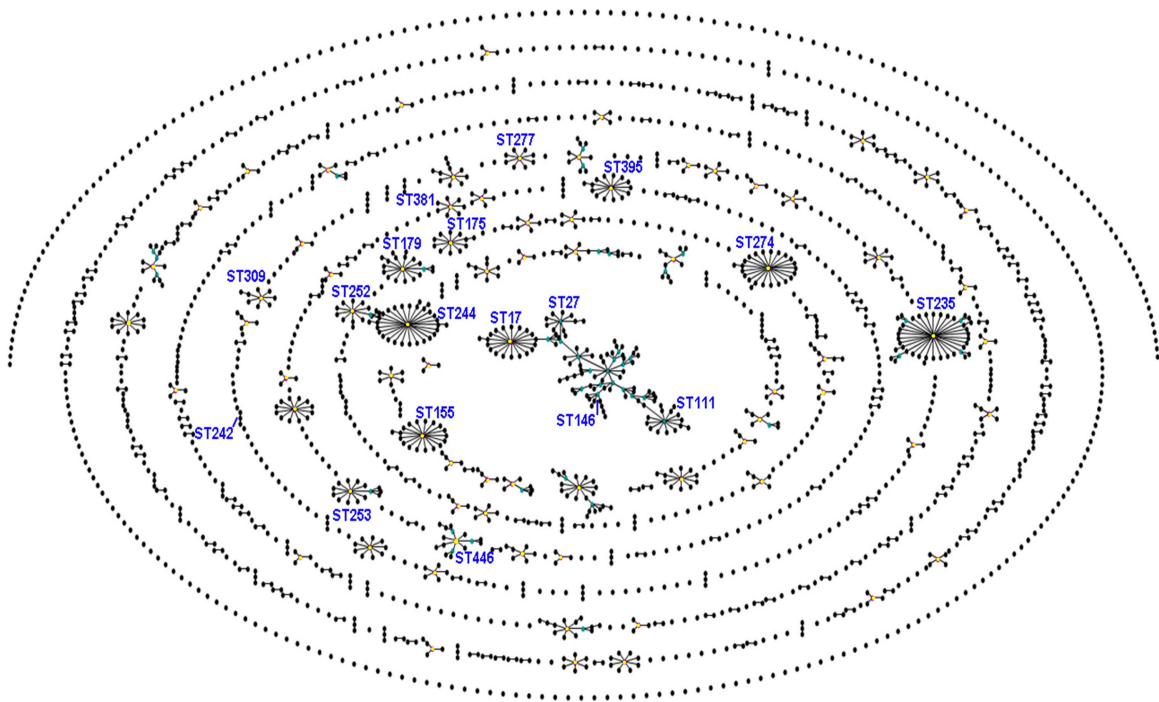


FIGURE 1.4 – Diagramme eBURST représentant les liens entre les différents STs de *P. aeruginosa* (2106 STs en juin 2015) [11]. Les STs représentés par un point jaune sont les fondateurs des complexes clonaux (CC).

Chaque gène peut avoir de nombreux allèles, en fonction des modifications de sa séquence nucléotidique. Pour chacun des gènes d'intérêt, on relève l'allèle présent. La combinaison des différents allèles d'une souche définit son ST [12]. Dans le cas de *P. aeruginosa*, les gènes utilisés pour la MLST sont au nombre de 7 : *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* et *trpE* [13]. Il existe de nombreuses bases de données de MLST, chacune étant spécifique d'une espèce. Elles contiennent les allèles des différents gènes de ménage, les différentes combinaisons d'allèles associées aux STs, et parfois des moteurs de recherches intégrés permettant de caractéri-

ser l'allèle à partir de la séquence des gènes. La base de données de MLST sur *P. aeruginosa* est PubMLST [14].

Contrairement au PFGE, la technique MLST a l'avantage d'être parfaitement reproductible d'un laboratoire à un autre.

La figure 1.4 représente les 2106 STs répertoriés en juin 2015 pour l'espèce *P. aeruginosa*. Cette figure permet de comprendre les liens entre les différents STs. Les STs partageant 5 allèles sur les 7 gènes de ménage sont reliés par un trait sur le graphe. Chaque regroupement de STs est appelé complexe clonal (CC) et le ST se trouvant au centre du CC (en rose sur la figure) en est le fondateur. La majorité des STs n'appartiennent pas à un CC, mais il y a quelques CCs importants comme ceux des ST235, ST274, ST244 et celui du ST17 qui est le CC le plus étendu.

WGS

Whole Genome Sequencing consiste en le séquençage complet d'un génome. L'utilisation du WGS en épidémiologie est assez récente (à partir des années 2000 [15]) et sert entre autre au suivi des maladies infectieuses bactériennes [16]. Les deux autres techniques citées précédemment sont moins discriminantes que la technique de WGS qui permet une analyse plus fine. Le développement du WGS permet la création d'une vraie carte d'identité propre à chaque bactérie en séparant des souches appartenant à un même pulsotype ou à un même ST.

Cette technique vise en effet à obtenir l'information contenue dans l'ensemble du génome observé ce qui en fait la technique la plus précise à ce jour. Elle est principalement utilisée afin d'identifier le génotype d'un isolat, de prédire ses propriétés telles que la résistance et la virulence mais aussi pour surveiller l'émergence et la diffusion de pathogènes [17, 18].

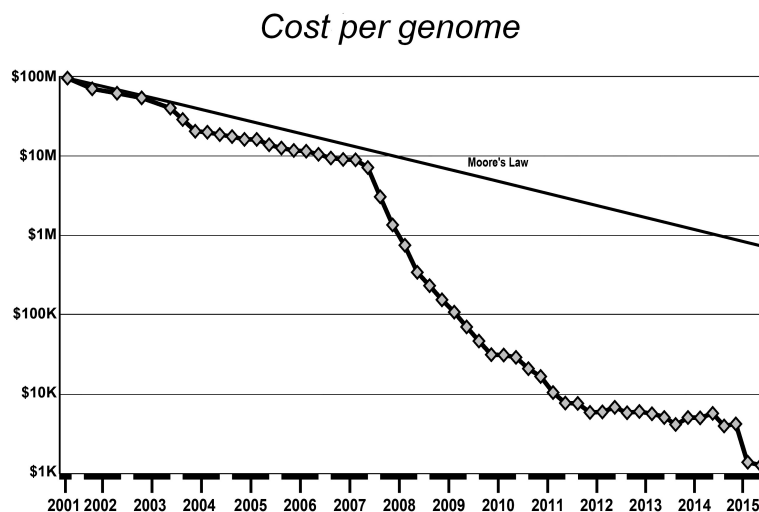


FIGURE 1.5 – Évolution du coût du séquençage pour un génome humain entre 2001 et 2015 [19]

L'apparition des techniques de séquençage haut-débit en 2007 puis le développement de ces techniques de NGS (*Next Generation Sequencing*) a permis une forte diminution du coût du séquençage (figure 1.5) et une augmentation des capacités ces dix dernières années. Le coût encore assez élevé ainsi que la difficulté d'analyse des résultats générés - nécessitant souvent l'intervention d'un bioinformaticien - empêche pour le moment une utilisation systématique de cette méthode. Pourtant, les techniques en constante évolution ainsi que la disponibilité croissante de logiciels d'analyses tendent vers une utilisation accrue du WGS en clinique.

1.2.2 Épidémiologie de *P. aeruginosa*

L'analyse de la composition génétique de *P. aeruginosa* démontre l'existence d'un *core genome* important (entre 4455 et 5316 gènes) associé à un génome accessoire de taille variable [7]. Celui-ci est fréquemment composé d'éléments externes que ce soient des plasmides ou encore des blocs d'ADN étrangers (venant d'autres espèces) intégrés dans le chromosome. Ce génome accessoire est en majorité composé d'éléments acquis par transferts horizontaux (souvent grâce à des bactériophages) [11].

La structure de la population de *P. aeruginosa*, avec un fort taux de recombinaison entre les isolats de l'espèce, indique une population panmictique [11]. La population de *P. aeruginosa* est considérée comme épidémique non-clonale [13]. Pourtant, les différents STs présentés précédemment (figure 1.4) peuvent être assimilés à des clones [11].

Parmi les différents STs, certains d'entre eux sont caractérisés comme des clones épidémiques à haut-risque, répandus ou internationaux car ils sont plus épidémiogènes que les autres et diffusent plus [11].

1.2.3 Clones à haut risque ayant une diffusion internationale

Parmi les différents STs de *P. aeruginosa*, 2500 profils existants en avril 2017, il y en a quelques uns que l'on retrouve de manière généralisée dans différents pays. Ces STs particuliers sont généralement multi-résistants aux antibiotiques et à l'origine de nombreuses épidémies hospitalières dans le monde.

Ils sont dénommés "clones à haut risque". Ces STs sont les ST111, ST175, ST235, ST244, ST253 et ST395 [11].

Le ST111 a une distribution mondiale bien qu'il soit tout de même moins fréquent que le ST235 [11]. Ce dernier est le plus répandu puisqu'il est retrouvé dans de nombreux pays sur tous les continents [20]. En revanche, le ST175 est très spécifique à l'Europe et hors de celle-ci, il est surtout retrouvé au Japon [11]. La présence du clone ST253 dans l'environnement est décrite en France, en Australie et au Royaume-Uni [21] mais également en milieu hospitalier [22, 23]. Le clone ST244 est retrouvé en Asie [24] mais également en Europe [25]. Plusieurs articles font état de la présence du ST395 en Europe [26, 27] où il est retrouvé dans les eaux usées urbaines et hospitalières [21, 28].

1.2.4 Caractéristiques des clones à haut risque épidémiques

En lien avec la dispersion des clones à haut risque dans le monde, il est intéressant de comprendre pourquoi ce sont ces clones qui sont majoritairement à l'origine d'épidémies. Le premier élément intéressant est l'observation de la résistance chez ces clones. En effet, en plus de la résistance naturelle, qui sera abordée dans le paragraphe 1.4, la grande majorité de ces clones sont multi-résistants aux antibiotiques par acquisition de matériel étranger ou par mutations.

C'est le cas en particulier du ST175 pour qui l'acquisition de résistance se fait surtout par de nombreuses mutations. Pour ce qui est des ST235 et ST111, on retrouve fréquemment l'acquisition d'une β -lactamase à large spectre [11]. Ces enzymes induisent une résistance aux céphalosporines de 3^{ème} génération et/ou aux carbapénèmes.

L'acquisition d'éléments mobiles par ces STs est également un facteur important dans l'augmentation de la résistance aux antibiotiques.

La diversité des résistances acquises ainsi que leur distribution géographique laisse à penser que bien que la dissémination des clones soit globale, l'acquisition de résistances spécifiques peut être locale [20, 29].

La virulence contribue à la pathogénicité des clones et peut être impliquée dans le succès des clones répandus. Une étude montre que certains des facteurs tels que la capacité à se déplacer (différents types de mobilités, que ce soit la capacité à nager, se déplacer sur une surface ou encore se déplacer au sein d'un groupe de bactéries) et la production de pigments sont fortement réduits chez les clones à haut-risque [30]. Ce phénotype est associé à une réduction de *fitness* lors d'expérience de compétition, c'est à dire que la capacité à se reproduire et à perdurer sur un milieu est moindre par rapport à des clones appartenant à des STs non répandus [30]. En revanche, ces clones à haut-risque ont une capacité accrue à former du biofilm permettant à la bactérie de survivre plus longtemps sur une surface tout en se protégeant contre les agressions externes. Le ST111 est capable d'acquérir plus facilement des mutations spontanées [30].

Lors de l'adaptation à un hôte, les souches de *P. aeruginosa* modulent leur virulence avec des réductions de la mobilité, une baisse de la production de pigment, une perte de *fitness*, une augmentation de la fréquence de mutation et une formation accrue de biofilm [31]. Les clones à haut risque sont également concernés par cette évolution.

Bien qu'il existe un grand nombre de génomes complets (94 en juillet 2017) déposés sur le NCBI, moins de la moitié sont décrits dans des publications scientifiques (35 articles), qui sont pour la plupart des *genome announcements* succincts. De plus lorsque c'est le cas, il est rarement question de la description du ST, seuls 4 d'entre eux indiquent celui de la souche décrite. Seules les caractéristiques des ST235, ST111 et ST175 ont été décrites [11, 20]).

1.2.5 Génomique d'épidémies hospitalières massives

Plusieurs études menées sur des épidémies de *P. aeruginosa* permettent de mieux comprendre l'origine, le mécanisme de transmission, l'évolution et les modifications génétiques influençant la modification du phénotype. Quelques épidémies de *P. aeruginosa* ont été analysées en utilisant une technique WGS. Les différentes approches utilisées par les auteurs sont détaillées dans le tableau 1.1.

Épidémie	Birmingham (UK) [28]	Allemagne [32]	UK (2 hôpitaux) [33]	UK [34]	UK (24 hôpitaux) [35]	Ontario (Canada) [36]	Danemark [37]	Liverpool (UK) et Ontario (Canada) [38]
Quand ?	300 jours	2009-2012	2005-2011, 2009-2010	2001-2007	2005-2014	/	1973-2008	1998-2007
Nombre d'isolats	141	41	9	5	87	24	45	7
ST	ST395	ST308	ST111	Non	ST111	Non	Non	Non
PFGE	Non	Non	Oui	Non	Non	Non	Non	Non
Analyse SNP	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Gain/Perte gènes	Non	Non	Non	Oui	Oui	Oui	Oui	Oui
Résistance	Oui	Oui	Oui	Oui	Oui	Oui	Non	Non
Virulence	Oui	Oui	Oui	Non	Non	Oui	Non	Oui
Phylogénie	Oui	Oui	Oui	Non	Oui	Oui	Oui	Non
tMRCA	Non	Oui	Non	Non	Oui	Non	Non	Non
Transmission	Oui	Oui	Non	Non	Non	Non	Non	Non

TABLEAU 1.1 – Bilan des épidémies de *P. aeruginosa* analysées par WGS

La plupart des épidémies décrites se trouvent sur le continent européen. Le nombre d'échantillons disponibles est très variable et peut être assez limité ou relativement conséquent (entre 5 et 141 isolats séquencés). La technique de PFGE est rarement utilisée pour typer les souches alors que la technique de MLST est utilisée dans la moitié des cas. Pour ce qui est de l'analyse génomique, une analyse de SNPs est toujours menée et est presque à chaque fois couplée à une analyse phylogénétique. Peu d'études utilisent les données d'évolution afin de dater un ancêtre commun. La composition en gènes est recherchée dans la majorité des cas parfois aidée ou remplacée par la recherche spécifique de gènes de résistance et de virulence. Celle-ci dans le cas d'absence de gènes est souvent associée à des tests phénotypiques. Pour finir, certaines études reconstruisent les événements de transmission de la souche au long de l'épidémie. Mais cette analyse n'est possible qu'à condition d'avoir toutes les informations nécessaires (liste exhaustive des souches de l'épidémie, informations sur les patients tels que chambre et durée d'hospitalisation).

Afin de mieux comprendre l'évolution de ces souches épidémiques, deux épidémies seront détaillées ci-après avec les différentes informations autour de la propagation et la modification de la bactérie. Il s'agit de deux épidémies ayant eu lieu respectivement en Angleterre [28] et en Allemagne [32] (épidémies 1 et 2 respectivement dans le tableau 1.1).

La première épidémie est décrite par *Quick et al.* en 2014 [28]. Il s'agit d'une souche de *P. aeruginosa* ST395 qui a diffusé dans le service des grands brûlés d'un hôpital anglais récemment ouvert, à Birmingham. Des échantillons prélevés sur le réseau d'eau et dans l'environnement des patients (86 isolats) sont comparés à ceux venant des patients (55 isolats venant de 5 patients).

L'ensemble des isolats cliniques appartient au même clone. La présence de celui-ci dans l'environnement proche des patients était attendu. En revanche, sa présence dans le réseau de plomberie prouve que l'eau est la source principale de contamination. L'utilisation de techniques de WGS a donc permis aux auteurs de conclure sur une contamination de l'hôpital avant même son ouverture indiquant une probable contamination du service des grands brûlés avant l'arrivée des patients.

Au cours de l'épidémie, deux plasmides ont été mis en évidence, bien qu'ils ne soient jamais présents en même temps dans une souche. Leur origine a pu être reliée à des chambres puisque le premier d'entre eux correspond à des isolats prélevés dans l'environnement d'une seule chambre tandis que le deuxième est présent dans les prélèvements venant de la douche d'une autre chambre. De plus, au sein des isolats de l'épidémie, des mutations liées à l'acquisition de résistance ont été démontrées (mutation T83I dans *gyrA* et H321Y dans *mexS* induisant une résistance à la ciprofloxacine, des indels dans *nalC* induisant une résistance au méropénème et des indels dans *oprD* conduisant à une résistance à l'imipénème et au méropénème).

La deuxième épidémie décrite a eu lieu en Allemagne [32] entre juillet 2009 et mars 2012. Un total de 26 patients est concerné par l'étude, cela correspond à 41 prélèvements associés à 8 prélèvements environnementaux. Deux services de l'hôpital ont été touchés, le premier est un service général contenant des chambres doubles alors que le second ne contient que des chambres simples et appartient au service de soins intensifs.

La technique WGS est utilisée afin de rechercher les informations propres à cette épidémie. La reconstruction du *core genome*, a permis d'identifier 24 SNPs. Parmi ces 24 nucléotides modifiés, 17 sont intra-géniques dont 11 sont non-silencieux.

Le typage par MLST a démontré l'appartenance du clone au ST308. Les différents isolats de l'épidémie possèdent une résistance particulière due à la présence d'une metallo- β -lactamase de type IMP-8, associée à une mutation (D106G) dans *oprD* conférant de la résistance à l'imipénème.

La reconstitution d'un arbre phylogénétique utilisant des données temporelles a permis de dater l'origine de la souche épidémique en juillet 2008, soit un an avant le cas index. La vitesse d'évolution de la souche a été évaluée à 1,18 SNPs par an. Les données concernant les patients (service et date d'hospitalisation) ont été utilisées afin de créer un arbre de transmission permettant de comprendre l'ensemble de l'épidémie. Les critères utilisés sont (1) la présence de deux patients dans le même service au même moment, (2) la présence dans la même chambre à moins de 14 jours d'écart et consécutivement l'un à l'autre, (3) la

présence dans la même chambre au même moment et (4) la proximité génétique entre les isolats bactériens.

1.2.6 Notion de vagues épidémiques : autres pathogènes

Il existe peu d'articles décrivant des épidémies causées par *P. aeruginosa*, et il est difficile d'avoir une vision complète de l'évolution du pathogène dans ces cas précis. Il existe en revanche chez d'autres micro-organismes des articles ayant une vision plus large de l'évolution permettant de mieux caractériser l'épidémie.

Il existe un grand nombre de pathogènes humains à l'origine d'épidémies, que l'on parle de bactéries ou de virus. Les épidémies hospitalières causées par ces deux types de micro-organismes peuvent avoir une évolution semblable, après une flambée épidémique initiale, on observe une extinction de la souche causée à la fois par des modifications génomiques et des mesures d'hygiène mises en place [39, 40].

Fièvre hémorragique à virus Ebola

Le virus à l'origine des épidémies d'Ebola est assez rare. Ce virus responsable de fièvre hémorragique a un taux de mortalité très élevé, entre 25 et 90% de décès [40]. Le virus possède un réservoir animal, aussi la contamination d'une seule personne par un animal peut entraîner une épidémie brutale. La contamination se fait par simple contact avec du sang, des sécrétions ou des liquides biologiques de personnes infectées ou de manière plus indirecte par le contact avec des surfaces ou matériaux contaminés tels que le linge de lit ou les vêtements. La grande diversité de contaminants conduit à une propagation rapide du virus dans un temps court. La mise en quarantaine est le moyen le plus simple de limiter la contamination. L'épidémie la plus longue et la plus complexe à ce jour est celle ayant eu lieu récemment (2014-2015) puisqu'elle comptabilise plus de 28.000 cas contre 425 pour la deuxième plus grosse épidémie (ayant eu lieu en Ouganda en 2000). Les pays principalement touchés sont la Guinée, le Libéria et le Sierra Leone. Le cas index est un enfant de 2 ans, mort le 6 décembre 2013 [41]. Une étude menée sur le génome de cette souche épidémique par *Ramaiah et al.* [42] a montré qu'au cours de l'épidémie, la souche a évolué pour minimiser sa détection par le système immunitaire en accumulant des mutations lui permettant de mieux survivre. Ils observent deux types de sélections naturelles : la sélection positive et la sélection négative. La première consiste en la diffusion de caractères qui apportent un avantage sélectif à l'organisme alors que la seconde correspond à l'inhibition de la propagation de caractères nuisibles [43]. En effet, en observant les acides aminés qui composent les protéines du virus, les auteurs ont montré qu'il y avait 11 acides aminés qui étaient positivement sélectionnés tandis que 14 sont négativement sélectionnés. Tous les sites positivement sélectionnés sont présents sur des protéines directement en rapport avec la résistance au système immunitaire de l'homme indiquant un effet sur l'adaptation de l'*ebolavirus* à l'homme et donc un avantage évolutif pour le virus. De plus, les protéines en question sont celles qui évoluent le plus rapidement chez le virus indiquant une adaptation de celui-ci à son hôte dès les premiers stades de l'épidémie.

Choléra

La bactérie responsable du choléra, *Vibrio cholerae*, a été à l'origine de grandes épidémies

dans le monde, aussi appelées pandémies. Depuis l'apparition de ce pathogène en 1817 dans le golfe du Bengale [44], 7 pandémies ont eu lieu causant un grand nombre de morts sur les différents continents. La dernière pandémie a commencé en 1961 et s'est propagée à travers les continents avant de s'installer de manière permanente sur le sol africain recensant le plus grand nombre de cas depuis [44].

Pourtant, c'est en Haïti que s'est déclarée la dernière épidémie en 2010 [45] bien qu'aucune mention de choléra n'ait été faite auparavant dans ce pays.

Au total entre octobre 2010 et janvier 2012, 604.634 cas ont été recensés avec 7.436 décès. Une étude menée en 2013 a analysé un total de 108 isolats, comprenant aussi 3 souches népalaises [46]. L'étude des protéines antigéniques de la souche a démontré son appartenance au sérotype O1. L'analyse des gènes communs à l'épidémie a révélé 566 SNPs entre tous les isolats. L'ancêtre commun a été daté à la fin du mois de septembre 2010 permettant d'appuyer l'hypothèse principale de l'origine de l'épidémie. En effet, début octobre, des troupes des Nations Unies venant du Népal sont venues en Haïti et aurait été les porteurs de la souche responsable de l'épidémie. De plus, il est fait état de conditions sanitaires mauvaises dans les camps des soldats. Les analyses sur les souches ont permis d'étayer cette hypothèse ainsi que la déclaration du cas index (le 12 octobre) seulement quelques jours après l'arrivée des troupes (le 9 octobre) [47]. La comparaison des différents génomes de l'épidémie avec la souche de référence (un des premiers génomes de l'épidémie) a permis de montrer qu'il n'y avait aucune acquisition de gènes durant l'épidémie. En revanche, des blocs de gènes (entre 10 et 100 kb) sont perdus par certains des isolats (9 sur les 180 séquencés). Les auteurs décrivent également des réarrangements chromosomiques chez certaines souches. Des inversions dans le génome sont en effet visibles chez 9 souches de l'épidémie. L'absence d'ajout de gènes dans le génome, est appuyée par la démonstration que les souches de l'épidémie ne sont pas transformables. Les auteurs précisent aussi qu'il n'y a aucune preuve que les souches environnementales aient joué un quelconque rôle dans l'évolution de la souche épidémique.

Les analyses utilisant le WGS permettent d'obtenir de nombreuses informations sur l'évolution des épidémies telles que l'origine de la transmission, la datation de la contamination et un aperçu de l'évolution génique, ce qui fait du WGS un outil majeur pour l'analyse de l'évolution des épidémies microbiennes.

1.3 Évolution génomique

1.3.1 Mécanismes d'évolution

Il existe deux grands types de modifications permettant aux bactéries d'évoluer. Le premier est l'apparition de mutations ponctuelles que sont les SNPs et les indels et le second est l'ajout de matériel génétique étranger par les *Mobile Genetic Elements* ou éléments mobiles (MGEs).

A. Mutations ponctuelles

Les mutations ponctuelles peuvent se produire dans n'importe quel endroit du génome et avoir ou non un impact sur la bactérie (figure 1.6), par modification ou perte de fonction spécifique.

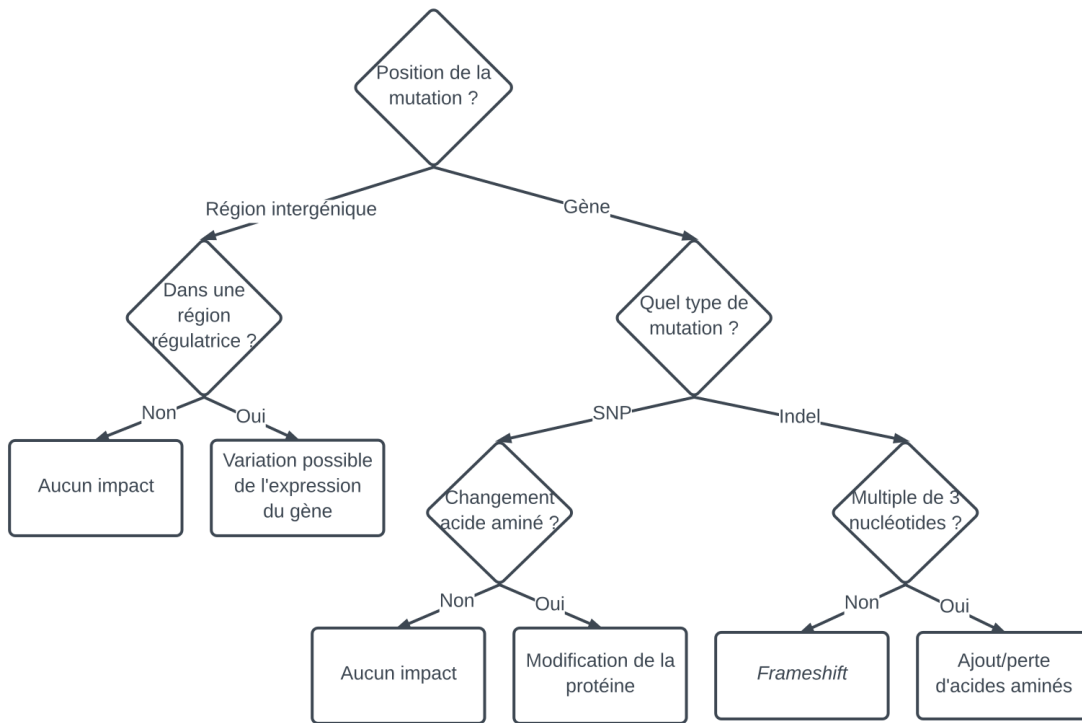


FIGURE 1.6 – Mutations possibles et leurs impacts

En effet, cela va dépendre de l'endroit où la mutation aura lieu et du type de mutation que l'on observe. Il peut y avoir un changement de nucléotide à un endroit précis du génome, un nucléotide échangé par un autre, ou bien la perte ou le gain d'un ou plusieurs nucléotides, on parle alors d'insertion/délétion ou indel. Ensuite, l'intérêt est de comprendre où se situe la mutation. Si elle est présente dans un gène ou dans une région intergénique. Dans le second cas, la mutation peut avoir un impact si elle se trouve dans une zone ayant un intérêt pour l'expression des gènes (site de fixation du ribosome ou promoteur), sinon il n'y aura pas d'effet visible sur le phénotype. En revanche, si la mutation intervient dans le gène et qu'elle est de type indel, il y aura obligatoirement une modification de la protéine codée par le gène, celle-ci pourra être plus courte, plus longue, non fonctionnelle ou encore non exprimée. Si la mutation entraîne une substitution d'acide aminé, alors cela dépendra de sa position dans le gène et du type de changement.

Lorsqu'un gène est exprimé, les nucléotides qui composent le gène sont transcrits en ARN puis traduits en acides aminés. Un acide aminé correspond à un triplet de nucléotides, appelé codon, visible sur la figure 1.7 qui représente l'ensemble du code génétique avec les différentes combinaisons de nucléotides formant les acides aminés. On observe donc qu'un même acide aminé peut être codé par plusieurs codons. Lorsqu'une mutation intervient, elle peut ne pas changer l'acide aminé si elle transforme un codon en un autre équivalent, donnant le même acide aminé. Dans ce cas, la mutation est dite synonyme et n'a aucune influence sur la protéine. En revanche, la modification d'un acide aminé peut avoir pas, peu ou beaucoup

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG } Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

FIGURE 1.7 – Code génétique standard : codons correspondants aux différents acides aminés

d'effets sur la protéine suivant l'endroit où il est placé. De plus, il est possible de transformer un acide aminé en codon stop, ou un codon stop en acide aminé conduisant respectivement à une protéine tronquée ou rallongée. On parle alors de mutations non synonymes qui ont généralement un impact sur la protéine et qui vont donc altérer une fonction biologique de la bactérie.

B. Acquisition/perte d'éléments mobiles

Plusieurs types d'éléments mobiles ou *Mobile Genetic Elements* (MGEs) existent [48]. Ils sont tous capables de s'intégrer au sein d'un génome et d'apporter du matériel génétique étranger qui peut contenir de l'information nouvelle et utile pour la bactérie. Cela peut être des gènes impliqués dans le métabolisme, la résistance aux antibiotiques ou aux métaux ou encore des facteurs de virulence. Ils contiennent également les gènes nécessaires à leur propre mobilisation.

Insertion Sequence ou IS.

Ce sont les plus petits éléments mobiles connus. Ils ne contiennent que les gènes responsables de leur transfert, une transposase et une protéine régulatrice qui sert à activer ou inhiber le déplacement (figure 1.8).

Ils peuvent s'insérer à diverses positions dans le génome. Ils sont fréquemment responsables d'apparition de résistances lorsqu'ils s'insèrent dans des gènes d'intérêt, exemple de la résistance aux carbapénèmes lors de l'insertion dans le gène *oprD* [49]. Les ISs sont souvent identifiables grâce aux séquences répétées qui les encadrent (figure 1.8). En effet, il existe deux types de répétitions, d'une part des séquences inversées répétées qui sont directement portées par l'IS et qui s'excisent et s'insèrent avec lui lors de ses déplacements, et d'autre part des séquences directement répétées qui peuvent se créer lors de l'insertion de l'IS, et sont en lien direct avec son mécanisme d'insertion. Dans la majorité des cas, les deux types

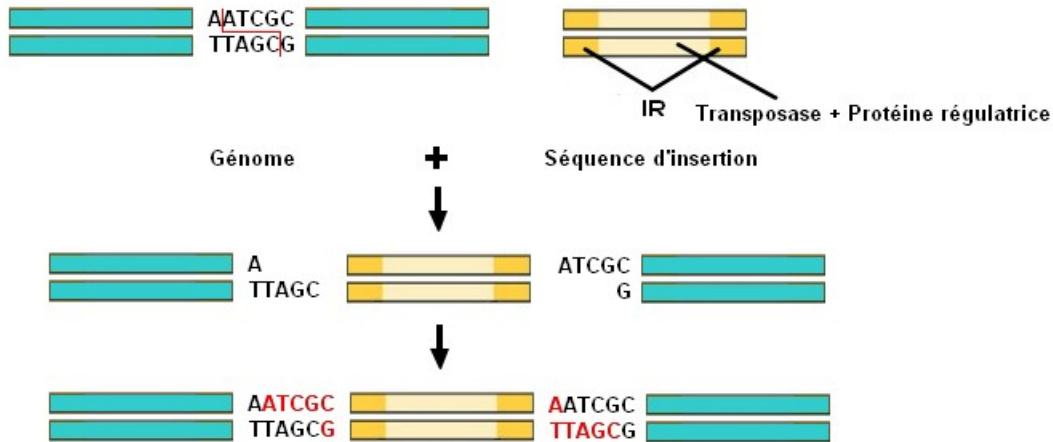


FIGURE 1.8 – Schéma descriptif de la structure d'un IS

de séquences répétées existent. Dans de rares cas, il peut y avoir soit l'un soit l'autre soit aucun des deux.

Il existe également des transposons composites qui sont composés de deux ISs (identiques) qui transportent des gènes entre eux (figure 1.9). Dans ce genre de cas, l'IS possède des séquences inversées répétées qui l'encadrent. Ce type d'élément mobile participe au gain de gènes au sein des génomes en transportant de l'information d'intérêt pour la bactérie.

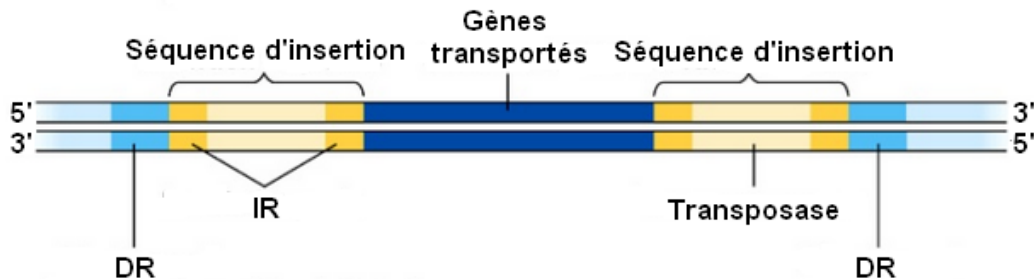


FIGURE 1.9 – Schéma descriptif de la structure d'un transposon composite [50]

Transposons.

Les transposons sont des éléments génétiques capables de se déplacer au sein d'un génome ou de plusieurs génomes suivant deux mécanismes [51]. Le premier est un couper-coller, où l'élément n'est pas conservé à la position de départ dans le génome. Dans le deuxième cas, c'est un copier-coller permettant de dupliquer la séquence de l'élément (figure 1.10).

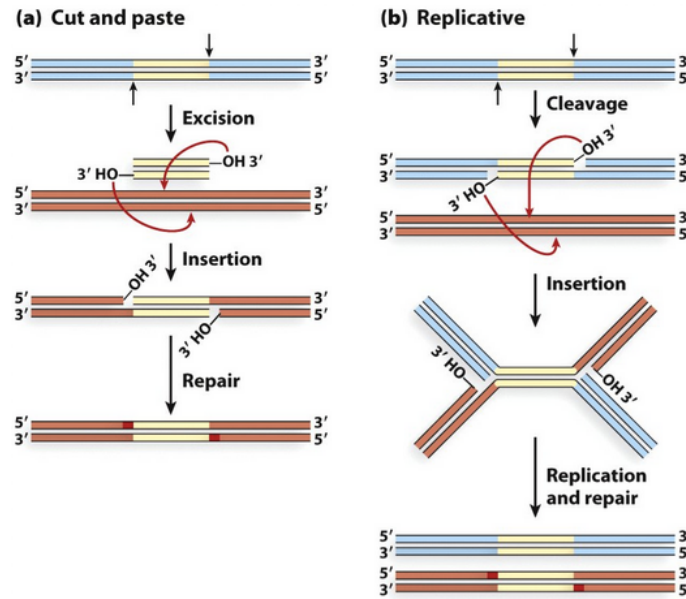


FIGURE 1.10 – Mécanisme de transfert d'un transposon [52] (a) Mécanisme de déplacement en couper/coller, (b) Mécanisme de duplication en copier/coller

Intégrons.

Un intégron possède une intégrase, enzyme qui permet l'intégration, et un site de recombinaison appelé *attI* où de nouveaux gènes contenus dans des cassettes pourront s'intégrer [53]. Le mécanisme d'intégration d'une cassette de gènes est visible sur la figure 1.11. Cet élément n'est pas mobile puisqu'il ne peut se déplacer à moins d'être localisé dans un autre élément mobile (type transposon). C'est l'ajout de cassette contenant des gènes d'intérêt (résistance et virulence entre autres) qui lui permettent de conférer un avantage évolutif à la bactérie.

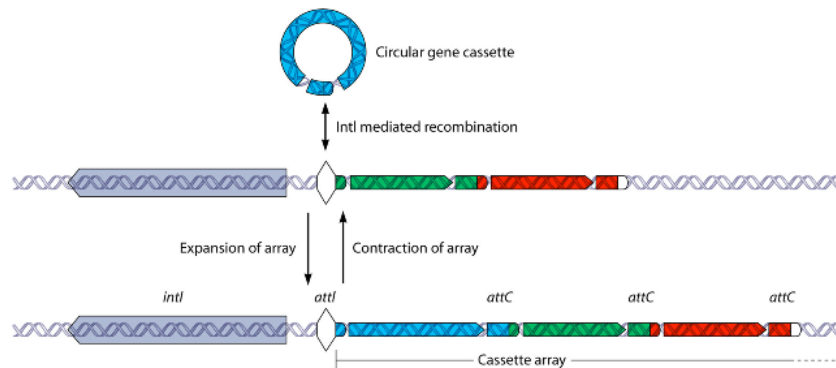


FIGURE 1.11 – Intégration d'une cassette dans un intégron [53]

ICEs, Integrative and Conjugative Elements.

Ce sont des structures qui sont intégrées dans des chromosomes bactériens mais qui peuvent suivant les conditions s'exciser et se transmettre par conjugaison. Le cycle de vie de l'ICE

est visible sur la figure 1.12. Il contient tous les gènes nécessaires à sa mobilisation et peut également contenir toutes sortes de gènes d'intérêt. C'est un élément de grande taille (entre 10 et 500 kb en fonction des espèces) qui ne peut pas se maintenir sous forme circulaire. En effet, s'il n'est pas rapidement intégré au chromosome bactérien, il est perdu, puisqu'il ne peut se répliquer et se transmettre dans les cellules filles [54].

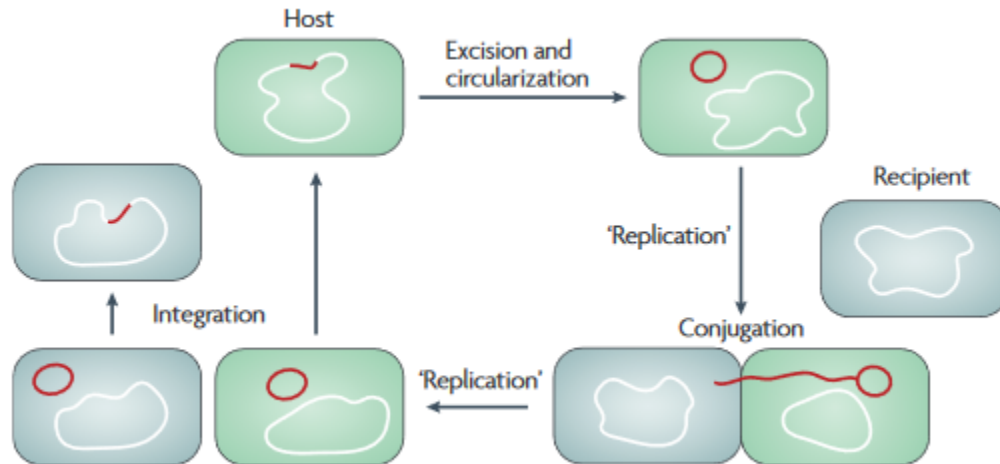


FIGURE 1.12 – Cycle de vie d'un ICE [54]

Plasmides. Ce sont des éléments indépendants du chromosome, qui sont capables d'auto-réplication [48]. Ils sont composés d'un double brin d'ADN circulaire d'une taille comprise entre 2 et 500 kb. Il n'y a pas de gènes codant pour des fonctions essentielles de la bactérie dans les plasmides. En revanche, il y a toute la machinerie cellulaire leur permettant d'avoir une réplication autonome, distincte de celle du chromosome. Il existe un phénomène d'incompatibilité de plasmide. En effet, deux plasmides ayant le même mécanisme de réplication ne peuvent coexister dans la même bactérie. Il existe 14 groupes d'incompatibilité chez le genre *Pseudomonas* [55].

L'origine de ces éléments mobiles est diverse [48]. Le transfert d'informations peut se faire aussi bien au sein d'une même espèce qu'entre espèces différentes. Aussi il est fréquent de retrouver les mêmes éléments mobiles chez plusieurs espèces bactériennes différentes.

Un des principaux marqueurs de l'intégration d'éléments étrangers est une variation dans leur taux de G-C. En effet, chaque espèce bactérienne va utiliser de manière différente les nucléotides pour coder les acides aminés, aussi appelé biais de codon. Aussi, chaque espèce est caractérisée par un taux de G-C spécifique, par exemple le G-C pourcent moyen de *P. aeruginosa* est de 66,2%, celui d'*E. coli* est de 50,6% alors que celui de *S. aureus* n'est que de 32,8%. Lors d'intégration d'éléments mobiles, le taux de G-C varie puisque l'élément inséré possède un taux de G-C différent du génome receveur. Ces biais ponctuels dans la composition en G-C permettent l'identification des anciens éléments mobiles intégrés dans le génome. C'est le cas notamment des *Genomics Islands* (GIs) qui sont d'anciens éléments intégrés de-

venus immobiles [48].

Plus un élément mobile a été inséré il y a longtemps dans un génome, plus sa variabilité en taux de G-C par rapport au génome receveur est faible. En effet, au cours du temps, il existe un phénomène de dérive génétique où le taux de G-C des éléments intégrés devient identique à celui du génome receveur [7].

1.3.2 Plasticité particulière du génome de *P. aeruginosa*

Comme vu précédemment, le génome du pathogène humain *P. aeruginosa* est capable de se réorganiser en modifiant son contenu génique par des mutations ou des gains et pertes de gènes (par acquisition ou perte d'éléments mobiles par exemple). Cette évolution au cours du temps se produit chez tous les génomes à plus ou moins grande vitesse. Bien que la partie conservée du génome soit grande, entre 4455 et 5316 gènes, il existe une variabilité non négligeable dans le génome accessoire montrant que *P. aeruginosa* a une grande capacité d'adaptation [7]. En effet, le génome comptant plus de 6000 gènes en moyenne, le nombre de gènes du génome accessoire est élevé, ce qui lui confère sa capacité à s'adapter à une grande variété de milieux.

En utilisant des mécanismes génétiques qui favorisent le mouvement et l'altération du matériel génétique, *P. aeruginosa* est capable de modifier son génome pour s'adapter aux exigences d'environnements hostiles. De plus, l'acquisition de nouveau matériel génétique, et donc de nouvelles caractéristiques, ne modifie pas celles qui sont présentes. L'organisme conserve toujours sa capacité à survivre dans beaucoup d'environnements différents, incluant les patients [56].

Il existe également une grande proportion de souches (environ 20% [57]) qui deviennent hypermutatrices dans les cas d'infections chroniques. La mutation d'un gène impliqué dans le mécanisme de réparation des erreurs de l'ADN (*mutS*, *mutL*, *mutH*, *mutU* ou *uvrD*) entraîne une augmentation significative du taux de mutations. Dans le cas d'une bactérie dont le mécanisme de réparation d'erreurs de l'ADN est intact, la fréquence de mutations n'est pas très élevée, avec une moyenne de 2.5×10^{-8} . Lorsque ce système est lésé, on parle de souches hypermutatrices puisque la fréquence de mutations est augmentée d'environ 100 fois, fréquence moyenne de 3.2×10^{-6} [58]. L'acquisition d'un phénotype enclin aux mutations représente un avantage évolutif pour la bactérie en condition de stress puisque cela lui permet de s'adapter plus rapidement comme c'est le cas lors d'infection chronique chez des patients atteints de mucoviscidose [57].

1.4 Épidémiogénicité : résistance et virulence

Afin de comprendre au mieux l'épidémiogénicité du pathogène humain qu'est *P. aeruginosa*, il est important de connaître les facteurs qui jouent un rôle dans sa diffusion et sa pathogénicité. Avec en particulier, la résistance aux antibiotiques, qu'elle soit innée ou acquise, et la virulence.

Les modifications génétiques acquises au cours de l'évolution d'une bactérie peuvent avoir un lien avec la modification des facteurs de virulence et des mécanismes de résistances aux antibiotiques. La mesure du niveau de résistance des bactéries est importante pour le traitement des patients. L'antibiogramme est un outil utile pour connaître les résistances des bactéries à une large gamme d'antibiotiques (figure 1.13).



FIGURE 1.13 – Antibiogrammes de souches de *P. aeruginosa* : sensible vs résistant

Naturellement, le génome de *P. aeruginosa* possède de nombreux facteurs impliqués dans la résistance aux antibiotiques. Tout d'abord, la membrane externe étant peu perméable, cette bactérie est intrinsèquement résistante à de nombreux agents antimicrobiens [8]. En effet, la majorité des antibiotiques utilisés contre *P. aeruginosa* doivent traverser la membrane pour agir. Malgré la faible perméabilité de cette membrane, il existe d'autres possibilités de pénétrer la cellule pour les antibiotiques. La présence de porine sur la membrane permet le passage d'une partie des antibiotiques. D'autres tels que la colistine ou les aminosides se lient aux LPS, composant de la membrane, ce qui leur permet de traverser la membrane externe pour arriver à la membrane cytoplasmique.

P. aeruginosa peut exprimer un répertoire varié de mécanismes de résistance. D'une part grâce à des mutations chromosomiques dans des gènes impliqués dans la régulation de la résistance mais également, grâce à l'acquisition de gènes de résistance d'autres organismes (d'une même espèce ou d'espèces différentes) via des plasmides ou des éléments transposables [8].

1.4.1 Généralités sur les antibiotiques

Afin de comprendre l'origine des mécanismes de résistance, la connaissance des grandes familles d'antibiotiques fréquemment utilisés en clinique est nécessaire [8]. Plusieurs grandes classes seront listées ici avec leurs actions sur la cellule bactérienne. Les aminosides sont des sucres ayant une fonction amine ciblant la petite sous-unité du ribosome afin de bloquer la traduction et donc inhiber la synthèse de protéines. Les quinolones inhibent les gyrases, des enzymes essentielles à la réplication de l'ADN ainsi qu'au maintien de la structure du chromosome dans la cellule. Les β -lactamines possèdent un cycle β -lactame et on retrouve dans cette famille les pénicillines, les carbapénèmes et les céphalosporines. Ces antibiotiques bloquent la synthèse de peptidoglycanes qui sont les composants majoritaires de la paroi des

cellules. Pour finir, les polymyxines sont des peptides cycliques avec une queue hydrophobe qui se lie aux phospholipides de la membrane cytoplasmique ou membrane interne et la détruisent lui faisant ainsi perdre son rôle de barrière.

P. aeruginosa est naturellement résistant aux cyclines et aux macrolides qui agissent au niveau du ribosome, ainsi qu'aux sulfamides qui ciblent la synthèse de l'acide folique (vitamine B9) et aux imidazoles qui inhibent la synthèse des acides nucléiques [8].

Trois grands mécanismes conduisent à une résistance aux antibiotiques, (1) une réduction de la perméabilité de la membranes [59], (2) l'inactivation enzymatique de l'antibiotique et (3), la modification de la cible de l'antibiotique.

1.4.2 Résistance naturelle

La membrane externe de la bactérie sert de première barrière à l'entrée des composés dans la cellule. Une partie des molécules est bloquée à cet endroit et ne peut atteindre leur but dans la bactérie. Certaines molécules peuvent en revanche entrer par l'intermédiaire des porines, qui sont des canaux traversant la membrane externe de la cellule bactérienne, par des canaux spécifiques ou encore simplement en traversant la bicouche lipidique constituant la membrane externe (figure 1.14).

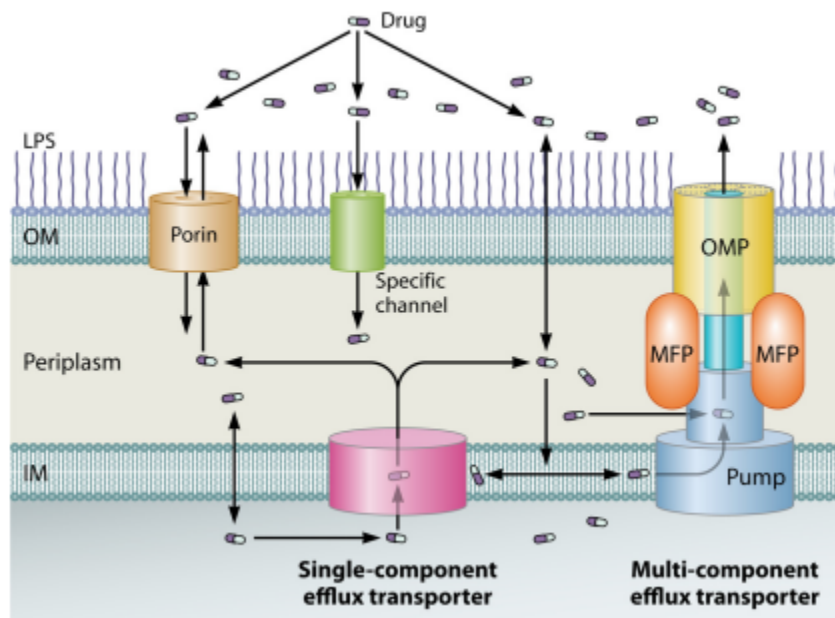


FIGURE 1.14 – Entrée et sortie des antibiotiques dans la cellule bactérienne [59]

Les systèmes d'efflux ont un rôle dans la résistance aux antibiotiques, grâce à leur capacité de relargage des composés néfastes à l'extérieur de la cellule. Chaque système d'efflux de type *Resistance Nodulation Division* (RND) est composé de trois protéines, une pompe localisée sur la membrane interne qui nécessite de l'énergie pour fonctionner, une porine qui sert de

canal et une protéine de fusion membranaire qui fait le lien entre les deux. Les systèmes d'intérêt majeur sont au nombre de 4 chez *P. aeruginosa* : MexAB-OprM, MexXY/OprM, MexCD-OprJ et MexEF-OprN [59] (tableau 1.2).

Pompe	Résistance	Surexpression (mutations)
MexAB-OprM	Quinolones, tétracyclines, macrolides, β -lactamines	<i>mexR</i> , <i>nalC</i> , <i>nalD</i>
MexXY/OprM	Quinolones, macrolides, tétracyclines, aminosides, céphalosporines, chloramphénicol	<i>mexZ</i> , protéines ribosomales (L1, L25, L21, L27) ou <i>parRS</i>
MexCD-OprJ	Quinolones, macrolides, tétracyclines, céphalosporines, chloramphénicol	<i>nfxB</i>
MexEF-OprN	Quinolones, chloramphénicol, carbapénèmes	<i>nfxC</i>

TABLEAU 1.2 – Les pompes majeures de *P. aeruginosa*, leurs résistances associées et les gènes dont la mutation induisent la surexpression.

Chaque système d'efflux possède un certain nombre de substrats, qui sont plus ou moins spécifiques et correspondent aux composés que la pompe peut prendre en charge.

Les pompes d'efflux MexAB-OprM et MexXY sont constitutivement exprimées et ont un rôle important dans la résistance naturelle [59]. En effet, ces deux pompes sont tout le temps actives dans la bactérie. De plus, la production basale d'AmpC, une céphalosporinase, participe à la résistance innée de *P. aeruginosa*.

MexAB-OprM

Elle a été la première pompe identifiée chez *P. aeruginosa*, pour son implication dans l'efflux de la pyoverdine [60]. Son implication dans la résistance aux antibiotiques a également été démontrée [61]. Les substrats de cette pompe sont les β -lactamines, les fluoroquinolones, les tétracyclines et les macrolides, ainsi que d'autres composés tels que les désinfectants et les détergents.

MexAB-OprM est aussi impliquée dans l'efflux de composés liés au *quorum sensing* et la signalisation cellulaire [62]. Le *quorum sensing* permet l'expression coordonnée de certains gènes contrôlés par des signaux émis par les bactéries elles-mêmes, comme c'est le cas lorsqu'elles sont sous forme de biofilm. La coordination inter-bactérienne se fait par la sécrétion des homosérine-lactones (HSL) acétylées. Chez *P. aeruginosa*, les opérons *las* (*lasR* et *lasI*) et *rhl* (*rhlR* et *rhlI*) entraînent la synthèse de deux HSL : N-3-oxododecanoyl HSL (C₁₂-HSL) et N-butyryl HSL (C₄-HSL) respectivement. Ces deux molécules activent l'expression de plusieurs gènes de virulence tels que l'élastase *lasB*, la protéase *lasA*, l'opéron *xcp* codant pour le système de sécrétion de type III (sécrétant les toxines synthétisées par la bactérie).

MexAB-OprM est directement impliquée dans l'efflux de C₁₂-HSL participant ainsi à la régulation du *quorum sensing* et à la virulence de la souche [63]. De plus, C₁₂-HSL active la réaction inflammatoire chez le patient infecté [64].

Cette pompe est également impliquée dans la capacité de *P. aeruginosa* à envahir des cellules épithéliales. En effet, Hirakata et *al.* [65, 66] démontrent que la perte du système d'efflux entraîne la diminution de la capacité d'invasion. L'expérience est menée sur des cellules de rein de chien (MDCK), cultivées *in vitro* et formant une couche représentant la barrière d'un organe. Les bactéries ne possédant pas la pompe MexAB-OprM, tout comme celles qui la surexpriment, envahissent moins les cellules que les bactéries dont l'efflux est intact. Ils concluent donc sur l'importance de ce système d'efflux dans l'export de composés servant à l'invasion des cellules de l'hôte.

MexXY (OprM)

Cette pompe est également constitutivement exprimée, et c'est la porine OprM qui s'associe avec MexXY [59]. L'extrusion des aminosides semble spécifique à cette pompe [67]. Elle joue également un rôle dans l'export des fluoroquinolones [68].

Céphalosporinase AmpC

Les β -lactamases sont des enzymes hydrolysant le cycle β -lactame des β -lactamines. Une céphalosporinase, codée par le gène *ampC*, est présente sur le chromosome de toutes les souches de *P. aeruginosa* [7]. C'est une enzyme ayant un rôle dans la résistance intrinsèque de *P. aeruginosa* aux pénicillines et à de nombreuses céphalosporines. L'enzyme est localisée dans le périplasme de la bactérie où elle va hydrolyser l'antibiotique (figure 1.15).

Parmi les régulateurs de *ampC*, on retrouve *ampR*, *ampG* et *ampD*. *ampR* est un régulateur transcriptionnel qui comme *ampG* (codant pour une protéine transmembranaire) est impliqué dans l'induction de *ampC*. *ampD*, en revanche code pour un répresseur de l'expression de *ampC* [69].

1.4.3 Résistance par mutation de gènes chromosomiques

A. Barrière naturelle et système d'efflux

MexAB-OprM

Comme vu précédemment, cette pompe est constitutivement exprimée. La surexpression de MexAB-OprM résulte de l'inactivation des répresseurs MexR, NalC ou NalD (figure 1.16) et entraîne une résistance accrue aux substrats de la pompe [70].

De plus, la surexpression de la pompe d'efflux entraîne une diminution de la production de pyocyanine et pyoverdine, deux pigments produits par la bactérie. Les élastases et protéases permettant la dégradation des protéines et servant de facteurs de virulence à la bactérie sont également moins produites lorsque la pompe est surexprimée [63, 71]. De plus, lors de la surexpression, l'équipe de Sanchez et *al.* [71] a montré que la survie du mutant dans l'eau était réduite par rapport au contrôle sauvage. Il en est de même sur une surface sèche. En

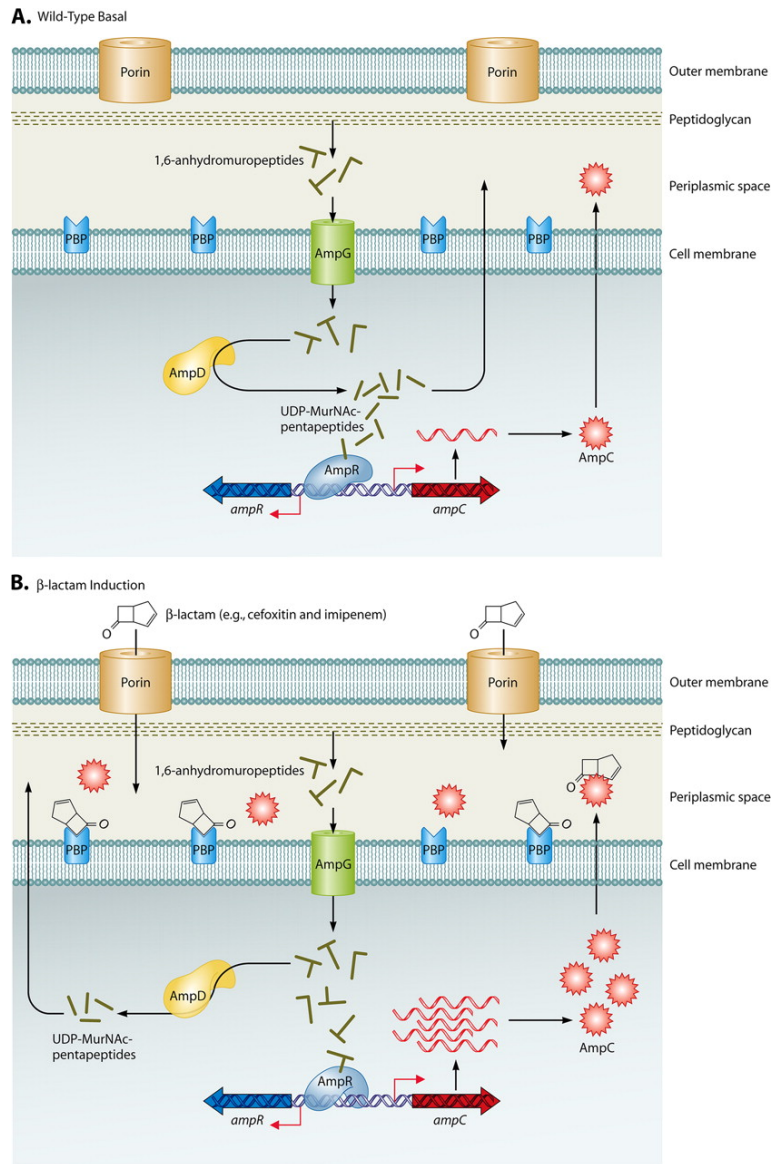


FIGURE 1.15 – Régulation de l'expression de AmpC [69]. (a) expression basale de *ampC*, l'interaction entre AmpR et le résultat du clivage de peptides par *ampD* limite l'expression de *ampC*. (b) induction de l'expression de *ampC* par les β -lactamines, l'afflux de peptides non clivés par *ampD* induit un changement de conformation de AmpR qui devient activateur de l'expression de *ampC* augmentant ainsi le niveau de AmpC dans le périplasme.

revanche, une augmentation de la production de biofilm est observée chez le mutant surproduisant MexAB-OprM. Les analyses menées par les auteurs sur la capacité de la bactérie à tuer *Caenorhabditis elegans*, un petit ver faisant environ 1 mm de longueur, montrent que les mutants *nalB* sont incapables de tuer les vers alors que la souche sauvage peut tuer jusqu'à 80% de la population de vers en 70 heures. Ceci démontre une baisse de la virulence chez les mutants surproduisant MexAB-OprM.

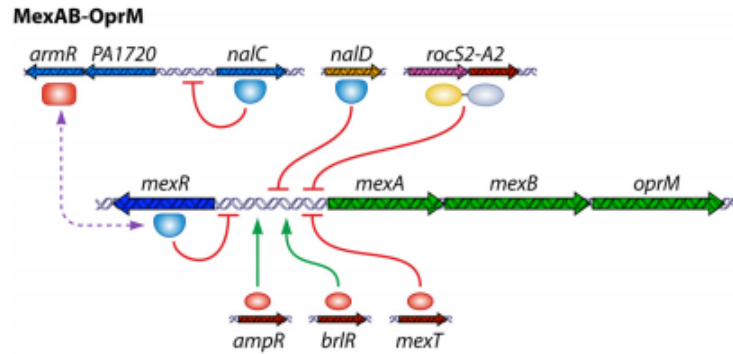


FIGURE 1.16 – Régulation de la production de MexAB-OprM [59]

MexXY

La surproduction de MexXY donne un niveau de résistance entre 2 et 16 fois supérieur à tous les substrats de la pompe, aminosides, fluoroquinolones, macrolides, tétracyclines et céphalosporines [59]. On observe trois types de mutants surproduisant MexXY (figure 1.17) : (1) avec une mutation sur *mexZ* (mutants *agrZ*) qui est un répresseur de la pompe [72], (2) avec des mutations sur les protéines ribosomales (L21 et L27, mutants *agrW1*) qui vont affecter la synthèse protéique [73] et (3) des mutations sur *parR* et *parS* (mutants *agrW2*) [74].

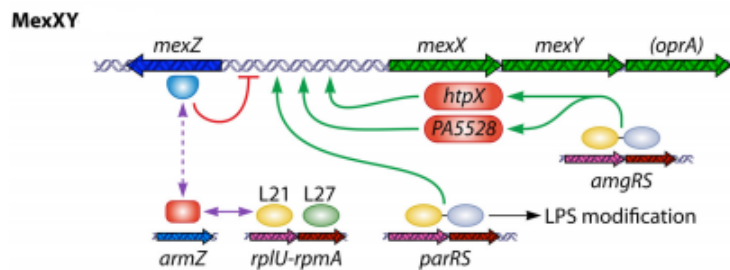


FIGURE 1.17 – Régulation de la production de MexXY [59]

MexCD-OprJ

Contrairement à MexAB-OprM et MexXY, MexCD-OprJ n'est pas exprimée constitutivement dans la cellule [59]. Cette pompe est inductible par des agents qui causent des dommages à la membrane tels que les détergents ou les solvants. La surexpression de cette pompe dans des mutants *nfxB* (figure 1.18) entraîne l'augmentation de la résistance à plusieurs familles d'antibiotiques que sont les fluoroquinolones, céphalosporines, chloramphénicol et tétracyclines, mais semble rendre sensible aux substrats de MexAB-OprM (β -lactamines) donnant l'impression que cette dernière est déficiente [75]. De plus, la surproduction de MexCD-OprJ induit une diminution de l'expression de MeXY-OprM et de MexAB-OprM ainsi que leurs résistances associées [76].

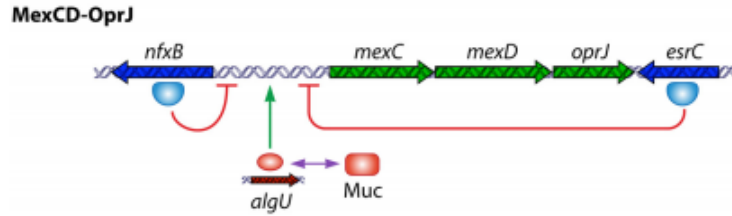


FIGURE 1.18 – Régulation de la production de MexCD-OprJ [59]

MexEF-OprN

La dernière pompe majeure est MexEF-OprN qui tout comme MexCD-OprJ n'est pas constitutivement exprimée [59]. Il y a une surproduction de MexEF-OprN dans les mutants *nfxC*, portant des mutations sur les gènes *mexS*, *mexT* ou *mvaT* [77]. Ce phénotype est rarement retrouvé chez des souches cliniques. Chez ces mutants, on observe une résistance aux fluoroquinolones, au chloramphénicol et aux carbapénèmes alors qu'ils sont hypersensibles aux aminosides et aux β -lactamines [78]. La régulation de la production de MexEF-OprN est présentée sur la figure 1.19.

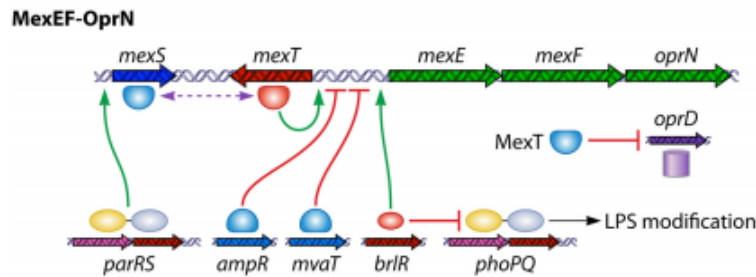


FIGURE 1.19 – Régulation de MexEF-OprN [59]

B. Modification du composé actif

La surproduction de la céphalosporinase AmpC peut subvenir à la suite de mutations dans un de ses gènes régulateurs (*ampD*, *ampDh2*, *ampDh3*, *ampR*, *ampG*, *ampO*, *ampP* et *dacB*) [11, 79, 80].

C. Modification de la cible des fluoroquinolones

Des mutations ayant lieu dans les cibles des antibiotiques peuvent les rendre résistantes à ceux-ci [11]. En effet, ces mutations ont pour but de conserver la fonction première de la protéine, qui est souvent vitale pour la cellule car impliquée dans son métabolisme, sa survie ou sa reproduction, tout en bloquant l'action des antibiotiques. C'est le cas en particulier des mutations dans le QRDR (*quinolone-resistance determining region*) de GyrA et GyrB (des gyrases, des enzymes liées à la structure du chromosome bactérien), mais également dans celui de ParC et ParE (des topoisomérases utiles à la ségrégation des chromosomes) [81].

Pompes d'efflux et virulence/*fitness* associé (Tableau 1.3)

Référence	Espèce	Système d'efflux	Changement d'état	Effet (virulence et <i>fitness</i>)
Pacheco et al. [82]	<i>P. aeruginosa</i>	Pompes RND	wt -> +++	Diminution de <i>fitness</i> contrebalancé par une modification de la respiration anaérobie (nitrate)
Cosson et al.	<i>P. aeruginosa</i>	<i>mexEF-oprN</i>	wt -> +++	Perte de la capacité à tuer des rats
Köhler et al. [83]	<i>P. aeruginosa</i>	<i>mexEF-oprN</i>	wt -> +++	Baisse de la production de facteurs de virulence extra-cellulaires (pyocyanine, élastase, rhamnolipides)
Sánchez et al. [71]	<i>P. aeruginosa</i>	<i>mexAB-oprM</i> , <i>mexCD-oprJ</i>	wt -> +++	Baisse de la survie dans l'eau, baisse de la production de phénazines et protéases et baisse de la virulence (tue moins <i>C. elegans</i>)
Hirakata et al. [65]	<i>P. aeruginosa</i>	<i>mexAB-oprM</i>	+++ -> wt/ Δ	Perte de la capacité d'invasion
Hirakata et al. [66]	<i>P. aeruginosa</i>	<i>mexAB-oprM</i>	wt -> Δ	Perte de la capacité d'invasion des cellules épithéliales, perte de la capacité à tuer des souris
Wang-Kan et al. [84]	<i>S. enterica</i> serovar Typhimurium	<i>acrAB-tolC</i>	wt -> Δ	Attenuation de virulence chez la souris et le papillon <i>Galleria mellonella</i> , réduction de la capacité d'invasion de cellules épithéliales
Ding et al. [85]	<i>S. aureus</i>	<i>norB</i>	wt -> Δ	Augmentation de la capacité de déplacement dans l'eau (<i>swimming motility</i>) Perte de <i>fitness</i> , baisse de la survie et croissance dans les abcès

TABLEAU 1.3 – Effet de la variation des pompes d'efflux sur la virulence et le *fitness* de la souche. wt : souche sauvage, Δ : pompe absente, +++ : surexpression de la pompe

1.4.4 Résistance aux antibiotiques par acquisition de matériel étranger

L'acquisition de résistance aux antibiotiques peut se faire également par intégration d'éléments mobiles dans le génome, éléments qui peuvent porter un ou plusieurs gènes de résistance, ainsi que des gènes de virulence ou d'autres qui peuvent conférer un avantage évolutif à la bactérie [86]. C'est le cas par exemple d'une souche de *P. aeruginosa* capable de vivre dans un environnement pollué par du pétrole, par acquisition de gènes lui permettant la dégradation du pétrole (des gènes de dégradation des alcanes et des phénols et des transporteurs des solvants organiques) [87]. On parle alors de transfert horizontal qui se fait d'une bactérie à une autre. Ces éléments mobiles sont de plusieurs types comme vu dans la section précédente.

Des gènes codant des β -lactamases sont fréquemment portés par ces éléments mobiles [11]. En effet, elles sont très fréquemment acquises par échange de matériel génétique entre les bactéries. Il existe différents types de β -lactamases que l'on peut retrouver dans les différentes souches de *P. aeruginosa*, souvent en lien direct avec la localisation géographique. Les β -lactamases sont classées en familles dans la classification d'Ambler : (A) les pénicillinases, (B) les métallo-enzymes, (C) les céphalosporinases et (D) les oxacillinases [88].

Des gènes impliqués dans la résistance aux aminosides sont aussi retrouvés sur des éléments transportés tels que les *aminoglycosides acetyltransferases* AAC(3') et AAC(6') [89], l'*aminoglycoside nucleotidyltransferase* ANT(2') [90] et l'aminoglycoside phosphoryltransférase APH(3') [91].

1.4.5 Facteurs de virulence

La pathogénicité bactérienne est directement liée à l'expression de facteurs de virulence. Il en existe de différents types : (1) les pigments servant à la multiplication et la persistance de la bactérie, (2) les facteurs membranaires tels que les pili de type IV, le flagelle et l'alginate qui jouent un rôle dans le mécanisme d'adhésion aux muqueuses et dans la formation de biofilm, (3) les facteurs extra-cellulaires excrétés tels que les protéases, les exotoxines, les hémolysines, la phospholipase et les toxines protéiques et (4) la création de biofilm et la communication intercellulaire appelé *quorum sensing* [2].

Les pigments. Il existe deux pigments majeurs produits par *P. aeruginosa*, la pyoverdine et la pyocyanine [2]. Le premier est un pigment jaune-vert fluorescent qui a un rôle dans la virulence de la souche bactérienne. En effet, il peut se lier au fer de l'hôte et le transporter à l'intérieur de la bactérie [92]. La pyocyanine est un pigment bleu caractéristique de *P. aeruginosa*. Ce pigment endommage les cellules endothéliales, retrouvées sur la face interne des vaisseaux sanguins, et joue un rôle important dans la nécrose et les dommages causés au tissu pulmonaire [93].

Les pili de type IV. Ils interviennent principalement pour la fixation de la bactérie sur les cellules ou sur une surface [2]. C'est la première étape de formation du biofilm. De plus, grâce à ces pili la bactérie peut se déplacer par *twitching motility*. C'est un système de déplacement

sur une surface qui se fait grâce à l'extension et la rétractation des pili [94].

Le flagelle. Le flagelle est ce qui permet à la bactérie de nager, ce que l'on appelle aussi *swimming motility*. Il a aussi une fonction importante dans la création de biofilm, puisque des mutations dans les gènes du flagelle bloquent la formation de biofilm [94].

Alginate. La production de ce composé est active en particulier chez les souches dites muqueuses. En règle générale, cette production est réversible mais peut devenir irréversible à la suite de mutations. Cette substance obstrue les voies respiratoires tout en protégeant la bactérie du système immunitaire et de l'action des antibiotiques [2].

LPS, Lipopolysaccharide. Le LPS est un des composants majeurs de la membrane externe et joue un rôle dans la pathogénicité [95] avec l'activité endotoxique du lipide A d'une part et la capacité de résister aux défenses naturelles de l'hôte d'autre part. Il sert de barrière contre l'extérieur en bloquant l'entrée de composés néfastes pour la cellule bactérienne dont certains antibiotiques.

Protéases. Ces enzymes dégradent les protéines tissulaires de l'hôte [2]. Plusieurs types existent parmi lesquels on retrouve la protéase alcaline, l'élastase (qui dégrade essentiellement le tissu pulmonaire) et la collagénase (dégradation du collagène) [2]. En plus de pouvoir plus facilement infecter les cellules, la dégradation des tissus met à disposition de la bactérie un grand nombre de nutriments qui favorise sa croissance.

Toxines. L'exotoxine A, produite par plus de 90% des souches de *P. aeruginosa* est létale car elle inhibe la synthèse protéique et donc le bon fonctionnement des cellules de l'hôte [2]. Les exoenzymes S, T, U et Y, sont sécrétées par le système de sécrétion de type III. Ce système ressemble à une seringue qui injecte les composés toxiques directement dans les cellules cibles. L'exoenzyme S permet de bloquer la réaction du système immunitaire en tuant les macrophages, cellules capables de détruire les bactéries. L'exoenzyme T a un fonctionnement relativement semblable à celui de l'exoenzyme S. Les exoenzymes S et U sont mutuellement exclusives, en effet si l'une est présente dans le génome, l'autre ne peut pas l'être. L'exoenzyme U détruit les cellules et est souvent associée à la destruction du tissu pulmonaire. La fonction exacte de l'exoenzyme Y est encore inconnue [96].

Hémolysines. Les hémolysines sont des protéines capables de détruire les globules rouges. Deux types sont produits par *P. aeruginosa*, d'une part un glycolipide et d'autre part une phospholipase C [2]. Le glycolipide est cytotoxique puisqu'il bloque les cils des cellules respiratoires. La phospholipase C produit une réaction inflammatoire.

Biofilm. Quatre étapes majeures sont nécessaires à la formation du biofilm [2]. La première est l'attachement de manière réversible des bactéries sur une surface. Ensuite, la formation de micro-colonies va rendre cet attachement irréversible. Le biofilm va se structurer par le déplacement des bactéries et la création d'une matrice externe.

Pour finir, le biofilm peut se détacher en partie et aller contaminer d'autres surfaces.

La grande résistance aux traitements physiques ou biochimiques (incluant les antibiotiques)

fait du biofilm un mode de croissance idéal en condition de stress externe [94]. En effet, la grande densité de cellules bactériennes rend difficile l'accès aux cellules du centre du biofilm, les protégeant ainsi des composés toxiques. De plus, la population très hétérogène que l'on retrouve à l'intérieur d'un biofilm permet d'avoir une grande variabilité de cellules. Certaines peuvent croître très vite tandis que la croissance d'autres sera retardée. Certaines bactéries seront résistantes aux antibiotiques par surproduction d'enzymes inactivatrices ou de pompes d'efflux. D'autres encore n'exprimeront rien de tout cela. Les processus physiologiques au sein du biofilm sont modifiés, certains processus métaboliques sont bloqués et des mécanismes de protection sont activés tel que le *quorum sensing* qui permet une communication entre les bactéries au sein du biofilm.

1.4.6 Évolution intrahôte

Au cours de l'évolution d'une bactérie, un grand nombre de modifications est observable, que ce soit par acquisition de gènes de résistance ou de virulence, intégration d'éléments mobiles ou encore mutations. Il existe un cas particulier d'évolution de la bactérie qui lui permet de s'adapter à un hôte unique. Cette évolution est observable dans le cas de patients infectés de manière chronique comme c'est le cas, par exemple des patients atteints de mucoviscidose. Il est très difficile voire impossible pour ces patients de les débarrasser des souches de *P. aeruginosa* qui les colonisent. En effet, la bactérie persiste dans leurs poumons grâce à la formation de biofilm souvent associé à un échec des traitements antibiotiques [97].

Ces souches sont très particulières puisqu'elles ne se propagent presque plus [98], en effet la transmission de souches adaptées est rare. La comparaison des souches prélevées dans les premiers stades de la maladie et celles plus tardives permet d'observer l'évolution dans un modèle chronique.

Le passage de l'état ubiquitaire de *P. aeruginosa* à un état spécialisé avec l'adaptation à une niche écologique particulière entraîne des modifications génomiques importantes. En effet, la bactérie va perdre des fonctions qui sont utiles à l'adaptation dans d'autres niches mais qui ne le sont pas dans celle-ci. On observe une modification du métabolisme des acides aminés dans les souches adaptées à la chronicité. L'auxotrophie de nombreux isolats issus de patients atteints de mucoviscidose a été démontrée [99]. Il existe également des modifications dans l'utilisation des acides aminés. Behrends et *al.* ont montré qu'il y a une relation significative entre la réponse adaptative pour un acide aminé et le coût métabolique de la fabrication de cet acide aminé, c'est à dire que les souches ont évolué pour faire une meilleure utilisation des acides aminés métaboliquement coûteux [100].

Il existe un grand nombre de modifications phénotypiques communes lors de l'adaptation de *P. aeruginosa* à l'infection chronique, ce sont les marqueurs de chronicité. Les fonctions touchées sont la production de facteurs de virulence qui est fortement diminuée, la modification du mode de vie en biofilm et l'augmentation de la résistance aux antibiotiques [101]. Le LPS joue également un rôle pendant la mise en place de l'infection chronique, puisqu'il est sensé permettre l'adhésion, la colonisation, l'adaptation au milieu ainsi que l'échappement du système immunitaire [95].

On retrouve également chez les souches isolées d'infections chroniques, une forte proportion de phénotype hypermutateur [31] souvent lié à l'inactivation de *mutS* ou *mutL*. Les isolats hypermutateurs ont plus d'évènements d'insertion-délétion, en particulier causés par insertions ou délétions de 1 à 4 nucléotides, alors que les non-hypermutateurs ont une plus grande proportion de substitution de nucléotides [102]. De plus, bien que la pompe d'efflux MexXY soit maintenue dans les isolats issus d'infections chroniques, une perte ou une inactivation de MexAB-OprM est observée dans la majorité des cas [103].

Les mutations dans les gènes de régulation du *quorum sensing*, en particulier *lasR*, sont communes [104]. La perte de mobilité avec modification dans les gènes de pili ainsi que la perte de sécrétion des toxines par mutation du système de sécrétion est aussi observée chez les souches de patients infectés de manière chronique [101]. AmpR qui a un rôle dans la résistance aux β -lactamines est fréquemment muté dans le cas d'infections chroniques [101]. D'autres gènes impliqués dans la résistance aux antibiotiques sont aussi mutés. Il s'agit en partie des gènes impliqués dans les pompes d'efflux, *mexA*, *mexB*, *mexR* et *mexZ*, ainsi que des gènes responsables de résistances lorsqu'ils sont mutés, *oprD*, *gyrA*, *gyrB*, *ampC* et *ampD* [103, 64].

L'adaptation de *P. aeruginosa* lors des premiers stades de l'infection chronique est caractérisée par la perte de large portions chromosomiques [37].

1.5 Outils d'analyse bioinformatique

La bioinformatique tient une place importante dans l'analyse des données génomiques. Aussi, la connaissance des outils informatiques utilisés dans le cas de techniques WGS permet de mieux comprendre les données obtenues et les traitements applicables. J'aborderai dans un premier temps le séquençage à haut débit et le type de données générées, puis les différentes techniques d'analyse que sont l'assemblage et l'alignement et, enfin le principe et les enjeux de la phylogénie.

1.5.1 Notion de séquençage haut débit

Les analyses WGS produisent la séquence complète du génome observé. L'obtention de ces informations génomiques passe par l'utilisation du séquençage à haut débit. Pour mieux comprendre cette technique, il faut d'abord repartir aux techniques historiques du séquençage.

Techniques de séquençage

Séquençage Sanger. Le premier séquençage disponible est le séquençage Sanger (du nom de son inventeur Frederick Sanger), publié en 1977 [105]. L'utilisation d'une ADN polymérase permet de synthétiser des nouveaux brins d'ADN complémentaires des brins existants.

L'hybridation d'une amorce sur la séquence à lire permet à l'ADN polymérase, en présence des 4 nucléotides, l'élongation de la séquence en cours de synthèse, en ajoutant le nucléotide complémentaire au brin d'ADN matrice. Des "terminateurs de chaîne" sont incorporés dans le milieu de synthèse, permettant de bloquer l'élongation. L'obtention de séquences de différentes tailles, chacune avec à son extrémité un terminateur de chaîne permet d'obtenir

l'ensemble de la séquence observée. Les différentes séquences sont séparées en électrophorèse capillaire de manière très précise puisque la séparation se fait à la base près. Un laser permet de révéler les agents fluorescents fixés sur les terminateurs de chaîne. Les signaux lumineux ainsi produits sont lus par ordinateur. Les profils obtenus sont visibles sur la figure 1.20.

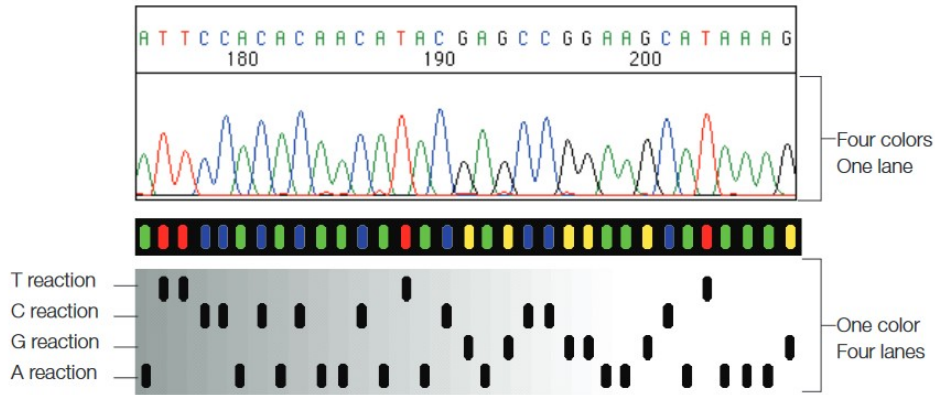


FIGURE 1.20 – Visualisation de résultats de séquençage Sanger [106]. Le schéma du haut représente le signal lumineux capté par le séquenceur. Chaque pic correspond à un nucléotide. La bande du milieu représente le résultat de la révélation dans un capillaire. La partie basse du schéma correspond à l'électrophorèse classique avec une piste par nucléotide.

La taille du fragment qui peut être séquençé est comprise entre 400 et 900 nucléotides. Les nucléotides lus sont peu fiables en début et en fin de séquence. Les limites principales de la technique sont son coût élevé par bp et son temps d'analyse long. Pour pallier à cela, les techniques dites à haut-débit ou NGS ont vu le jour (tableau 1.4). Le principe est la parallélisation de la lecture qui permet de lire un grand nombre de fragments d'ADN en un temps court.

Technique (Fournisseur)	Longueur de reads (bp)	Reads	Précision	Temps (1 run)	Coût (\$/Mbp)
Sanger	400-900	1	99,99 %	20 min à 3h	2400
454 (Roche)	500-700	1 million	99,90 %	24h	10
Ion Torrent (Life Technologies)	< 400	80 millions	98 %	2h	1
MiSeq (Illumina)	50-300	6 milliards	98 %	1-10 jours	0,05-0,15
SMRT (PacBio)	10.000-15.000	0,5-1 milliard	87 %	30 min à 4h	0,13-0,60

TABLEAU 1.4 – Caractéristiques des principales techniques de NGS [107, 108]

Le tableau 1.4 résume les caractéristiques des différentes techniques permettant d'avoir une vision globale de chacune avec leurs avantages et inconvénients.

Il existe différentes techniques de séquençage haut-débit disponibles à l'heure actuelle et toutes ne seront pas décrites ici.

La plupart de ces techniques haut-débit reposent sur l'utilisation de la PCR décrite ci-dessous.

PCR. Le but de cette technique est de dupliquer un grand nombre de fois un même fragment d'ADN. Après séparation des deux brins d'ADN, une amorce s'hybride sur le début de la séquence à amplifier. Une polymérase va ensuite copier le brin d'ADN qui sert de modèle. Après, un nouveau cycle commence : séparation des brins, fixation de l'amorce, élongation (figure 1.21). Après plusieurs dizaines de cycles, la séquence d'intérêt est présente des millions de fois dans le milieu.

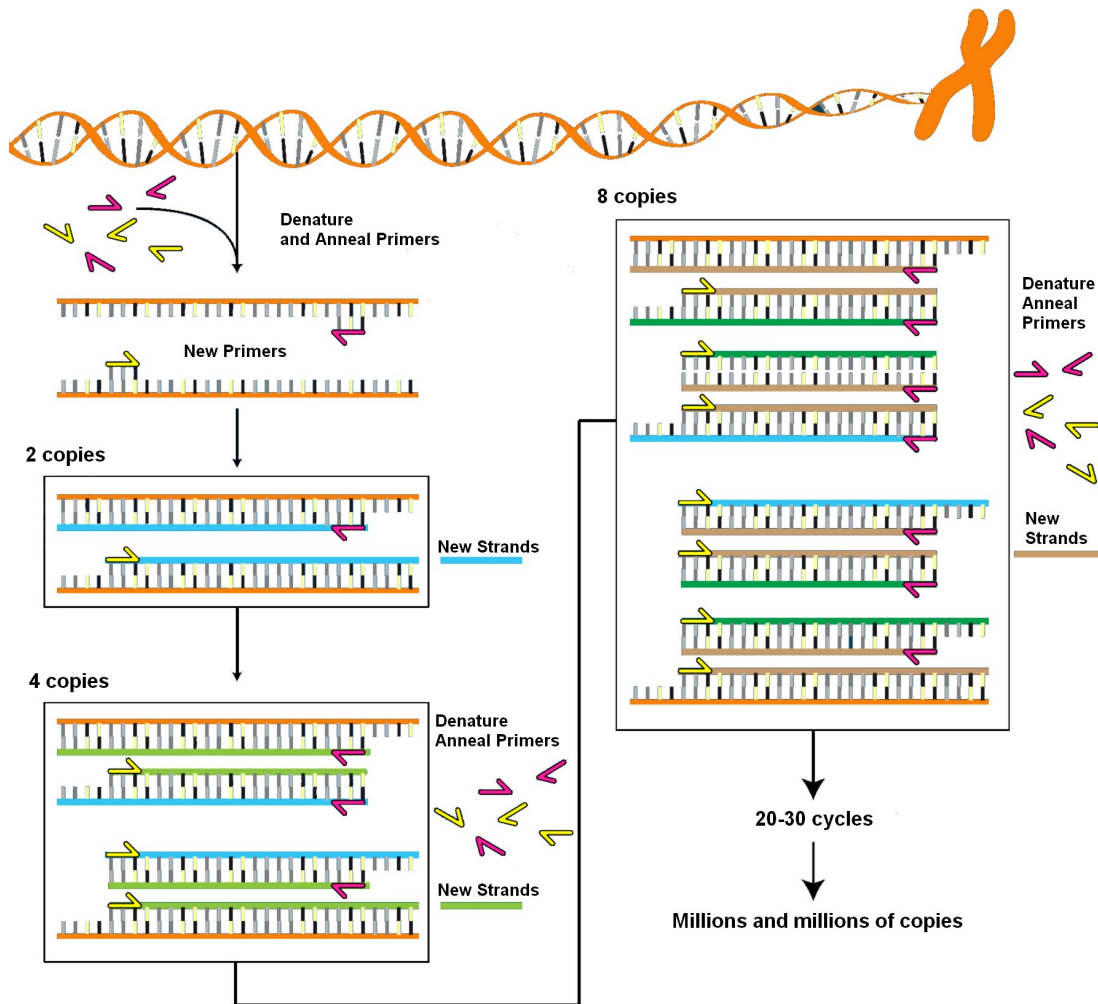


FIGURE 1.21 – Schéma des différentes étapes de la technique de PCR [109]

Le séquençage haut débit se caractérise par la création d'une librairie correspondant à une

collection de fragments d'ADN auxquels sont ajoutés des adaptateurs, permettant de fixer les fragments sur une surface.

Parmi les techniques de séquençage haut-débit, les quatre plus connues sont le 454 de Roche, l'Ion Torrent de Life Technologies, le HiSeq et MiSeq d'Illumina ou encore la technique SMRT de Pacific Biosciences (PacBio). Les caractéristiques de chacune sont répertoriées dans le tableau 1.4.

Roche 454. Après création de la librairie, les fragments d'ADN sont fixés sur une bille (un seul fragment par bille). Une PCR amplifie chacun des fragments. Celle-ci se fait en émulsion dans une goutte d'huile. Cette amplification permet l'obtention d'un signal plus élevé lors de la lecture des bases. Ensuite, chaque bille est déposée dans un puits (figure 1.22).

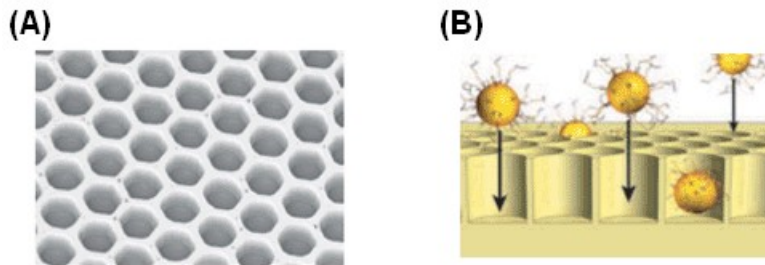


FIGURE 1.22 – Séquençage haut débit, création de librairie : A. Puits de dépôt des billes pour PCR en émulsion [110], B. Billes déposées dans les puits sur lesquelles sont visibles les fragments d'ADN après amplification [110]

Il s'agit d'une technique de synthèse séquentielle. Après la préparation de la librairie, les 4 bases sont ajoutées de manières séquentielles dans le milieu. La première est ajoutée, le milieu est lavé puis une autre est ajoutée et ainsi de suite. Lorsqu'un ou plusieurs nucléotides sont incorporés, un signal lumineux (émis par une luciférase) est enregistré. L'avantage de cette technique tient à des *reads* assez longs et à sa rapidité, 24h pour une expérience. En revanche, la difficulté à discriminer les signaux lumineux (figure 1.23) quand il y a plus d'un nucléotide incorporé en une seule étape tend à produire des erreurs d'homopolymère, par exemple 'TAAA' facilement confondu avec 'TAAAA'. Ce séquenceur a été lancé en 2005, et les kits de réactifs pour cette machine ne sont plus disponibles depuis l'été 2016.

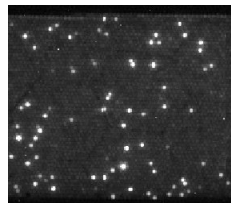


FIGURE 1.23 – Signal lumineux produit par l'appareil Roche 454

Ion Torrent. La technique d'Ion Torrent ressemble au 454, après création de la librairie, les

fragments sont liés à des billes puis amplifiés par PCR en émulsion. De plus, les bases sont également ajoutées de manière séquentielle dans le milieu mais cette fois, l'incorporation de nucléotide(s) est mesurée par une variation de pH. Quand un nucléotide est incorporé, il libère un proton dans le milieu qui va faire varier le pH. C'est cette variation qui est mesurée. Tout comme pour le 454, la méthode est peu discriminante lorsqu'il s'agit de connaître exactement le nombre de fois qu'un même nucléotide est incorporé, ce qui génère fréquemment des erreurs d'homopolymère. L'avantage de cette technique est tout de même le coût moindre de l'équipement par rapport aux autres (Illumina et PacBio) et sa rapidité (2h pour un *run*), voir tableau 1.4.

Illumina. Après création de la librairie, les fragments d'ADN sont fixés sur une plaque avant d'être amplifiés par une PCR en colonie qui va générer des clusters de fragments d'ADN identiques (figure 1.24).

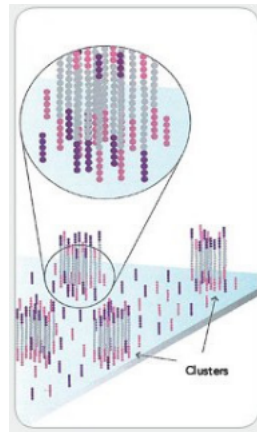


FIGURE 1.24 – Fixation de l'ADN sur plaque et création de cluster par PCR [111]

Cette dernière technique repose sur un principe différent des deux précédentes. En effet, il ne s'agit pas ici d'une synthèse séquentielle. Les nucléotides ajoutés dans le milieu sont des terminateurs de chaînes. Les nucléotides sont incorporés un à la fois ce qui élimine le problème de détection de nucléotides répétés. Lorsque la séquence est lue par fluorescence (figure 1.25), le terminateur est enlevé du nucléotide ce qui permet la continuation de l'élongation. Cette technique est la moins chère, avec un coût situé entre 0,05 et 0,15\$ pour 1 million de bases, et avec le rendement le plus élevé (jusqu'à 6 milliards de *reads* par *run*), voir tableau 1.4.

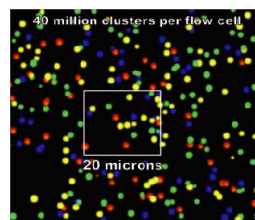


FIGURE 1.25 – Signal fluorescent produit par l'appareil Illumina [112]

PacBio. Une autre technique a vu le jour plus récemment développée par *Pacific Biosciences* [113]. Le séquençage se fait cette fois en temps réel, sans aucune amplification préalable (séquençage SMRT, *Single Molecule Real Time*). Après extraction, l'ADN doit être intact c'est à dire, peu fragmenté. Dans le cas contraire, les résultats seront de mauvaise qualité avec des *reads* courts.

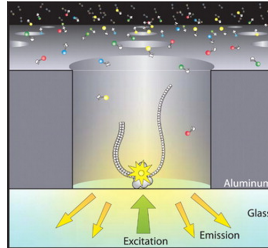


FIGURE 1.26 – Schéma de la technique de séquençage PacBio [114]

Contrairement aux autres techniques où la molécule d'ADN est fixée sur un support (lors de la création de la librairie), ici c'est l'enzyme qui est fixe (voir figure 1.26). En présence de la molécule d'ADN et de nucléotides fluorescents, l'enzyme va synthétiser le nouveau brin d'ADN. Lors de l'incorporation d'un nucléotide, l'enzyme libère le marqueur fluorescent détecté par l'appareil. Cela permet de générer des *reads* qui sont beaucoup plus longs ce qui facilitera les étapes de traitement par la suite. En revanche, la précision est légèrement moindre que celle des autres appareils. Il n'existe pour le moment que peu d'appareils disponibles dans le monde du à son coût relativement élevé.

Multiplexage. Lors de l'utilisation de techniques de séquençage haut-débit, un grand nombre de séquences sont lues simultanément. La couverture produite par ces techniques NGS correspond au nombre de fois qu'une base sera théoriquement lue. Elle est calculée en fonction du nombre total de bases lues dans un *run* (un séquençage) et de la taille du génome à séquencer. Le nombre de bases qui peuvent être lues en un seul *run* dépasse souvent le nombre de bases voulues (environ 80X pour 80 lectures de chaque bases). Il est donc avantageux de séquencer plusieurs échantillons/génomes en même temps. Pour cela, on va utiliser la technique de multiplexage qui consiste à ajouter un *tag* (une séquence nucléotidique spécifique) à chacun des fragments d'ADN au moment de la création de la librairie. Chaque *tag* est spécifique d'un échantillon. Lors de l'analyse, les *reads* seront triés grâce au *tag* qu'ils portent.

Données

Les données brutes récupérées après séquençage sont généralement des fichiers au format fastq qui contiennent les *reads*, c'est à dire les lectures de chacun des fragments d'ADN. Ce sont des fichiers très volumineux puisqu'il y a plusieurs millions de *reads* lus au cours d'un *run* de séquençage. Chacun des *reads* présent dans le fichier correspond à 4 lignes. La première, commençant obligatoirement par un '@' correspond au nom de la séquence. C'est souvent un nom générique donné par l'appareil. La deuxième ligne correspond à la séquence nucléotidique du *read*. La troisième ligne ne contient que le symbole '+' qui sert de séparateur. En effet la dernière ligne correspondant au score de qualité peut contenir les lettres présentes dans la séquence. Le '+' sert donc bien ici de séparateur. Le score de qualité ou *phred score*

donne la précision avec laquelle la base a été lue. Il est codé par 40 caractères variant en fonction des séquenceurs (par exemple, pour le score de qualité Illumina 1.8+ Phred+33 / vaut 15, A vaut 33 et E vaut 37). Plus le score est élevé plus la certitude d’avoir lue la bonne base est forte. Un exemple est présenté en figure 1.27.

```

read 1
@NS500311:174:HFJJWBGXY:1:11101:20150:1048 1:N:0:TAAGGCGA+NTAGAGAG
GCGTCAGGCCATGACGCATGGCGCCGGCTTCGGCGAAATGGCGGAACAGCAGCGGGATGTCTTCGCCGCGCTCGCGCAGC
GGTGGAAATGCGCAGGCTGGCGACGTTGAGGCGGTAATAGAGGTCGGCGCGGAAGCGGCCCTGGTCGGCGG
+
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
read 2
@NS500311:174:HFJJWBGXY:1:11101:8051:1050 1:N:0:TAAGGCGA+NTAGAGAG
GGCGAAGCGCACGCAGTAGATCGCGCGCAGGTTGGCGGAGAGTGCTGCCAGTCTCCCCGCCGTTCTCGCTGAGCCACA
GGCGCCGGTGGTCGAGCCATGGCCAGGCGCTCGCCGCTGGGGTCGATGTCCAGGCCGTGGCGGTAGA
+
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA</
read 3
@NS500311:174:HFJJWBGXY:1:11101:14107:1052 1:N:0:TAAGGCGA+NTAGAGAG
GCGCAGCACCGACGCCAGCGTCTATTGGTGGTGAAGGCAGCGGCGTGGCGCAGATCGCGCGCCGCCCTTCGCTTCG
AAGCCAAGGAACTGTAAGTCGTGCCCTCCTGGGCCGAGATGCGCTGGAAGCCGGCGCCACCGACTACGT
+
AAAAAEEEA/EE/EAE/E/EE/A/EEAAEAEEA<EAAE<AA/EE6//EE/E/EEEE/EE/<EAE//E/EA/E/EE/
E///EE/<E/A6E6/6/EA<6/EE6EE<///<EA/EA//6<//EA//EE<AA<A//EAAA/EE/////
    
```

FIGURE 1.27 – Extrait d’un fichier au format fastq. La première ligne correspond au nom de la séquence (commençant par @NS500311), vient ensuite la séquence du *read*, un ‘+’ qui sert de séparateur et le score de qualité (commençant par AAAAAEEEEE, codé en Illumina 1.8+ Phred+33). La ligne suivante indique le début d’un nouveau *read*. Ici, les séquences et scores de qualités de trois *reads* sont présentés.

Ce format s’oppose au format fasta qui contient des séquences nucléotidiques ou protéiques et non plus des *reads*. Le format fasta ne comporte que deux lignes, la première commence par un ‘>’ qui est suivi par le nom de la séquence, qui peut être un nom de chromosome par exemple. La deuxième ligne correspond à la séquence elle-même (nucléotidique ou protéique), exemple en figure 1.28.

```

>NC_002516.2 Pseudomonas aeruginosa PAO1 chromosome, complete genome
TTTAAAGAGACCGGCGATTCTAGTGAAATCGAACGGCAGGTCAATTTCCAACCAGCGATGACGTAATAG
ATAGATACAAGGAAGTCATTTTTCTTTAAAGGATAGAAAACGGTTAATGCTCTTGGGACGGCCTTTTTCT
GTGCATAACTCGATGAAGCCCAGCAATTGCGTGTTTCTCCGGCAGGCAAAAGGTTGTGAGAACCGGTGT
CGAGGCTGTTTCTTCTGAGCGAAGCCTGGGGATGAACGAGATGTTTATCCACAGCGTTTTTCCACA
CGGCTGTGCGCAGGGATGTACCCCTTCAAAGCAAGGGTTATCCACAAAGTCCAGGACGACCGTCCGTCCG
GCCTGCCTGTTTTTAAAGGCTTGATTTGCTTGGGGCCTCAGCGCATGGCATGTGGATAAGTCCGGC
CCGTCCGGCTACAATAGGCGCTTATTTGTTGTGCCGCTTTCCAATCTTTGGGGGATATCCGTGTCCGT
GGAATTTGGCAGCAGTGGTGGATCTTCTCCGGATGAGTGCCGTCCCAACAATTCAACACCTGGATC
CGTCCCTTGCAAGTGAAGCCGAAGGCGACGAATTGCGTGTGTATGCACCCAACCGTTTTCGTCTCGATT
GGGTGAACGAGAAATACCTCGGTCGGCTTCTGGAAGTCTCGGTGAACGCGGCGAGGGTCAGTTGCCCGC
GCTTTCTTAAATAGGCAGCAAGCGTAGCCGTACGCCGCGCGCCATCGTCCCATCGCAGACCCAC
GTGGCTCCCCCGCTCCGTTGCTCCGCCCGCGGCCAGTGACGCCGGTATCGGCCGCGCCCGTGGTAG
TGCCACGTGAAGAGTGCAGCCAGTGACGACGGCTCCAGCGTGTGAGCGATCCCTACGAGCCGGAAGA
    
```

FIGURE 1.28 – Extrait d’un fichier au format fasta. La première ligne correspond au nom de la séquence (commençant par >), puis la séquence elle-même sur une ou plusieurs lignes. La séquence présentée correspond au début du génome de *P. aeruginosa* PAO1.

Elle ne contient en général que les lettres A, T, G et C mais peut également contenir la lettre

N qui correspond à un nucléotide non défini, soit A, T, G ou C. Il existe également d'autres lettres symbolisant les combinaisons possibles de chaque bases (tableau 1.5).

Code	Nucléotides
R	A ou G
Y	C ou T
S	G ou C
W	A ou T
K	G ou T
M	A ou C
B	C ou G ou T
D	A ou G ou T
H	A ou C ou T
V	A ou C ou G
N	N'importe quelle base

TABLEAU 1.5 – Code IUPAC (*International Union of Pure and Applied Chemistry*)

1.5.2 Techniques d'assemblage et d'alignement

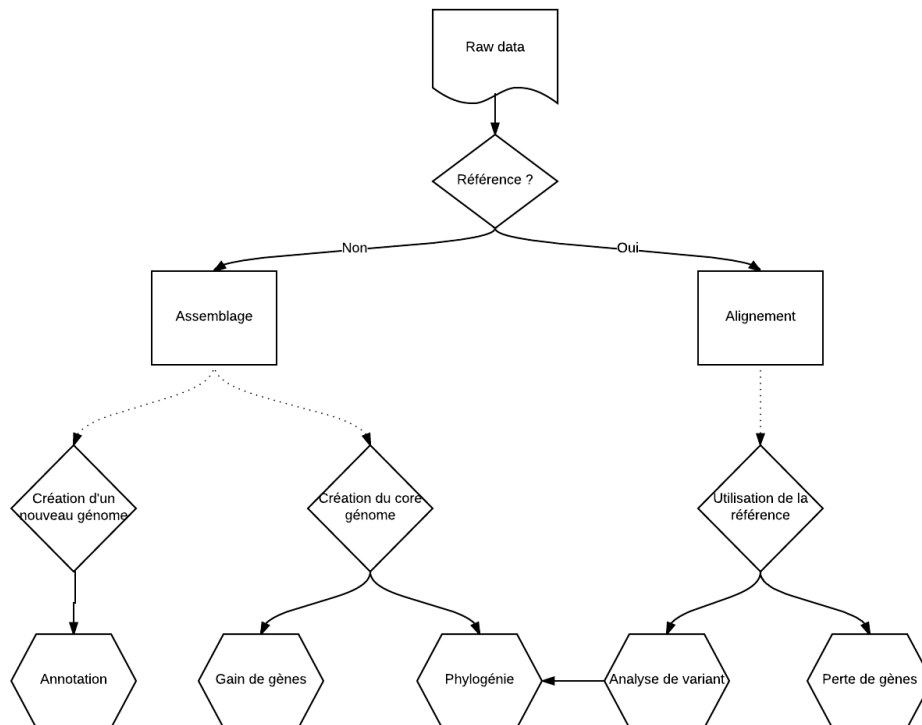


FIGURE 1.29 – Schéma des stratégies utilisées en bioinformatique pour l'analyse des données

A partir des données brutes (*raw data*) générées par le séquençage, il existe plusieurs possibilités d'analyse de ces données suivant ce que l'on recherche (figure 1.29). La composition exhaustive du génome ainsi que la place des gènes les uns par rapport aux autres nécessite la reconstruction du génome par une technique nommée assemblage. En revanche, la comparaison du ou des génome(s) séquençé(s) ainsi que la recherche de SNPs ou d'indels par rapport à une référence nécessite une autre technique dite d'alignement.

Assemblage

La technique d'assemblage est la reconstruction du génome de manière la plus complète possible en rassemblant les *reads* qui se chevauchent pour obtenir des séquences plus longues que l'on appelle des contigs (figure 1.30).

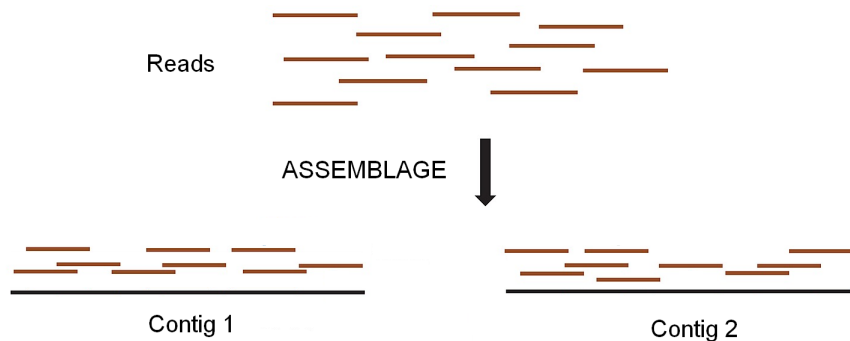


FIGURE 1.30 – Assemblage de *reads* en contigs

En utilisant une technique classique d'assemblage à partir de séquençage Ion Torrent ou Illumina, il est pratiquement impossible d'obtenir le génome complet. Bien que les logiciels d'assemblage soient performants, la complexité du génome à reconstruire (avec plus ou moins de séquences répétées et des séquences peu ou mal couvertes) ainsi que la qualité du séquençage influent sur le nombre et la qualité des contigs qui seront produits. De plus, le génome séquençé doit avoir une couverture globale suffisante pour permettre l'assemblage, une moyenne de 60X ou 80X étant généralement utilisée.

Il existe un grand nombre de logiciels disponibles permettant d'assembler des génomes. Trois d'entre eux seront présentés ici : Ray [115], MIRA [116] et SPAdes [117]. Chaque logiciel est optimisé pour certains types de séquenceur. En fonction du séquenceur qui a produit les données, il existe un certain biais, qui correspond à des erreurs générées lors du séquençage. Les logiciels permettent de nettoyer les données avant assemblage (conservation des *reads* de bonne qualité, erreurs systématiques de séquençage enlevées, *etc.*). Ray est spécialisé pour l'assemblage de données Illumina. MIRA, en fonction des arguments spécifiés peut assembler des données de 454, d'Ion Torrent ou d'Illumina. SPAdes gère les données 454, Ion Torrent, Illumina et PacBio.

Étant donné que chaque logiciel possède ses propres spécificités, type de données pris en entrée mais également algorithme utilisé, il est important de comparer leurs performances

afin d'obtenir le meilleur assemblage possible pour le génome et les données brutes que l'on possède.

Alignement

L'autre technique utilisée, l'alignement, est surtout utile pour comparer des séquences à un génome de référence. Il s'agit d'aligner les *reads* sur la référence et d'observer la répartition de la couverture comme visible sur la figure 1.31. Le logiciel BWA [118] est un des plus utilisé pour aligner les *reads*.

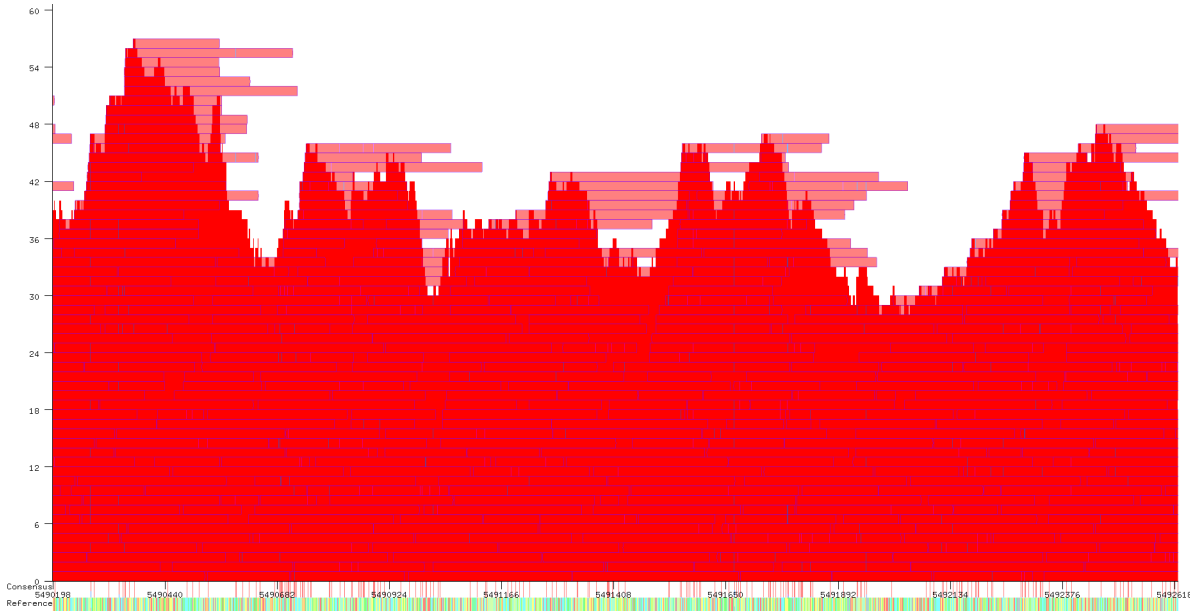


FIGURE 1.31 – Alignement des *reads* sur une référence. Visualisation à l'aide du logiciel samscope [119]. En bas se trouve la séquence de référence et les *reads* (rectangles rouges) sont positionnés sur celle-ci. Des variations de couverture (nombre de *reads* sur l'axe des Y) induisent des pics pour les zones les plus lues.

L'alignement donne une indication sur les zones manquantes du génome (figure 1.32 visualisée avec le logiciel Samscope [119]). En effet, une zone non couverte par les *reads* indique une zone qui n'existe pas dans le génome séquencé. Par contre, il est impossible de visualiser des zones présentes chez le génome séquencé et absente de la référence. Lorsque l'alignement est fait, les *reads* qui ne sont pas *mappés* ou positionnés sur la référence n'apparaissent pas lorsque l'on visualise le résultat, bien qu'ils puissent être récupérés depuis le fichier d'alignement. Les fichiers d'alignement sont au format SAM ou BAM (format compressé), et contiennent l'ensemble des *reads*, qu'ils soient ou non alignés sur la référence.

L'alignement indique la présence de SNPs ou indels qui peuvent avoir un intérêt majeur dans l'étude des fonctions biologiques modifiées (figure 1.33). Avec la visualisation, il est possible de voir les différents SNPs présents sur le génome. En effet, il y a des cas où les *reads* alignés indiquent un certain nucléotide alors que la référence en indique un autre. C'est le marqueur d'une variation ponctuelle de séquence. Ces variations sont en général utilisées pour la création d'arbres phylogénétiques qui permettent d'observer la parenté entre différents génomes.

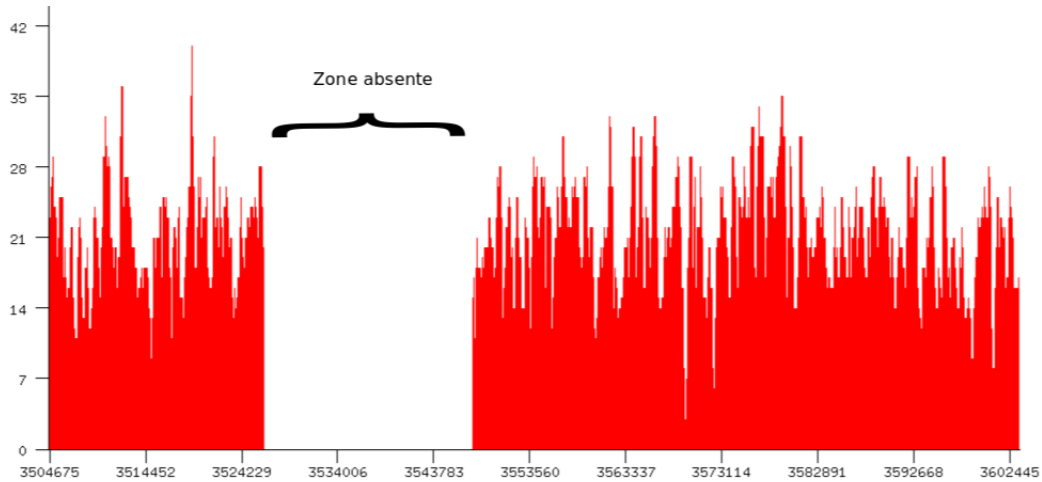


FIGURE 1.32 – Zone absente du génome mais présente dans la référence. Visualisation à l’aide du logiciel samscope [119]. L’axe des Y donne la couverture du génome.



FIGURE 1.33 – Visualisation d’un SNP dans la séquence observée avec le logiciel samscope [119]. La présence du SNP est indiquée par la présence du nucléotide A (en vert) en remplacement du nucléotide G présent sur la référence. La validité du SNP est certifiée par la présence de la mutation sur la majorité des *reads*.

Choix de la technique

En fonction des informations que l’on recherche et de la qualité du séquençage, on utilisera préférentiellement l’une ou l’autre de ces techniques. Les avantages et inconvénients de chacune sont présentés dans le tableau 1.6.

Alignement	Assemblage
Visualisation de données manquantes Analyse de SNPs	Visualisation de données spécifiques absentes de la référence Notions de synténie (enchaînement des gènes sur un chromosome)
Fait avec une couverture de séquençage faible (moins de 40X)	Se fait <i>de novo</i> , sans référence préalable

TABLEAU 1.6 – Comparaison des techniques d’assemblage et d’alignement

1.5.3 Analyse de données

A. Analyse de variants

Lors de l’analyse de génomes, la recherche de variants est une étape importante. Le logiciel *freebayes* [120] permet, à partir du fichier d’alignement d’obtenir tous les variants existants entre la souche alignée et la référence. Une probabilité est disponible pour chaque variant (SNP ou indel) ainsi que le nombre de *reads* qui couvrent la position. Ce logiciel permet également de comparer un grand nombre de génomes. En fournissant plusieurs alignements au programme, il est possible de voir les allèles de chacune des souches pour chaque position mutée.

B. Comptage de *reads*, recherche de gènes

La mesure de la couverture peut être un moyen efficace de déterminer la présence de gènes dans un génome aligné. Si elle est suffisante sur toute la longueur du gène, alors on peut considérer que le gène est présent. A noter que puisqu’il s’agit d’alignement, un gène absent de la référence ne peut être recherché.

Le logiciel *featureCounts* [121] sert à compter le nombre de *reads* qui couvrent une séquence donnée. Pour cela, les paramètres utilisés sont le fichier contenant l’alignement des *reads* sur la référence et un fichier de type BED, qui correspond à un tableau avec les positions des différents gènes à rechercher. *featureCounts* produit un fichier précisant le nombre de *reads* couvrant chaque gène.

C. Core genome et phylogénie

L’ensemble des séquences présentes dans tous les génomes que l’on souhaite étudier correspond au *core genome*. Celui-ci peut être utilisé pour la construction d’arbres phylogénétiques servant à estimer la distance génétique entre les souches étudiées représenté par le nombre et la longueur des branches. En effet, leur longueur est proportionnelle au nombre de SNPs différenciant deux souches. L’arbre final (aussi appelé arbre consensus) est généré à partir de ces données et des paramètres utilisés par le logiciel de phylogénie.

Création du core genome. Celui-ci peut être défini de plusieurs manières (cf. figure 1.29). Premièrement, par extraction de séquences communes à tous les génomes. Pour cela, un logiciel de comparaison (pour les génomes assemblés) ou d’alignement (pour les génomes non assem-

blés) est utilisé. Les séquences communes (gènes ou régions intergéniques) sont extraites pour chaque génome puis concaténées et alignées. Un exemple d'alignement multiple est visible sur la figure 1.34.

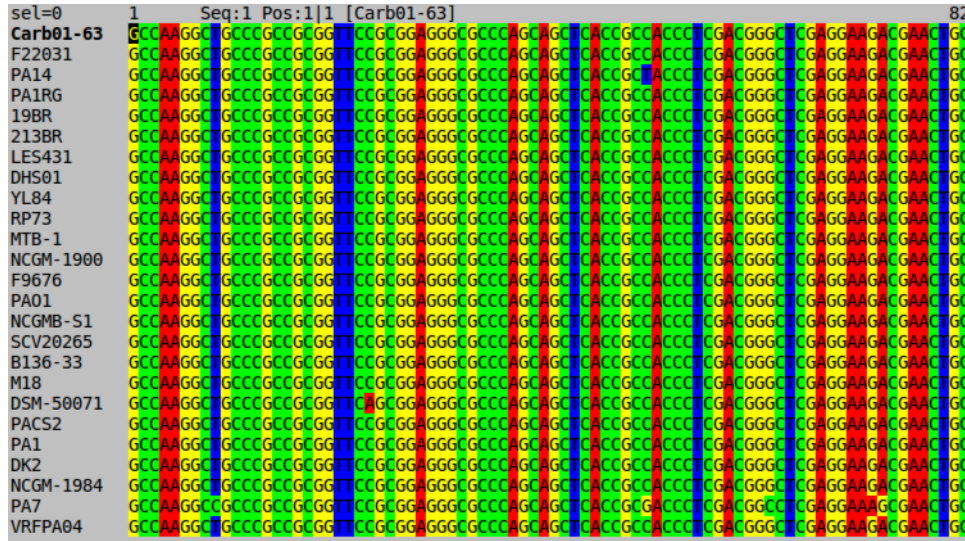


FIGURE 1.34 – Visualisation d'un alignement multiple avec le logiciel seaview [122]

La deuxième stratégie de définition du *core genome* repose sur l'extraction des gènes. Chacun des gènes annotés d'un des génomes du panel va être recherché dans tous les autres. L'existence des ces gènes (avec des pourcentages de couverture et d'identité suffisants par rapport à la référence) dans chacun des génomes indique leur appartenance au *core genome*. Ces gènes sont donc extraits, concaténés et alignés. L'alignement multiple généré (comme précédemment) est utilisé pour la création de l'arbre phylogénétique.

Quelle que soit la méthode utilisée, les *gaps* dans la séquence (souvent liés à des indels) peuvent poser des problèmes lors de la création de l'arbre phylogénétique puisqu'ils peuvent être assimilés à une séquence manquante par le logiciel de phylogénie.

L'arbre phylogénétique peut être calculé directement à partir des SNPs. Pour cela, seuls les SNPs présents dans chacun des génomes sont extraits et concaténés. La séquence finale qui est comparée est plus petite que celle obtenue par les deux méthodes précédentes.

Les avantages et inconvénients de ces trois méthodes sont listés dans le tableau 1.7.

Phylogénie. Il existe un certain nombre de logiciels de phylogénie à l'heure actuelle. La plupart d'entre eux sont accessibles librement et gratuitement. C'est le cas de PhyML [123] et RaxML [124] utilisant des méthodes de maximum de vraisemblance, mais aussi de Mr-Bayes [125] et BEAST [126] qui utilisent une méthode bayésienne pour le calcul des arbres. Ce sont deux méthodes probabilistes basées sur un concept de vraisemblance. La première méthode est la recherche de l'arbre ayant la meilleure probabilité de conduire aux données observées tandis que l'autre méthode consiste à fixer des probabilités *a priori* aux arbres, avant

	Séquences communes	Gènes	SNPs
Avantages	Facile à définir	Très informatif	Rapide et simple à obtenir
Inconvénients	Peu informatif	Nécessite la liste des gènes	Ne contient pas les séquences conservées (calcul du GC impossible)

TABLEAU 1.7 – Comparaison des stratégies de définition du *core genome*

de les inférer et de calculer les hypothèses *a posteriori*. La méthode du maximum de vraisemblance est plus rapide que la méthode bayésienne tandis que celle-ci possède une mesure de confiance intégrée et permet de prendre en compte un plus grand nombre de paramètres. Un exemple d'arbre généré par MrBayes est visible sur la figure 1.35.

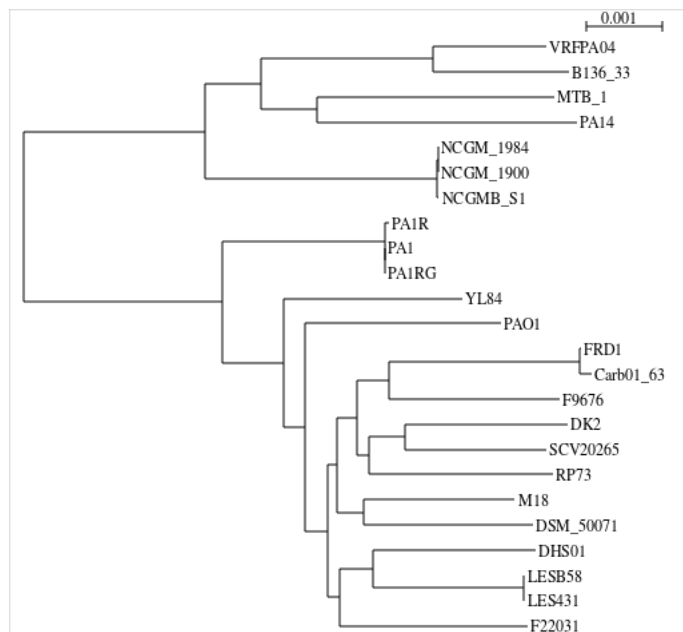


FIGURE 1.35 – Arbre phylogénétique de génomes de *P. aeruginosa* généré à partir de MrBayes [125]

Most Recent Common Ancestor (MRCA). Les logiciels de phylogénie peuvent également dater un potentiel ancêtre commun. Il peut être calculé de deux façons différentes. Soit en utilisant les dates des feuilles (par exemple date de prélèvement des souches) soit en fixant un point de calibration connu (existence d'un ancêtre commun connu à une date précise ou connaissance d'une date de divergence entre deux taxons). Ainsi, il est possible d'estimer la date d'existence d'un ancêtre d'une épidémie par exemple pour comprendre quand l'hôpital, la ville ou le pays a été contaminé. Un exemple est présenté sur la figure 1.36.

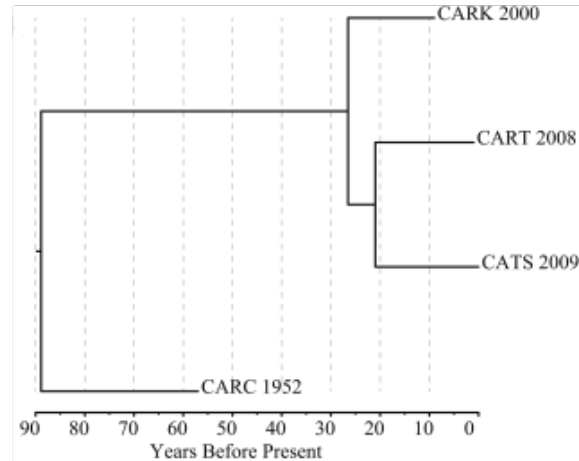


FIGURE 1.36 – Arbre phylogénétique de génomes de *Salmonella enterica* avec données temporelles [127] générées à partir de MrBayes [125]. L'ancêtre commun à ces quatre souches a existé il y a 90 ans environ.

1.6 Axe de recherche

L'axe de recherche de cette thèse est la comparaison génomique des bacilles à Gram négatifs pathogènes résistants aux antibiotiques.

1.6.1 Sujet principal

Il concerne une souche épidémique de *P. aeruginosa* ST395 ayant diffusé pendant 11 ans dans l'hôpital universitaire de Besançon. La première partie concerne la description de la souche index ainsi que la caractérisation du clone ST395 alors que la deuxième partie analyse l'évolution génotypique et phénotypique du clone au cours de l'épidémie.

Chacune des parties fait l'objet d'un article :

- 1 Petitjean, M., Martak, D., Silvant, A., Bertrand, X., Valot, B., & Hocquet, D. (2017). Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395. *Microbial Genomics*.
- 2 Petitjean, M., Juarez, P., Meunier, A., Bertrand, X., Valot, B., & Hocquet, D. (2017). Phylogenetic analysis of a *Pseudomonas aeruginosa* epidemic lineage reveals parallel evolutions associated with the outbreak decline. En écriture.

Présentation des résultats en congrès

1. *European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) 2016*, Amsterdam (Pays-Bas)
Présentation poster => The rise and fall of an epidemic clone of *Pseudomonas aeruginosa* : 17 years of evolution during its spread in a hospital
2. Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM) 2016, Lyon (France)

Présentation poster => Analyse de l'évolution d'une épidémie bactérienne par séquençage haut-débit

3. Réunion Interdisciplinaire de Chimiothérapie Anti Infectieuse (RICAI) 2016, Paris (France)

Présentation orale => Évolution d'un clone épidémique de *Pseudomonas aeruginosa* : 17 ans de diffusion hospitalière

4. *International Conference on Pseudomonas* 2017, Liverpool (Royaume-Uni)
Présentation poster et orale => The rise and fall of an epidemic clone of *Pseudomonas aeruginosa* : 11 years of evolution during its spread in a hospital

Vulgarisation scientifique

1. Participation aux Fêtes de la Science et Nuit des Chercheurs (2015-2017)
2. Participation au concours international MT180, Finaliste Bourgogne - Franche-Comté 2016, représentante de la région lors de la finale nationale
3. Création d'une BD vulgarisée "Marie et son Python" par le service Sciences, arts et culture de l'université de Franche-Comté en association avec la Mission Culture Scientifique de l'université de Bourgogne. BD disponible en annexe E.

1.6.2 Travaux annexes

***Enterobacter cloacae* CTX-M-15, en collaboration avec Marissa Haenni (Unité Antibiorésistance et Virulence Bactériennes, Lyon)**

Analyse d'*Enterobacter cloacae* producteur de β -lactamase à spectre étendu, comparaison de leur distribution clonale chez l'animal et chez l'homme.

Apport personnel : La génération d'un diagramme eBURST représentant les données de MLST d'*E. cloacae* permet de visualiser la distribution clonale de cette bactérie. L'article est disponible en Annexe B.

***P. aeruginosa* bruns, avec Didier Hocquet**

Caractérisation de *P. aeruginosa* mutants, producteurs de pyomélanine. Recherche de caractéristiques géniques et en particulier des gènes absents chez les mutants bruns. Des tests phénotypiques, tels que des tests de croissance sur différents milieux de culture, des tests de *fitness* ou encore des tests de production de facteurs de virulence, permettent d'évaluer les caractéristiques des mutants.

Apport personnel : La comparaison des génomes des paires (*P. aeruginosa* mutants bruns/non mutés) a permis d'identifier les délétions génomiques des mutants bruns. La position exacte de chacune des délétions a été validée par PCR. Celles-ci ont également permis de valider l'existence de régions répétées en bordure des délétions. L'article est disponible en Annexe C.

***Escherichia coli* ST131 H22/H30, en collaboration avec Marie-Hélène Nicolas-Chanoine (Service de Microbiologie, Hôpital Beaujon, AP-HP, Clichy)**

Ce travail de recherche concerne la caractérisation de souches d'*Escherichia coli* ST131 retrouvées dans le microbiote intestinal. Afin de comprendre la prédominance de certaines souches, un isolat représentatif de chacun des sous-clones H30 et H22 sont étudiés et leurs

caractéristiques phénotypiques recherchées. Pour cela, les génomes des isolats JJ1886 et S250 sont comparés à celui de l'isolat SE15 appartenant au sous-clone H41.

Apport personnel : Comparaisons génomiques des souches d'*E. coli* ST131 H22, H30 et H41 avec la recherche des gènes de virulence, de ceux impliqués dans le métabolisme des sucres ainsi que ceux liés à la formation de biofilm. L'article est disponible en Annexe D.

Chapitre 2

Contexte de l'étude et objectifs

2.1 Description de l'épidémie

Pendant 11 ans, une souche de *P. aeruginosa* a infecté ou colonisé plus de 300 patients au sein du Centre Hospitalier Régional Universitaire (CHRU) de Besançon [128]. Une analyse en champ pulsé (PFGE) menée sur les souches prélevées entre 1997 et 2008 a permis d'identifier les souches proches c'est à dire semblables à 80%, ce qui constitue la première étape de caractérisation des isolats de l'épidémie. Ensuite, un typage par MLST a permis d'associer la souche épidémique au ST395.

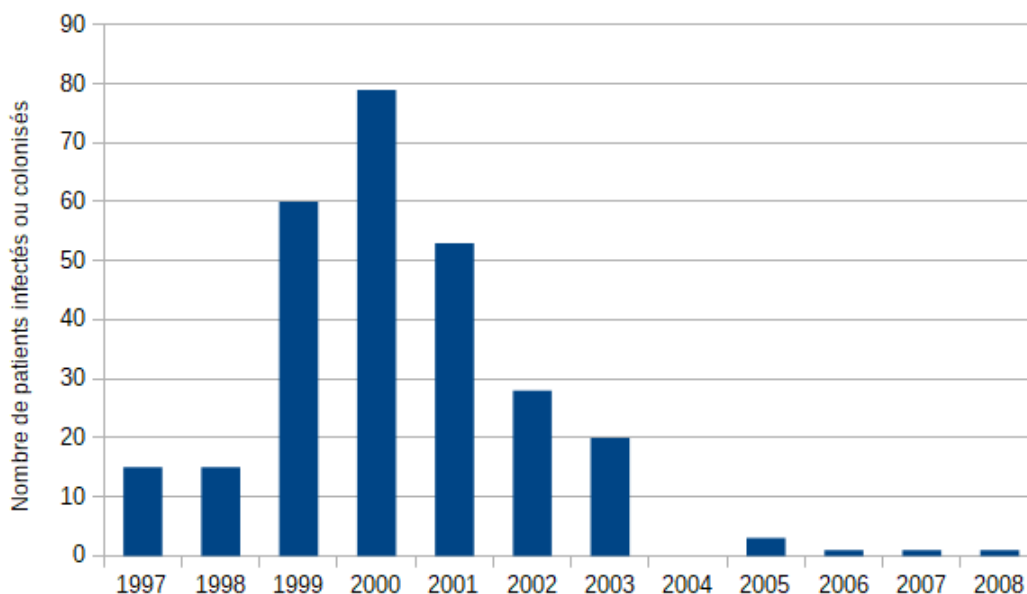


FIGURE 2.1 – Évolution du nombre de patients atteints par an (typage par PFGE), données épidémiologiques issues du laboratoire d'hygiène hospitalière de Besançon

Des mesures d'hygiène ont été mises en place dans les services d'hospitalisation concernés afin de limiter la transmission. Des modifications phénotypiques ont pu être observées pendant les 11 ans d'épidémie, telles qu'une modification de la résistance aux antibiotiques. De plus,

le nombre de nouveaux patients infectés ou colonisés augmente entre 1997 et 2000, alors que dans les années suivantes, le nombre de nouveaux cas diminue jusqu'en 2008 avec le dernier isolat typé comme appartenant à l'épidémie (figure 2.1).

Après 2003, les phénotypes de la souche étant trop divers, il devenait difficile de connaître exactement le nombre de patients atteints. Cependant, le typage des souches isolées des patients hospitalisés a permis d'identifier 6 nouveaux cas après 2004 avec un dernier cas en avril 2008 signant la fin probable de l'épidémie.

2.2 Hypothèse

L'hypothèse principale de cette thèse concerne l'adaptation de la bactérie à un environnement particulier représenté par les patients hospitalisés. L'évolution de la bactérie serait semblable à celle observée lors d'infections chroniques, où l'adaptation à une niche particulière (le patient) entraîne la perte de fonctions inutiles conduisant à l'épuisement de la souche et à l'arrêt de l'épidémie.

La mise en place de mesures d'hygiène couplées à l'évolution parallèle des différents isolats conduiraient à de multiples culs-de-sac évolutifs marquant ainsi la fin de l'épidémie.

2.3 Matériel et Méthodes

2.3.1 Collection d'isolats

La collection conservée à l'hôpital contient une partie des isolats de l'épidémie (80 isolats), qui n'ont pas tous été séquencés pour des raisons budgétaires.

La sélection des souches utilisées, a été faite de manière à conserver la représentativité de l'échantillon initial. En effet, les échantillons choisis sont répartis tout au long de l'épidémie de manière à couvrir toutes les années de diffusion de la bactérie. Les analyses préalables (clustering par PFGE) ont permis de sélectionner des isolats ayant des profils identiques, correspondant à un clone unique.

Cinquante-quatre isolats ont été sélectionnés pour l'analyse. Les dates de prélèvement, le service d'hospitalisation du patient, ainsi que le type de prélèvement de la bactérie sont disponibles dans le tableau 2.1.

Tous les isolats de l'étude seront nommés par la suite BES pour *Besançon Epidemic Strain* sauf le premier isolat nommé DHS01 ou BES-1. Celui-ci, prélevé en mai 1997, sert de référence pour toutes les études décrites ci-après.

Le dernier isolat, nommé DHS30, et prélevé en août 2014 a été retiré de l'étude. Il s'agit d'un ST395 mais les analyses génétiques menées sur cet isolat ont indiqué une grande dissemblance avec les autres isolats de l'épidémie. Il sera utilisé par la suite comme un *outgroup* local.

Souche	Date	Service	Prélèvement
BES-1/DHS01	05-Mai-1997	Réanimation chirurgicale	Nez
BES-2	27-Mai-1997	Réanimation chirurgicale	Aspiration trachéale
BES-3	11-Août-1997	Réanimation chirurgicale	Nez
BES-4	29-Sept-1997	Réanimation chirurgicale	Rectum
BES-5	20-Oct-1997	Réanimation chirurgicale	Rectum
BES-6	22-Nov-1997	Réanimation chirurgicale	Aspiration trachéale
BES-7	26-Jan-1998	Réanimation chirurgicale	Rectum
BES-8	06-Fév-1998	Réanimation chirurgicale	Aspiration trachéale
BES-9	24-Avr-1998	Réanimation chirurgicale	Drain
BES-10	26-Mai-1998	Réanimation chirurgicale	Urines
BES-11	17-Août-1998	Réanimation chirurgicale	Cathéter
BES-12	04-Nov-1998	Réanimation chirurgicale	Lavage broncho-alvéolaire
BES-13	06-Déc-1998	Réanimation chirurgicale	Nez
BES-14	11-Jan-1999	Réanimation chirurgicale	Rectum
BES-15	29-Mars-1999	Réanimation chirurgicale	Urines
BES-16	26-Avr-1999	Réanimation chirurgicale	Selles
BES-17	25-Mai-1999	Réanimation chirurgicale	Rectum
BES-18	14-Juin-1999	Réanimation chirurgicale	Nez
BES-19	19-Juil-1999	Réanimation chirurgicale	Nez
BES-20	06-Août-1999	Réanimation chirurgicale	Aspiration trachéale
BES-21	27-Sept-1999	Réanimation chirurgicale	Rectum
BES-22	03-Nov-1999	Traumatologie	Cathéter
BES-23	24-Déc-1999	Gériatrie 1	Ostéo-arthrite
BES-24	18-Jan-2000	Neurochirurgie	Nez
BES-25	26-Jan-2000	Réhabilitation 1	Crachat
BES-26	01-Fév-2000	Réanimation médicale	Rectum
BES-27	30-Nov-2000	Réanimation médicale	Nez
BES-28	01-Fév-2001	Réanimation médicale	Aspiration trachéale
BES-29	19-Juin-2001	Réanimation médicale	Sang
BES-30	07-Juil-2001	Réhabilitation 1	Urines
BES-31	22-Août-2001	Réhabilitation 1	Urines
BES-32	21-Sept-2001	Réhabilitation 2	Plaie
BES-33	01-Oct-2001	Orthopédie	Plaie
BES-34	03-Nov-2001	Réanimation médicale	Nez
BES-35	21-Déc-2001	Urologie	Urines
BES-36	17-Jan-2002	Dermatologie	Plaie
BES-37	07-Mars-2002	Diabétologie/Endocrinologie	Plaie
BES-38	10-Juin-2002	Néphrologie	Urines
BES-39	02-Sept-2002	Gastro-entérologie	Urines
BES-40	12-Oct-2002	Gériatrie 2	Selles
BES-41	07-Nov-2002	Chirurgie vasculaire	Urines
BES-42	24-Fév-2003	Réanimation médicale	Aspiration trachéale
BES-43	15-Mai-2003	Pneumologie	Sang

Souche	Date	Service	Prélèvement
BES-44	24-Juin-2003	Neurochirurgie	Urines
BES-45	13-Juil-2003	Réanimation chirurgicale	Lavage broncho-alvéolaire
BES-46	02-Août-2003	Hépatologie	Selles
BES-47	08-Août-2003	Réhabilitation 2	Urines
BES-48	01-Oct-2003	Neurochirurgie	Aspiration trachéale
BES-49	01-Jan-2005	Réhabilitation 1	Sang
BES-50	01-Sept-2005	Hépatologie	Cathéter
BES-51	01-Déc-2005	Réanimation médicale	Urines
BES-52	01-Mars-2006	Réanimation chirurgicale	Aspiration trachéale
BES-53	01-Déc-2007	Urologie	Urines
BES-54	01-Avr-2008	Réanimation médicale	Urines

TABLEAU 2.1 – Données sur les isolats issus de la collection BES

2.3.2 Objectifs et méthodologie générale

Afin de comprendre en détail la présence à l'hôpital, la persistance et la disparition de la souche épidémique, une approche WGS a été utilisée.

Quelle est la structure et les particularités du ST395 ?

La souche DHS01, premier isolat de l'épidémie, est séquencée en Illumina et en PacBio, et la composition du génome est investiguée afin de comprendre ce qui favorise son épidémiogénicité. Ensuite l'analyse est étendue au ST395 en utilisant d'autres isolats ST395 et non-ST395 pour comprendre les particularités de ce clone.

Comment la souche épidémique a évolué au cours de l'épidémie, et quels sont les éléments ayant participé à son déclin ?

Le deuxième objectif de cette thèse est l'analyse de 54 génomes d'isolats épidémiques pour comprendre l'évolution du contenu en gènes, les mutations ponctuelles sélectionnées et les modifications phénotypiques associées. La datation de l'ancêtre commun à partir de données phylogénétiques a permis de déterminer l'origine de la souche épidémique.

Chapitre 3

La caractérisation génomique d'un clone épidémique local de *P. aeruginosa* révèle des éléments spécifiques du clone répandu ST395

3.1 Objectifs et méthodologie

Afin de caractériser le premier isolat de l'épidémie du clone *P. aeruginosa* ST395 au CHRU de Besançon, DHS01, le génome complet a été obtenu par séquençage PacBio. La souche avait précédemment été séquencée en Illumina [129], et les *reads* ainsi obtenus ont été utilisés afin de valider l'assemblage PacBio. L'annotation automatique générée par le NCBI permet de connaître la composition en gènes. Cette annotation est complétée par la recherche précise des gènes impliqués dans la résistance et la virulence de l'isolat. L'intégration d'éléments mobiles tels que les îlots génomiques et les ICEs sont recherchés, tout comme la présence du système CRISPR-Cas qui semble jouer un rôle dans l'intégration d'éléments mobiles.

Malgré le grand nombre de génomes de *P. aeruginosa* disponibles sur le site du NCBI (2437 génomes déposés dont 94 génomes complets en juillet 2017), il n'y a aucun génome de référence disponible pour le clone ST395. La recherche des éléments génomiques spécifiques du ST395 par comparaison avec des isolats non-ST395 permet la caractérisation du ST. Pour cela, 5 isolats ST395 épidémiques dans d'autres hôpitaux et 8 isolats non-ST395 épidémiques ont été séquencés. Une phylogénie utilisant les génomes complets de *P. aeruginosa* permet de situer le ST395 vis à vis des autres STs.

3.2 Résultats : Article publié dans *Microbial Genomics* (Juillet 2017)

Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395

Marie Petitjean^{1,2}, Daniel Martak¹, Alicia Silvant¹, Xavier Bertrand^{1,2}, Benoit Valot², and Didier Hocquet^{1,2,3}

¹Laboratoire d'Hygiène Hospitalière, Centre Hospitalier Régional Universitaire, Besançon, France

²UMR CNRS 6249, Chrono-environnement, Université de Bourgogne Franche-Comté, Besançon, France

³Centre de Ressources Biologiques - Filière Microbiologique de Besançon

Key words : *Pseudomonas aeruginosa*, multidrug resistance, copper, outbreak, high-risk clone

3.2.1 Abstract

Pseudomonas aeruginosa is a ubiquitous opportunistic pathogen with several clones being frequently associated with outbreaks in hospital settings. ST395 is among these so-called 'international' clones. We aimed here to define the biological features that could have helped the implantation and spread of the clone ST395 in hospital settings. The complete genome of a multidrug resistant index isolate (DHS01) of a large hospital outbreak was analyzed. We identified DHS01-specific genetic elements, among which were identified those shared with a panel of six ST395 independent isolates responsible for outbreaks in other hospitals. DHS01 has the fifth largest chromosome genome of the species (7.1 Mpb) with most of its 1,555 accessory genes borne by either genomic islands (GIs, $n=48$) or Integrative and Conjugative Elements (ICEs, $n=5$). DHS01 was multidrug resistant mostly due to chromosomal mutations. It displayed signatures of adaptation to chronic infection in part due to the loss of a 131-kbp chromosomal fragment. Four GIs were specific to the clone ST395 and contained genes involved in metabolism (GI-4), in virulence (GI-6), and in resistance to copper (GI-7). GI-7 harbored an array of 6 copper transporters and was shared with non-pathogenic *Pseudomonas* sp. retrieved from copper-contaminated environments. Copper resistance was confirmed phenotypically in all other ST395 isolates and possibly accounted for the spreading capability of the clone in hospital outbreaks, where water networks have been incriminated. This suggests that genes transferred from copper-polluted environments may have favored the implantation and spread of the international clone *P. aeruginosa* ST395 in hospital settings.

3.2.2 Impact Statement

The opportunist pathogen *Pseudomonas aeruginosa* has a panmictic population structure, with a few international clones being frequently associated with outbreaks in hospitals world-

wide. The biological features that favor the implantation and spread of these clones in hospital settings are largely unknown. ST395, one of these clones, has been retrieved from polluted environments and responsible for outbreaks in hospitals. The genomic analysis of a collection of independent ST395 isolates identified four specific genomic islands (GIs) that contained genes involved in metabolism, in virulence, and in the resistance to copper. An array of six copper transporters was borne on a single GI that was shared with non-pathogenic *Pseudomonas* sp. found in copper-contaminated environments. The systematic resistance to copper of ST395 isolates possibly accounted for their spreading capability in hospital outbreaks, where water networks have been incriminated. This observation could help implement infection control measures to tackle the proliferation of copper-resistant pathogenic bacteria in water networks.

3.2.3 Introduction

Pseudomonas aeruginosa is an opportunistic human pathogen responsible for outbreaks in hospital setting [130]. This bacterial pathogen infects injured, burned, immunodeficient, immunocompromised and cystic fibrosis patients [131]. Despite the non-clonal epidemic population structure of *P. aeruginosa*, several multi-locus sequence types have been observed to spread worldwide and are frequently associated with epidemics where multidrug resistance confounds treatment. ST111, ST177, ST235, and ST395 are among these so-called ‘international’ or ‘high-risk’ clones [11]. The biological features that make these clones more epidemic than other STs are largely unknown.

We focused here on the high-risk clone ST395, which spread has been documented within hospitals in France or in the UK [132, 28]. From 1997 to 2008, a strain of *P. aeruginosa* ST395 has infected or colonized more than 300 patients in the University Hospital of Besançon [128]. Similarly, this clone (ST395) was responsible for an outbreak in a UK hospital. The authors suggested that *P. aeruginosa* ST395 rapidly colonized the plumbing system of a new hospital, before its commissioning, and was transmitted to burn patients with the water [28]. Besides, the clone ST395 has also been retrieved in a plant from a metal-contaminated estuary, and harbored resistance determinants to mercury, arsenic and copper, together with antibiotic resistance genes [27].

P. aeruginosa strains adapt to environments through the gain or loss of blocks of genes [i.e. horizontal gene transfers of Genomic Islands (GIs) or Integrative and Conjugative Elements (ICEs)] and modifications at the nucleotide level. Acquired genes originated from either clinical or environmental species [56].

The virulence determinants are mostly part of the core genome of the species, accounting for the ability of *P. aeruginosa* to inhabit a wide range of environments, including patients [133]. Moreover, the co-selection of antibiotic resistance genes with metal resistance genes can promote the ability of strains adapted to toxic environments to thrive in hospital settings, where the antibiotic consumption is high [5].

We aimed here at defining the features that could have helped the implantation and spread

of the clone *P. aeruginosa* ST395 in hospital settings. We first thoroughly analyzed the complete genome of the multidrug resistant index isolate of an outbreak in France (isolate DHS01), and identified genetic elements specific to this isolate, paying a particular attention to the resistome, the virulome and mobile genetic elements. Among these specific elements, we further identified those shared with a panel of ST395 independent isolates responsible for outbreaks in other hospitals.

3.2.4 Methods

Genome sequencing of the index isolate. The isolate DHS01, the first isolate of the epidemics, was isolated in May 1997 from the nose of a 31-year old male patient in the surgical intensive care unit in the University Hospital of Besançon (Eastern France). In our hospital, *P. aeruginosa* is searched weekly in patients hospitalized in intensive care unit. Microbiological records revealed that the patient was already colonized by *P. aeruginosa* 25 days before isolation of DHS01. DHS01 was identified at the species level by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) with a Microflex LT (Bruker Daltonik GmbH, Bremen, Germany), according to the manufacturer’s procedures. Bacterial DNA was isolated from an overnight culture on Mueller-Hinton agar using the Genomic-tip kit (Qiagen) and further sequenced with Single Molecule Real-Time technology (Pacific Biosciences) [113]. The reads were assembled using Hgap3 into a circular complete genome [134]. Genome was annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAT) [135]. The genes were classed as core or accessory genes of the species *P. aeruginosa* by comparing to previous pangenome analysis [7].

Dataset. We compared the DHS01 genome against two datasets. The dataset 1 included 82 *P. aeruginosa* isolates which complete genomes were available on NCBI at the time of the study (June 2017; Table 3.7, available in the Supplementary Material). The dataset 2 included the raw reads of 6 ST395 isolates of *P. aeruginosa* and 10 non-ST395 isolates of *P. aeruginosa* isolated during outbreaks in French or UK hospitals (Table 3.1) sequenced with the Illumina NextSeq technology with an average depth of 200X coverage per genome. The absence of close relatedness between the ST395 isolates ($n=7$) was determined by pulsed-field gel electrophoresis as previously described (Fig. 3.1) [136].

Strain	Origin	Availability	ST	Reference
PT B	Hospital of Besançon, France	BioSample SAMN06627436	395	This study
PT C	Hospital of Besançon, France	BioSample SAMN06627437	395	This study
PT D	Hospital of Lyon, France	BioSample SAMN06627442	395	This study
PT E	Hospital of Lyon, France	BioSample SAMN06627446	395	This study

Strain	Origin	Availability	ST	Reference
PT F (Birmin- gham 903)	Hospital of Birmin- gham, UK	European Nu- cleotide Archive ERP006058	395	Quick J et al. BMJ Open. 2014 Nov 4;4(11) :e006278.
PT G	Hospital of Bor- deaux, France	BioSample SAMN06627443	395	This study
ST111	Hospital of Besan- çon, France	BioSample SAMN06651179	111	This study
ST175a	Hospital of Besan- çon, France	BioSample SAMN06627435	175	This study
ST175b	Hospital of Besan- çon, France	BioSample SAMN06651181	175	This study
ST233	Hospital of Besan- çon, France	BioSample SAMN06651183	233	This study
ST235a	Hospital of Besan- çon, France	BioSample SAMN06651178	235	This study
ST235b	Hospital of Besan- çon, France	BioSample SAMN06651180	235	This study
ST348	Hospital of Nancy, France	BioSample SAMN06651182	348	This study
PAO1	Melbourne, Austra- lia	NCBI Refe- rence Sequence : NC_002516.2	549	Stover CK et al. Nature. 2000 Aug 31 ;406(6799) :959- 64.
ST1342	Hospital of Laribois- sière, Paris, France	BioSample SAMN06627441	1342	This study
ST1602	Hospital of Lyon, France	BioSample SAMN06627438	1602	This study

TABLEAU 3.1 – Informations about raw data of *P. aeruginosa* ST395 and non-ST395 strains used for characterisation of specific genetic material, dataset 2.

Phylogenetic analysis. We determined the sequence type (ST) of DHS01 with the Center for Genomic Epidemiology website [137]. Then DHS01 genome was compared with the genomes of dataset 1 (Table 3.7, available in the Supplementary Material) with MUMmer [138]. The length of the core genome alignment was 1,770,651 bp, representing 1,973 genes with 118,840 variable positions. The phylogenetic tree was determined using PhyML software [123] with the following parameters : GTR substitution model with gamma-distributed rate variation across sites and a proportion of invariable sites.

Mobile genetic elements. Genomic islands and CRISPR-Cas systems were located with IslandViewer and CRISPRfinder, respectively [139, 140]. Integrative and conjugative elements (ICEs) were retrieved from gene annotation (i.e. containing an integrase-encoding gene at the

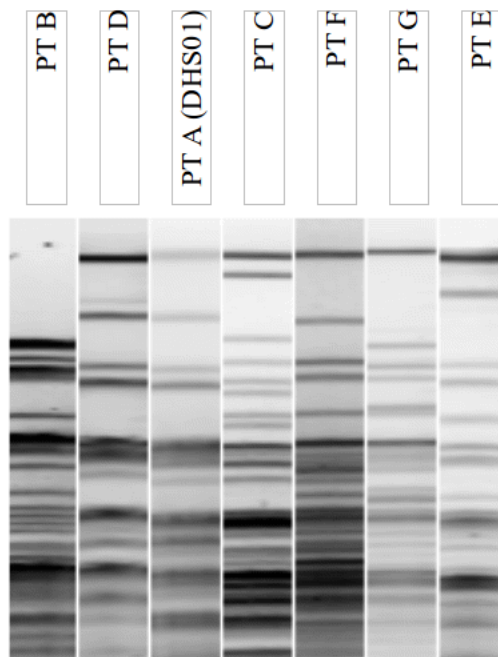


FIGURE 3.1 – Pulsed-field gel electrophoresis profiles of *DraI*-digested DNA from *P. aeruginosa* ST395 isolates (n=7).

5'-end, genes encoding the full conjugation machinery *traG* and *trbBCDEFGHI*, and a relaxase). Free circular forms of the ICEs were detected in the DNA extracted from SOS-induced bacteria by nested PCR using primers that overlapped the *attI* sequence (Table 3.2). The SOS response was induced by mitomycin as previously described [141]. To identify ST395-specific GIs and ICEs, we mapped the Illumina reads from a collection of ST395 and non-ST395 isolates (Table 3.7, available in the Supplementary Material) on the genome of DHS01. The presence of GIs and ICEs' gene was determined with featureCount [121].

Identification of resistance and virulence determinants. The susceptibility to the major antipseudomonal agents (ticarcillin, piperacillin, cefepime, ceftazidime, aztreonam, imipenem, amikacin, gentamicin, tobramycin, and ciprofloxacin) were tested using the EUCAST recommendations with the reference agar dilution method with the reference strain PAO1 as a control [142]. Resistance genes were searched with ResFinder [143]. We searched chromosomal mutations, in comparison with the genome of PAO1, responsible for resistance to antibiotics in the gene *oprD*, those encoding the main efflux systems (MexAB-OprM, MexXY, MexCD-OprJ, MexEF-OprN) and their regulators (MexR, MexZ, ArmZ, NfxB, NalC, NalD, ParS, ParR, MexS, and MexT), the quinolone-resistance determining regions (QRDR; in *gyrA*, *gyrB*, *parC*, *parE*), the gene encoding the AmpC cephalosporinase (*ampC*) and its regulators (*dacB*, *ampR*, *ampD*, *ampDh2*, *ampDh3*, *ampG*, *ampO*, and *ampP*), the regulators of polymyxin resistance (*colR*, *colS*, *cprR*, *cprS*, *pmrA*, and *pmrB*) and the *phoP-phoQ* genes associated with colistin resistance. Virulence genes were searched using Virulence Factor DataBase (VFDB) [144]. Mutations in resistance and virulence genes were detected using BLAST [145]. We first determined the serotype of DHS01 in silico using PAST online tool [146]

ICE number	PCR	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
ICE-1	Initial	GGATCTGCAAGGCCAGGTTG	GCAGGCTGAGAGCAGGATAG
	Nested	GCGAACTGATCGCCGGG	TCGTTGTTCGGAGAGGGTGTA
ICE-2	Initial	GTGGATGGGGATCTACTGCC	TGAGCTGGTTGTTCTGGTCCG
	Nested	CATTGCTACGCCTGCAGAAT	CATTGCTACGCCTGCAGAAT
ICE-3	Initial	TGACTTGATCGGCGTAACCC	CGCATGTATCTTCCCCGGTT
	Nested	TGTACGCCCTTCTTGACGAG	ATTTCCCCGCCATAGGTAGC
ICE-4	Initial	CCTGTTCTTCGCTGGATCGT	ATGACGGGCATCTTTCAGGG
	Nested	CTCCCTTCTGATCGCAGTG	AGGTCTGAAAGCATGAGCCC
ICE-5	Initial	GATACTGAAGGGCCGGCTAC	GAAGTGAAGTTCCTCGGCCA
	Nested	CGCCTGCAGAATGGTTTCAG	CCGGTTCCTGGTAACGTCTG

TABLEAU 3.2 – Primers used for identification of circular forms of the ICEs using nested PCR.

and confirmed the result with specific antiserum (Biorad, Marnes-La-Coquette, France).

Survival in copper solution. The bacterial survival in 7.5 mM of CuSO₄ has been determined by the inoculation of cation-calibrated Mueller-Hinton broth (Becton-Dickinson) with 10⁷ CFU/ml of stationary-phase cells and incubation during 48 h at 37°C under shaking (150 rpm). The initial and final concentrations of living bacteria were determined by counting on MH agar plates. Experiments were performed at least three times independently.

3.2.5 Results and Discussion

Genome assembly. The sequencing of DHS01 genome generated 98,987 passed filter reads, with a mean read length of 12 kb and a mean coverage of 134X. The quality of the data allowed the assembly of the DHS01 genome in a complete circular chromosome of 7,055,752 bp (Fig. 3.2).

Plasmids were absent from the genome. PGAT annotation predicted 6,524 coding genes, 67 tRNAs, 5 ncRNAs and 12 rRNAs (4 5S, 4 16S and 4 23S). The GC content of the DHS01 genome (65.8 %) was low compared with that usually found in the species (\simeq 66.6%). The genes acquired by horizontal transfer have generally a lower GC content than that of the core genome. The size of the core genome of *P. aeruginosa* has been evaluated at 5,234 genes [7]. By clustering with CD-HIT, we found that 1,555 genes formed the accessory genome of DHS01. We found that 866 out of 1,555 accessory genes (56 %) were borne by either GIs or ICEs confirming that these mobile genetic elements were major source of accessory genome and genetic diversity [56, 147]. The distribution of core and accessory genes is indicated in the Fig. 3.2.

Phylogenetic analysis. DHS01 belonged to ST395 and was predicted of serotype O6. DHS01 has the fifth largest chromosome when compared with those of other isolates of

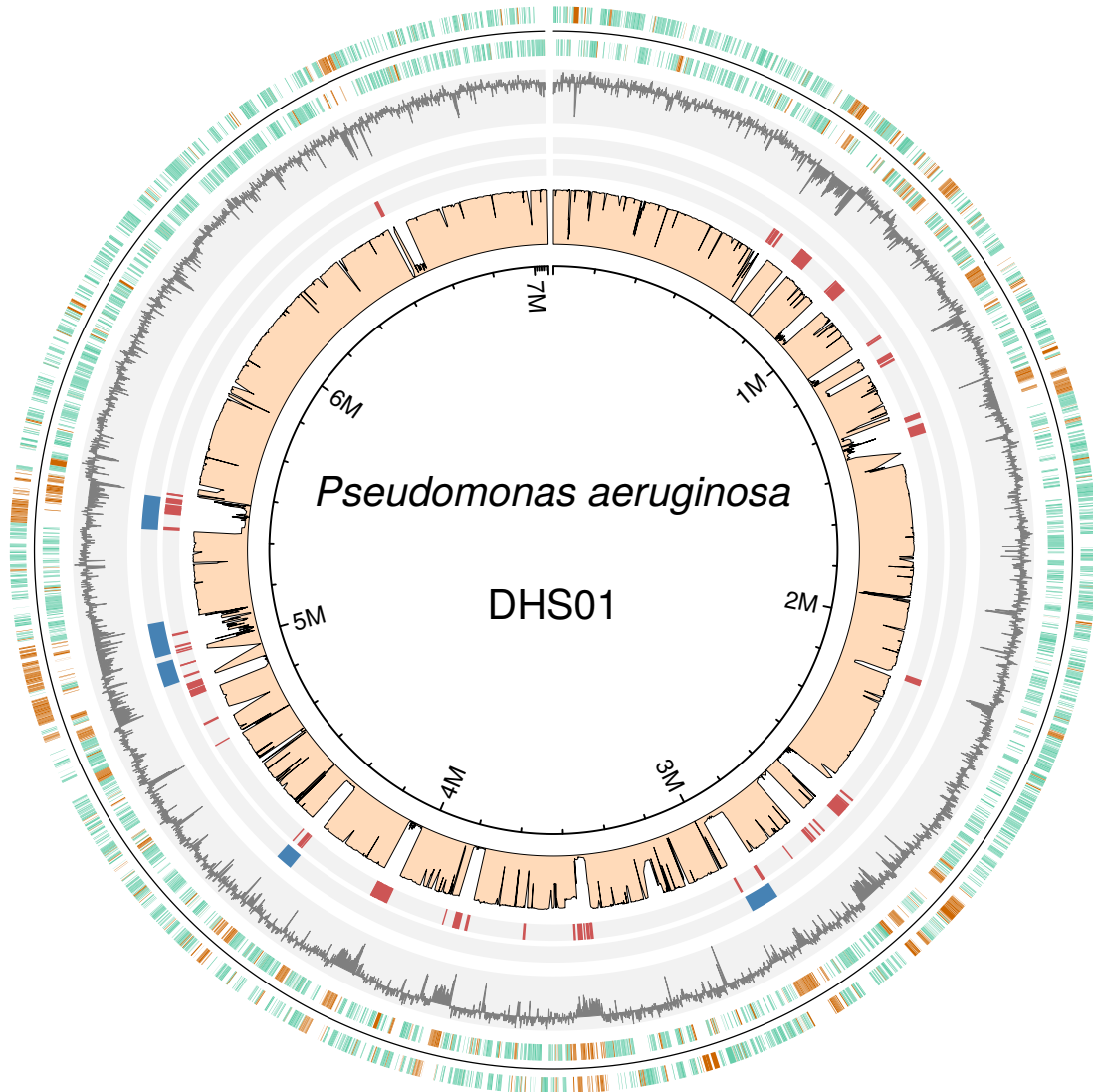


FIGURE 3.2 – Circle plot of complete genome of the epidemic *P. aeruginosa* DHS01 (ST395). First and second outer circles indicate CDS on the plus and the minus strand, respectively. Genes in green are part of the core genome of the species and genes in brown are accessory genes. The third circle display the GC content (deviation from average) of the DHS01 genome. The two next circles show the location of GIs (in red) and ICEs (in blue). The pink innermost circle displays the numbers of genomes (of the 82 genomes of the dataset 1) containing this sequence. We drew the circle plot with R software.

the species, suggesting the recent integration of genetic material in its genome. The three largest chromosomes of the species belonged to ST111 isolates, the fourth largest belonged to a ST357 isolate, and the sixth largest belonged to a ST235 isolates. Interestingly, all these STs are responsible for large outbreaks in hospital settings. This observation was in line with recent data showing that the genomes of ST111, ST395, and ST235 isolates were of 7.05, 7.00, and 6.81 Mbp, respectively, while the average genome of the species was 6.72 Mbp in size [148].

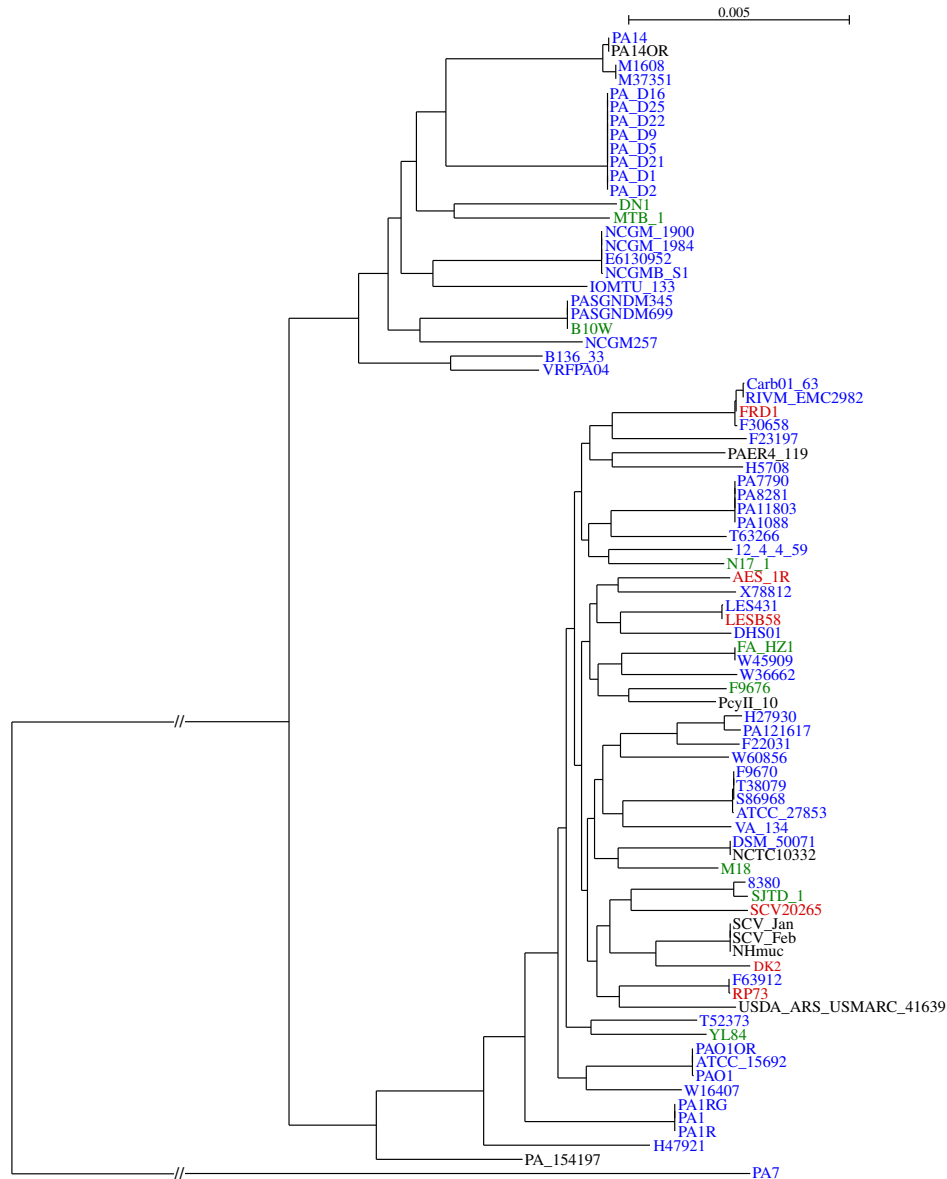


FIGURE 3.3 – Phylogenetic tree of 83 isolates of *Pseudomonas aeruginosa* with complete genome, including DHS01. The phylogenetic tree is well supported with the majority of branch support values higher than 98. The isolates cultured from cystic fibrosis patients are colored red, non-cystic fibrosis clinical isolates are colored blue, and isolates of environmental origin are written in green, and isolates of other origin are written in black.

The phylogenetic analysis of 83 *P. aeruginosa* isolates (DHS01 and those of dataset 1) confirmed the split into two clusters containing either reference strains PAO1 or PA14 (Fig. 3.3) [149]. When compared to PAO1 genome, DHS01 genome expectedly displayed the major chromosomal inversion between two copies of ribosomal RNA reported in PA14 and all the genomes tested by Lee et al. [149]. DHS01 had 5,833 and 5,837 SNPs with its closest isolates LESB58 and LES431, respectively, that belonged to the highly virulent epidemic strain (LES, Liverpool Epidemic Strain) [38].

DHS01 underwent a 131-kb chromosomal deletion. DHS01 was the first isolate of a 11-year outbreak in the University Hospital of Besançon and was then considered as an ancestor from which the late isolates derived [128]. However, when compared with the genome of DHS29, a late isolate of the outbreak, DHS01 genome underwent a 131-kb chromosomal deletion [129]. This segment was intact in the genomes of five of six ST395 independent isolates (dataset 2, Table 3.1). The 131-kb gene block included the genomic regions PA1246-PA1292 and PA1381-PA1450 in the reference strain PAO1, encompassed the virulence genes encoding the alkaline protease AprA (that degrades components of the host immune system) and its secretion system AprDEFI. The deletion also included the genes encoding the biofilm dispersion regulator BdlA, the quorum-sensing master regulator LasR, the LasI enzyme and the quorum sensing repressor RsaL, and the flagellar operon region including *fliKLMNOPQR* and *fliB* [150, 151]. Based on published phenotypic observations, one can hypothesize that the loss of these genes certainly affect the biofilm formation, the virulence, the adherence and the motility of DHS01 [94, 152].

DHS01 virulence factors. We compared the virulome of DHS01 with that of the reference strain PAO1 using the Virulence Factor DataBase (VFDB). The 347 virulence genes found in DHS01 were also retrieved PAO1. Unexpectedly, 36 virulence genes present in PAO1 and part of the core genome of the species were missing from DHS01 genome [129]. The 131-kb chromosomal deletion described above included 18 of these genes. Other genes missing from the DHS01 chromosome were implicated in the bacterial adherence and the biofilm formation. Hence, 8 virulence genes contiguous in PAO1 (*flgK*, *flgL*, *fgtA*, *fliC*, *flaG*, *fliD*, *fliS*, and *fleP*) and forming a block of 14,846 nucleotides were missing from DHS01, probably as the result of a deletion ‘en bloc’. Besides, the genes encoding the chaperonins CupE3 and CupC1 displayed 6-nt and 5-nt deletions, respectively, when compared with PAO1. Genes implicated in the synthesis of type IV pili were missing (*pilA*, only 17% of the sequence were retrieved) or mutated (*pilB* and *pilC*). The deletion of the type IV pili-related *pilA* gene has been shown to impair the biofilm formation [94].

DHS01 surprisingly harbored less virulence genes than the reference strain PAO1 suggesting a lower virulence in vivo. This contrasts with its status of international epidemic clone. DHS01 was isolated after 25 days of patient colonization and displayed genomic signatures of strain adapted to chronic colonization (i.e. loss of flagellum, deficiency in twitching motility, in the quorum sensing, and in biofilm formation) [94, 104]. Although being the first isolate of the epidemics, DHS01 was likely an epidemiological dead end and was not representative of the strain that has spread to more than 300 patients [153].

DHS01 multidrug resistance is mostly due to chromosomal mutations. DHS01 was multidrug resistant with a high level of resistance to gentamicin (MIC > 512 mg/l) and to tobramycin (MIC, 32 mg/l), to ciprofloxacin (MIC, 32 mg/l), and to β -lactam compounds (ticarcillin MIC, 64 mg/l; piperacillin MIC, 128 mg/l). It was of intermediate resistance to aztreonam (MIC, 16 mg/l) and remained susceptible to cefepime (MIC, 8 mg/l), ceftazidime (MIC, 8 mg/l), imipenem (MIC, 4 mg/l), and amikacin (MIC, 2 mg/l). The acquired genes and mutations in regulators leading to antibiotic resistance are detailed in Table 3.3. We had

previously shown that DHS01 overproduced AmpC (650-times more than PAO1), and overexpressed *mexA* and *mexX* (1.85- and 3.70-times more than PAO1, respectively) [154]. Briefly, DHS01 produced a 2'-aminoglycoside nucleotidyl-transferase, overproduced an extended-spectrum cephalosporinase AmpC (ESAC; *ampR* and *ampD* mutant), presumably overproduced MexAB-OprM (*nalB* mutant), MexXY (*agrZ* mutant), MexEF-OprN (*mexS* mutant), and displayed canonical mutations in the QRDRs of GyrA and ParC.

Consequence	Antibiotic(s) affected	Gene or mutation (in comparison with PAO1)
2'-aminoglycoside nucleotidyl-transferase production	Gentamicin, tobramycin	<i>ant(2'')-Ia</i> (borne by ICE-5)
MexXY efflux pump overproduction	Aminoglycosides, fluoroquinolones	8 nt-deletion in the regulator-encoding gene <i>mexZ</i>
Quinolone target mutation	Fluoroquinolones	T83I in GyrA and S87L in ParC
AmpC cephalosporinase overproduction	Ticarcillin, piperacillin, cefepime, ceftazidime, aztreonam	M1L in the regulator AmpR and A134V and T139M in AmpD
MexAB-OprM efflux pump overproduction	Ticarcillin, cefepime, aztreonam	H107P in the regulator MexR
MexEF-OprN efflux pump overproduction	Imipenem	R48C in the oxidoreductase MexS
Reduced production of the porin OprD	Imipenem	R48C in the oxidoreductase MexS
Production of the extended-spectrum (ESAC) AmpC	Imipenem	A105 in AmpC

TABLEAU 3.3 – Acquired genes and mutations in regulators leading to antibiotic resistance in the ST395 *P. aeruginosa* isolate DHS01.

Integrative and Conjugative Element (ICE) in DHS01. ICEs are mobile genetic elements integrated into a genome and able to excise and transfer by conjugation in another genome [54]. Genome analysis predicted 5 ICEs in the genome of DHS01 (Fig. 3.4).

As expected, each ICE was composed of 2 to 4 GIs (Table 3.4 and Table 3.5). Hence, 13 GIs were integrated into the DHS01 genome via the integration of ICEs. ICE-1, ICE-2, ICE-3, ICE-4 and ICE-5 were 78 kb, 47 kb, 66 kb, 94 kb, and 93 kb in size, respectively. None of these ICEs were retrieved in the other ST395 isolates (dataset 2) with the exception of ICE-4 shared by the strain PT F from the hospital of Birmingham, UK [28]. The Fig. 3.4 details the structure, the gene content, and the site of insertion of the 5 ICEs within DHS01 chromosome. The 5 integrated ICEs provided DHS01 with genes involved in virulence and resistance antibiotics and heavy metals that could have favored its installation and spread into an hospital. Hence, ICE-4 contained the pili type IVa operon (*pilLNOXQVTM*) also present in PAPI-1 of strain PA14 and involved in the transfer of this pathogenicity island to other

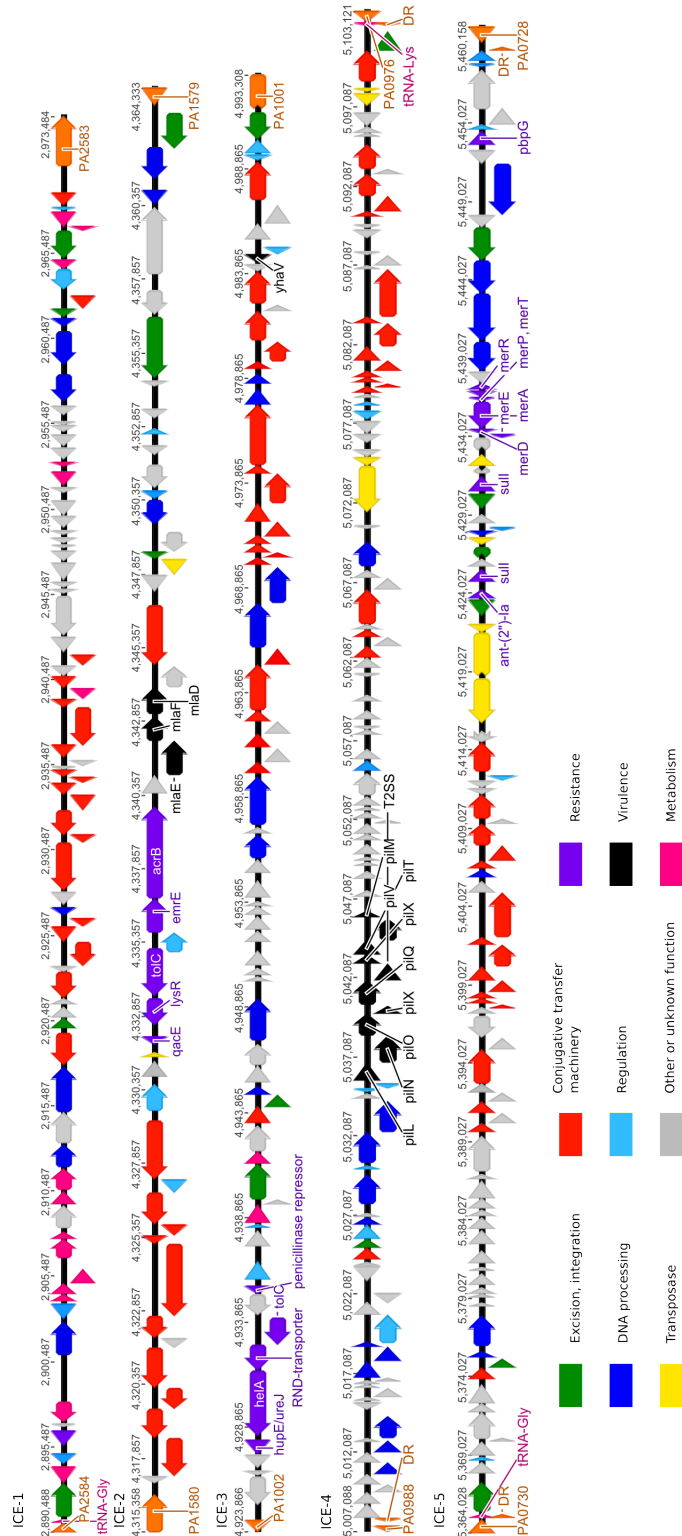


FIGURE 3.4 – Composition of the five ICEs of the isolate DHS01. tRNA, direct repeat (DR) and PAO1 gene name were indicated at the boundary of ICEs. Only genes associated with antibiotic resistance, virulence and metabolism were named.

strains [155]. ICE-5 harbored genes conferring resistance to aminoglycosides (*ant(2'')-Ia*), to sulfonamide (*sulI*) and to mercury (*merRTPADE*). ICE-1 and ICE-5 were integrated into the 3'-ends of two different tRNA_{Gly} while ICE-4 was integrated into the 3'-end of tRNA_{Lys}. No tRNA were found in the vicinity of ICE-2 and ICE-3. The activity of the ICE-4 and ICE-5 was confirmed by the existence of free circular forms after SOS stress (Fig. 3.5).

Name	Start position	Stop position	Length (bp)	Putative species of origin	Encoded proteins and their biological functions
GI-1	684,629	697,340	12,712	-	RNA helicase and DNA methylase (DNA processing)
GI-2	701,883	709,761	7,879	-	Unknown
GI-3	2,569,006	2,615,647	46,642	-	RsmA (carbon storage regulator)
GI-4	2,671,220	2,693,490	22,270	<i>Pseudomonas putida</i> KT2440 (soil)	AlpA (regulation of prophage excision and pathogenicity), TetR (transcriptional regulator), Betaine-aldehyde dehydrogenase (choline metabolism), Flavin oxidoreductase (iron uptake)
GI-5	2,705,853	2,711,674	5,822	<i>Acidithiobacillus ferrooxidans</i> SS3 (82% identity and 96% coverage)	HsdS (restriction-modification)
GI-6	2,713,010	2,723,680	10,671	-	Heat-labile enterotoxin
GI-7	3,430,733	3,447,684	16,951	<i>Pseudomonas putida</i> H8234 (93% of coverage and 99% identity)	2 cupredoxins, copper ABC transporter ATPase, CopZ, CopA, and CopB (copper transport) Helicase UvrD and endonuclease (DNA processing)
GI-8	3,813,161	3,834,228	21,067	-	(Phage protein)

TABLEAU 3.4 – Characteristics of the 8 genomic islands (GIs) specific of the genome of the *P. aeruginosa* ST395 isolate DHS01. These 10 GIs were not retrieved in any of the 82 complete genomes of *P. aeruginosa* (Table 3.7). The three first columns gives the genomic coordinates (start and stop) and the length of the GI, respectively. Fourth column gives the name of the non-*P. aeruginosa* strain which genome contains an identical nucleotidic sequence (with > 99% of length and identity, otherwise specified). The major functions encoded by the genes harboured on the GIs were presented in the last column. The last update of the genome collection revealed 2 clinical ST253 isolates (M1608 and M37351) had acquired the GI-7.

Name	Start	Stop	Length (bp)	Genomes (<i>n</i>)	Putative species of origin	Functions
GI-9	676,826	683,715	6,890	65	-	(Phage protein)
GI-10	775,197	816,878	41,682	84	-	DNA processing, defense mechanisms, secondary metabolites biosynthesis, transport and catabolism.
GI-11	676,826	683,715	6,890	10	-	(Phage protein)
GI-12	775,197	816,878	41,682	16	-	DNA processing, defense mechanisms, secondary metabolites biosynthesis, transport and catabolism.
GI-13	909,813	914,334	4,522	1	-	Unknown
GI-14	914,744	948,980	34,237	4	-	(Phage protein, regulation)
GI-15	1,112,226	1,117,419	5,194	2	<i>Pseudomonas</i> phage MP22 / Phage F_HA0480sp / Pa1651	(Phage protein)
GI-16	1,117,436	1,123,713	6,278	1	<i>Pseudomonas</i> phage JBD93	(Phage protein)
GI-17	1,171,634	1,186,830	15,197	11	-	DNA processing
GI-18	1,190,408	1,198,422	8,015	4	<i>Pseudomonas fluorescens</i> / <i>Pseudomonas balearica</i>	DNA processing
GI-19	1,363,454	1,379,352	15,899	3	<i>Pseudomonas</i> phage JBD25	(Phage protein)
GI-20	1,395,087	1,425,084	29,998	4	-	(Phage protein)
GI-21	2,151,362	2,169,893	18,532	15	-	Sugar metabolism, cell wall/membrane/envelope biogenesis
GI-22	2,551,027	2,557,727	6,701	1	-	Unknown
GI-23	2,793,073	2,797,178	4,106	26	-	Intracellular trafficking, secretion and vesicular transport
GI-24	2,893,400	2,903,896	10,497	5	<i>Pseudomonas misse- lii</i>	Metabolism (ICE-1)
GI-25	2,966,826	2,974,108	7,283	18	-	Metabolism (ICE-1)
GI-26	3,420,507	3,429,999	9,493	6	-	DNA processing
GI-27	3,451,116	3,466,146	15,031	9	-	DNA processing, lipid and copper transport
GI-28	3,472,217	3,479,103	6,887	20	-	Metabolism

Name	Start	Stop	Length (bp)	Genomes (<i>n</i>)	Putative species of origin	Functions
GI-29	3,620,322	3,628,120	7,799	8	-	DNA processing, amino acid transport and metabolism
GI-30	3,789,125	3,798,230	9,106	2	-	Unknown
GI-31	3,860,394	3,864,454	4,061	6	-	Unknown
GI-32	4,033,634	4,087,975	54,342	2	-	(Phage protein)
GI-33	4,326,868	4,331,361	4,494	7	<i>K. pneumoniae</i> plasmid / <i>E. cloacae</i> plasmid	Conjugation machinery (ICE-2)
GI-34	4,331,871	4,344,728	12,858	3	<i>K. pneumoniae</i> plasmid / <i>E. cloacae</i> plasmid	Resistance and virulence (ICE-2)
GI-35	4,358,031	4,362,326	4,296	5	<i>K. pneumoniae</i> plasmid / <i>E. cloacae</i> plasmid	DNA processing (ICE-2)
GI-36	4,713,679	4,717,717	4,039	11	-	Virulence (Pyocin S2)
GI-37	4,782,199	4,789,393	7,195	6	-	Metabolism
GI-38	4,871,972	4,891,972	20,001	2	-	Unknown
GI-39	4,893,999	4,913,052	19,054	1	-	(Phage protein)
GI-40	5,004,422	5,008,882	4,461	23	-	Metabolism, repair system
GI-41	5,010,115	5,015,280	5,166	11	-	DNA processing (ICE-4)
GI-42	5,020,832	5,027,255	6,424	13	-	Regulation, DNA processing (ICE-4)
GI-43	5,055,166	5,062,300	7,135	5	<i>P. aeruginosa</i> genomic island PGI-5	Unknown (ICE-4)
GI-44	5,365,376	5,375,164	9,789	9	<i>Klebsiella oxytoca</i>	Unknown (ICE-4)
GI-45	5,411,217	5,440,266	29,050	4	-	Resistance, DNA processing (ICE-5)
GI-46	5,443,133	5,459,224	16,092	1	-	Regulation, DNA processing (ICE-5)
GI-47	5,467,263	5,473,009	5,747	1	-	Unknown
GI-48	6,538,381	6,550,792	12,412	6	-	DNA processing

TABLEAU 3.5 – Characteristics of the 40 genomic islands (GIs) non-specific of the genome of the *P. aeruginosa* ST395 isolate DHS01. The four first columns give the GI number, its genomic coordinates (start and stop) and length (bp), respectively. The fifth column gives the number of complete genomes (Table 3.7) sharing the nucleotidic sequence. The sixth column gives the name of the non-*P. aeruginosa* strain which genome contains a similar nucleotidic sequence (with > 99% of length and identity). The last column gives the functions and the name of the proteins encoded.

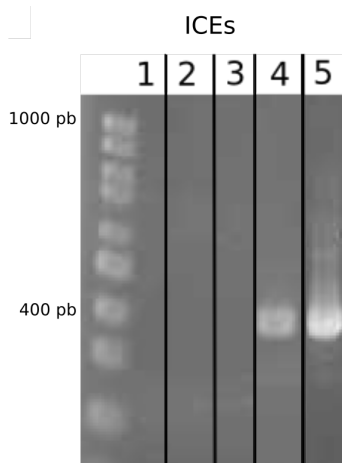


FIGURE 3.5 – Agarose gel analyses of fragments of circular ICE form after PCR. ICE excision was induced via SOS-response using mitomycin 1 mg/l. The ICE numbers are indicated.

Four genomic islands (GIs) are ST395-specific. The large genome of DHS01 (7.1 Mpb) and its low GC content suggested the presence of numerous horizontally acquired GIs. We retrieved 48 GIs in the genome of DHS01 (Table 3.5), among which 8 GIs were specific to DHS01 since not retrieved in any other strain of *P. aeruginosa* of the dataset 1 (Table 3.4). Two of these specific GIs shared identity with sequences found in environmental proteobacteria species (*Pseudomonas putida* and *Acidithiobacillus ferrivorans*). Four of the eight DHS01-specific GIs (GI-4, GI-5, GI-6, and GI-7, Fig. 3.6) were also found in all the 6 ST395 *P. aeruginosa* strains tested (dataset 2), possibly identifying these 4 GIs as markers of the spreading capability of *P. aeruginosa* ST395 [28]. In these four ST395-specific GIs, we specifically searched elements that could help the implantation and the spread of this international clone in hospitals. GI-4 harbored genes involved in the metabolism of choline, which is present in large amount in the outer leaflet of eukaryotic plasma membranes [156]. GI-6 harbored a gene encoding a protein that shares 58% of identity with the heat-labile enterotoxin (LT) of *E. coli* which production favors the colonization of the small intestine by enterotoxigenic *E. coli* [157] (Table 3.4). GI-7 contained an array of 6 genes encoding copper transporters : 2 cupredoxins, a copper ABC transporter ATPase, CopZ, CopA, and CopB (Fig. 3.6).

Cupredoxins are involved in copper homeostasis, as reported for CopC in *Pseudomonas syringae* [158]. The transporter CopZ is a copper resistance determinant in the extremophile *Acidithiobacillus ferrooxidans* [159]. CopA and CopB are sufficient to render the cells copper resistant [160]. Hence, CopA is an inducible efflux P-type ATPase located in the inner cell membrane that export copper from the periplasm. It also shares homology with multicopper oxidases that convert Cu^+ into the less toxic Cu^{2+} . Therefore, CopA has been thought to play a role in copper resistance through its multicopper oxidase activity and its high copper-binding capacity [161]. The gene *copB* encodes an inducible heavy metal ATPase that extrudes copper from the cell when it reaches intracellular toxic concentrations [162]. The source of the DHS01 within the University Hospital of Besançon is still under investiga-

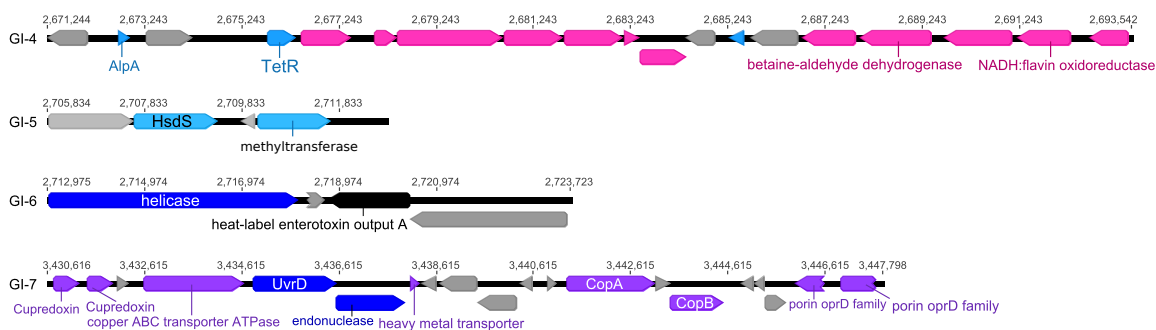


FIGURE 3.6 – Structure of the four GIs specific to ST395 clone. Only relevant genes are cited.

tion, however, the role of a water network made of copper pipes has been recently suggested in an ST395 outbreak [28]. We then hypothesized that the bacterial resistance to copper helped the installation of ST395 in the hospital water network and assessed the survival of a panel of ST395 isolates in copper-containing media. All the 7 ST395 tested behaved similarly, with 4.10^3 - 8.10^5 CFU surviving after 48 h of incubation in Mueller Hinton broth containing 7.5 mM CuSO_4 (Fig. 3.7).

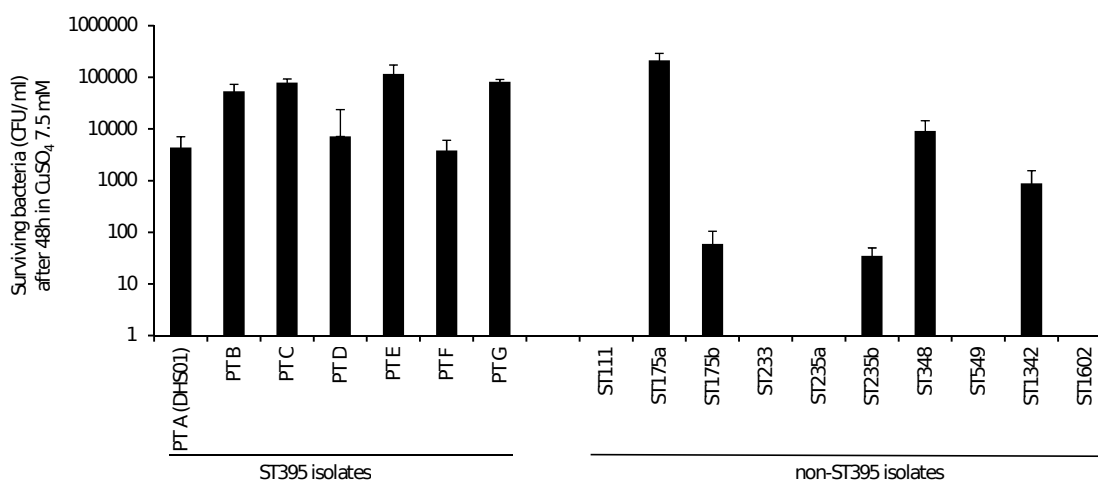


FIGURE 3.7 – Survival of ST395 and non-ST395 isolates of *P. aeruginosa* in copper-containing media. The resistance to copper was determined in cation-adjusted Mueller-Hinton broth supplemented with 7.5 mM of CuSO_4 . Strains were described in dataset 2 (Table 3.1). Ratio of means \pm SEM from at least 3 independent experiments.

This contrasted with the variations of survival of the 10 non-ST395 tested. Hence, 5 isolates (belonging to ST175, ST235, ST342, and ST348) could survive in copper solution while 5 other isolates (belonging to ST111, ST233, ST235, ST549, and ST1602) did not. The ST395 survival in copper is presumably due to the presence of *copA* and *copB* in all ST395 isolates but other mechanisms are implicated in copper resistance in non-ST395 strains. As expected, GI-7 containing *copA* and *copB* genes was also found in environmental *P. aeruginosa* ST395 strain isolated from a metal-contaminated aquatic environment [27]. GI-7 was retrie-

ved (with 99% of identity and 93% of coverage) in a strain of *Pseudomonas putida* which typically inhabits soil and water [163]. More generally, genes *copA* and *copB* are found in soil, commensal and plant pathogenic bacterial species and may be favored by the long term use of spray of copper-based bactericides on crops [164]. ST395 can contaminate hospitalized patients and those afflicted with cystic fibrosis [132, 28, 165] but is also frequently retrieved in polluted environments (estuaries, rivers, hospital wastewaters, water treatment plants) in which copper resistance could have been favored [27, 165, 26, 21, 166]. Our data indicate that an ST395 ancestor has horizontally acquired GI-7 from a single environmental source.

ST395-specific CRISPR-Cas system. CRISPR-Cas system is thought to play a role in controlling horizontal gene transfer [148]. The high number of GIs integrated into the genome of DHS01 led us to hypothesize the absence of active CRISPR-Cas system in this epidemic strain. However, CRISPR-Finder detected two CRISPR arrays and the associated Cas proteins which correspond to the type I-E (Fig. 3.8).

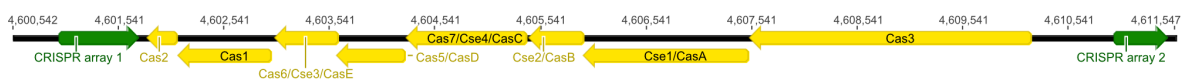


FIGURE 3.8 – Structure of the two type I-E CRISPR arrays and Cas proteins associated in the genomes of *P. aeruginosa* ST395 isolates.

It appears to be specific of the ST395 sublineage since retrieved from all the ST395 isolates tested (dataset 2) and absent from non-ST395 (dataset 1 and 2) [148]. Table 3.6 details the 12 and 8 spacers of the CRISPR arrays 1 and 2, respectively. Interestingly, the phage JBD25 was among the targets of one spacer of the CRISPR-system while genes of this phage were integrated into the GI-19 (Table 3.5). The GC content of JBD25 genes (62.5 %) was much lower than that of the rest of the genome (65.8 %) and the prophage nucleotidic sequence was highly similar (99 % of identity) with that of the JDB25 phage. Such an integration was also found in all the other ST395, excepted in isolate PT B. This suggests a recent integration of JDB25 phage despite a dedicated CRISPR-Cas spacer, and that the system may be inactive. However, no anti-CRISPR protein I-E were retrieved in the DHS01 strain, which suggests that this system was inactivated by another mechanism, implicating possibly phage-encoded anti-CRISPR proteins [167, 168].

Conclusion. We describe here the large genome (7.1 Mbp) of the first isolate of the international clone of *P. aeruginosa* ST395 that further spread into our University hospital during at least 11 years. The majority of the 1,555 accessory genes were brought by the many GIs (n=48) and ICEs (n=5) integrated into the DHS01 chromosome. The multidrug resistance of the isolate was mostly due to mutations in antibiotic targets and in regulators that trigger the overproduction of cephalosporinase and efflux pumps. DHS01 was retrieved from a chronically-colonized patient and displayed genomic signatures of adaptation to chronic infection mostly due to the deletion of a 131-kbp chromosomal fragment. Four GIs were seemingly specific of the widespread clone ST395. They harbored genes involved in metabolism, in virulence, and in the resistance to copper which was confirmed phenotypically in a panel of independent ST395 isolates. Resistance to copper possibly accounts for the spreading capability of ST395 in hospital settings in which water network may be incriminated [28]. This

CRISPR array	Spacer number	Phage name	GenBank accession number
1	3	<i>Pseudomonas</i> phage vB_PaeP_Tr60_Ab31	HG798806.1
	5	Bacteriophage D3112	AY394005.1
		<i>Pseudomonas</i> phage H70	M233689.1
		<i>Pseudomonas</i> phage PaMx73	JQ067085.2
		<i>Pseudomonas</i> phage JBD24	JX434031.1
<i>Pseudomonas</i> phage JBD25		JX495042.1	
6	8	<i>Pseudomonas</i> phage JBD26	JN811560.1
		<i>Pseudomonas</i> phage DMS3	DQ631426.1
		Bacteriophage D3112	AY394005.1
		<i>Pseudomonas</i> phage DMS3	DQ631426.1
2	6	<i>Pseudomonas</i> phage vB_PaeS_PAO1_Ab30	LN610590.1
		<i>Pseudomonas</i> phage JBD24	JX434031.1
		<i>Pseudomonas</i> phage MP38	EU272037.1
		<i>Pseudomonas</i> phage DO4	KM389210.1
2	2	Bacteriophage D3112	AY394005.1

TABLEAU 3.6 – Correlation between spacers in CRISPR and bacteriophages. Only spacers with correlation with known phages are cited in the table.

suggests that genes selected in environmental bacterial species by the use of heavy metals can be transferred to pathogens that further spread into hospital settings.

Supplementary material

Strain	Origin	Length (bp)	Number of genes	ST	NCBI accession number
Carb01-63	Clinical (unknown)	7497593	6910	111	NZ_CP011317.1
RIVM-EMC2982	Clinical (unknown)	7380063	6916	111	NZ_CP016955.1
F30658	Clinical (unknown)	7273258	6717	111	NZ_CP008857.1
NCGM257	Clinical (urine)	7090694	6712	357	NZ_AP014651.1
E6130952	Clinical (sputum)	7040952	6631	235	NZ_CP02603.1
PA7790	Clinical (tracheal aspirate)	7018690	6645	277	NZ_CP014999.1
PA11803	Clinical (blood-stream)	7006578	6637	277	NZ_CP015003.1
PASGNM699	Clinical (sputum)	6985102	6545	308	NZ_CP020704.1
S86968	Clinical (unknown)	6934277	6399	155	NZ_CP008865.2

Strain	Origin	Length (bp)	Number of genes	ST	NCBI accession number
PA8281	Clinical (tracheal aspirate)	6928736	6461	277	NZ_CP015002.1
M37351	Clinical (unknown)	6897231	6387	253	NZ_CP008863.1
IOMTU-133	Clinical (urinary catheter)	6897018	6338	1047	NZ_AP017302.1
W60856	Clinical (unknown)	6896700	6376	959	NZ_CP008864.2
PASGNDM345	Clinical (sputum)	6893164	6432	308	NZ_CP020703.1
FA-HZ1	Environmental (wastewater)	6866790	6209	27	NZ_CP017353.1
NCGM-1984	Clinical (urinary catheter)	6850954	6257	235	NZ_AP014646.1
H47921	Clinical (unknown)	6836415	6306	1105	NZ_CP008861.1
ATCC-27853	Clinical (unknown)	6827737	6376	155	NZ_CP015117.1
VRFPA04	Clinical (eye)	6818030	5841	823	NZ_CP008739.1
NCGM-1900	Clinical (urinary catheter)	6814936	6206	235	NZ_AP014622.1
W16407	Clinical (unknown)	6808844	6234	244	NZ_CP008869.2
T38079	Clinical (unknown)	6795741	6176	155	NZ_CP008866.2
F9670	Clinical (unknown)	6794354	6028	155	NZ_CP008873.1
W36662	Clinical (unknown)	6791731	6286	17	NZ_CP008870.2
W45909	Clinical (unknown)	6777566	6299	27	NZ_CP008871.2
NCGM2-S1	Clinical (urine)	6764661	6140	235	NC_017549.1
SCV20265	Clinical (cystic fibrosis sputum)	6725183	6131	299	NC_023149.1
B10W	Environmental (wastewater)	6723378	6216	308	NZ_CP017969.1
PA1088	Clinical (urine)	6721480	6247	277	NZ_CP015001.1
FRD1	Clinical (cystic fibrosis sputum)	6712339	5871	111	NZ_CP010555.1
PA-D25	Clinical (sputum)	6682304	6143	1971	NZ_CP012584.1
PA-D5	Clinical (sputum)	6681992	6153	1971	NZ_CP012579.1
PA-D22	Clinical (sputum)	6681981	6146	1971	NZ_CP012583.1
PA-D16	Clinical (sputum)	6681975	6150	1971	NZ_CP012581.1
PA-D9	Clinical (sputum)	6645477	6119	1971	NZ_CP012580.1
PA-D1	Clinical (sputum)	6643823	6120	1971	NZ_CP012585.1
PA-D2	Clinical (sputum)	6642996	6120	1971	NZ_CP012578.1
DN1	Environmental (soil)	6641902	6535	316	NZ_CP017099.1
PA-D21	Clinical (sputum)	6639108	6116	1971	NZ_CP012582.1
F63912	Clinical (unknown)	6618768	6186	198	NZ_CP008858.2
8380	Clinical (gut)	6613260	6150	Unknown	NZ_AP014839.1
F22031	Clinical (unknown)	6603099	5968	485	NZ_CP007399.1

Strain	Origin	Length (bp)	Number of genes	ST	NCBI accession number
LESB58	Clinical (cystic fibrosis sputum)	6601757	6005	146	NC_011770.1
PA7	Clinical (unknown)	6588339	5932	1195	NC_009656.1
MTB-1	Environmental	6580038	5953	Unknown	NC_023019.1
H27930	Clinical (unknown)	6568228	5982	389	NZ_CP008860.1
LES431	Clinical (unknown)	6550070	5925	146	NC_023066.1
PA14OR	Unknown	6541482	6038	253	NZ_LT608330.1
PA14	Clinical (burn)	6537648	5873	253	NC_008463.1
F23197	Clinical (unknown)	6517340	5941	1295	NZ_CP008856.2
PAER4-119	Unknown	6504659	5984	260	NZ_CP013113.1
PA1RG	Clinical (unknown)	6500439	5917	782	NZ_CP012679.1
PA1	Clinical (respiratory tract infection)	6498072	5910	782	NC_022808.2
M1608	Clinical (unknown)	6460023	5934	253	NZ_CP008862.2
T63266	Clinical (unknown)	6456866	5925	132	NZ_CP008868.1
PA-154197	Unknown	6445239	5961	550	NZ_CP014866
YL84	Environmental	6433441	5785	(CC169)	NZ_CP007147.1
12-4-4-59	Clinical (blood culture of burn patient)	6431911	5912	152	NZ_CP013696.1
PA121617	Clinical (sputum)	6430493	6367	389	NZ_CP016214.1
B136-33	Clinical (unknown)	6421010	5771	1024	NC_020912.1
DK2	Clinical (cystic fibrosis sputum)	6402658	5788	386	NC_018080.1
VA-134	Clinical (skin wound of burn patient)	6400418	5850	Unknown	NZ_CP013245.1
N17-1	Environmental (soil)	6370730	5899	2362	NZ_CP014948.1
F9676	Environmental	6368008	5721	167	NZ_CP012066.1
USDA-ARS-USMARC-41639	Nasopharyngeal swab of Bos Taurus	6364583	5856	Unknown	NZ_CP013989.1
X78812	Clinical (unknown)	6348761	5807	257	NZ_CP008872.2
AES-1R	Clinical (cystic fibrosis sputum)	6343337	5888	649	NZ_CP013680.1
RP73	Clinical (cystic fibrosis sputum)	6342034	5724	198	NC_021577.1
H5708	Clinical (unknown)	6334378	5829	Unknown	NZ_CP008859.2
M18	Environmental	6327754	5717	1239	NC_017548.1
T52373	Clinical (unknown)	6322459	5691	Unknown	NZ_CP008867.1
DSM-50071	Clinical (unknown)	6317050	5709	Unknown	NZ_CP012001.1
NCTC10332	Unknown	6316979	5761	Unknown	NZ_LN831024.1
PA1R	Clinical (unknown)	6309305	5752	782	NC_022806.1
PcyII-10	Unknown	6288645	5770	1233	NZ_LT673656.1

Strain	Origin	Length (bp)	Number of genes	ST	NCBI accession number
PAO1OR	Orsay variant of PAO1 strain	6276469	5797	549	NZ_LN871187.1
ATCC-15692	Clinical (wound)	6276434	5782	549	NZ_CP017149.1
PAO1	Clinical (wound)	6264404	5572	549	NC_002516.2
SJTD-1	Environmental (soil)	6243825	5739	Unknown	NZ_CP015877.1
Nhmuc	Unknown	6213276	5708	387	NZ_CP013479.1
SCVJan	Murine model	6213029	5704	387	NZ_CP013478.1
SCVFeb	Murine model	6213026	5704	387	NZ_CP013477.1

TABLEAU 3.7 – Characteristics of the 82 complete genomes extracted from NCBI, dataset 1 (June 2017).

Chapitre 4

Évolution génotypique et phénotypique d'un clone de *P. aeruginosa* au cours d'une épidémie hospitalière de 11 ans

4.1 Objectifs et méthodologie

Bien que DHS01 ne soit probablement pas le cas index de l'épidémie puisqu'il s'agit d'une souche déjà adaptée au patient, elle sera utilisée comme référence pour la suite des analyses. Toutes les informations relatives à DHS01 sont citées dans l'article correspondant.

Afin de comprendre l'origine du clone, nous avons établi la phylogénie représentant la proximité entre les différents isolats épidémiques. De plus, les dates de prélèvement ont permis d'estimer la date du plus proche ancêtre commun, celui de l'épidémie et celui de l'apparition du clone ST395 à l'hôpital. Puis, afin d'identifier l'évolution et la disparition de ce clone, les modifications génomiques telles que le gain et la perte de blocs de gènes mais aussi les déplacements inter-chromosomiques ont été étudiés. Pour finir, nous avons analysé les SNPs pour identifier les fonctions biologiques sélectionnées positivement (diffusion de caractères avantageux) et négativement (inhibition de la propagation de caractères nuisibles) au cours de l'épidémie.

4.2 Méthodes

Seules les méthodes non décrites dans l'article seront détaillées ici.

4.2.1 Choix de l'outil d'assemblage

Principe

Un bon assemblage permet d'obtenir un génome le plus proche possible de la réalité avec peu de contigs (un seul dans le cas idéal). Lors d'un mauvais assemblage, il est possible que des portions de génomes soient mal positionnées et que des gènes soient absents ou tronqués. Il est

donc important de trouver le meilleur logiciel et les meilleurs paramètres pour un assemblage optimal.

Ici, trois logiciels d'assemblage que sont Ray [115], MIRA [116] et SPAdes [117] sont comparés. Le génome de DHS01 qui sert de référence est complet et circulaire. Il a été re-séquencé en Illumina NextSeq, de la même manière que les autres génomes. Les différents assemblages de DHS01 générés par les logiciels et paramètres testés seront comparés au génome complet PacBio. Le meilleur assembleur et les paramètres associés seront utilisés pour l'assemblage des 53 autres génomes.

Choix des paramètres

Les données de départ possédant une profondeur d'environ 200X, le premier paramètre qui va être testé est l'effet de la profondeur de séquençage sur l'assemblage. Pour cela, des sous-échantillonnages à 60X, 80X et 100X vont être testés. En effet, une trop faible couverture ne permet pas d'assembler les génomes alors qu'une trop forte couverture rend difficile l'assemblage avec des temps de calculs longs, beaucoup de mémoire requise, des fichiers volumineux créés, plus d'erreurs générées qui rendent les graphes (utilisés par les logiciels d'assemblage) difficiles à résoudre.

Le logiciel SPAdes inclut un module de correction des *reads* (automatique pour ce logiciel) dont l'efficacité sera testée pour les deux autres logiciels.

Un dernier paramètre est spécifique à Ray, et correspond à une taille de k-mer (mots de longueurs k correspondant à k paires de bases) utilisée pour l'assemblage. Trois tailles de k-mer différentes (61, 77 et 101 bp) seront testées. Au total, 3 tests seront faits avec SPAdes, 6 avec MIRA et 18 avec Ray.

Outils de comparaison

QUAST [169] est utilisé pour évaluer les assemblages. Il compare la taille et le nombre de contigs, la présence de contigs mal assemblés, le nombre de mauvais assemblages et d'indels pour 100 bp, le nombre de contigs non alignés sur la référence (contigs créés par l'assemblage mais n'existant pas réellement) et le nombre de gènes de la référence retrouvés intacts (complets sur un contig unique).

Résultats des tests

Les résultats de comparaison obtenus par QUAST sont disponibles dans le tableau 4.1).

On observe une faible proportion de longs contigs (≥ 50000 bp), mais représentant une grande part du génome assemblé par Ray et SPAdes, alors que les plus longs contigs de MIRA correspondent à moins de la moitié du génome assemblé. De plus, MIRA génère plus de contigs, pour la plupart de taille inférieure à 25000 bp.

Lors de l'assemblage par Ray, plusieurs contigs sont mal assemblés, presque 1 Mbp pour le sous-échantillonnage à 100X, entre 350 kbp et 1 Mbp pour celui à 80X et entre 299 kbp et 640 kbp pour celui à 60X. En revanche, aucun contig mal assemblé n'est retrouvé dans les assemblages générés par SPAdes. On retrouve par contre pour ce dernier, une portion du

	# congs (0bp)	# congs (2500bp)	# congs (5000bp)	Longueur (0bp)	Longueur (2500bp)	Longueur (5000bp)	# congs mal assemblés	Longueur des congs mal assemblés	Mis hors congs	# congs non alignés	# bases non alignées	Ratio de décalage	# méappariement pour 100 kbp	# indels pour 100 kbp	# gènes	
mira_100X	867	92	41	7,24	5,74	3,90	43	80516	220089	60954	10 + 73 part	7660	1,032	4,25	0,39	6313 + 203 part
mira_100X_corr	1124	92	33	7,30	4,84	2,66	78	107130	205982	40003	19 + 100 part	17068	1,039	4,74	0,39	6219 + 296 part
mira_60X	716	103	41	7,18	5,57	3,30	23	46425	131669	49552	5 + 49 part	5603	1,024	3,99	0,43	6294 + 221 part
mira_60X_corr	892	100	35	7,22	4,83	2,56	36	21047	131765	35429	10 + 65 part	10364	1,028	3,81	0,43	6226 + 287 part
mira_80X	828	87	45	7,22	5,38	3,84	36	87089	240499	58916	11 + 51 part	12050	1,029	3,43	0,40	6289 + 226 part
mira_80X_corr	1025	104	33	7,26	5,12	2,59	42	56288	158831	39041	15 + 70 part	11004	1,035	4,43	0,59	6245 + 269 part
ray_100X_101	157	48	40	7,48	7,33	7,02	7	968281	529667	234761	77 + 0 part	10880	1,082	2,11	0,38	6372 + 33 part
ray_100X_61	157	48	40	7,48	7,33	7,02	7	968281	529667	234761	77 + 0 part	10880	1,082	2,11	0,38	6372 + 33 part
ray_100X_77	157	48	40	7,48	7,33	7,02	7	968281	529667	234761	77 + 0 part	10880	1,082	2,11	0,38	6372 + 33 part
ray_100X_101_corr	156	44	38	7,23	7,13	6,89	5	906134	489072	232965	84 + 0 part	11964	1,058	3,21	0,26	6307 + 36 part
ray_100X_61_corr	156	44	38	7,23	7,13	6,89	5	906134	489072	232965	84 + 0 part	11964	1,058	3,21	0,26	6307 + 36 part
ray_100X_77_corr	156	44	38	7,23	7,13	6,89	5	906134	489072	232965	84 + 0 part	11964	1,058	3,21	0,26	6307 + 36 part
ray_60X_101	132	66	47	7,20	6,83	6,21	3	299937	365426	126053	25 + 0 part	3527	1,025	1,71	0,17	6464 + 58 part
ray_60X_61	132	66	47	7,20	6,83	6,21	3	299937	365426	126053	25 + 0 part	3527	1,025	1,71	0,17	6464 + 58 part
ray_60X_77	132	66	47	7,20	6,83	6,21	3	299937	365426	126053	25 + 0 part	3527	1,025	1,71	0,17	6464 + 58 part
ray_80X_101_corr	127	59	47	7,12	6,85	6,40	4	643332	404825	130480	29 + 0 part	4110	1,014	3,21	0,27	6464 + 49 part
ray_80X_61_corr	126	59	47	7,12	6,85	6,40	3	635947	404825	130480	29 + 0 part	4110	1,015	3,21	0,27	6464 + 49 part
ray_80X_77_corr	126	59	47	7,12	6,85	6,40	2	549156	404825	130480	29 + 0 part	4110	1,014	3,18	0,27	6465 + 49 part
ray_80X_101	151	62	44	7,16	6,96	6,30	6	396680	370255	130745	53 + 0 part	7509	1,026	2,87	0,49	6433 + 46 part
ray_80X_61	149	61	44	7,13	6,83	6,30	4	360386	370255	130738	53 + 0 part	7509	1,021	2,81	0,49	6433 + 46 part
ray_80X_77	151	62	44	7,16	6,96	6,30	5	387763	370255	130745	53 + 0 part	7509	1,025	2,81	0,49	6433 + 46 part
ray_80X_101_corr	149	55	42	7,24	7,06	6,55	6	1055447	370259	145296	59 + 0 part	8290	1,037	3,42	0,21	6437 + 52 part
ray_80X_61_corr	151	56	42	7,12	6,93	6,38	6	882598	370259	145296	59 + 0 part	8290	1,019	3,27	0,20	6429 + 54 part
ray_80X_77_corr	152	56	42	7,29	7,08	6,57	6	814845	370259	145296	59 + 0 part	8290	1,044	3,15	0,16	6436 + 53 part
spades_100X	293	37	32	7,05	6,78	6,41	0	0	716770	306671	187 + 3 part	44776	1,001	2,02	0,23	6445 + 52 part
spades_60X	197	47	37	7,02	6,81	6,41	0	0	423370	218796	86 + 5 part	20231	1,001	2,70	0,20	6447 + 51 part
spades_80X	242	39	33	7,04	6,82	6,58	0	0	487556	263288	143 + 3 part	35208	1,001	2,92	0,24	6448 + 50 part

TABLEAU 4.1 – Optimisation de paramètres d’assemblage des *reads* Illumina. Les noms de lignes réfèrent aux noms du logiciel suivi de la profondeur (60X, 80X ou 100X), puis s’il y a correction des *reads* par *corr* et enfin, pour Ray de la valeur de k-mer.

génom non aligné sur la référence.

C'est donc SPAdes qui est utilisé pour l'assemblage des 53 génomes de *P. aeruginosa* ST395 épidémique. Le sous-échantillonnage choisi est 80X puisque cela permet d'avoir un compromis entre un bon N50 (correspondant à la taille du contig pour laquelle la somme de tous les contigs de taille supérieure représente au moins la moitié de la taille du génome), un nombre correct de gènes entiers détectés et un faible nombre de contigs non alignés sur la référence.

4.2.2 Construction d'une base de données

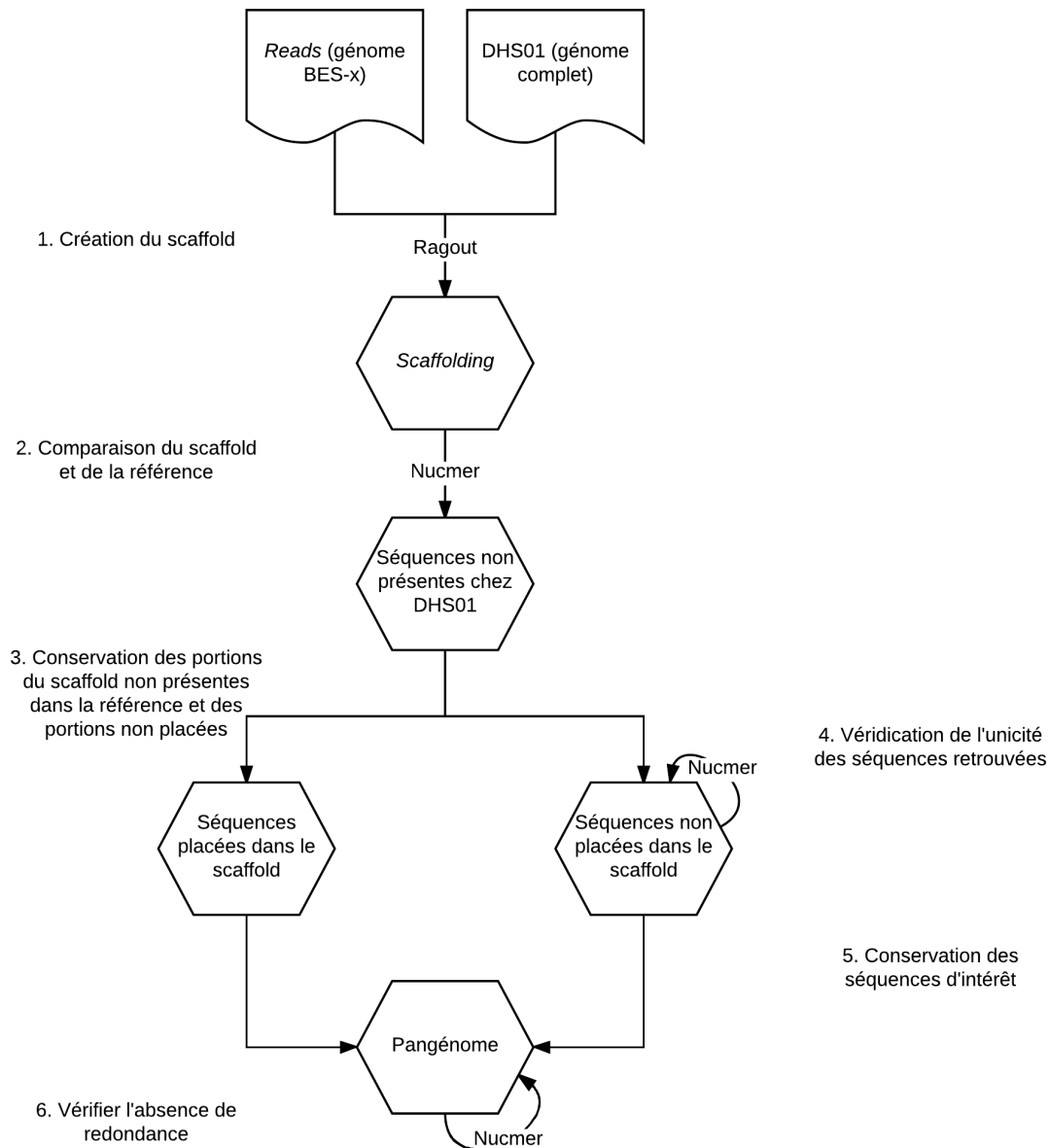


FIGURE 4.1 – Schéma des étapes utilisées pour la création du pangénomme.

Pour comprendre l'évolution de la bactérie au cours du temps, et pouvoir comparer les génomes, le pangénome de la collection a été créé. Cela correspond à l'ensemble des séquences présentes dans au moins un génome. En d'autres termes, c'est une base de données regroupant tous les gènes présents dans le clone au cours de l'épidémie. Ensuite, en utilisant cette base de données comme référence, il est possible de dire si chacun des gènes est présent ou non dans les différents génomes et de voir l'apparition ou la disparition de blocs de gènes au cours du temps. Il est également possible d'observer les mutations présentes chez chaque isolat.

Le pangénome va contenir à la fois le génome de DHS01 complet ainsi que toutes les portions venant des autres génomes mais n'appartenant pas à DHS01. Le logiciel Ragout [170] permet à partir des contigs d'un génome et d'une référence de faire du *scaffolding* (regroupement de contigs ordonnés séparés par des *gaps* de taille connue). Il va ensuite produire si possible un génome circulaire comportant des Ns. Les contigs non placés par rapport à la référence (donc non inclus dans le scaffold) seront dans un fichier séparé.

Les étapes 1 à 6 sont appliquées à chacun des génomes (figure 4.1) : (1) utilisation de Ragout pour obtenir le scaffold, (2) on compare le scaffold obtenu avec la référence (alignement du scaffold et de DHS01 avec nucmer [171]) et (3) on ne conserve que les portions non présentes dans la référence. Pour les séquences non placées, (4) on vérifie (à l'aide de nucmer) qu'elles n'existent pas déjà dans le génome de référence et si ce n'est pas le cas, (5) on les conserve. La dernière étape (6) est l'alignement du pangénome sur lui même (avec nucmer) pour vérifier qu'il n'y a pas de redondance.

4.3 Résultats : Article en rédaction (Août 2017)

Phylogenetic analysis of a *Pseudomonas aeruginosa* epidemic lineage reveals parallel evolutions associated with outbreak decline

Marie Petitjean^{1,2}, Paulo Juarez^{2,3}, Alexandre Meunier¹, Xavier Bertrand^{1,2}, Benoit Valot², and Didier Hocquet^{1,2,4}

¹Laboratoire d'Hygiène Hospitalière, Centre Hospitalier Régional Universitaire, Besançon, France

²UMR CNRS 6249, Chrono-environnement, Université de Bourgogne Franche-Comté, Besançon, France

³Laboratoire de Bactériologie, Centre Hospitalier Régional Universitaire, Besançon, France

⁴Centre de Ressources Biologiques - Filière Microbiologique de Besançon

Key words : *Pseudomonas aeruginosa*, high-risk clone, outbreak, WGS, parallel evolution

4.3.1 Abstract

Pseudomonas aeruginosa is an opportunistic pathogen responsible for outbreaks in hospital settings. ST395 is among the 'high-risk' clones that were the most widespread. An epidemic strain of *P. aeruginosa* ST395 was spreaded in University Hospital of Besançon between 1997 and 2008. The genome of 54 multidrug resistant isolate of this outbreak were analysed to identify the origin of the outbreak, the features that could have helped its implantation in our hospital and those associated with the end of the epidemics. Phylogenetic analysis indicated two parallel outbreaks and the ancestor of the ST395 clone possibly contaminated our hospital during its construction in 1979. SNPs and phenotypic data showed that late isolates displayed independent genomic signatures of chronic adaptation in patients, resulting in altered LPS, mutated OprD porin and extinction of MexAB-OprM overproduction. The parallel in host evolution during the spread of the pathogen along with infection control measures are associated with the decline of the outbreak.

4.3.2 Introduction

Pseudomonas aeruginosa is a Gram-negative bacteria that has the capacity to adapt to a wide range of environment. Its metabolic diversity and its ability to form biofilm makes it a remarkable opportunistic pathogen causing nosocomial infections mainly in critically ill and immunocompromised patients [130]. During outbreaks in hospital settings, transmission can occur by two principal ways to infect patients. First, water environment plays an important role in the transmission of *P. aeruginosa* [172, 173]. Second, patient-to-patient transmission is an important source of cross-transmission [174].

Among the panmictic population of *P. aeruginosa*, denominated 'high-risk' clones (e.g. ST111, ST235, ST175 and ST395) that were predominantly responsible for worldwide hospital outbreaks [11]. ST395 clone were responsible for hospital outbreaks in Hungary, France, Spain and United Kingdom [26, 132, 165, 28]. In the hospital of Birmingham (UK), this clone was thought to contaminate the burn unit via the water network [28]. ST395 clone has also been isolated in an aquatic environment polluted with high level of metals [27]. We have previously showed that the clone ST395 specifically harboured a genomic island with genes encoding 6 copper transporters, associated with the resistance to copper in solution [175].

Between 1997 and 2008, an epidemic strain of *P. aeruginosa* ST395 had colonized or infected more than 300 patients in the University Hospital of Besançon (France). We explored the genomic evolution of this epidemic strain with the analysis of the phylogenetic distribution of epidemic isolates, the identification of genetic elements that have moved during the outbreak (gain or loss of genomic content and movement of gene blocks). Then, specific mutations selected during the extinction of the outbreak were identified and linked to phenotypic data.

4.3.3 Material and Methods

Isolate collection and sequencing. Between May 1997 and April 2008, more than 300 patients hospitalized in the University Hospital of Besançon (France) have been contaminated

with the *P. aeruginosa* strain BES (BES stands for Besançon Epidemic Strain), with isolates sharing the same band profile after pulsed field gel electrophoresis (PFGE) [10, 128]. We selected 54 isolates (BES-1 to BES-54) evenly distributed over time for sequencing (Table 4.1 available in the supplementary material) and considered that they were representative of the outbreak. The first 27 isolates were isolated in the early stages of the outbreak (before 2001) and the others were considered as late isolates (Figure 4.2 and Table 4.1 available in the supplementary material). Bacterial DNAs were isolated from overnight cultures on Mueller-Hinton agar using the Genomic-tip kit (Qiagen) and paired-end sequenced with Illumina NextSeq at 2*150bp. Reads were subsampled before assembly to limit the coverage at 80X. The contigs were built with SPAdes [117]. Local and external outgroup isolates of *P. aeruginosa* of the same sequence type (ST) were used to root the phylogenetic tree (Table 4.1) [28]. We previously sequenced BES-1, the first isolate identified in May 1997, with the PacBio technology and used it as a reference [175]. STs were determined from the assembled genomes with MLST 1.8 tool [176] (Larsen et al. 2012).

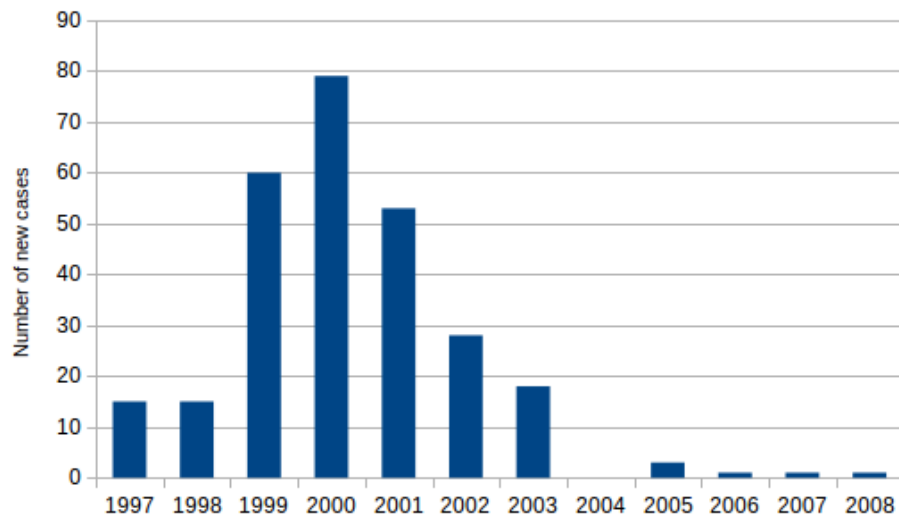


FIGURE 4.2 – Number of new cases by year during the eleven years of the outbreak.

Epidemic strain pangenome. We built the BES pangenome by combining BES-1 circular genome with non-redundant contigs found in all the other BES isolates. For this purpose, the assemblies of each BES isolates were aligned against BES-1 genome with Mummer [171]. Reads from all isolates were then aligned against the BES pangenome with BWA and used for further analysis [118].

Gene annotation and specificity, variant calling. Genes of the BES pangenome were detected using Prodigal and then annotated with BLAST [177, 145]. Resistance and virulence genes were searched with ResFinder and Virulence Factor DataBase, respectively [143, 144]. We tested the presence or absence of each gene with FeatureCounts [121]. SNPs were identified with freebayes with a minimum coverage of 10 reads per isolate and a quality of 30 [120]. Non synonymous SNPs and SNPs in promoting regions (RBS and sequence at -10 and -35) were further sorted. SNPs, gained or lost gene blocks were clustered in relation to the phy-

logenetic tree with MixOmics [178].

Phylogeny and Most Recent Common Ancestor (MRCA) calculation. For phylogenetic analysis, we concatenate the 265 core genes of BES isolates that harbored at least one SNP producing a 397,062-bp alignment with 285 SNPs. The maximum clade credibility tree was generated with MrBayes [125]. The isolation dates of BES and outgroup isolates were used as nodes calibration to calculate the age of the MRCA.

Insertion sequence. IS insertion events were detected using the panISa software [179]. The boundaries of each potential IS were searched into DHS01 genome to exclude chromosomal rearrangement. These boundaries were then searched into the BES genomes to assess the length of the potential IS. The presence of the potential IS was confirmed by PCR and sequencing with specific primers (5'-CCTCCATGTGACTACCCACG-3' and 5'-CGCGACTATGATCCGGAGAC-3').

RT-qPCR experiments. The expression of *mexB* was assessed in all BES isolates with RT-qPCR as described previously [103].

Phenotype determination. The susceptibility to antipseudomonal compounds was tested using agar diffusion as recommended by the EUCAST [142]. We determined the serotype of all isolates with specific antiserum (BioRad).

Ongoing experiments. Survival in distilled water, in the water of the distribution network, resistance to the dessication, analysis of the LPS (SDS-PAGE, composition in collaboration with LipSTIC facility), macrophage response to LPS (in collaboration with UMR1098, Besançon).

4.3.4 Results & discussion

Population structure and MRCA of P. aeruginosa BES. Phylogenetic analysis generated with 285 SNPs in the core genome showed that BES isolates fell into the 2 major clusters 1 and 2, with cluster 1 being divided into 4 sub-clusters 1a to 1d (Figure 4.3). There was no clear link between date of isolation or hospital ward and position in the phylogenetic tree. Although no environmental sampling were made during the outbreak, the cluster 1d in the phylogenetic tree probably gathered isolates from patients contaminated from a common environmental reservoir. Beside, stair-shaped branches (e.g. cluster 1c) indicated probable patient-to-patient transmission. Although the PFGE pattern of BES isolates was constant, WGS analysis revealed two parallel outbreaks occurring simultaneously in the same hospital (Table 4.1, available in the supplementary material). Bayesian analysis could date the putative common ancestors that link all the isolates of the outbreak and all the ST395 in our hospital. A common ST395 ancestor of BES isolates dated from 1990 (95% confidence interval between 1987 and 1993), 7 years before the first isolation in a patient in May 1997. With the addition of a non-BES ST395 outgroup (Table 4.1, available in the supplementary material), we estimated the emergence of the ST395 common ancestor in our hospital 19 years ago (in 1978, with a confidence interval at 95% between 1970 and 1985). Interestin-

gly, the common ancestor emerged at the same time as the beginning of the construction of the hospital. We had previously showed that ST395 specifically harboured a genomic island with 6 genes encoding copper transporters and constantly survived in copper solution after 48 hours [175]. The role of a water network (made of copper pipes) had been suggested in another ST395 outbreak with a similar hypothesis of contamination before the opening of the hospital [28]. All this suggests that the resistance to copper of ST395 could have favored its installation in the hospital.

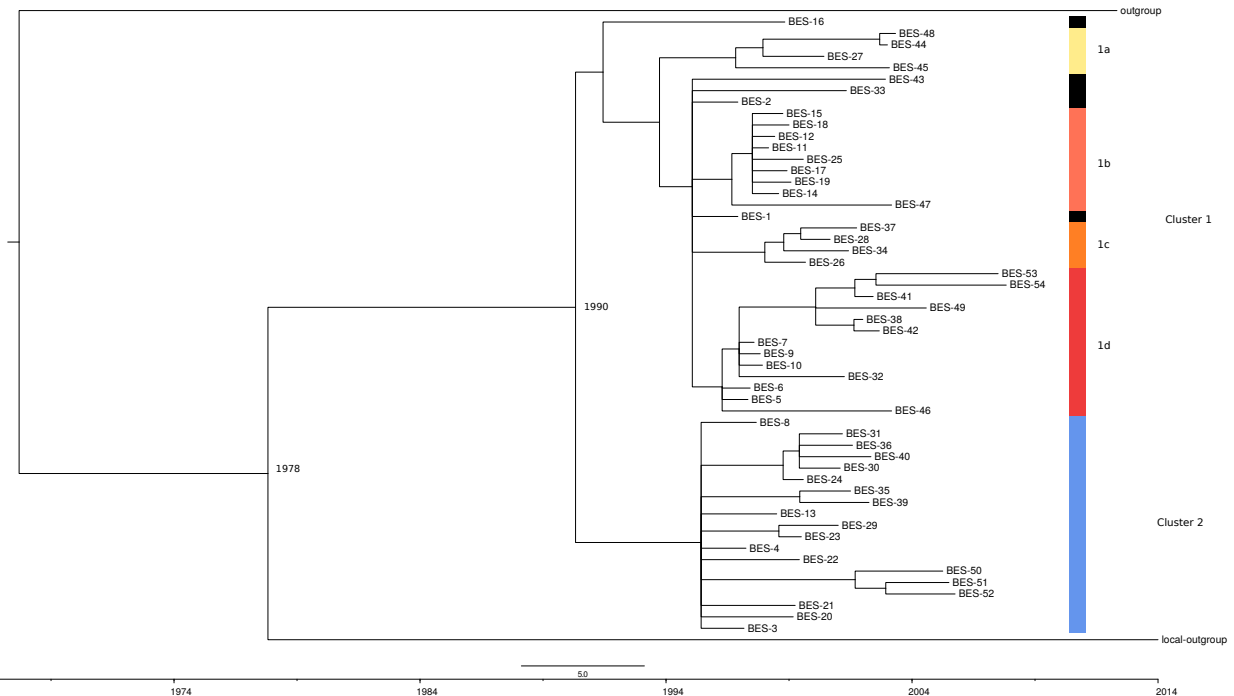


FIGURE 4.3 – Phylogenetic tree and temporal data of ST395 epidemic strain in Besançon University Hospital. The tree was performed with MrBayes using the following parameters : GTR substitution model, number of cycles for the MCMC algorithm is 1,000,000 with sample of the chain every 1,000 cycles. We defined 2 majors clusters 1 and 2, with cluster 1 divided into 4 sub-clusters cluster 1a to 1d. The isolates in black was considered out of cluster.

Gene block insertion or deletion during BES spread. *P. aeruginosa* can also evolve through the acquisition or loss of genomic fragments [56]. Accessory gene profiling shows clear differences in accessory gene content among these isolates with genes blocks specific of some clusters (Figure 4.4). Fifteen gene blocks (with size ranging from 5 to 278 kb) were gained or lost during the outbreak. The block 2 (38-kb in size, absent in the genomes of the outgroups) was gained by all the isolates of cluster 1 and by three isolates of cluster 2 (BES-29, BES-35, BES-39). Block 2 contains genes encoding hypothetical proteins and genes associated with phages. The isolate BES-52 underwent the deletion of the 278-kb block 3 that encompassed *hmgA*, which deletion lead to an easily recognizable brown colony phenotype due to homogentisate accumulation [180]. Block 11 was gained by 4 isolates of the cluster 2 (BES-23, BES-30, BES-31, and BES-40) and harbored the whole genome of the bacteriophage

PAN70 (GenBank Accession KJ959591.1). Interestingly, the distant isolates (BES-1 isolated in May 1997 and BES-28 isolates in February 2001) underwent a large deletion of 131-kb (block 12). This block included the genomic regions pa1246-pa1292 and pa1381-pa1450 in the reference strain PAO1 that encoded aprADEFI (alkaline protease system), biofilm dispersion regulator bdIA, quorum-sensing regulators lasI, lasR and rsaL, and the flagellar operon region fliKLMNOPQR [175].

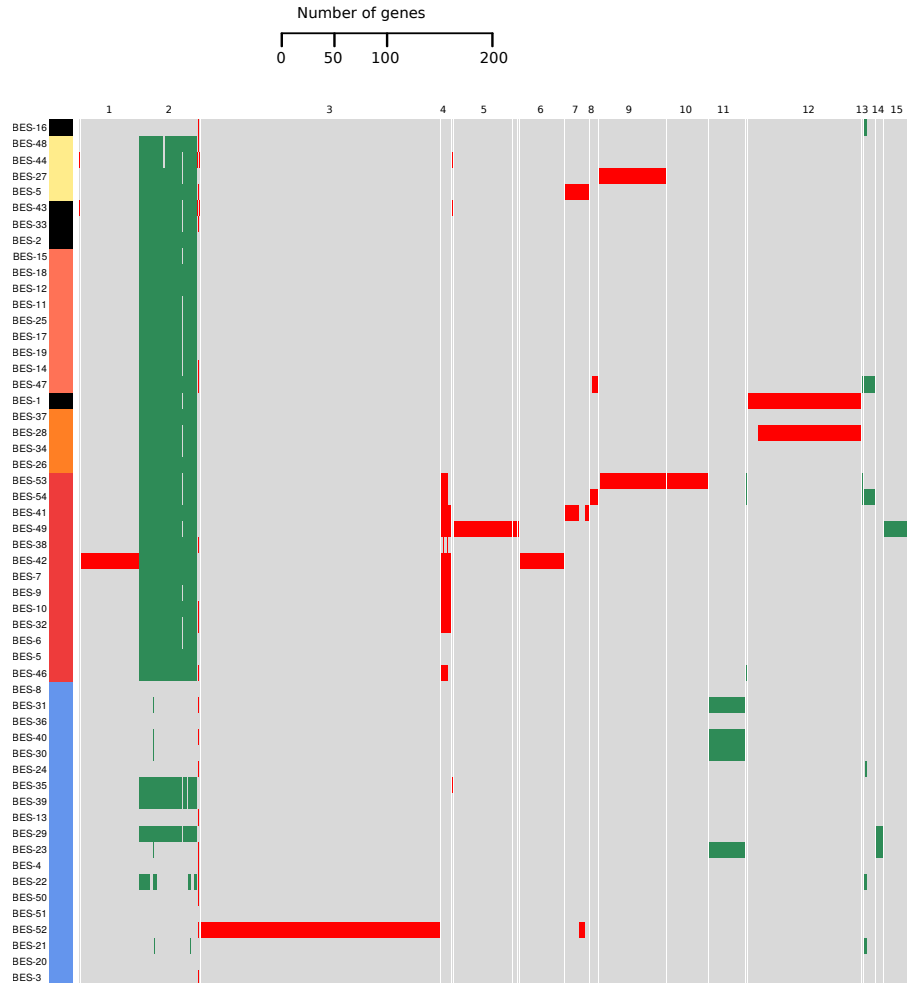


FIGURE 4.4 – Acquisition or loss of gene blocks during the spread of *P. aeruginosa* ST395 Besançon Epidemic Strain. The isolates are in the same order as in the Figure 4.3. Lost and gained gene blocks are indicated in red and green, respectively. FeatureCounts was set as follows : (i) threshold at 10 reads per 100 bp of length, (ii) read quality at 20, and (iii) at least 75 bp aligned on the annotated gene. Blocks-1, 2, 4, 5, 6, 7, 8, 9, 10, 13, 14, and 15, which were gained or lost by several strains, harbor a high number of genes coding for hypothetical proteins or associated with bacteriophages. Block-3 harbored the hmgA gene, the absence of which is responsible for brown mutant phenotype [180]. Block-11 harbor the whole genome of the bacteriophage PAN70 and Block-12 corresponds to the 131-kb specific block lost by DHS01 and described previously [175].

Despite a relatively high plasticity observed all along the outbreak, the genome size from all isolates was relatively constant suggesting that keeping the same size of genome could represent an evolutive advantage. Although the adaptation of *P. aeruginosa* in early stages of chronic adaptation was characterized by the loss of large genomic region [37], we found no loss of large genomic block at the beginning or at the end of the outbreak (Figure 4.5). Mobile genetic elements were the same as those described previously excepted for isolate BES-49 which integrated two copies of the ISPa74 [175].

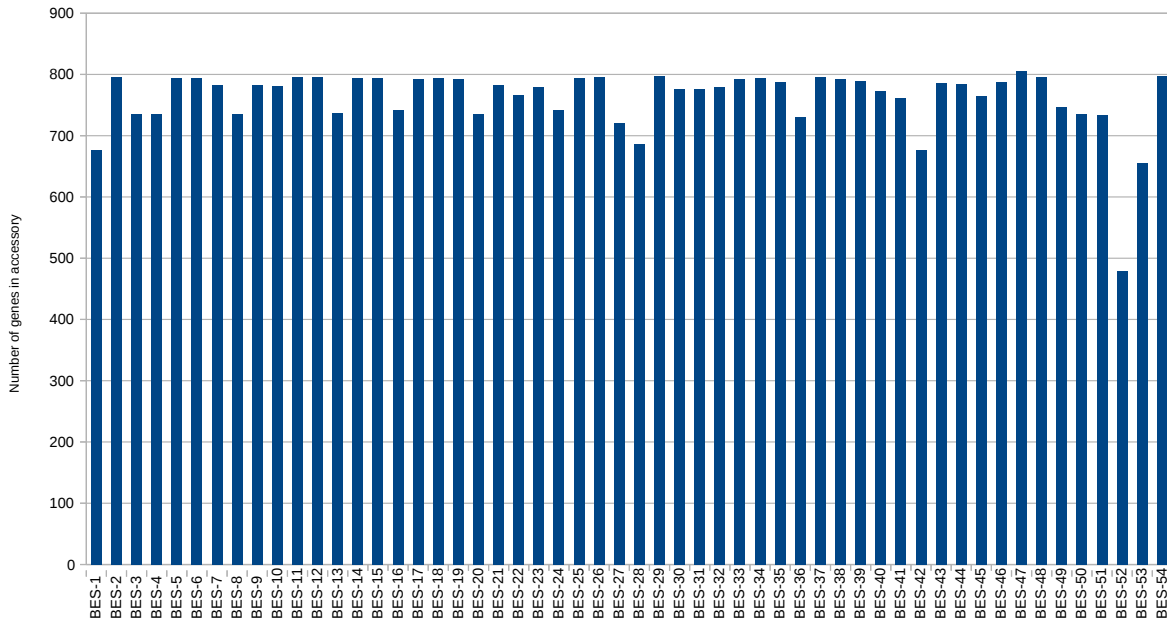


FIGURE 4.5 – Size of accessory genome in all the 54 strains involved in 11-year outbreak. Only genes in accessory genome were counted, excluding gene present in all genomes.

Chromosomal rearrangements. We also searched chromosomal inversions and rearrangements which may confer intra-population diversity, and found the rearrangements of 2 gene blocks. One 40-kb block moved out of its original position in the genomes of cluster 1 isolates (between *pa3825* and an integrase-encoding gene) to insert in the gene *pa5185* (encoding a thioesterase) in all the genomes of cluster 2 isolates. Its ancestral position is presumably that observed in cluster 1 and in the local ST395 outgroup which diverged in 1978. This rearrangement confirms our observation of two parallel outbreak in phylogenetic analysis (Figure 4.3). The same 40-kb block was inserted into *pa0813* (coding for hypothetical protein) in the genome of the BES-32. Another 67-kb block moved in *pa3879* (*narL*) in the genomes of BES-44 and BES-48. $\Delta narL$ mutant displays a hyper-swarming phenotype as a consequence of the overproduction of rhamnolipid, which decrease its ability to form a biofilm [181].

Biological functions selected during the 11-year spread of BES. We hypothesized that biological features have been selected or lost during the *P. aeruginosa* outbreak. The variant calling retrieved 458 SNPs across the BES population, among which 220 were non-synonymous or present in the promoting regions of genes (Figure 4.6). We searched the func-

tion of the modified proteins and found it for 55 proteins, corresponding to 78 mutations. The proportion of early (before 2001) and late (after 2000) isolates mutated were compared for each function. We saw a parallel evolution in late isolates, where four major functions were selected (chisq2 test ; $p < 0.05$) : amino-acid metabolism, resistance to carbapenems associated with *oprD* mutations, loss of overexpression of *mexAB-oprM* efflux operon and modification in LPS biosynthesis coding genes.

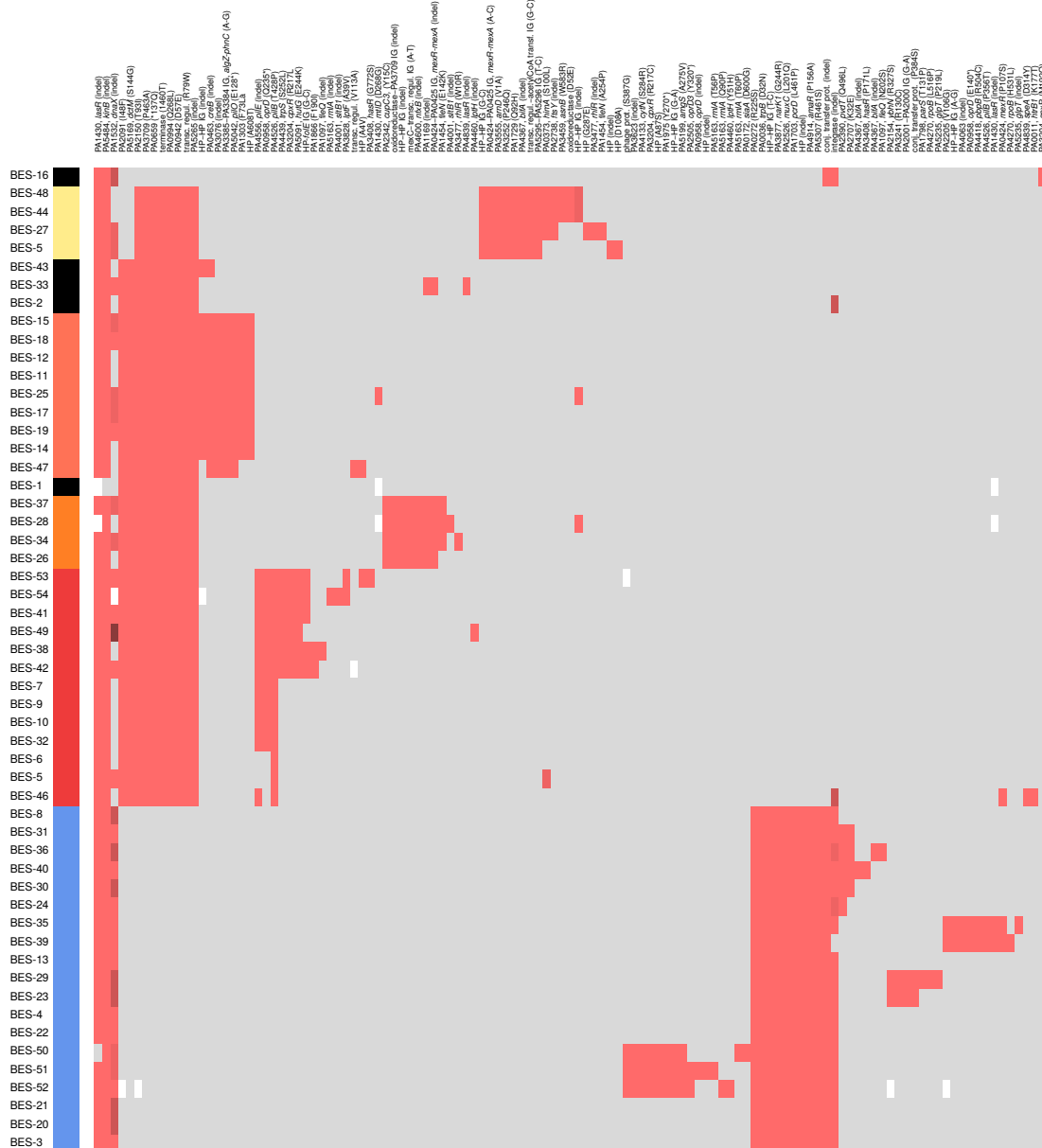


FIGURE 4.6 – SNPs variation related to phylogenetic clustering. The strains (in row) were in the same order as in the phylogenetic tree and colors of clusters were indicated.

Modification of amino-acid metabolism. Specific modifications occurred in amino-acid metabolism genes, particularly in late isolates (from 2001 to 2008). Two genes were mu-

tated in 6 isolates in the 1d cluster (BES-38, 41, 42, 49, 53 and 54) : *hutG* (Glu244Lys) and *trpS* (Ser252Leu). Mutations in *asnB* (Trp583Arg, in BES-44 and BES-48) and in *trpB* (Asp32Asn in BES-29, 30, 31, 35, 36, 39, 40, 50, 51 and 52) also occurs. Adaptation of *P. aeruginosa* during persistence in cystic fibrosis (CF) induced metabolism modifications. CF isolates were often auxotrophic for various amino acid [182]. Due to the cost of metabolization of each amino acid, the probability that the strain evolved towards a more efficient use of metabolites could account for the modification in the amino acid metabolism [183].

Evolution of resistance and selection of *oprD* mutations. All the isolates of the epidemic strain were multidrug resistant (high level of resistance to gentamicin, tobramycin, ciprofloxacin and to β -lactam compounds) mostly by chromosomal mutations. Early isolates displayed an identical resistance genotype than the index isolate BES-1 [175]. Briefly, they produced a 2"-aminoglycoside nucleotidyl-transferase, overproduced an extended-spectrum cephalosporinase AmpC, presumably overproduced the efflux pumps MexAB–OprM due to mutation in *nalB*, MexXY and MexEF–OprN, and displayed canonical mutations in the QRDRs of GyrA and ParC. No new resistance genes were acquired in BES genomes. However, chromosomal mutations occurred in *rpoB* leading to the resistance to rifampicin (His₅₃₁Leu and Leu₅₁₆Pro in BES-29 and BES-39, respectively) [184]. The gene *glpT* was mutated in isolates BES-29 (frameshift mutation) and BES-29 (Pro₂₁₉Leu) to probably lead to fosfomycin resistance [185, 186].

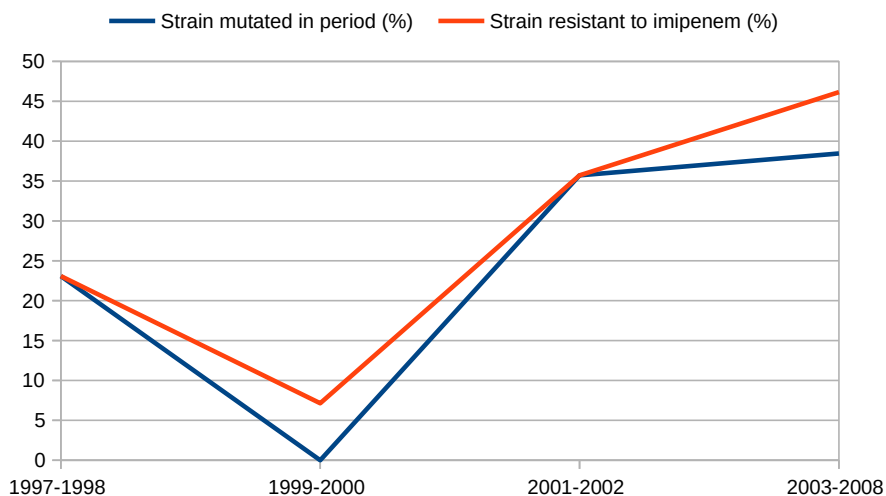


FIGURE 4.7 – Variation of mutated or deleted genes implicated in chronic infection.

We clustered the isolates by date of isolation (1997-1998, n=13; 1999-2000, n=14; 2001-2002, n=14 and 2003-2008, n=13). Blue curve were for genotype, percentage of strain carrying mutated *oprD* gene and red curve represent the phenotype of resistance to imipenem.

Three different mutations in *oprD* responsible for frameshift or codon stop apparition were retrieved in 3 early isolates (BES-7, 9 and 10) and 10 late isolates (BES-32, 35, 38, 39, 41, 42, 49, 51, 53 and 54). All mutations found in *oprD* were responsible for resistance to imipenem (single nucleotide change resulting in premature stop codon or insertion or deletion resulting

in frameshift) [187]. We found a significant statistical correlation ($p < 0.05$) between *oprD* mutations and imipenem resistance (Figure 4.7). We hypothesized that this represents a positive selection because selecting variant that improve survival under imipenem pressure [188].

Restoration of mexAB-OprM basal expression. *mexR*, *cpxR*, and *nalD* are among the genes that regulate the expression of the efflux operon mexAB-*oprM* [189, 59]. Early isolates overproduced MexAB-OprM compared to PAO1 reference strain as the result of the alteration of the repressor protein MexR (His₁₀₇Pro) [154]. In contrast, mutations in MexAB-OprM regulating regions occurred in 19 out of the 27 late isolates. Twelve BES isolates carried non-synonymous mutations in at least one of these genes (Pro₁₀₇Ser in *mexR*, Arg₂₁₇Leu, Arg₂₁₇Cys and Leu₁₉₂Gln in *cpxR*, and a 13-nt deletion in *nalD*). Moreover, 7 other isolates displayed mutated promoter region of *mexAB-oprM* operon (sequence at -35 and RBS regulatory regions). To verify the impact of these mutations on the *mexAB-oprM* expression, we measured the expression of *mexB* in all the BES isolates using RT-qPCR.

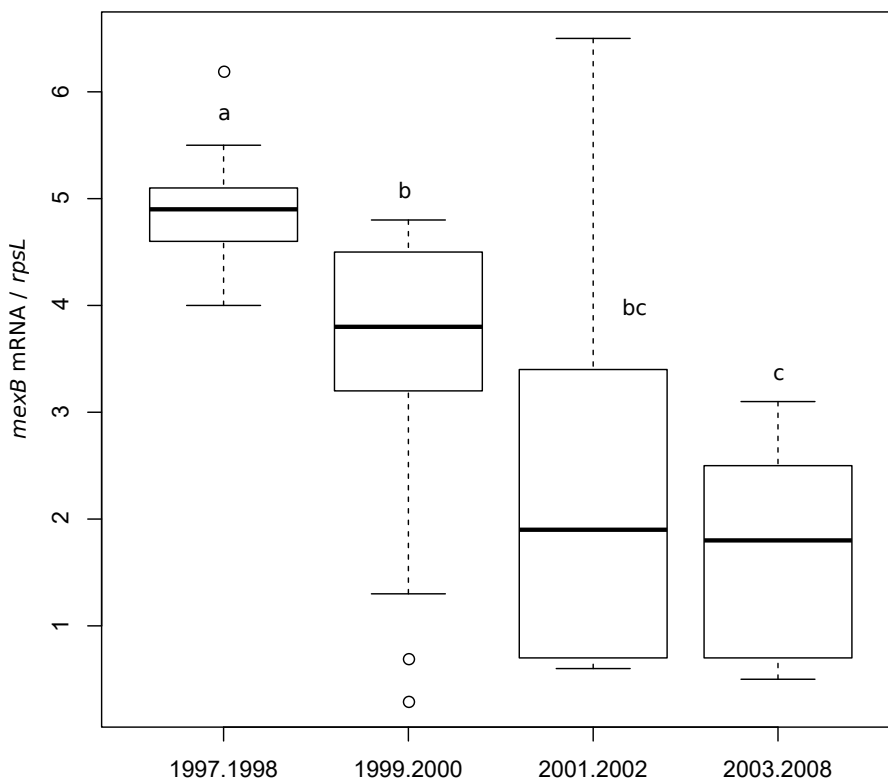


FIGURE 4.8 – Variation of mutated or deleted genes implicated in chronic infection. We clustered the isolates by date of isolation (1997-1998, $n=13$; 1999-2000, $n=14$; 2001-2002, $n=14$ and 2003-2008, $n=13$). Boxplot represent expression of *mexB* transcripts in RT-qPCR. Mean gene expression values were calculated from two independent bacterial cultures, each assayed in duplicate and normalized with the expression of *rpsL*. Strain PT629 was used as a positive control for the overexpression of gene *mexB* [190]. a, b and c represent statistical differences, $p < 0.05$

We confirmed that the majority (83%) of isolates with the above-mentioned genes or region identical to those of BES-1 mostly overexpressed *mexAB-oprM*. As expected, most (84%) isolates harbouring additional mutations in *mexAB-oprM* regulating regions retrieved a basal expression of this efflux pump (Figure 4.8). Hence, *mexAB-oprM* expression significantly decreased in late isolates when compared to early isolates.

MexAB-OprM plays an important role in virulence as the deletion of this efflux system from a wild-type isolate or a *nalD* mutant significantly reduced the bacterial invasiveness [66, 65]. Overproduction of MexAB-OprM also affects *P. aeruginosa* production of the quorum sensing regulated virulence factors pyocyanin, protease, and elastase. This indicates that quorum-sensing is restored when *mexAB-oprM* expression level returned to a basal level (like in PAO1) [63]. The overexpression of *mexAB-oprM* reduces the survival of *P. aeruginosa* in water or on a dry surface [71]. However, the restoration of basal expression of *mexAB-oprM* in late isolates could reduce the resistance level to the substrates of this pump [191]. This modulation could have masked the associated resistance mechanism. The overexpression of *mexAB-oprM* induced change in the anaerobic nitrate respiratory pathway and a decrease in intracellular pathway to compensate the loss of fitness due to overexpression [82]. During long-term colonization, *P. aeruginosa* population tend to abolish MexAB-OprM which means that the loss of this pump confer an advantage for survival in the hostile environment of CF airways [103]. It was difficult to conclude about positive or negative selection in this case. First, positive selection was linked to conservation of advantage like restored quorum-sensing, better survival in water or on a dry surface, while loss of overexpression inducing diminution of multidrug resistance seems to be link to negative selection.

Loss of LPS expression. The LPS biosynthesis pathway is encoded by many genes responsible for lipid A biosynthesis, O-antigen biosynthesis and transport [192]. Eleven isolates carried at least one mutation in the LPS biosynthesis pathway (four mutations in *rmlA*, frameshift mutation in BES-54, Gln₉₀Pro in BES-50, Thr₅₆Pro in BES-51 and Thr₆₀Pro in BES-52; *lptF*, Ala₃₉Val in BES-53 and BES-54; two mutations in *lptH*, frameshift mutation in BES-49 and Tyr₅₁His in BES-52; *htrB1*, Lys₁₇₇Thr in BES-46; and *arnD*, Val₁Ala in BES-27, 44, 45 and 48) from which 10 were late isolates. O-antigen serotyping confirmed that 7 of these isolates were polyagglutinable, meaning that these isolates were deficient in LPS like shown by Fomsgaard et al. [193], while all other isolates were of serotype O6 (Figure 4.9). Significant correlation ($p < 0.05$) was found between genetic mutations and serotype. During chronic lung infection in CF patients, synthesis of various lipid A structures alters host immune response and promote bacterial persistence [194]. We found here that LPS synthesis pathway was altered by independent evolutionary pathways in late isolates of the BES collection. A normal LPS synthesis is then negatively selected during the outbreak.

Fitness. Ongoing experiments. Main idea : the resistance in water and the resistance to desiccation (necessary for a efficient transmission of the pathogen) is altered in late isolates which displayed altered LPS and basal *mexAB-oprM* expression.

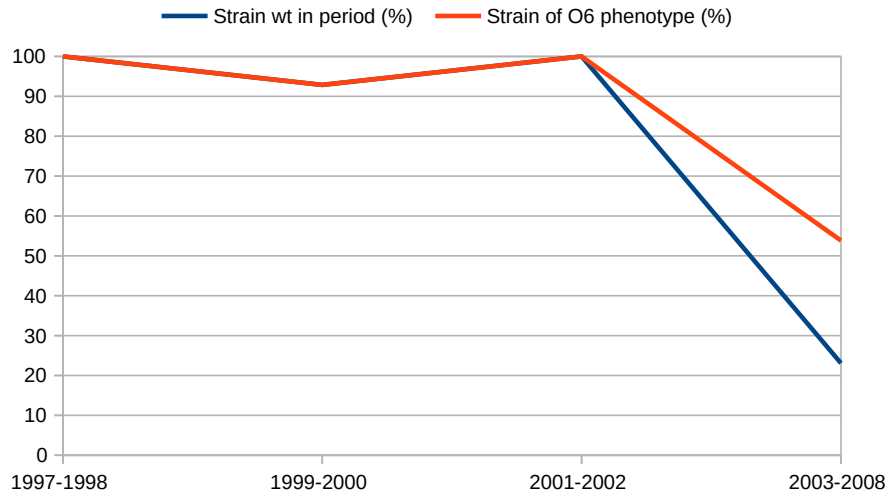


FIGURE 4.9 – Variation of mutated or deleted genes implicated in chronic infection. We clustered the isolates by date of isolation (1997-1998, n=13 ; 1999-2000, n=14 ; 2001-2002, n=14 and 2003-2008, n=13). The red curve indicates evolution of phenotype of LPS determined to O6 at the beginning of the outbreak, the blue curve corresponds to modification of genotype with the percentage of strain wild type for LPS genes (*rmlA*, *lptF*, *lptH*, *arnD* and *htrB1*).

Conclusion. Whole genome phylogenetic analysis of the outbreak reveals two independent parallel spreads in hospital with the installation of ST395 clone at the construction of the hospital. Isolates retrieved from patients before or during the peak of outbreak displayed similar genotypes and phenotypes. The prevalence of patients affected by the epidemic clone ST395 started to decrease in 2001, concomitantly with the parallel evolution of amino-acid metabolism pathways, loss of overexpression of the efflux pump operon *mexAB-oprM*, alteration of the LPS, and the emergence of carbapenem resistance. We hypothesize that the parallel adaptation of the ST395 isolates to patients decreased its resistance in the environment and thereby its patient to patient transmissibility. This, along with the implemented infection control measures, could have participated to the extinction of this epidemic pathogen in hospital setting.

Supplementary material

Isolate name (alternate name)	Isolation date	Isolation location	Hospitalisation unit ^a	Accession number (NCBI or ENA)
BES-1/DHS01	May 1997	Nose	SICU	NZ_CP013993.1
BES-2	May 1997	Tracheal aspirate	SICU	SRR5949766
BES-3	August 1997	Nose	SICU	SRR5949767
BES-4	September 1997	Rectum	SICU	SRR5949768
BES-5	October 1997	Rectum	SICU	SRR5949769

Isolate name (alternate name)	Isolation date	Isolation location	Hospitalisation unit ^a	Accession number (NCBI or ENA)
BES-6	November 1997	Tracheal aspirate	SICU	SRR5949774
BES-7	January 1998	Rectum	SICU	SRR5949777
BES-8	February 1998	Tracheal aspirate	SICU	SRR5949772
BES-9	April 1998	Drain	SICU	SRR5949773
BES-10	May 1998	Urinary tract infection	SICU	SRR5949763
BES-11	August 1998	Catheter	SICU	SRR5949764
BES-12	November 1998	Bronchoalveolar lavage	SICU	SRR5949809
BES-13	December 1998	Nose	SICU	SRR5949810
BES-14	January 1999	Rectum	SICU	SRR5949811
BES-15	March 1999	Urinary tract infection	SICU	SRR5949812
BES-16	April 1999	Stool	SICU	SRR5949805
BES-17	May 1999	Rectum	SICU	SRR5949806
BES-18	June 1999	Nose	SICU	SRR5949807
BES-19	July 1999	Nose	SICU	SRR5949808
BES-20	August 1999	Tracheal aspirate	SICU	SRR5949803
BES-21	September 1999	Rectum	SICU	SRR5949804
BES-22	November 1999	Catheter	Trauma unit	SRR5949779
BES-23	December 1999	Osteoarthritis	Geriatrics 1	SRR5949778
BES-24	January 2000	Nose	Neurosurgery	SRR5949781
BES-25	January 2000	Sputum	Rehabilitation 1	SRR5949780
BES-26	February 2000	Rectum	MICU	SRR5949783
BES-27	November 2000	Nose	MICU	SRR5949782
BES-28	February 2001	Tracheal aspirate	MICU	SRR5949785
BES-29	June 2001	Blood culture	MICU	SRR5949784
BES-30	July 2001	Urinary tract infection	Rehabilitation 1	SRR5949776
BES-31	August 2001	Urinary tract infection	Rehabilitation 1	SRR5949775
BES-32	September 2001	Wound	Rehabilitation 2	SRR5949798
BES-33	October 2001	Wound	Orthopaedics	SRR5949799
BES-34	November 2001	Nose	MICU	SRR5949796
BES-35	December 2001	Urinary tract infection	Urology	SRR5949797
BES-36	January 2002	Wound	Dermatology	SRR5949770

Isolate name (alternate name)	Isolation date	Isolation location	Hospitalisation unit ^a	Accession number (NCBI or ENA)
BES-37	March 2002	Wound	Diabetes - Endocrinology	SRR5949771
BES-38	June 2002	Urinary tract infection	Nephrology	SRR5949800
BES-39	September 2002	Urinary tract infection	Gastroenterology	SRR5949801
BES-40	October 2002	Stool	Geriatrics 2	SRR5949765
BES-41	November 2002	Urinary tract infection	Vascular surgery	SRR5949802
BES-42	February 2003	Tracheal Aspi-rate	MICU	SRR5949795
BES-43	May 2003	Blood culture	Pulmonary	SRR5949794
BES-44	June 2003	Urinary tract infection	Neurosurgery	SRR5949793
BES-45	July 2003	Bronchoalveolar lavage	SICU	SRR5949792
BES-46	August 2003	Stool	Hepatology	SRR5949791
BES-47	August 2003	Urinary tract infection	Rehabilitation 2	SRR5949790
BES-48	October 2003	Tracheal aspi-rate	Neurosurgery	SRR5949789
BES-49	January 2005	Blood culture	Rehabilitation 1	SRR5949788
BES-50	September 2005	Catheter	Hepatology	SRR5949787
BES-51	December 2005	Urinary tract infection	MICU	SRR5949786
BES-52	March 2006	Tracheal aspi-rate	SICU	SRR5949760
BES-53	December 2007	Urinary tract infection	Urology	SRR5949761
BES-54	April 2008	Urinary tract infection	MICU	SRR5949762
ST395 local outgroup	June 2014	Rectum	MICU	
ST395 external outgroup	2010	Water outlet	Burn unit	ERP006058 (European Nucleotide Archive)

TABLEAU 4.1 – Information about the 54 BES isolates and the outgroups used for phylogeny.
^aMICU, Medical Intensive Care Unit ; SICU, Surgical Intensive Care Unit.

Chapitre 5

Conclusion et perspectives

5.1 Méthodologie

5.1.1 Choix de l'outil WGS

La technique de WGS est à l'heure actuelle la technique d'analyse de génome la plus discriminante puisqu'elle permet de déterminer plus ou moins complètement la séquence nucléotidique d'un génome et d'en analyser finement la composition.

L'utilisation des techniques NGS dans l'analyse des épidémies a permis ces dernières années de mieux comprendre l'évolution de pathogènes au cours d'épidémie [18]. De plus, l'analyse des génomes complets permettent de mieux comprendre comment et pourquoi les micro-organismes pathogènes se propagent [195]. De nombreuses variations ponctuelles sont visibles, il est en effet possible de voir des mutations rares qui ne sont pas observables en MLST ou en PFGE, et qui peuvent avoir un impact sur l'évolution de la souche. Il est également possible d'observer plusieurs épidémies parallèles ayant lieu en même temps, comme c'est le cas pour cette étude.

5.1.2 Limites de l'approche WGS

Choix du séquenceur et difficultés rencontrées.

L'évolution des techniques de séquençage ainsi que la diminution des coûts a facilité le recours aux analyses WGS en microbiologie. A l'heure actuelle, la technologie Illumina reste la plus performante pour le séquençage d'un génome complet puisqu'elle représente un compromis entre le coût du séquençage et la qualité obtenue. Cette technique génère ainsi des données de bonne qualité (score de qualité élevé des *reads*), une couverture suffisante du génome ainsi qu'un faible taux d'erreur dans les séquences [108]. De plus l'évolution des technologies a modifié nos conclusions. Par exemple, au début de l'analyse de l'épidémie de *P. aeruginosa* ST395, deux génomes avaient été séquencés en Illumina, DHS01 (BES-1) ainsi que BES-54. Un bloc de 131 kb avait été identifié chez BES-54 mais absent de BES-1. Nous avons initialement suggéré que le génome de BES-54 avait gagné le bloc au cours de l'évolution [129]. Hors, les analyses utilisant les données Illumina des 54 isolats couplées à celles générées par le séquençage PacBio, ont permis de démontrer que ce bloc était uniquement absent chez BES-1.

De plus, les autres génomes avaient d'abord été séquencés sur un appareil PGM de Ion Torrent (Life Technologies). Notre choix s'était porté sur cette technique car la plateforme de séquençage SFR-FED de l'université venait d'acquérir un de ces appareils. Les contraintes financières nous avaient conduit à choisir une couverture de 100X pour 10 souches et de 25X pour les autres. Avec ces données, le taux de SNPs relativement élevé détecté lors des premières analyses était majoritairement constitué d'erreur de séquençage. Lors de recherche de ces SNPs par PCR, très peu d'entre eux était vérifiés. Le re-séquençage de ces isolats en Illumina (NextSeq) a permis l'obtention de données de meilleure qualité permettant une analyse fiable de l'évolution du clone épidémique.

Choix et optimisation des outils.

Les outils bioinformatiques choisis pour l'analyse doivent permettre d'éviter des biais dans les résultats obtenus. Les logiciels d'assemblage ou d'alignement doivent donner des résultats pertinents, tous comme les différents outils d'analyse (recherche de variants, comparaison de séquences, *etc.*). La mise en place de nouvelles méthodes d'analyses avec par exemple, la création d'un pangénome mais également l'évaluation des outils avec les tests de différents logiciels d'assemblage avec des paramètres variables a permis de limiter les problèmes liés aux outils utilisés.

Critiques de la méthode "pangénome".

Une technique innovante a été mise en place, la création d'un pangénome, correspondant à une base de données de gènes présents chez au moins un génome de la collection. Cette approche permet l'analyse exhaustive des pertes et gains de gènes. L'utilisation de cette technique dépend de la qualité de l'assemblage, puisqu'un mauvais assemblage va impacter le nombre de gènes détectés. Il est donc possible que certaines informations soient manquantes dans le pangénome. Afin de pallier à ce genre de problème, un séquençage PacBio pourrait conduire à l'obtention de tous les génomes complets de l'épidémie et ainsi permettre une analyse plus complète. Cela est pour le moment difficile compte-tenu du prix élevé de cette technique de séquençage (1300 euros environ par génome contre 200 euros en moyenne pour un séquençage Illumina).

5.1.3 Limite de l'analyse

Collection réduite.

Plus de 300 patients ont été infectés ou colonisés lors de l'épidémie entre 1997 et 2008, pourtant la collection complète d'isolats épidémiques ne comprend que 80 échantillons. Lors de la sélection initiale des isolats pour l'étude, un total de 60 génomes possédant un profil PFGE similaire avait été séquencé. Après des premières vérifications (nombre de SNPs par souches en fonction de la référence et nombre de bases alignées sur la référence), les tests ont montré que 5 d'entre eux n'appartenaient pas au ST395 et ne faisaient donc pas partie de l'épidémie observée. Le dernier isolat prélevé en 2014 possédait une divergence génomique trop élevée pour appartenir au clone épidémique. Ainsi, l'analyse ne concerne que 54 génomes. De ce fait, bien que l'échantillonnage soit représentatif des isolats épidémiques (nombre d'isolats sélectionnés par an proportionnel au nombre de cas), l'analyse ne peut être réellement exhaustive.

De plus, nous avons vu précédemment que la modification de l'échantillonnage (2 génomes séquencés, puis 54) ainsi que les différentes techniques de séquençage utilisées (Illumina pour les deux premiers isolats, Ion Torrent pour les 54 en 100X et 25X et enfin Illumina avec la même couverture pour tous les isolats) avait modifié nos conclusions. Il est donc tout à fait possible que si l'on ajoutait des échantillons à notre analyse (parmi les 80 de la collection), nos résultats soient en partie modifiés en apportant des informations complémentaires ou nouvelles.

5.2 Compréhension de la biologie du pathogène pendant sa diffusion dans une structure hospitalière

Début de l'épidémie.

L'analyse des 54 génomes séquencés de l'épidémie de *P. aeruginosa* a permis de montrer la présence d'un îlot génomique contenant des gènes de résistance au cuivre, spécifique du clone ST395. La datation d'un ancêtre commun au ST395 dans l'hôpital de Besançon suggérait une contamination probable de celui-ci lors de sa construction en 1978. En effet, la capacité de survie dans un milieu contenant du cuivre aurait permis à la bactérie de survivre dans les tuyaux du réseau d'eau avant d'arriver au contact des patients.

Évolution.

La taille constante du génome de la souche ST395 épidémique au cours du temps, avec peu de gains et de pertes de gènes, indique un rôle faible de l'*accessory genome* lors de la propagation du pathogène. De plus, aucun gène de virulence ou résistance n'a pu être mis en évidence hormis les mutations d'*oprD* et celles modifiant l'expression de l'opéron *mexAB-oprM*. Les éléments mobiles décrits chez DHS01 (îlots génomiques spécifiques ainsi que les cinq ICEs) sont retrouvés chez tous les génomes de l'épidémie. Seul le génome de BES-49 a intégré une IS (ISPa74) que l'on retrouve en deux copies.

Dernières années de l'épidémie.

L'analyse des mutations ponctuelles acquises en fin d'épidémie a montré une accumulation de mutations dans certaines fonctions biologiques. Deux types de sélections peuvent être mises en évidence. La première, la sélection positive concerne la conservation de caractères apportant un avantage à l'organisme. La seconde en revanche, la sélection négative, est impliquée dans la disparition de caractères biologiques inutiles ou délétères dans la niche écologique considérée [188].

Au delà de ces sélections, on observe une évolution parallèle. Celle-ci correspond à l'acquisition d'un même phénotype, chez des organismes vivant dans le même milieu, par des voies différentes. Cela signifie que les fonctions mutées, positivement ou négativement sélectionnées, confèrent un phénotype identique à la majorité des isolats de la fin de l'épidémie.

Dans le cas du clone ST395, ces deux phénomènes sont visibles sur les souches prélevées après 2000 indiquant un lien probable avec la décroissance du nombre de patients nouvellement colonisés ou infectés. Les fonctions touchées sont également modifiées lors d'infections chro-

niques, notamment chez les patients atteints de mucoviscidose. En effet, on observe chez les isolats colonisant ces patients une modification de la porine *oprD*, un changement de structure du LPS ainsi qu'une perte de la pompe d'efflux MexAB-OprM [104, 95, 103]. Chez les isolats issus de patients atteints de mucoviscidose, il s'agit souvent de mutations abolissant la fonction [103]. Toutefois ici, nous avons montré chez le clone ST395 un retour à une expression basale.

La perte de surexpression de cet opéron pourrait être le résultat d'une sélection négative. En effet, la surexpression de *mexAB-oprM* entraîne une diminution de *fitness* de la souche, puisqu'elle survit moins dans l'eau ainsi que sur des surfaces sèches, qu'elle est moins virulente et qu'elle a perdu sa capacité à envahir des cellules [63, 65, 71]. En parallèle, le gain de *fitness* lié à la modification d'expression pourrait indiquer un avantage évolutif sélectionné positivement. La perte de LPS (conférant une certaine invisibilité vis à vis du système immunitaire) pourrait être due à une sélection positive mais également négative puisque la bactérie perd une fonction qui est délétère dans cet environnement. Pour ce qui est du gain de résistance à l'imipénème (lié à la mutation d'*oprD*) et de la modification du métabolisme des acides aminés (entraînant une meilleure utilisation des acides aminés coûteux à métaboliser), l'avantage conféré à la bactérie laisse à penser qu'il s'agit dans tous les cas de sélection positive.

Extinction de l'épidémie.

L'évolution de la souche épidémique vers un phénotype de chronicité tend à valider l'hypothèse principale en démontrant que l'évolution d'une souche épidémique dans un hôpital est comparable à ce que l'on observe lors d'une infection chronique chez un patient unique.

L'accumulation de culs-de-sac épidémiologiques dans les dernières années de diffusion favoriserait l'extinction de l'épidémie probablement liée à une moindre transmissibilité des souches de patients à patients. Des expériences sont en cours afin d'estimer la survie des isolats tardifs dans l'environnement (dans l'eau distillée, l'eau du robinet, résistance à la dessiccation).

Transmissibilité des souches.

Certaines souches retrouvées chez des patients atteints de mucoviscidose possèdent la capacité de se transmettre de patients à patients. C'est le cas en particulier de la souche LES qui possède une virulence accrue [196]. Pourtant, il semble que ce ne soit pas cette virulence qui soit à l'origine de la capacité à se transmettre. En effet, une souche Danoise présentant une virulence très faible possède toujours sa capacité de transmission [197]. Les facteurs impliqués dans la capacité de la souche à se transmettre ne sont pas encore connus. Des expériences menées sur un modèle animal permettraient d'observer une éventuelle modification de la capacité de transmission du clone épidémique BES.

Si le déclin de l'épidémie est lié en partie à la perte de transmissibilité d'une partie des isolats, doivent-ils tous devenir non/moins transmissible pour conduire à la fin de l'épidémie ? Étant donné qu'il existe peu d'analyse menée sur les bactéries, l'exemple considéré est relatif aux virus. Une modélisation menée sur le virus de la grippe, observe l'effet de la vaccination (assimilable à une baisse de la transmissibilité) sur le nombre de nouveaux cas ayant lieu en une année dans un hôpital. L'expérience contrôlée se base sur des données réelles de vaccination

aux États-Unis en 2009-2010. Cela correspond à la vaccination de 38% de la population et l'apparition de 305 cas en 1 an dans l'hôpital. La vaccination de la population communautaire a un effet important sur le nombre de cas. En effet, lorsque 50% de la population est vaccinée, le nombre de cas diminue de 34%. Et l'effet est encore plus marqué si 60% de la population est vaccinée puisque le nombre de nouveaux cas diminue de 79% [198]. Le modèle mathématique utilisé permet de calculer l'effet de plusieurs valeurs de vaccination de la population (40, 50, 60, 80 et 100%). La distribution du nombre de cas n'est pas linéaire et la première diminution réelle observée est lorsque 50% de la population est vaccinée. De plus, il semble y avoir un seuil entre 50 et 60%, puisque toutes les valeurs supérieures entraînent un effet important sur la diminution du nombre de cas.

Afin de faire le lien entre la fin de l'épidémie et une perte de *fitness* et de transmissibilité tel que décrite chez les virus, il est important de connaître le nombre de souches non transmissibles en fin d'épidémie. Pour cela, des tests permettant d'évaluer le *fitness* (survie dans l'eau et résistance à la dessiccation) des 54 isolats de l'épidémie sont en cours.

Limites de l'analyse génomique et perspectives.

L'analyse des génomes, en particulier ceux séquencés en haut-débit, permet d'observer un grand nombre de paramètres que ce soit des pertes, gains ou déplacement de portions de chromosome ou encore des mutations ponctuelles sélectionnées. Pourtant la relation entre les gènes manquants ou mutés et le phénotype associé est rarement évidente et ne permet pas de faire un lien direct entre une modification génique et une fonction modifiée ou perdue. En effet, l'observation du génotype seul ne permet pas de prédire à coup sûr le phénotype. Par exemple, c'est le cas pour les modifications observées en rapport avec l'opéron *mexAB-oprM*. A la fin de l'épidémie, il est observé une diminution significative de l'expression de l'opéron chez 23 isolats, pourtant, seuls 17 d'entre eux présentent des mutations dans les gènes régulateurs connus ou dans les régions promotrices. Il est possible qu'il existe d'autres régulateurs non connus qui influent sur l'expression de l'opéron et qui permettraient d'expliquer sa diminution. A l'inverse, il existe 2 isolats mutés dans ces régions surexprimant encore *mexAB-oprM*. Dans ce cas, les mutations étant visiblement sans effet sur l'expression de *mexAB-oprM* peuvent avoir été compensées par une autre mutation dans un gène non identifié, être sans effet sur la protéine ou encore peuvent avoir un effet sur d'autres voies biologiques non observées.

Des tests phénotypiques complémentaires sont nécessaires afin d'approfondir l'analyse. Des tests de *fitness* entre les différents isolats de l'épidémie ainsi que la caractérisation du LPS (modifié en fin d'épidémie) sont actuellement en cours, pour mieux comprendre le lien entre les modifications génotypiques et l'extinction de l'épidémie. Ces dernières visent à étudier l'effet de la modification du LPS sur la reconnaissance de la bactérie par les macrophages. Cela permettra de connaître le lien entre la perte du LPS et l'interaction avec le système immunitaire.

Le contexte hospitalier dans lequel se trouvaient les patients porteur de la bactérie *P. aeruginosa* a très probablement influencé l'évolution de la souche épidémique. L'utilisation de données cliniques supplémentaires (traitements antibiotiques, durée et chambre d'hospita-

lisation, date de colonisation du patient, *etc.*) pourrait être un atout pour une meilleure compréhension de l'épidémie.

De plus, les différentes mesures d'hygiène mises en place au cours des 11 années de l'épidémie, doivent être également prise en compte. Au début de l'épidémie, ayant eu lieu uniquement en réanimation chirurgicale, des nouveaux protocoles d'hygiène ont été mis en place (utilisation de solution hydroalcoolique ainsi que fermeture d'un service pendant un temps) ce qui a permis de limiter la contamination à ce seul service pendant un certain temps. A partir de la fin de l'année 1999, l'épidémie se propage dans différents services. La pression de sélection antibiotique exercée sur la souche épidémique est alors moindre que dans le service de réanimation.

L'analyse de l'impact de l'environnement hospitalier sur la souche épidémique combinée aux modifications génomiques observées au cours du temps permet d'avoir une vision globale de l'évolution du pathogène pendant les 11 années de l'épidémie.

Il existe peu d'exemples d'évolution de pathogènes bactériens pendant une épidémie. Pourtant les quelques exemples disponibles, ainsi que ceux en lien avec les épidémies virales, décrivent une adaptation individuelle du pathogène à l'hôte [199]. Celle-ci est souvent en lien avec la résistance aux antibiotiques ou la capacité à échapper au système immunitaire. Dans le premier cas, on retrouve plusieurs exemples de gain de résistance au cours d'épidémie comme par exemple chez *P. aeruginosa* [28] et chez *V. cholerae* [46]. Chez ces deux espèces, ce sont des mutations qui sont à l'origine des phénotypes de résistance. Dans le second cas, de nombreux virus gagnent la capacité à échapper au système immunitaire pour prolonger leur survie dans le corps de l'hôte. Par exemple, lors de la dernière épidémie causée par le virus responsable de la maladie Ebola, plusieurs mutations ont été sélectionnées au cours de l'épidémie. Parmi celles-ci, certaines servent à créer un leurre pour les anticorps de l'hôte tandis que d'autres permettent de bloquer la réponse du système immunitaire en inhibant la production d'interféron et ainsi rendant le virus plus pathogène [42]. L'adaptation liée à la résistance aux antibiotiques et à l'échappement au système immunitaire semble être un élément commun à tous les pathogènes que l'on retrouve également ici chez le clone ST395.

L'adaptation individuelle modèle la courbe globale de l'épidémie, puisqu'elle joue un rôle dans l'adaptation globale de la souche à son environnement. Il est également possible d'observer une évolution parallèle de certaines fonctions conférant des phénotypes communs à la majorité ou à la totalité des micro-organismes en fin d'épidémie. La compréhension des mécanismes conduisant à la sélection de ces fonctions particulières permettrait une meilleure gestion des épidémies liées à des micro-organismes pathogènes.

Bibliographie

- [1] C Potera. Common bacterium induces histamine production in neutrophils. *Environmental Health Perspectives*, 120(5) :a190, 2012.
- [2] J Freney, F Renaud, R Leclercq, and P Riegel. *Précis de bactériologie clinique*. Ed. Eska, 2007.
- [3] MW Silby, C Winstanley, SAC Godfrey, SB Levy, and RW Jackson. *Pseudomonas* genomes : diverse and adaptable. *FEMS microbiology reviews*, 35(4) :652–680, 2011.
- [4] CK Stover, XQ Pham, AL Erwin, SD Mizoguchi, P Warrenner, MJ Hickey, FSL Brinkman, WO Hufnagle, DJ Kowalik, M Lagrou, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 406(6799) :959–964, 2000.
- [5] A Alonso, P Sanchez, and JL Martinez. Environmental selection of antibiotic resistance genes. *Environmental Microbiology*, 3(1) :1–9, 2001.
- [6] JM Thiolet, S Vaux, M Lamy, A Gauthier, AS Barret, L Leon, et al. Enquête nationale de prévalence des infections nosocomiales et des traitements anti-infectieux en établissements de santé, France, mai-juin 2012. institut de veille sanitaire. 2013, 2016.
- [7] B Valot, C Guyeux, JY Rolland, K Mazouzi, X Bertrand, and D Hocquet. What it takes to be a *Pseudomonas aeruginosa* ? The core genome of the opportunistic pathogen updated. *PloS one*, 10(5) :e0126468, 2015.
- [8] PA Lambert. Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa*. *Journal of the Royal Society of Medicine*, 95(Suppl 41) :22, 2002.
- [9] DC Schwartz and CR Cantor. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37(1) :67–75, 1984.
- [10] D Talon, V Cailleaux, M Thouverez, and Y Michel-Briand. Discriminatory power and usefulness of pulsed-field gel electrophoresis in epidemiological studies of *Pseudomonas aeruginosa*. *Journal of Hospital Infection*, 32(2) :135–145, 1996.
- [11] A Oliver, X Mulet, C López-Causapé, and C Juan. The increasing threat of *Pseudomonas aeruginosa* high-risk clones. *Drug Resistance Updates*, 21 :41–59, 2015.
- [12] MCJ Maiden, JA Bygraves, E Feil, G Morelli, JE Russell, R Urwin, Q Zhang, J Zhou, K Zurth, DA Caugant, 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*, 95(6) :3140–3145, 1998.
- [13] B Curran, D Jonas, H Grundmann, T Pitt, and CG Dowson. Development of a multi-locus sequence typing scheme for the opportunistic pathogen *Pseudomonas aeruginosa*. *Journal of Clinical Microbiology*, 42(12) :5644–5649, 2004.

- [14] PubMLST. <https://pubmlst.org/>. Mise en ligne en 2005.
- [15] TD Read, SL Salzberg, M Pop, M Shumway, L Umayam, L Jiang, E Holtzapple, JD Busch, KL Smith, JM Schupp, et al. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science*, 296(5575) :2028–2033, 2002.
- [16] MJ Struelens and S Brisse. From molecular to genomic epidemiology : transforming surveillance and control of infectious diseases. *Euro Surveillance*, 18(4) :e20386, 2013.
- [17] X Didelot, R Bowden, DJ Wilson, TEA Peto, and DW Crook. Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, 13(9) :601–612, 2012.
- [18] NJ Croucher and X Didelot. The application of genomics to tracing bacterial pathogen transmission. *Current Opinion in Microbiology*, 23 :62–67, 2015.
- [19] DNA Sequencing Costs : Data from the NHGRI Genome Sequencing Program (GSP). <https://www.genome.gov/sequencingcostsdata/>.
- [20] P Treepong, VN Kos, C Guyeux, DS Blanc, X Bertrand, B Valot, and D Hocquet. Global emergence of the widespread *Pseudomonas aeruginosa* ST235 clone. *Clinical Microbiology and Infection*, in press, 2017.
- [21] C Slekovec, J Plantin, P Cholley, M Thouverez, D Talon, X Bertrand, and D Hocquet. Tracking down antibiotic-resistant *Pseudomonas aeruginosa* isolates in a wastewater network. *PLoS One*, 7(12) :e49300, 2012.
- [22] M Gomila, M del Carmen Gallegos, V Fernández-Baca, A Pareja, M Pascual, P Díaz-Antolín, E García-Valdés, and J Lalucat. Genetic diversity of clinical *Pseudomonas aeruginosa* isolates in a public hospital in Spain. *BMC Microbiology*, 13(1) :138, 2013.
- [23] M Koutsogiannou, E Drougka, A Liakopoulos, E Jelastopulu, E Petinaki, ED Anastassiou, I Spiliopoulou, and M Christofidou. Spread of multidrug-resistant *Pseudomonas aeruginosa* clones in a university hospital. *Journal of Clinical Microbiology*, 51(2) :665–668, 2013.
- [24] Y Chen, M Sun, M Wang, Y Lu, and Z Yan. Dissemination of IMP-6-producing *Pseudomonas aeruginosa* ST244 in multiple cities in China. *European Journal of Clinical Microbiology & Infectious Diseases*, 33(7) :1181–1187, 2014.
- [25] M Maatallah, J Cheriaa, A Backhrouf, A Iversen, H Grundmann, T Do, P Lanotte, M Mastouri, F Elghmati, MSand Rojo, et al. Population structure of *Pseudomonas aeruginosa* from five mediterranean countries : evidence for frequent recombination and epidemic occurrence of CC235. *PloS one*, 6(10) :e25617, 2011.
- [26] B Libisch, B Balogh, and M Füzi. Identification of two multidrug-resistant *Pseudomonas aeruginosa* clonal lineages with a countrywide distribution in Hungary. *Current Microbiology*, 58(2) :111–116, 2009.
- [27] P Teixeira, M Tacão, A Alves, and I Henriques. Antibiotic and metal resistance in a ST395 *Pseudomonas aeruginosa* environmental isolate : A genomics approach. *Marine Pollution Bulletin*, 110(1) :75–81, 2016.

- [28] J Quick, N Cumley, CM Wearn, M Niebel, C Constantinidou, CM Thomas, MJ Pallen, NS Moïemen, A Bamford, B Oppenheim, et al. Seeking the source of *Pseudomonas aeruginosa* infections in a recently opened hospital : an observational study using whole-genome sequencing. *BMJ open*, 4(11) :e006278, 2014.
- [29] DM Livermore. Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa* : our worst nightmare? *Clinical Infectious Diseases*, 34(5) :634–640, 2002.
- [30] X Mulet, G Cabot, AA Ocampo-Sosa, MA Domínguez, L Zamorano, C Juan, F Tubau, C Rodríguez, B Moyà, C Peña, et al. Biological markers of *Pseudomonas aeruginosa* epidemic high-risk clones. *Antimicrobial Agents and Chemotherapy*, 57(11) :5527–5535, 2013.
- [31] A Oliver, A Mena, and MD Maciá. Evolution of *Pseudomonas aeruginosa* pathogenicity : from acute to chronic infections. In *Evolutionary biology of bacterial and fungal pathogens*, pages 433–444. American Society of Microbiology, 2008.
- [32] M Willmann, D Bezdan, L Zapata, H Susak, W Vogel, K Schröppel, J Liese, C Weidenmaier, IB Autenrieth, S Ossowski, et al. Analysis of a long-term outbreak of XDR *Pseudomonas aeruginosa* : a molecular epidemiological study. *Journal of Antimicrobial Chemotherapy*, 70(5) :1322–1330, 2015.
- [33] AA Witney, KA Gould, CF Pope, F Bolt, NG Stoker, MD Cubbon, CR Bradley, A Fraise, AS Breathnach, PD Butcher, et al. Genome sequencing and characterization of an extensively drug-resistant sequence type 111 serotype O12 hospital outbreak strain of *Pseudomonas aeruginosa*. *Clinical Microbiology and Infection*, 20(10) :O609–O618, 2014.
- [34] LA Snyder, NJ Loman, LJ Faraj, K Levi, J Weinstock, TC Boswell, Mark J Pallen, and DA Ala’Aldeen. Epidemiological investigation of ‘*Pseudomonas aeruginosa*’ isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing. *Eurosurveillance*, 18(42), 2013.
- [35] JF Turton, L Wright, A Underwood, AA Witney, YT Chan, A Al-Shahib, C Arnold, M Doumith, B Patel, TD Planche, et al. High-resolution analysis by whole-genome sequencing of an international lineage (sequence type 111) of *Pseudomonas aeruginosa* associated with metallo-carbapenemases in the United Kingdom. *Journal of Clinical Microbiology*, 53(8) :2622–2631, 2015.
- [36] JR Dettman, N Rodrigue, SD Aaron, and R Kassen. Evolutionary genomics of epidemic and nonepidemic strains of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences*, 110(52) :21065–21070, 2013.
- [37] MH Rau, RL Marvig, GD Ehrlich, S Molin, and L Jelsbak. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environmental Microbiology*, 14(8) :2200–2211, 2012.
- [38] J Jeukens, B Boyle, I Kukavica-Ibrulj, MM Ouellet, SD Aaron, SJ Charette, JL Fothergill, NP Tucker, C Winstanley, and RC Levesque. Comparative genomics of isolates of a *Pseudomonas aeruginosa* epidemic strain associated with chronic lung infections of cystic fibrosis patients. *PloS one*, 9(2) :e87611, 2014.

- [39] D Pasetto, F Finger, A Rinaldo, and E Bertuzzo. Real-time projections of cholera outbreaks through data assimilation and rainfall forecasting. *Advances in Water Resources*, *in press*, 2016.
- [40] Maladie à virus Ebola, OMS. <http://www.who.int/mediacentre/factsheets/fs103/fr/>.
- [41] D Gatherer. The 2014 Ebola virus disease outbreak in West Africa. *Journal of General Virology*, 95(8) :1619–1624, 2014.
- [42] A Ramaiah and V Arumugaswami. Ebolavirus evolves in human to minimize the detection by immune cells by accumulating adaptive mutations. *VirusDisease*, 27(2) :136–144, 2016.
- [43] JH Zhang. Positive Darwinian selection in gene evolution. In *Darwin’s Heritage Today : Proceedings of the Darwin 200 Beijing International Conference : 24–26 October 2009 ; Beijing*, pages 288–309. Citeseer, 2010.
- [44] RV Tauxe, ED Mintz, and RE Quick. Epidemic cholera in the new world : translating field epidemiology into new prevention strategies. *Emerging Infectious Diseases*, 1(4) :141, 1995.
- [45] CS Chin, J Sorenson, JB Harris, WP Robins, RC Charles, RR Jean-Charles, J Bullard, DR Webster, A Kasarskis, P Peluso, et al. The origin of the Haitian cholera outbreak strain. *New England Journal of Medicine*, 364(1) :33–42, 2011.
- [46] LS Katz, A Petkau, J Beaulaurier, S Tyler, ES Antonova, MA Turnsek, Y Guo, S Wang, EE Paxinos, F Orata, et al. Evolutionary dynamics of *Vibrio cholerae* o1 following a single-source introduction to Haiti. *MBio*, 4(4) :e00398–13, 2013.
- [47] RR Frerichs, PS Keim, R Barrais, and R Piarroux. Nepalese origin of cholera epidemic in Haiti. *Clinical Microbiology and Infection*, 18(6) :E158–E163, 2012.
- [48] LS Frost, R Leplae, AO Summers, and A Toussaint. Mobile genetic elements : the agents of open source evolution. *Nature Reviews Microbiology*, 3(9) :722–732, 2005.
- [49] RC Fowler and ND Hanson. Emergence of carbapenem resistance due to the novel insertion sequence ISPa8 in *Pseudomonas aeruginosa*. *PloS one*, 9(3) :e91299, 2014.
- [50] Campbell’s Biology, 5th Edition. Figure 18.17, page 336.
- [51] T Van Opijnen and A Camilli. Transposon insertion sequencing : a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, 11(7) :435–442, 2013.
- [52] Molecular Biology : principles and practice. 2012, WH Freeman and company.
- [53] MR Gillings. Integrons : past, present, and future. *Microbiology and Molecular Biology Reviews*, 78(2) :257–277, 2014.
- [54] RAF Wozniak and MK Waldor. Integrative and conjugative elements : mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology*, 8(8) :552–563, 2010.
- [55] M Shintani, ZK Sanchez, and K Kimbara. Genomics of microbial plasmids : classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in microbiology*, 6 :242, 2015.
- [56] K Mathee, G Narasimhan, C Valdes, X Qiu, JM Matewish, M Koehrsen, A Rokas, CN Yandava, R Engels, E Zeng, et al. Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proceedings of the National Academy of Sciences*, 105(8) :3100–3105, 2008.

- [57] A Oliver, F Baquero, and journal=Molecular Microbiology volume=43 number=6 pages=1641–1650 year=2002 publisher=Wiley Online Library Blázquez, J. The mismatch repair system (*mutS*, *mutL* and *uvrD* genes) in *Pseudomonas aeruginosa* : molecular characterization of naturally occurring mutants.
- [58] A Oliver, R Cantón, P Campo, F Baquero, and journal=Science volume=288 number=5469 pages=1251–1253 year=2000 publisher=American Association for the Advancement of Science Blázquez, J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection.
- [59] XZ Li, P Plésiat, and H Nikaido. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clinical Microbiology Reviews*, 28(2) :337–418, 2015.
- [60] K Poole, K Krebes, C McNally, and S Neshat. Multiple antibiotic resistance in *Pseudomonas aeruginosa* : evidence for involvement of an efflux operon. *Journal of Bacteriology*, 175(22) :7363–7372, 1993.
- [61] XZ Li, H Nikaido, and K Poole. Role of *mexA-mexB-oprM* in antibiotic efflux in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 39(9) :1948–1953, 1995.
- [62] JP Pearson, C Van Delden, and Barbara H Iglewski. Active efflux and diffusion are involved in transport of *Pseudomonas aeruginosa* cell-to-cell signals. *Journal of Bacteriology*, 181(4) :1203–1210, 1999.
- [63] K Evans, L Passador, R Srikumar, E Tsang, J Nezezon, and K Poole. Influence of the MexAB-OprM multidrug efflux system on quorum sensing in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 180(20) :5443–5447, 1998.
- [64] RS Smith, SG Harris, R Phipps, and B Iglewski. The *Pseudomonas aeruginosa* quorum-sensing molecule N-(3-oxododecanoyl) homoserine lactone contributes to virulence and induces inflammation *in vivo*. *Journal of Bacteriology*, 184(4) :1132–1139, 2002.
- [65] Y Hirakata, A Kondo, K Hoshino, H Yano, K Arai, A Hirotsu, H Kunishima, N Yamamoto, M Hatta, M Kitagawa, et al. Efflux pump inhibitors reduce the invasiveness of *Pseudomonas aeruginosa*. *International Journal of Antimicrobial Agents*, 34(4) :343–346, 2009.
- [66] Y Hirakata, R Srikumar, K Poole, N Gotoh, T Suematsu, S Kohno, S Kamihira, REW Hancock, and DP Speert. Multidrug efflux systems play an important role in the invasiveness of *Pseudomonas aeruginosa*. *Journal of Experimental Medicine*, 196(1) :109–118, 2002.
- [67] N Masuda, E Sakagawa, S Ohya, N Gotoh, H Tsujimoto, and T Nishino. Substrate specificities of MexAB-OprM, MexCD-OprJ, and MexXY-OprM efflux pumps in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 44(12) :3322–3327, 2000.
- [68] JR Aires, T Köhler, H Nikaido, and P Plésiat. Involvement of an active efflux system in the natural resistance of *Pseudomonas aeruginosa* to aminoglycosides. *Antimicrobial Agents and Chemotherapy*, 43(11) :2624–2628, 1999.
- [69] PD Lister, DJ Wolter, and ND Hanson. Antibacterial-resistant *Pseudomonas aeruginosa* : clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clinical Microbiology Reviews*, 22(4) :582–610, 2009.

- [70] ML Sobel, D Hocquet, L Cao, P Plesiat, and K Poole. Mutations in PA3574 (*nalD*) lead to increased MexAB-OprM expression and multidrug resistance in laboratory and clinical isolates of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 49(5) :1782–1786, 2005.
- [71] P Sánchez, JF Linares, B Ruiz-Díez, E Campanario, A Navas, F Baquero, and JL Martínez. Fitness of *in vitro* selected *Pseudomonas aeruginosa nalB* and *nfxB* multidrug resistant mutants. *Journal of Antimicrobial Chemotherapy*, 50(5) :657–664, 2002.
- [72] D Hocquet, A Muller, K Blanc, P Plésiat, D Talon, DL Monnet, and X Bertrand. Relationship between antibiotic use and incidence of MexXY-OprM overproducers among clinical isolates of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 52(3) :1173–1175, 2008.
- [73] CHF Lau, S Fraud, M Jones, SN Peterson, and K Poole. Reduced expression of the *rplU-rpmA* ribosomal protein operon in mexXY-expressing pan-aminoglycoside-resistant mutants of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 56(10) :5171–5179, 2012.
- [74] D Fournier, C Richardot, E Müller, M Robert-Nicoud, C Llanes, P Plésiat, and K Jeannot. Complexity of resistance mechanisms to imipenem in intensive care unit strains of *Pseudomonas aeruginosa*. *Journal of Antimicrobial Chemotherapy*, 68(8) :1772–1780, 2013.
- [75] K Poole, N Gotoh, H Tsujimoto, Q Zhao, A Wada, T Yamasaki, S Neshat, J Yamagishi, XZ Li, and Takeshi Nishino. Overexpression of the *mexC–mexD–oprJ* efflux operon in *nfxB*-type multidrug-resistant strains of *Pseudomonas aeruginosa*. *Molecular Microbiology*, 21(4) :713–725, 1996.
- [76] K Jeannot, S Elsen, T Köhler, I Attree, C Van Delden, and P Plésiat. Resistance and virulence of *Pseudomonas aeruginosa* clinical strains overproducing the MexCD-OprJ efflux pump. *Antimicrobial Agents and Chemotherapy*, 52(7) :2455–2462, 2008.
- [77] C Llanes, T Köhler, I Patry, B Dehecq, C Van Delden, and P Plésiat. Role of the efflux system MexEF-OprN in low level resistance of *Pseudomonas aeruginosa* to ciprofloxacin. *Antimicrobial Agents and Chemotherapy*, pages AAC.00101–11, 2011.
- [78] M Michéa-Hamzehpour, U Henze, N Gotoh, L Kocjancic Curty, JC Pechère, et al. Characterization of MexE–mexF–OprN, a positively regulated multidrug efflux system of *Pseudomonas aeruginosa*. *Molecular Microbiology*, 23(2) :345–354, 1997.
- [79] C Juan, B Moyá, J L Pérez, and A Oliver. Stepwise upregulation of the *Pseudomonas aeruginosa* chromosomal cephalosporinase conferring high-level β -lactam resistance involves three AmpD homologues. *Antimicrobial Agents and Chemotherapy*, 50(5) :1780–1787, 2006.
- [80] KF Kong, A Aguila, L Schneper, and K Mathee. *Pseudomonas aeruginosa* β -lactamase induction requires two permeases, AmpG and AmpP. *BMC Microbiology*, 10(1) :328, 2010.
- [81] JK Lee, YS Lee, YK Park, and BS Kim. Alterations in the GyrA and GyrB subunits of topoisomerase II and the ParC and ParE subunits of topoisomerase IV in ciprofloxacin-resistant clinical isolates of *Pseudomonas aeruginosa*. *International Journal of Antimicrobial Agents*, 25(4) :290–295, 2005.

- [82] JO Pacheco, C Alvarez-Ortega, MA Rico, and JL Martínez. Metabolic compensation of fitness costs is a general outcome for antibiotic-resistant *Pseudomonas aeruginosa* mutants overexpressing efflux pumps. *mBio*, 8(4) :e00500–17, 2017.
- [83] T Köhler, C van Delden, LK Curty, MM Hamzehpour, and J-C Pechere. Overexpression of the MexEF-OprN multidrug efflux system affects cell-to-cell signaling in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 183(18) :5213–5222, 2001.
- [84] X Wang-Kan, JMA Blair, B Chirullo, J Betts, RM La Ragione, A Ivens, V Ricci, TJ Opperman, and LJV Piddock. Lack of AcrB efflux function confers loss of virulence on *Salmonella enterica* serovar typhimurium. *mBio*, 8(4) :e00968–17, 2017.
- [85] Y Ding, Y Onodera, JC Lee, and DC Hooper. NorB, an efflux pump in *Staphylococcus aureus* strain MW2, contributes to bacterial fitness in abscesses. *Journal of Bacteriology*, 190(21) :7123–7129, 2008.
- [86] X Qiu, BR Kulasekara, and S Lory. Role of horizontal gene transfer in the evolution of *Pseudomonas aeruginosa* virulence. In *Microbial Pathogenomics*, volume 6, pages 126–139. Karger Publishers, 2009.
- [87] D Das, R Baruah, A Sarma Roy, AK Singh, HPD Boruah, J Kalita, and TC Bora. Complete genome sequence analysis of *Pseudomonas aeruginosa* n002 reveals its genetic adaptation for crude oil degradation. *Genomics*, 105(3) :182–190, 2015.
- [88] BG Hall and M Barlow. Revised Ambler classification of β -lactamases. *Journal of Antimicrobial Chemotherapy*, 55(6) :1050–1051, 2005.
- [89] V Dubois, L Poirel, C Marie, C Arpin, P Nordmann, and C Quentin. Molecular characterization of a novel class 1 integron containing blaGES-1 and a fused product of aac(3)-Ib/aac(6"-Ib" gene cassettes in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 46(3) :638–645, 2002.
- [90] D Girlich, T Naas, A Leelaporn, L Poirel, M Fennewald, and P Nordmann. Nosocomial spread of the integron-located *veb-1*—like cassette encoding an extended-spectrum β -lactamase in *Pseudomonas aeruginosa* in thailand. *Clinical Infectious Diseases*, 34(5) :603–611, 2002.
- [91] K Poole. Aminoglycoside resistance in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 49(2) :479–487, 2005.
- [92] JM Meyer, A Neely, A Stintzi, C Georges, and IA Holder. Pyoverdinin is essential for virulence of *Pseudomonas aeruginosa*. *Infection and Immunity*, 64(2) :518–523, 1996.
- [93] GW Lau, H Ran, F Kong, DJ Hassett, and D Mavrodi. *Pseudomonas aeruginosa* pyocyanin is critical for lung infection in mice. *Infection and Immunity*, 72(7) :4275–4278, 2004.
- [94] M Klausen, A Heydorn, P Ragas, L Lambertsen, A Aaes-Jørgensen, S Molin, and T Tolker-Nielsen. Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Molecular Microbiology*, 48(6) :1511–1524, 2003.
- [95] RF Maldonado, I Sá-Correia, and MA Valvano. Lipopolysaccharide modification in gram-negative bacteria during chronic infection. *FEMS Microbiology Reviews*, 40(4) :480–493, 2016.

- [96] J Engel and P Balachandran. Role of *Pseudomonas aeruginosa* type III effectors in disease. *Current Opinion in Microbiology*, 12(1) :61–66, 2009.
- [97] O Ciofu, LF Mandsberg, H Wang, and N Høiby. Phenotypes selected during chronic lung infection in cystic fibrosis patients : implications for the treatment of *Pseudomonas aeruginosa* biofilm infections. *FEMS Immunology & Medical Microbiology*, 65(2) :215–225, 2012.
- [98] K Cheng, RL Smyth, JRW Govan, C Doherty, C Winstanley, N Denning, DP Heaf, H van Saene, and CA Hart. Spread of β -lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. *The Lancet*, 348(9028) :639–642, 1996.
- [99] RFH Taylor, ME Hodson, and TL Pitt. Auxotrophy of *Pseudomonas aeruginosa* in cystic fibrosis. *FEMS Microbiology Letters*, 92(3) :243–246, 1992.
- [100] V Behrends, B Ryall, JEA Zlosnik, DP Speert, JG Bundy, and HD Williams. Metabolic adaptations of *Pseudomonas aeruginosa* during cystic fibrosis chronic lung infections. *Environmental Microbiology*, 15(2) :398–408, 2013.
- [101] C Winstanley, S O’Brien, and MA Brockhurst. *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. *Trends in Microbiology*, 24(5) :327–337, 2016.
- [102] A Mena, EE Smith, JL Burns, DP Speert, SM Moskowitz, JL Perez, and A Oliver. Genetic adaptation of *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients is catalyzed by hypermutation. *Journal of Bacteriology*, 190(24) :7910–7917, 2008.
- [103] L Vettoretti, P Plésiat, C Muller, F El Garch, G Phan, I Attrée, A Ducruix, and C Llanes. Efflux unbalance in *Pseudomonas aeruginosa* isolates from cystic fibrosis patients. *Antimicrobial Agents and Chemotherapy*, 53(5) :1987–1997, 2009.
- [104] M Hogardt and journal=International Journal of Medical Microbiology volume=300 number=8 pages=557–562 year=2010 publisher=Elsevier Heesemann, J. Adaptation of *Pseudomonas aeruginosa* during persistence in the cystic fibrosis lung.
- [105] F Sanger, S Nicklen, and AR Coulson. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12) :5463–5467, 1977.
- [106] *DNA Sequencing by Capillary Electrophoresis*, ThermoFisher. <https://www.thermofisher.com/fr/fr/home/life-science/sequencing/sanger-sequencing/sanger-dna-sequencing.html>.
- [107] L Liu, Y Li, S Li, N Hu, Y He, R Pong, D Lin, L Lu, and M Law. Comparison of next-generation sequencing systems. *BioMed Research International*, 2012, 2012.
- [108] MA Quail, M Smith, P Coupland, TD Otto, SR Harris, TR Connor, A Bertoni, HP Swerdlow, and Y Gu. A tale of three next generation sequencing platforms : comparison of Ion Torrent, Pacific Biosciences and Illumina Miseq sequencers. *BMC Genomics*, 13(1) :341, 2012.
- [109] *National Human Genome*. <https://www.genome.gov/10000207/>.
- [110] JM Rothberg and JH Leamon. The development and impact of 454 sequencing. *Nature Biotechnology*, 26(10) :1117–1124, 2008.
- [111] Illumina sequencing technology. http://www.openwetware.org/images/d/de/Illumina_sequencing.pdf.

- [112] Illumina inc. <http://slideplayer.com/slide/10633510/>.
- [113] J Eid, A Fehr, J Gray, K Luong, J Lyle, G Otto, P Peluso, D Rank, P Baybayan, B Bettman, et al. Real-time dna sequencing from single polymerase molecules. *Science*, 323(5910) :133–138, 2009.
- [114] TP Niedringhaus, D Milanova, MB Kerby, MP Snyder, and AE Barron. Landscape of next-generation sequencing technologies. *Analytical chemistry*, 83(12) :4327–4341, 2011.
- [115] S Boisvert, F Laviolette, and J Corbeil. Ray : simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *Journal of Computational Biology*, 17(11) :1519–1533, 2010.
- [116] B Chevreux, T Wetter, S Suhai, et al. Genome sequence assembly using trace signals and additional sequence information. In *German conference on bioinformatics*, volume 99, pages 45–56. Hanover, Germany, 1999.
- [117] A Bankevich, S Nurk, D Antipov, AA Gurevich, M Dvorkin, AS Kulikov, VM Lesin, SI Nikolenko, S Pham, AD Prjibelski, et al. SPAdes : a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5) :455–477, 2012.
- [118] H Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv :1303.3997*, 2013.
- [119] K Pependorf and Y Sakakibara. SAMSCOPE : an OpenGL-based real-time interactive scale-free SAM viewer. *Bioinformatics*, 28(9) :1276–1277, 2012.
- [120] E Garrison and G Marth. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv :1207.3907*, 2012.
- [121] Y Liao, GK Smyth, and W Shi. featureCounts : an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7) :923–930, 2013.
- [122] M Gouy, S Guindon, and O Gascuel. Seaview version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2) :221–224, 2009.
- [123] S Guindon, JF Dufayard, V Lefort, M Anisimova, W Hordijk, and O Gascuel. New algorithms and methods to estimate maximum-likelihood phylogenies : assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3) :307–321, 2010.
- [124] A Stamatakis. RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9) :1312–1313, 2014.
- [125] F Ronquist, M Teslenko, P Van Der Mark, DL Ayres, A Darling, S Höhna, B Larget, L Liu, MA Suchard, and JP Huelsenbeck. Mrbayes 3.2 : efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3) :539–542, 2012.
- [126] AJ Drummond and A Rambaut. BEAST : Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1) :214, 2007.
- [127] JB Pettengill. The time to most recent common ancestor does not (usually) approximate the date of divergence. *PloS one*, 10(8) :e0128407, 2015.

- [128] X Bertrand, P Bailly, G Blasco, P Balvay, A Boillot, and D Talon. Large outbreak in a surgical intensive care unit of colonization or infection with *Pseudomonas aeruginosa* that overexpressed an active efflux pump. *Clinical Infectious Diseases*, 31(4) :e9–e14, 2000.
- [129] B Valot, L Rohmer, MA Jacobs, SI Miller, X Bertrand, and D Hocquet. Comparative genomic analysis of two multidrug-resistant clinical isolates of ST395 epidemic strain of *Pseudomonas aeruginosa* obtained 12 years apart. *Genome Announcements*, 2(3) :e00515–14, 2014.
- [130] AJ Morrison Jr and RP Wenzel. Epidemiology of infections due to *Pseudomonas aeruginosa*. *Reviews of Infectious Diseases*, 6(Supplement_3) :S627–S642, 1984.
- [131] JA Driscoll, SL Brody, and MH Kollef. The epidemiology, pathogenesis and treatment of *Pseudomonas aeruginosa* infections. *Drugs*, 67(3) :351–368, 2007.
- [132] P Cholley, M Thouverez, D Hocquet, N Van Der Mee-Marquet, D Talon, and X Bertrand. The majority of multi-drug resistant *Pseudomonas aeruginosa* isolates from hospitals in eastern France belongs to a few clonal types. *Journal of clinical microbiology*, pages JCM.00102–11, 2011.
- [133] MC Wolfgang, BR Kulasekara, X Liang, D Boyd, K Wu, Q Yang, CG Miyada, and S Lory. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences*, 100(14) :8484–8489, 2003.
- [134] CS Chin, DH Alexander, P Marks, AA Klammer, J Drake, C Heiner, A Clum, A Copeland, J Huddleston, AE Eichler, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, 10(6) :563–569, 2013.
- [135] MJ Brittnacher, C Fong, HS Hayden, MA Jacobs, M Radey, and L Rohmer. PGAT : a multistrain analysis resource for microbial genomes. *Bioinformatics*, 27(17) :2429–2430, 2011.
- [136] D Talon, MJ Dupont, J Lesne, M Thouverez, and Y Michel-Briand. Pulsed-field gel electrophoresis as an epidemiological tool for clonal identification of *Aeromonas hydrophila*. *Journal of Applied Microbiology*, 80(3) :277–282, 1996.
- [137] Center for Genomic Epidemiology, MLST. <https://cge.cbs.dtu.dk//services/MLST/>.
- [138] AL Delcher, A Phillippy, J Carlton, and SL Salzberg. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, 30(11) :2478–2483, 2002.
- [139] BK Dhillon, TA Chiu, MR Laird, MGI Langille, and FSL Brinkman. IslandViewer update : improved genomic island discovery and visualization. *Nucleic Acids Research*, 41(W1) :W129–W132, 2013.
- [140] I Grissa, G Vergnaud, and C Pourcel. CRISPRFinder : a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 35(suppl_2) :W52–W57, 2007.
- [141] D Hocquet, C Llanes, M Thouverez, HD Kulasekara, X Bertrand, P Plésiat, D Mazel, and SI Miller. Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS pathogens*, 8(6) :e1002778, 2012.

- [142] R Leclercq, R Cantón, DFJ Brown, CG Giske, P Heisig, AP MacGowan, JW Mouton, P Nordmann, AC Rodloff, GM Rossolini, et al. EUCAST expert rules in antimicrobial susceptibility testing. *Clinical Microbiology and Infection*, 19(2) :141–160, 2013.
- [143] E Zankari, H Hasman, S Cosentino, M Vestergaard, S Rasmussen, O Lund, FM Aarestrup, and MV Larsen. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, 67(11) :2640–2644, 2012.
- [144] L Chen, J Yang, J Yu, Z Yao, L Sun, Y Shen, and Q Jin. VFDB : a reference database for bacterial virulence factors. *Nucleic Acids Research*, 33(suppl_1) :D325–D328, 2005.
- [145] SF Altschul, W Gish, W Miller, EW Myers, and DJ Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3) :403–410, 1990.
- [146] Center for Genomic Epidemiology, PAST. <https://cge.cbs.dtu.dk//services/Past-1.0/>.
- [147] M Juhas, JR van der Meer, M Gaillard, RM Harding, DW Hood, and DW Crook. Genomic islands : tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, 33(2) :376–393, 2009.
- [148] A van Belkum, LB Soriaga, MC LaFave, S Akella, JB Veyrieras, EM Barbu, D Shortridge, B Blanc, G Hannum, G Zambardi, et al. Phylogenetic distribution of CRISPR-cas systems in antibiotic-resistant *Pseudomonas aeruginosa*. *MBio*, 6(6) :e01796–15, 2015.
- [149] DG Lee, JM Urbach, G Wu, NT Liberati, RL Feinbaum, S Miyata, LT Diggins, J He, M Saucier, E Déziel, et al. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biology*, 7(10) :R90, 2006.
- [150] R Morgan, S Kohn, SH Hwang, DJ Hassett, and K Sauer. BdlA, a chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 188(21) :7335–7343, 2006.
- [151] P Williams and M Cámara. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa* : a tale of regulatory networks and multifunctional signal molecules. *Current Opinion in Microbiology*, 12(2) :182–191, 2009.
- [152] M Hentzer, H Wu, JB Andersen, K Riedel, TB Rasmussen, N Bagge, N Kumar, MA Schembri, Z Song, P Kristoffersen, et al. Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. *The EMBO journal*, 22(15) :3803–3815, 2003.
- [153] L Rohmer, D Hocquet, and SI Miller. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends in Microbiology*, 19(7) :341–348, 2011.
- [154] D Hocquet, X Bertrand, T Köhler, D Talon, and P Plésiat. Genetic and phenotypic variations of a resistant *Pseudomonas aeruginosa* epidemic clone. *Antimicrobial Agents and Chemotherapy*, 47(6) :1887–1894, 2003.
- [155] X Qiu, AU Gurkar, and S Lory. Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences*, 103(52) :19830–19835, 2006.
- [156] MJ Wargo. Homeostasis and catabolism of choline and glycine betaine : lessons from *Pseudomonas aeruginosa*. *Applied and Environmental Microbiology*, 79(7) :2112–2120, 2013.

- [157] KP Allen, MM Randolph, and JM Fleckenstein. Importance of heat-labile enterotoxin in colonization of the adult mouse small intestine by human enterotoxigenic *Escherichia coli* strains. *Infection and Immunity*, 74(2) :869–875, 2006.
- [158] F Arnesano, L Banci, I Bertini, and AR Thompsett. Solution structure of CopC : a cupredoxin-like protein involved in copper homeostasis. *Structure*, 10(10) :1337–1347, 2002.
- [159] CA Navarro, D von Bernath, C Martínez-Bussenius, RA Castillo, and CA Jerez. Cytoplasmic CopZ-like protein and periplasmic rusticyanin and AcoP proteins as possible copper resistance determinants in *Acidithiobacillus ferrooxidans* ATCC 23270. *Applied and Environmental Microbiology*, 82(4) :1015–1022, 2016.
- [160] JS Cha and DA Cooksey. Copper hypersensitivity and uptake in *Pseudomonas syringae* containing cloned components of the copper resistance operon. *Applied and Environmental Microbiology*, 59(5) :1671–1674, 1993.
- [161] GM Teitzel, A Geddie, K Susan, MJ Kirisits, M Whiteley, and MR Parsek. Survival and growth in the presence of elevated copper : transcriptional profiling of copper-stressed *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 188(20) :7242–7256, 2006.
- [162] K Bondarczuk and Z Piotrowska-Seget. Molecular basis of active copper resistance mechanisms in Gram-negative bacteria. *Cell Biology and Toxicology*, 29(6) :397–405, 2013.
- [163] L Molina, P Bernal, Z Udaondo, A Segura, and JL Ramos. Complete genome sequence of a *Pseudomonas putida* clinical isolate, strain H8234. *Genome Announcements*, 1(4) :e00496–13, 2013.
- [164] DA Cooksey. Genetics of bactericide resistance in plant pathogenic bacteria. *Annual Review of Phytopathology*, 28(1) :201–219, 1990.
- [165] A Fernández-Olmos, M García-Castillo, JM Alba, MI Morosini, A Lamas, B Romero, JC Galán, R del Campo, and R Cantón. Population structure and antimicrobial susceptibility of both non-persistent and persistent *Pseudomonas aeruginosa* isolates recovered in cystic fibrosis patients. *Journal of Clinical Microbiology*, pages JCM.00802–13, 2013.
- [166] K Martin, B Baddal, N Mustafa, C Perry, A Underwood, C Constantidou, N Loman, DT Kenna, and JF Turton. Clusters of genetically similar isolates of *Pseudomonas aeruginosa* from multiple hospitals in the UK. *Journal of Medical Microbiology*, 62(7) :988–1000, 2013.
- [167] J Bondy-Denomy and AR Davidson. To acquire or resist : the complex biological effects of CRISPR–cas systems. *Trends in Microbiology*, 22(4) :218–225, 2014.
- [168] A Pawluk, J Bondy-Denomy, VHW Cheung, KL Maxwell, and AR Davidson. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR–cas system of *Pseudomonas aeruginosa*. *MBio*, 5(2) :e00896–14, 2014.
- [169] A Gurevich, V Saveliev, N Vyahhi, and G Tesler. Quast : quality assessment tool for genome assemblies. *Bioinformatics*, 29(8) :1072–1075, 2013.
- [170] M Kolmogorov, B Raney, B Paten, and S Pham. Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*, 30(12) :i302–i309, 2014.

- [171] S Kurtz, A Phillippy, AL Delcher, M Smoot, M Shumway, C Antonescu, and SL Salzberg. Versatile and open software for comparing large genomes. *Genome Biology*, 5(2) :R12, 2004.
- [172] A-M Rogues, H Boulestreau, A Lashéras, A Boyer, D Gruson, C Merle, Y Castaing, CM Bébear, and J-P Gachie. Contribution of tap water to patient colonisation with *Pseudomonas aeruginosa* in a medical intensive care unit. *Journal of Hospital Infection*, 67(1) :72–78, 2007.
- [173] M Trautmann, C Bauer, C Schumann, P Hahn, M Höher, M Haller, and PM Lepper. Common RAPD pattern of *Pseudomonas aeruginosa* from patients and tap water in a medical intensive care unit. *International Journal of Hygiene and Environmental Health*, 209(4) :325–331, 2006.
- [174] M Halwani, M Solaymani-Dodaran, H Grundmann, C Coupland, and R Slack. Cross-transmission of nosocomial pathogens in an adult intensive care unit : incidence and risk factors. *Journal of Hospital Infection*, 63(1) :39–46, 2006.
- [175] M Petitjean, D Martak, A Silvant, X Bertrand, B Valot, and D Hocquet. Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395. *Microbial Genomics*, 2017.
- [176] MV Larsen, S Cosentino, S Rasmussen, C Friis, H Hasman, RL Marvig, L Jelsbak, TS Pontén, DW Ussery, FM Aarestrup, et al. Multilocus sequence typing of total genome sequenced bacteria. *Journal of Clinical Microbiology*, pages JCM-06094, 2012.
- [177] D Hyatt, GL Chen, PF LoCascio, ML Land, FW Larimer, and LJ Hauser. Prodigal : prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1) :119, 2010.
- [178] F Rohart, B Gautier, A Singh, and K-A Le Cao. mixOmics : an R package for omics feature selection and multiple data integration. *bioRxiv*, page 108597, 2017.
- [179] panISa. <https://github.com/panisa>. Mise en ligne en 2017.
- [180] D Hocquet, M Petitjean, L Rohmer, B Valot, HD Kulasekara, E Bedel, X Bertrand, P Plésiat, T Köhler, A Pantel, et al. Pyomelanin-producing *Pseudomonas aeruginosa* selected during chronic infections have a large chromosomal deletion which confers resistance to pyocins. *Environmental Microbiology*, 18(10) :3482–3493, 2016.
- [181] NE Van Alst, KF Picardo, BH Iglewski, and CG Haidaris. Nitrate sensing and metabolism modulate motility, biofilm formation, and virulence in *Pseudomonas aeruginosa*. *Infection and Immunity*, 75(8) :3780–3790, 2007.
- [182] SR Thomas, A Ray, ME Hodson, and TL Pitt. Increased sputum amino acid concentrations and auxotrophy of *Pseudomonas aeruginosa* in severe cystic fibrosis lung disease. *Thorax*, 55(9) :795–797, 2000.
- [183] MA Oberhardt, JB Goldberg, M Hogardt, and JA Papin. Metabolic network analysis of *Pseudomonas aeruginosa* during chronic cystic fibrosis lung infection. *Journal of Bacteriology*, 192(20) :5534–5548, 2010.
- [184] T Jatsenko, A Tover, R Tegova, and M Kivisaar. Molecular characterization of Rif^r mutations in *Pseudomonas aeruginosa* and *Pseudomonas putida*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 683(1) :106–114, 2010.

- [185] A Castañeda-García, A Rodríguez-Rojas, JR Guelfo, and J Blázquez. The glycerol-3-phosphate permease GlpT is the only fosfomycin transporter in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 191(22) :6968–6974, 2009.
- [186] VN De Groote, M Fauvart, CI Kint, N Verstraeten, A Jans, P Cornelis, and J Michiels. *Pseudomonas aeruginosa* fosfomycin resistance mechanisms affect non-inherited fluoroquinolone tolerance. *Journal of Medical Microbiology*, 60(3) :329–336, 2011.
- [187] J-Y Lee and KS Ko. OprD mutations and inactivation, expression of efflux pumps and AmpC, and metallo- β -lactamases in carbapenem-resistant *Pseudomonas aeruginosa* isolates from South Korea. *International Journal of Antimicrobial Agents*, 40(2) :168–172, 2012.
- [188] EK Karlsson, DP Kwiatkowski, and PC Sabeti. Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6) :379–393, 2014.
- [189] Z-X Tian, X-X Yi, A Cho, F O’Gara, and Y-P Wang. CpxR activates MexAB-OprM efflux pump expression and enhances antibiotic resistance in both laboratory and clinical *nalB*-type isolates of *Pseudomonas aeruginosa*. *PLoS Pathogens*, 12(10) :e1005932, 2016.
- [190] C Llanes, D Hocquet, C Vogne, D Benali-Baitich, C Neuwirth, and P Plésiat. Clinical strains of *Pseudomonas aeruginosa* overproducing mexab-oprm and mexxy efflux pumps simultaneously. *Antimicrobial Agents and Chemotherapy*, 48(5) :1797–1802, 2004.
- [191] L Adewoye, A Sutherland, R Srikumar, and K Poole. The *mexR* repressor of the *mexAB-oprM* multidrug efflux operon in *Pseudomonas aeruginosa* : characterization of mutations compromising activity. *Journal of Bacteriology*, 184(15) :4308–4312, 2002.
- [192] JD King, D Kocíncová, EL Westman, and JS Lam. Lipopolysaccharide biosynthesis in *i*. *Innate Immunity*, 15(5) :261–312, 2009.
- [193] ANDERS Fomsgaard, RS Conrad, C Galanos, GH Shand, and N Høiby. Comparative immunochemistry of lipopolysaccharides from typable and polyagglutinable *Pseudomonas aeruginosa* strains isolated from patients with cystic fibrosis. *Journal of Clinical Microbiology*, 26(5) :821–826, 1988.
- [194] SM Moskowitz and RK Ernst. The role of *Pseudomonas* lipopolysaccharide in cystic fibrosis airway infection. In *Endotoxins : Structure, Function and Recognition*, pages 241–253. Springer, 2010.
- [195] P Tang, MA Croxen, MR Hasan, WWL Hsiao, and LM Hoang. Infection control in the new age of genomic epidemiology. *American Journal of Infection Control*, 45(2) :170–179, 2017.
- [196] P Salunkhe, CHM Smart, JAW Morgan, S Panagea, MJ Walshaw, CA Hart, R Gefers, B Tümmler, and C Winstanley. A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance. *Journal of Bacteriology*, 187(14) :4908–4920, 2005.
- [197] L Jelsbak, HK Johansen, AL Frost, R Thøgersen, LE Thomsen, O Ciofu, L Yang, JAJ Haagenen, N Høiby, and S Molin. Molecular epidemiology and dynamics of *Pseudomonas aeruginosa* populations in lungs of cystic fibrosis patients. *Infection and Immunity*, 75(5) :2214–2224, 2007.

- [198] N Blanco, MC Eisenberg, T Stillwell, and B Foxman. What transmission precautions best control influenza spread in a hospital? *American Journal of Epidemiology*, 183(11) :1045–1054, 2016.
- [199] S Kundu, CG Faulkes, AG Greenwood, CG Jones, P Kaiser, OD Lyne, SA Black, A Chowrimootoo, and JJ Groombridge. Tracking viral evolution during a disease outbreak : the rapid and complete selective sweep of a circovirus in the endangered echo parakeet. *Journal of Virology*, 86(9) :5221–5229, 2012.

Annexe A

Article DHS01

Petitjean, M., Martak, D., Silvant, A., Bertrand, X., Valot, B., & Hocquet, D. (2017). Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread ST395. *Microbial Genomics*.

Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395

Marie Petitjean,^{1,2} Daniel Martak,^{1,2} Alicia Silvant,¹ Xavier Bertrand,^{1,2} Benoit Valot² and Didier Hocquet^{1,2,3,*}

Abstract

Pseudomonas aeruginosa is a ubiquitous opportunistic pathogen with several clones being frequently associated with outbreaks in hospital settings. ST395 is among these so-called 'international' clones. We aimed here to define the biological features that could have helped the implantation and spread of the clone ST395 in hospital settings. The complete genome of a multidrug resistant index isolate (DHS01) of a large hospital outbreak was analysed. We identified DHS01-specific genetic elements, among which were identified those shared with a panel of six independent ST395 isolates responsible for outbreaks in other hospitals. DHS01 has the fifth largest chromosome of the species (7.1 Mbp), with most of its 1555 accessory genes borne by either genomic islands (GIs, $n=48$) or integrative and conjugative elements (ICEs, $n=5$). DHS01 is multidrug resistant mostly due to chromosomal mutations. It displayed signatures of adaptation to chronic infection in part due to the loss of a 131 kbp chromosomal fragment. Four GIs were specific to the clone ST395 and contained genes involved in metabolism (GI-4), in virulence (GI-6) and in resistance to copper (GI-7). GI-7 harboured an array of six copper transporters and was shared with non-pathogenic *Pseudomonas* sp. retrieved from copper-contaminated environments. Copper resistance was confirmed phenotypically in all other ST395 isolates and possibly accounted for the spreading capability of the clone in hospital outbreaks, where water networks have been incriminated. This suggests that genes transferred from copper-polluted environments may have favoured the implantation and spread of the international clone *P. aeruginosa* ST395 in hospital settings.

DATA SUMMARY

1. This whole-genome sequencing project has been deposited at GenBank/EMBL/DDBJ; accession number: CP013993.1 (url – <https://www.ncbi.nlm.nih.gov/nucore/CP013993.1>).
2. The collections of ST395 and non-ST395 isolates have been deposited at GenBank/EMBL/DDBJ; Bio-Projects: PRJNA379554 (url – <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA379554>) and PRJNA380885 (url – <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA380885>) (Table S2, available in the online Supplementary Material).
3. The 82 *Pseudomonas aeruginosa* isolates for which complete genomes were available from the National Center for Biotechnology Information at the time of the study (June 2017) were included in the study.

INTRODUCTION

Pseudomonas aeruginosa is an opportunistic human pathogen responsible for outbreaks in hospital settings [1]. This bacterial pathogen infects injured, burned, immunodeficient, immunocompromised and cystic fibrosis patients [2]. Despite the non-clonal epidemic population structure of *P. aeruginosa*, several multi-locus sequence types (STs) spread worldwide and are frequently associated with epidemics where multidrug resistance confounds treatment. ST111, ST175, ST235 and ST395 are among these so-called 'international' or 'high-risk' clones [3]. The biological features that make these clones more epidemic than other STs are largely unknown.

We focused here on the high-risk clone ST395, for which spread has been documented within hospitals in France and

Received 12 April 2017; Accepted 23 July 2017

Author affiliations: ¹Hygiène Hospitalière, Centre Hospitalier Régional Universitaire de Besançon, Besançon, France; ²UMR CNRS 6249 Chrono-Environnement, Université de Bourgogne Franche-Comté, Besançon, France; ³Centre de Ressources Biologiques - Filière Microbiologique, Centre Hospitalier Universitaire de Besançon, 3 Boulevard Fleming, Besançon, France.

*Correspondence: Didier Hocquet, dhocquet@chu-besancon.fr

Keywords: *Pseudomonas aeruginosa*; multidrug resistance; copper; outbreak; high-risk clone.

Abbreviations: CRISPR-Cas, clustered regularly interspaced short palindromic repeats-CRISPR associated; GI, genomic island; ICE, integrative and conjugative element; NCBI, National Center for Biotechnology Information; ncRNA, non-coding RNA; QRDR, quinolone-resistance-determining region; ST, sequence type.

Datastatement: We confirm all supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary tables and four supplementary figures are available with the online Supplementary Material.

in the UK [4, 5]. From 1997 to 2008, a strain of *P. aeruginosa* ST395 has infected or colonized more than 300 patients in the University Hospital of Besançon (Besançon, France) [6]. Similarly, the clone ST395 was responsible for an outbreak in a UK hospital. The authors suggested that *P. aeruginosa* ST395 rapidly colonized the plumbing system of a new hospital, before its commissioning, and was transmitted to burn patients with the water [5]. Besides this, the clone ST395 has also been retrieved in a plant from a metal-contaminated estuary, and harboured resistance determinants to mercury, arsenic and copper, together with antibiotic-resistance genes [7].

P. aeruginosa strains adapt to environments through the gain or loss of blocks of genes [i.e. horizontal gene transfers of genomic islands (GIs) or integrative and conjugative elements (ICEs)] and modifications at the nucleotide level. Acquired genes originate from either clinical or environmental species [8].

The virulence determinants are mostly part of the core genome of the species, accounting for the ability of *P. aeruginosa* to inhabit a wide range of environments, including patients [9]. Moreover, the co-selection of antibiotic-resistance genes with metal-resistance genes can promote the ability of strains adapted to toxic environments to thrive in hospital settings, where the antibiotic consumption is high [10].

We aimed here to define the features that could have helped the implantation and spread of the clone *P. aeruginosa* ST395 in hospital settings. We first thoroughly analysed the complete genome of the multidrug-resistant index isolate of an outbreak in France (isolate DHS01), and identified genetic elements specific to this isolate, paying particular attention to the resistome, the virulome and mobile genetic elements. Among these specific elements, we further identified those shared with a panel of independent ST395 isolates responsible for outbreaks in other hospitals.

METHODS

Genome sequencing of the index isolate

The isolate DHS01, the first isolate of the epidemic, was isolated in May 1997 from the nose of a 31-year-old male patient in the surgical intensive care unit in the University Hospital of Besançon (Eastern France). In our hospital, *P. aeruginosa* is searched for weekly in patients hospitalized in the intensive care unit. Microbiological records revealed that the patient was already colonized by *P. aeruginosa* 25 days before isolation of DHS01. DHS01 was identified at the species level by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) with a Microflex LT (Bruker Daltonik), according to the manufacturer's procedures. Bacterial DNA was isolated from an overnight culture on Mueller-Hinton (MH) agar using a Genomic-tip kit (Qiagen) and further sequenced with Single Molecule Real-Time technology (Pacific Biosciences) [11]. The reads were assembled using Hgap3 into

IMPACT STATEMENT

The opportunist pathogen *Pseudomonas aeruginosa* has a panmictic population structure, with a few international clones being frequently associated with outbreaks in hospitals worldwide. The biological features that favour the implantation and spread of these clones in hospital settings are largely unknown. ST395, one of these clones, has been retrieved from polluted environments and found to be responsible for outbreaks in hospitals. The genomic analysis of a collection of independent ST395 isolates identified four specific genomic islands (GIs) that contained genes involved in metabolism, virulence and the resistance to copper. An array of six copper transporters was borne on a single GI that was shared with non-pathogenic *Pseudomonas* sp. found in copper-contaminated environments. The systematic resistance to copper of ST395 isolates possibly accounted for their spreading capability in hospital outbreaks, where water networks have been incriminated. This observation could help implement infection control measures to tackle the proliferation of copper-resistant pathogenic bacteria in water networks.

a circular complete genome [12]. The genome was annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) [13]. The genes were sorted as core or accessory genes of the species *P. aeruginosa* by comparing to previous pangenome analysis [14].

Dataset

We compared the DHS01 genome against two datasets. Dataset 1 included 82 *P. aeruginosa* isolates whose complete genomes were available on NCBI at the time of the study (June 2017; Table S1). Dataset 2 included the raw reads of 6 ST395 isolates of *P. aeruginosa* and 10 non-ST395 isolates of *P. aeruginosa* isolated during outbreaks in French or UK hospitals (Table S2) sequenced with the Illumina NextSeq technology with a mean depth of 200X coverage per genome. The absence of close relatedness between the ST395 isolates ($n=7$) was determined by pulsed-field gel electrophoresis, as described elsewhere (Fig. S1) [15].

Phylogenetic analysis

We determined the ST of DHS01 with the Center for Genomic Epidemiology website (2017; <https://cge.cbs.dtu.dk/services/MLST/>). Then, the DHS01 genome was compared with the genomes of dataset 1 (Table S1) with MUMmer [16]. The length of the core genome alignment was 1770 651 bp, representing 1973 genes with 118 840 variable positions. The phylogenetic tree was determined using PhyML software [17] with the following parameters: GTR (general time reversible) substitution model with gamma-distributed rate variation across sites and a proportion of invariable sites.

Mobile genetic elements

GIs and CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) systems were located with IslandViewer and CRISPRfinder, respectively [18, 19]. ICEs were retrieved from gene annotation (i.e. containing an integrase-encoding gene at the 5'-end, genes encoding the full conjugation machinery *traG* and *trbBCDEFGHI*, and a relaxase). Free circular forms of the ICEs were detected in the DNA extracted from SOS-induced bacteria by nested PCR using primers that overlapped the *attI* sequence (Table S3). The SOS response was induced by mitomycin, as previously described [20]. To identify ST395-specific GIs and ICEs, we mapped the Illumina reads from a collection of ST395 and non-ST395 isolates (Table S1) on the genome of DHS01. The presence of GIs and ICEs was determined with featureCount.

Identification of resistance and virulence determinants

The susceptibility to the major antipseudomonal agents (ticarcillin, piperacillin, cefepime, ceftazidime, aztreonam, imipenem, amikacin, gentamicin, tobramycin and ciprofloxacin) was tested using the European Committee on Antimicrobial Susceptibility Testing recommendations with the reference agar dilution method, with the reference strain PAO1 as a control [21]. Resistance genes were searched for with ResFinder [22]. We searched for chromosomal mutations, in comparison with the genome of PAO1, responsible for resistance to antibiotics in the gene *oprD*, those encoding the main efflux systems (MexAB-OprM, MexXY, MexCD-OprJ, MexEF-OprN) and their regulators (MexR, MexZ, ArmZ, NfxB, NalC, NalD, ParS, ParR, MexS and MexT), the quinolone-resistance-determining regions (QRDRs; in *gyrA*, *gyrB*, *parC*, *parE*), the gene encoding the AmpC cephalosporinase (*ampC*) and its regulators (*dacB*, *ampR*, *ampD*, *ampDh2*, *ampDh3*, *ampG*, *ampO* and *ampP*), the regulators of polymyxin resistance (*colR*, *cols*, *cprR*, *cprS*, *pmrA* and *pmrB*) and the *phoP-phoQ* genes associated with colistin resistance. Virulence genes were searched for using the virulence factor database (VFDB) [23]. Mutations in resistance and virulence genes were detected using BLAST [24]. We first determined the serotype of DHS01 *in silico* using the PAST (*P. aeruginosa* serotyper) online tool from the Center for Genomic Epidemiology (2015; <https://cge.cbs.dtu.dk/services/Past-1.0/>) and confirmed the result with specific antiserum (BioRad).

Survival in copper solution

The bacterial survival in 7.5 mM CuSO₄ was determined by the inoculation of cation-calibrated MH broth (Becton-Dickinson) with 10⁷ c.f.u. stationary-phase cells ml⁻¹ and incubation for 48 h at 37°C under shaking conditions (150 r.p.m.). The initial and final concentrations of living bacteria were determined by counting on MH agar plates. Experiments were performed at least three times independently.

RESULTS AND DISCUSSION

Genome assembly

The sequencing of the DHS01 genome generated 98 987 passed filter reads, with a mean read length of 12 kb and a mean coverage of 134×. The quality of the data allowed the assembly of the DHS01 genome in a complete circular chromosome of 7 055 752 bp (Fig. 1). Plasmids were absent from the genome. PGAT annotation predicted 6524 coding genes, 67 tRNAs, 5 non-coding RNAs (ncRNAs) and 12 rRNAs (4 5S, 4 16S and 4 23S). The G+C content of the DHS01 genome (65.8 mol%) was low compared with that usually found in the species (~66.6 mol%). The genes acquired by horizontal transfer have generally a lower G+C content than that of the core genome. The size of the core genome of *P. aeruginosa* has been evaluated at 5234 genes [14]. By clustering with CD-HIT (cluster database at high identity with tolerance), we found that 1555 genes formed the accessory genome of DHS01. We found that 866 out of 1555 accessory genes (56%) were borne by either GIs or ICEs, confirming that these mobile genetic elements were a major source of accessory genome and genetic diversity [8, 25]. The distribution of core and accessory genes is indicated in the Fig. 1.

Phylogenetic analysis

DHS01 belonged to ST395 and was predicted to be of serotype O6. DHS01 has the fifth largest chromosome when compared with those of other isolates of the species, suggesting the recent integration of genetic material in its genome. The three largest chromosomes of the species belonged to ST111 isolates, the fourth largest belonged to a ST357 isolate and the sixth largest belonged to a ST235 isolate. Interestingly, all these STs are responsible for large outbreaks in hospital settings. This observation was in line with recent data showing that the genomes of ST111, ST395 and ST235 isolates were of 7.05, 7.00 and 6.81 Mbp, respectively, while the mean genome of the species was 6.72 Mbp in size [26].

The phylogenetic analysis of 83 *P. aeruginosa* isolates (DHS01 and those of dataset 1) confirmed the split into two clusters containing either reference strains PAO1 or PA14 (Fig. 2) [27]. When compared to the PAO1 genome, the DHS01 genome expectedly displayed the major chromosomal inversion between two copies of ribosomal RNA reported in PA14 and all the genomes tested by Lee *et al.* [27]. DHS01 had 5833 and 5837 SNPs with its closest isolates LESB58 and LES431, respectively, that belonged to the highly virulent epidemic strain (LES, Liverpool Epidemic Strain) [28].

DHS01 has undergone a 131 kb chromosomal deletion

DHS01 was the first isolate of a 11 year outbreak in the University Hospital of Besançon and was then considered as an ancestor from which the late isolates derived [6]. However, when compared with the genome of DHS29, a

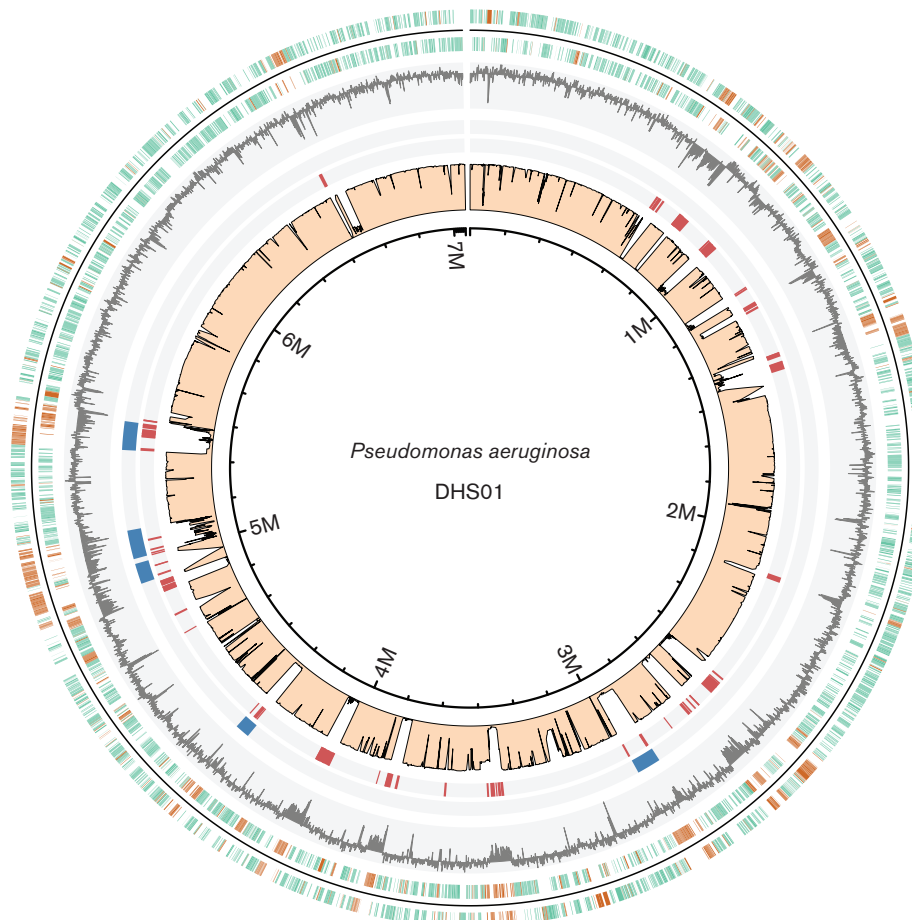


Fig. 1. Circle plot of the complete genome of the epidemic *P. aeruginosa* DHS01 (ST395). The first and second outer circles indicate the coding DNA sequences on the plus and the minus strand, respectively. Genes in green are part of the core genome of the species and genes in brown are accessory genes. The third circle displays the G+C content (deviation from the mean) of the DHS01 genome. The two next circles show the location of GIs (in red) and ICEs (in blue). The pink innermost circle displays the numbers of genomes (of the 82 genomes of the dataset 1) containing this sequence. We drew the circle plot with R software.

late isolate of the outbreak, the DHS01 genome had undergone a 131 kb chromosomal deletion [29]. This segment was intact in the genomes of five of six independent ST395 isolates (dataset 2, Table S2). The 131 kb gene block included the genomic regions PA1246–PA1292 and PA1381–PA1450 in the reference strain PAO1, and encompassed the virulence genes encoding the alkaline protease AprA (that degrades components of the host immune system) and its secretion system AprDEFI. The deletion also included the genes encoding the biofilm dispersion regulator BdlA, the quorum-sensing master regulator LasR, the LasI enzyme and the quorum sensing repressor RsaL, and the flagellar operon region including *fliKLMNOPQR* and *flhB* [30, 31]. Based on published phenotypic observations, one can hypothesize that the loss of these genes certainly affects the biofilm formation, the virulence, the adherence and the motility of DHS01 [32, 33].

DHS01 virulence factors

We compared the virulome of DHS01 with that of the reference strain PAO1 using the virulence factor database (VFDB). The 347 virulence genes found in DHS01 were also retrieved from PAO1. Unexpectedly, 36 virulence genes present in PAO1 and part of the core genome of the species were missing from the DHS01 genome [14]. The 131 kb chromosomal deletion described above included 18 of these genes. Other genes missing from the DHS01 chromosome were implicated in bacterial adherence and biofilm formation. Hence, eight virulence genes contiguous in PAO1 (*flgK*, *flgL*, *fgtA*, *fliC*, *flaG*, *fliD*, *fliS* and *fleP*) and forming a block of 14 846 nt were missing from DHS01, probably as the result of a deletion ‘*en bloc*’. Besides this, the genes encoding the chaperonins CupE3 and CupC1 displayed 6 nt and 5 nt deletions, respectively, when compared with PAO1. Genes implicated in the synthesis of type IV pili were missing (*pilA*, only 17 % of the sequence was retrieved)

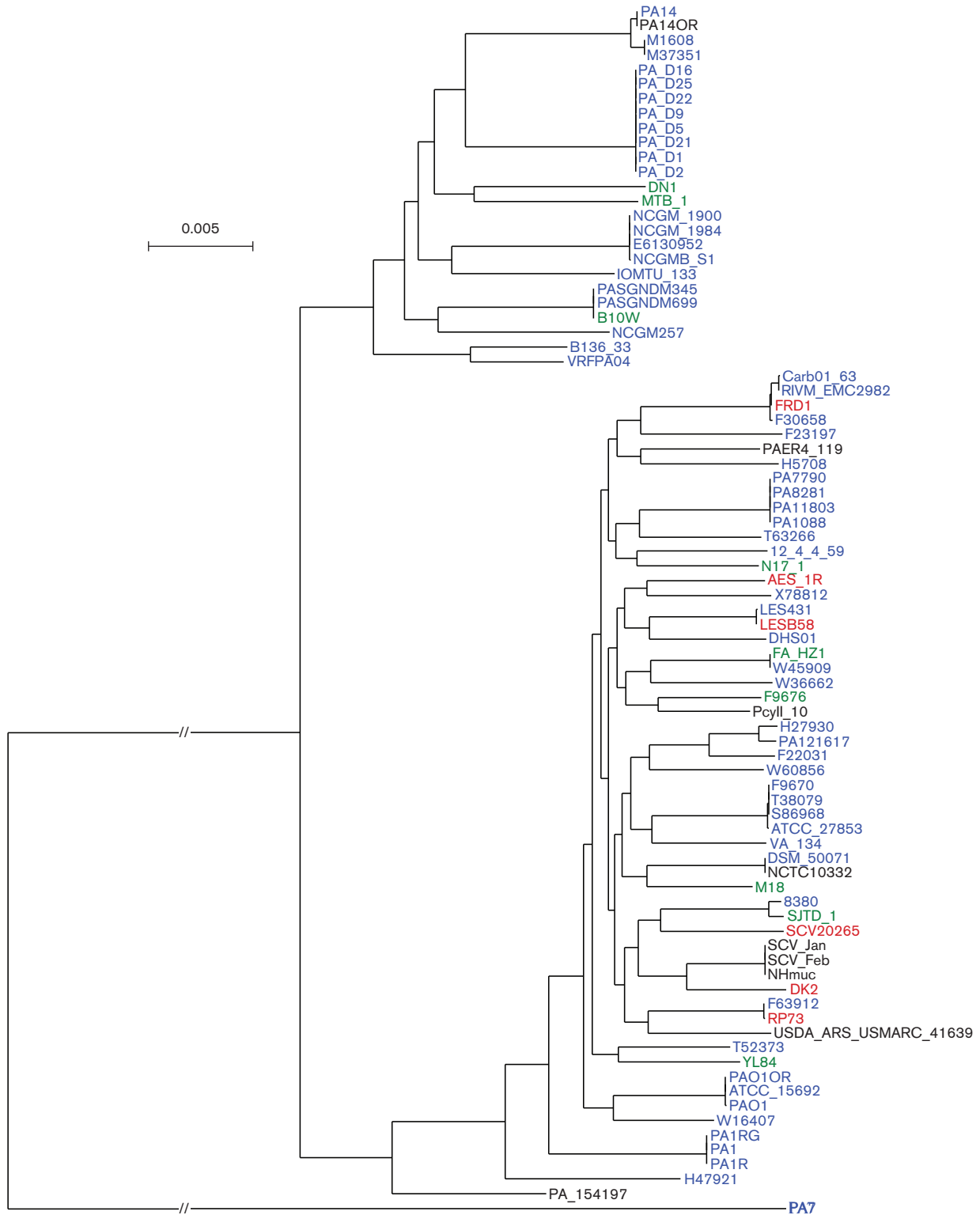


Fig. 2. Phylogenetic tree of 83 isolates of *P. aeruginosa* with complete genomes, including DHS01. The phylogenetic tree is well supported, with the majority of branch support values higher than 98. The isolates cultured from cystic fibrosis patients are coloured red, non-cystic fibrosis clinical isolates are coloured blue, isolates of environmental origin are written in green and isolates of other origin are written in black. Bar, 0.005 substitution per site.

or mutated (*pilB* and *pilC*). The deletion of the type IV pili-related *pilA* gene has been shown to impair the biofilm formation [32].

DHS01 surprisingly harboured less virulence genes than the reference strain PAO1, suggesting a lower virulence *in vivo*. This contrasts with its status of international epidemic clone. DHS01 was isolated after 25 days of patient colonization and displayed the genomic signatures of a strain adapted to chronic colonization (i.e. loss of flagellum, and deficiency in twitching motility, in the quorum sensing and in biofilm formation) [32, 34]. Although being the first isolate of the epidemic, DHS01 was likely an epidemiological dead end and was not representative of the strain that has spread to more than 300 patients [35].

DHS01 multidrug resistance is mostly due to chromosomal mutations

DHS01 was multidrug resistant, with a high level of resistance to gentamicin (MIC >512 mg l⁻¹), to tobramycin (MIC 32 mg l⁻¹), to ciprofloxacin (MIC 32 mg l⁻¹) and to β -lactam compounds (ticarcillin MIC 64 mg l⁻¹; piperacillin MIC 128 mg l⁻¹). It was of intermediate resistance to aztreonam (MIC 16 mg l⁻¹), and remained susceptible to cefepime (MIC 8 mg l⁻¹), ceftazidime (MIC 8 mg l⁻¹), imipenem (MIC 4 mg l⁻¹) and amikacin (MIC 2 mg l⁻¹). The acquired genes and mutations in regulators leading to antibiotic resistance are detailed in Table 1. We had previously shown that DHS01 overproduced AmpC (650-times more than PAO1), and overexpressed *mexA* and *mexX* (1.85- and 3.70-times more than PAO1, respectively) [36]. Briefly, DHS01 produced a 2'-aminoglycoside nucleotidyl-transferase, overproduced an extended-spectrum AmpC cephalosporinase (*ampR* and *ampD* mutant), overproduced MexAB-OprM (*nalB* mutant) and MexXY (*agrZ* mutant), presumably overproduced MexEF-OprN (*mexS* mutant), and displayed canonical mutations in the QRDRs of GyrA and ParC.

ICEs in DHS01

ICEs are mobile genetic elements integrated into a genome, and able to excise and transfer by conjugation to another genome [37]. Genome analysis predicted five ICEs in the genome of DHS01 (Fig. S2). As expected, each ICE was

composed of 2 to 4 GIs (Tables 2 and S4). Hence, 13 GIs were integrated into the DHS01 genome via the integration of ICEs. ICE-1, ICE-2, ICE-3, ICE-4 and ICE-5 were 78, 47, 66, 94 and 93 kb in size, respectively. None of these ICEs were retrieved in the other ST395 isolates (dataset 2) with the exception of ICE-4 shared by the strain PT F from a hospital in Birmingham, UK [5]. Fig. S2 details the structure, the gene content and the site of insertion of the five ICEs within the DHS01 chromosome. The five integrated ICEs provided DHS01 with genes involved in virulence and resistance to antibiotics and heavy metals that could have favoured its installation and spread into hospitals. Hence, ICE-4 contained the pili type IVa operon (*pilLNOXQVTM*) also present in the PAPI-1 of strain PA14 and involved in the transfer of this pathogenicity island to other strains [38]. ICE-5 harboured genes conferring resistance to aminoglycosides [*ant(2')-Ia*], to sulfonamide (*sulI*) and to mercury (*merRTPADE*). ICE-1 and ICE-5 were integrated into the 3'-ends of two different tRNA^{Gly}, while ICE-4 was integrated into the 3'-end of tRNA^{Lys}. No tRNAs were found in the vicinity of ICE-2 and ICE-3. The activity of the ICE-4 and ICE-5 was confirmed by the existence of free circular forms after SOS stress (Fig. S3).

Four GIs are ST395 specific

The large genome of DHS01 (7.1 Mbp) and its low G+C content suggested the presence of numerous horizontally acquired GIs. We retrieved 48 GIs in the genome of DHS01 (Table S4), among which 8 GIs were specific to DHS01 since they were not retrieved in any other strain of *P. aeruginosa* of dataset 1 (Table 2). Two of these specific GIs shared identity with sequences found in environmental proteobacteria species (*Pseudomonas putida* and *Acidithiobacillus ferrivorans*). Four of the eight DHS01-specific GIs (GI-4, GI-5, GI-6 and GI-7; Fig. 3) were also found in all the six ST395 *P. aeruginosa* strains tested (dataset 2), possibly identifying these four GIs as markers of the spreading capability of *P. aeruginosa* ST395 [5]. In these four ST395-specific GIs, we specifically searched for elements that could help the implantation and the spread of this international clone in hospitals. GI-4 harboured genes involved in the metabolism of choline, which is present in large amount in the outer

Table 1. Acquired genes and mutations in regulators leading to antibiotic resistance in the ST395 *P. aeruginosa* isolate DHS01

Consequence	Antibiotic(s) affected	Gene or mutation (in comparison with PAO1)
2'-Aminoglycoside nucleotidyl-transferase production	Gentamicin, tobramycin	<i>ant(2')-Ia</i> (borne by ICE-5)
MexXY efflux pump overproduction	Aminoglycosides, fluoroquinolones	8 nt deletion in the regulator-encoding gene <i>mexZ</i>
Quinolone target mutation	Fluoroquinolones	T83I in GyrA and S87L in ParC
AmpC cephalosporinase overproduction	Ticarcillin, piperacillin, cefepime, ceftazidime, aztreonam	M1L in the regulator AmpR, and A134V and T139M in AmpD
MexAB-OprM efflux pump overproduction	Ticarcillin, cefepime, aztreonam	H107P in the regulator MexR
MexEF-OprN efflux pump overproduction	Imipenem	R48C in the oxidoreductase MexS
Reduced production of the porin OprD	Imipenem	R48C in the oxidoreductase MexS
Production of extended-spectrum AmpC (ESAC)	Imipenem	A105 in AmpC

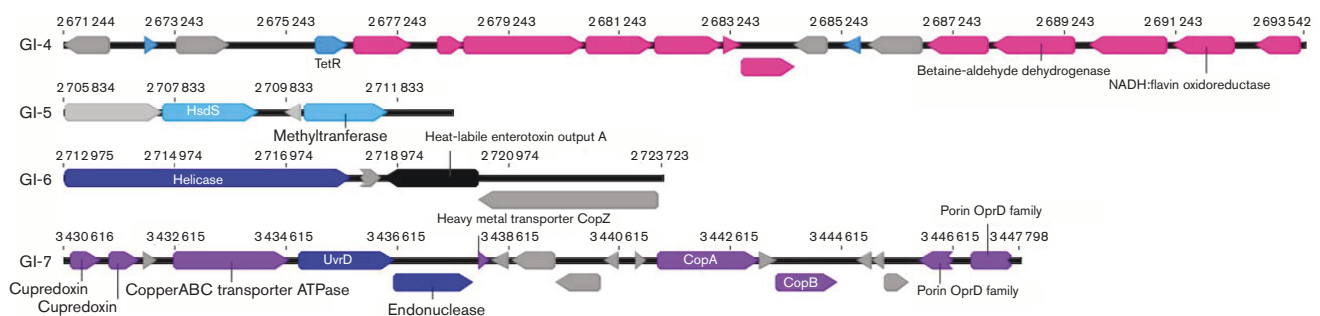
Table 2. Characteristics of the eight GIs specific of the genome of the *P. aeruginosa* ST395 isolate DHS01

These eight GIs were not retrieved in any of the 82 complete genomes of *P. aeruginosa* (Table S1). GIs in bold were specific of the clone ST395. Columns 2, 3 and 4 give the genomic coordinates (start and stop) and the length of the GI, respectively. The fifth column gives the name of the non-*P. aeruginosa* strain for which the genome contains an identical nucleotide sequence (with >99 % length coverage and identity, unless otherwise specified). The major functions encoded by the genes harboured on the GIs are presented in the last column. The last update of the genome collection revealed that two clinical ST253 isolates (M1608 and M37351) had acquired the GI-7.

Name	Start position	Stop position	Length (bp)	Putative species of origin	Encoded protein and its biological function
GI-1	684 629	697 340	12 712	-	RNA helicase and DNA methylase (DNA processing)
GI-2	701 883	709 761	7 879	-	Unknown
GI-3	2 569 006	2 615 647	46 642	-	RsmA (carbon storage regulator)
GI-4	2 671 220	2 693 490	22 270	<i>Pseudomonas putida</i> KT2440 (soil)	AlpA (regulation of prophage excision and pathogenicity) TetR (transcriptional regulator) Betaine-aldehyde dehydrogenase (choline metabolism) Flavin oxidoreductase (iron uptake) HsdS (restriction-modification)
GI-5	2 705 853	2 711 674	5 822	<i>Acidithiobacillus ferrovarans</i> SS3 (96 % coverage and 82 % identity)	HsdS (restriction-modification)
GI-6	2 713 010	2 723 680	10 671	-	Heat-labile enterotoxin
GI-7	3 430 733	3 447 684	16 951	<i>Pseudomonas putida</i> H8234 (93 % coverage and 99 % identity)	Two cupredoxins, copper ABC transporter ATPase, CopZ, CopA and CopB (copper transport) Helicase UvrD and endonuclease (DNA processing) (Phage protein)
GI-8	3 813 161	3 834 228	21 067	-	

leaflet of eukaryotic plasma membranes [39]. GI-6 harboured a gene encoding a protein that shares 58 % identity with the heat-labile enterotoxin (LT) of *Escherichia coli* for which production favours the colonization of the small intestine by enterotoxigenic *E. coli* [40] (Table 2). GI-7 contained an array of six genes encoding copper transporters: two cupredoxins, a copper ABC transporter ATPase, CopZ, CopA and CopB (Fig. 3). Cupredoxins are involved in copper homeostasis, as reported for CopC in *Pseudomonas syringae* [41]. The transporter CopZ is a copper-resistance determinant in the extremophile *Acidithiobacillus ferrooxidans* [42]. CopA and CopB are sufficient to render the cells copper resistant [43]. Hence, CopA is an inducible efflux P-type ATPase located in the inner cell membrane that exports copper from the periplasm. It also shares homology with multicopper oxidases that convert Cu^+ into the less toxic Cu^{2+} . Therefore, CopA has been thought to play a role in copper resistance through its multicopper oxidase activity and its high copper-binding capacity [44]. The gene *copB*

encodes an inducible heavy metal ATPase that extrudes copper from the cell when it reaches toxic intracellular concentrations [45]. The source of the DHS01 within the University Hospital of Besançon is still under investigation; however, the role of a water network made of copper pipes has been recently suggested in a ST395 outbreak [5]. We then hypothesized that the bacterial resistance to copper helped the installation of ST395 in the hospital water network and assessed the survival of a panel of ST395 isolates in copper-containing media. All the seven ST395 tested behaved similarly, with 4.10^3 – 8.10^5 c.f.u. surviving after 48 h incubation in MH broth containing 7.5 mM CuSO_4 (Fig. 4). This contrasted with the variations of survival of the 10 non-ST395 tested. Hence, five isolates (belonging to ST175, ST235, ST342 and ST348) could survive in copper solution, while five other isolates (belonging to ST111, ST233, ST235, ST549 and ST1602) did not. The ST395 survival in copper is presumably due to the presence of *copA* and *copB* in all ST395 isolates, but other mechanisms are implicated in

**Fig. 3.** Structure of the four GIs specific to the ST395 clone. Only relevant genes are cited.

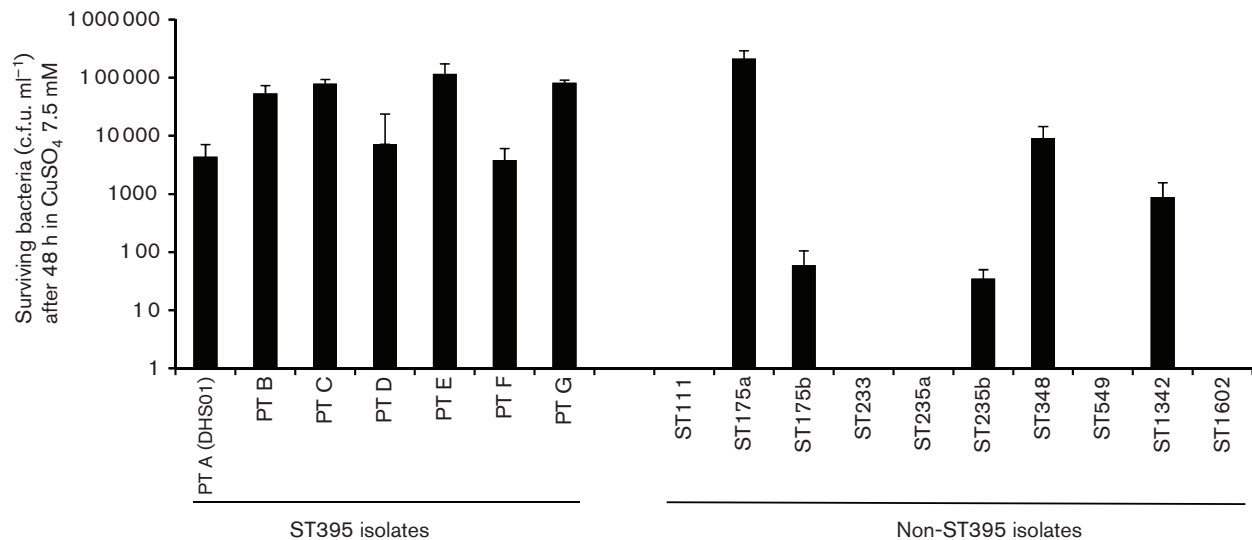


Fig. 4. Survival of ST395 and non-ST395 isolates of *P. aeruginosa* in copper-containing media. The resistance to copper was determined in cation-adjusted MH broth supplemented with 7.5 mM CuSO₄. Strains were described in dataset 2 (Table S2). Values are the ratio of means±SEM from at least three independent experiments.

copper resistance in non-ST395 strains. As expected, GI-7 containing *copA* and *copB* genes was also found in an environmental *P. aeruginosa* ST395 strain isolated from a metal-contaminated aquatic environment [7]. GI-7 was retrieved (with 99% identity and 93% length coverage) from a strain of *P. putida* that typically inhabits soil and water [46]. More generally, genes *copA* and *copB* are found in soil, commensal and plant pathogenic bacterial species, and may be favoured by the long-term use of sprays of copper-based bactericides on crops [47]. ST395 can contaminate hospitalized patients and those afflicted with cystic fibrosis [4, 5, 48], but is also frequently retrieved in polluted environments (estuaries, rivers, hospital wastewaters, water treatment plants) in which copper resistance could have been favoured [7, 48–51]. Our data indicate that an ST395 ancestor has horizontally acquired GI-7 from a single environmental source.

ST395-specific CRISPR-Cas system

The CRISPR-Cas system is thought to play a role in controlling horizontal gene transfer [26]. The high number of GIs integrated into the genome of DHS01 led us to hypothesize the absence of an active CRISPR-Cas system in this epidemic strain. However, CRISPR-Finder detected two CRISPR arrays and the associated Cas proteins that correspond to the type I-E (Fig. S4). It appears to be specific to the ST395 sublineage, since it was retrieved from all the ST395 isolates tested (dataset 2) and absent from non-ST395 (datasets 1 and 2) [26]. Table S5 details the 12 and 8 spacers of the CRISPR arrays 1 and 2, respectively. Interestingly, the phage JBD25 was among the targets of one spacer of the CRISPR system, while genes of this phage were integrated into the GI-19 (Table S4). The G+C content of

JBD25 genes (62.5 mol%) was much lower than that of the rest of the genome (65.8 mol%) and the prophage nucleotide sequence was highly similar (99% of identity) with that of the JDB25 phage. Such an integration was also found in all the other ST395, excepted in isolate PT B. This suggests a recent integration of JDB25 phage despite a dedicated CRISPR-Cas spacer, and that the system may be inactive. However, no anti-CRISPR protein I-E was retrieved in the DHS01 strain, which suggests that this system was inactivated by another mechanism, implicating possibly phage-encoded anti-CRISPR proteins [52, 53].

Conclusion

We describe here the large genome (7.1 Mbp) of the first isolate of the international clone of *P. aeruginosa* ST395 that further spread into our university hospital during at least 11 years. The majority of the 1555 accessory genes were brought by the many GIs ($n=48$) and ICEs ($n=5$) integrated into the DHS01 chromosome. The multidrug resistance of the isolate was mostly due to mutations in antibiotic targets, and in regulators which mutations trigger the overproduction of cephalosporinase and efflux pumps. DHS01 was retrieved from a chronically-colonized patient and displayed the genomic signatures of adaptation to chronic infection mostly due to the deletion of a 131 kbp chromosomal fragment. Four GIs were seemingly specific to the widespread clone ST395. They harboured genes involved in metabolism, in virulence and in the resistance to copper, which was confirmed phenotypically in a panel of independent ST395 isolates. Resistance to copper possibly accounts for the spreading capability of ST395 in hospital settings in which the water network may be incriminated [5]. This suggests that genes selected in environmental bacterial species by the

use of heavy metals can be transferred to pathogens that further spread into hospital settings.

Funding information

M. P. was supported by a grant from the Région de Franche-Comté and from the University Hospital of Besançon, France. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Dr Beryl Oppenheim from the Queen Elizabeth Hospital, Birmingham, UK, for the gift of ST395 isolate 903. Computations were performed at the Mésocentre de Calculs de Franche-Comté (Besançon, France).

Conflicts of interest

The authors declare that there are no conflicts of interest.

Data bibliography

1. Valot B, Rohmer L, Jacobs MA, Miller SI, Bertrand X et al. NCBI accession number CP013993.1 (2016).
2. Rogues AM, Cholley P, Petitjean M, Bertrand X, Hocquet D. NCBI BioProject PRJNA379554 (2017)
3. Bertrand X, Sauget M, Petitjean M, Hocquet D. NCBI BioProject PRJNA380885 (2017).
4. Multiple authors, sees Table S1 and S2. (June 2017).

References

1. Morrison AJ, Wenzel RP. Epidemiology of Infections due to *Pseudomonas aeruginosa*. *Clin Infect Dis* 1984;6:S627–S642.
2. Driscoll JA, Brody SL, Kollef MH. The epidemiology, pathogenesis and treatment of *Pseudomonas aeruginosa* infections. *Drugs* 2007; 67:351–368.
3. Oliver A, Mulet X, López-Causapé C, Juan C. The increasing threat of *Pseudomonas aeruginosa* high-risk clones. *Drug Resist Updat* 2015;21–22:41–59.
4. Cholley P, Thouverez M, Hocquet D, van der Mee-Marquet N, Talon D et al. Most multidrug-resistant *Pseudomonas aeruginosa* isolates from hospitals in Eastern France belong to a few clonal types. *J Clin Microbiol* 2011;49:2578–2583.
5. Quick J, Cumley N, Wearn CM, Niebel M, Constantinidou C et al. Seeking the source of *Pseudomonas aeruginosa* infections in a recently opened hospital: an observational study using whole-genome sequencing. *BMJ Open* 2014;4:e006278.
6. Bertrand X, Bailly P, Blasco G, Balvay P, Boillot A et al. Large outbreak in a surgical intensive care unit of colonization or infection with *Pseudomonas aeruginosa* that overexpressed an active efflux pump. *Clin Infect Dis* 2000;31:e9–e14.
7. Teixeira P, Tacão M, Alves A, Henriques I. Antibiotic and metal resistance in a ST395 *Pseudomonas aeruginosa* environmental isolate: a genomics approach. *Mar Pollut Bull* 2016;110:75–81.
8. Mathee K, Narasimhan G, Valdes C, Qiu X, Mawlish JM et al. Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc Natl Acad Sci USA* 2008;105:3100–3105.
9. Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K et al. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 2003;100:8484–8489.
10. Alonso A, Sánchez P, Martínez JL. Environmental selection of antibiotic resistance genes. *Environ Microbiol* 2001;3:1–9.
11. Eid J, Fehr A, Gray J, Luong K, Lyle J et al. Real-time DNA sequencing from single polymerase molecules. *Science* 2009;323: 133–138.
12. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013;10:563–569.
13. Brittnacher MJ, Fong C, Hayden HS, Jacobs MA, Radey M et al. PGAT: a multistrain analysis resource for microbial genomes. *Bioinformatics* 2011;27:2429–2430.
14. Valot B, Guyeux C, Rolland JY, Mazouzi K, Bertrand X et al. What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. *PLoS One* 2015; 10:e0126468.
15. Talon D, Dupont MJ, Lesne J, Thouverez M, Michel-Briand Y. Pulsed-field gel electrophoresis as an epidemiological tool for clonal identification of *Aeromonas hydrophila*. *J Appl Bacteriol* 1996;80:277–282.
16. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 2002;30:2478–2483.
17. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
18. Dhillon BK, Chiu TA, Laird MR, Langille MG, Brinkman FS. IslandViewer update: improved genomic island discovery and visualization. *Nucleic Acids Res* 2013;41:W129–W132.
19. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007;35:W52–W57.
20. Hocquet D, Llanes C, Thouverez M, Kulasekara HD, Bertrand X et al. Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS Pathog* 2012; 8:e1002778.
21. European Committee on Antimicrobial Susceptibility Testing. *Breakpoint Tables for Interpretation of MICs and Zone Diameters, version 7.1*. Basel: ESCMID; 2017. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_7.1_Breakpoint_Tables.pdf
22. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67:2640–2644.
23. Chen L, Yang J, Yu J, Yao Z, Sun L et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005;33: D325–D328.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
25. Juhas M, Van der Meer JR, Gaillard M, Harding RM, Hood DW et al. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 2009;33:376–393.
26. Van Belkum A, Soriaga LB, Lafave MC, Akella S, Veyrieras JB et al. Phylogenetic distribution of CRISPR-Cas systems in antibiotic-resistant *Pseudomonas aeruginosa*. *MBio* 2015;6:e01796–15.
27. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL et al. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol* 2006;7:R90.
28. Jeukens J, Boyle B, Kukavica-Ibrulj I, Ouellet MM, Aaron SD et al. Comparative genomics of isolates of a *Pseudomonas aeruginosa* epidemic strain associated with chronic lung infections of cystic fibrosis patients. *PLoS One* 2014;9:e87611.
29. Valot B, Rohmer L, Jacobs MA, Miller SI, Bertrand X et al. Comparative genomic analysis of two multidrug-resistant clinical isolates of ST395 epidemic strain of *Pseudomonas aeruginosa* obtained 12 years apart. *Genome Announc* 2014;2:e00515–14.
30. Morgan R, Kohn S, Hwang SH, Hassett DJ, Sauer K. BdlA, a chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. *J Bacteriol* 2006;188:7335–7343.
31. Williams P, Cámara M. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. *Curr Opin Microbiol* 2009;12: 182–191.
32. Klausen M, Heydorn A, Ragas P, Lambertsen L, Aaes-Jørgensen A et al. Biofilm formation by *Pseudomonas aeruginosa* wild type,

- flagella and type IV pili mutants. *Mol Microbiol* 2003;48:1511–1524.
33. Hentzer M, Wu H, Andersen JB, Riedel K, Rasmussen TB et al. Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. *EMBO J* 2003;22:3803–3815.
 34. Hogardt M, Heesemann J. Adaptation of *Pseudomonas aeruginosa* during persistence in the cystic fibrosis lung. *Int J Med Microbiol* 2010;300:557–562.
 35. Rohmer L, Hocquet D, Miller SI. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol* 2011;19:341–348.
 36. Hocquet D, Bertrand X, Köhler T, Talon D, Plésiat P. Genetic and phenotypic variations of a resistant *Pseudomonas aeruginosa* epidemic clone. *Antimicrob Agents Chemother* 2003;47:1887–1894.
 37. Wozniak RA, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 2010;8:552–563.
 38. Qiu X, Gurkar AU, Lory S. Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 2006;103:19830–19835.
 39. Wargo MJ. Homeostasis and catabolism of choline and glycine betaine: lessons from *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 2013;79:2112–2120.
 40. Allen KP, Randolph MM, Fleckenstein JM. Importance of heat-labile enterotoxin in colonization of the adult mouse small intestine by human enterotoxigenic *Escherichia coli* strains. *Infect Immun* 2006;74:869–875.
 41. Arnesano F, Banci L, Bertini I, Thompsett AR. Solution structure of CopC: a cupredoxin-like protein involved in copper homeostasis. *Structure* 2002;10:1337–1347.
 42. Navarro CA, Von Bernath D, Martínez-Bussenius C, Castillo RA, Jerez CA. Cytoplasmic CopZ-Like protein and periplasmic rusticyanin and AcoP proteins as possible copper resistance determinants in *Acidithiobacillus ferrooxidans* ATCC 23270. *Appl Environ Microbiol* 2015;82:1015–1022.
 43. Cha JS, Cooksey DA. Copper hypersensitivity and uptake in *Pseudomonas syringae* containing cloned components of the copper resistance operon. *Appl Environ Microbiol* 1993;59:1671–1674.
 44. Teitzel GM, Geddie A, de Long SK, Kirisits MJ, Whiteley M et al. Survival and growth in the presence of elevated copper: transcriptional profiling of copper-stressed *Pseudomonas aeruginosa*. *J Bacteriol* 2006;188:7242–7256.
 45. Bondarczuk K, Piotrowska-Seget Z. Molecular basis of active copper resistance mechanisms in gram-negative bacteria. *Cell Biol Toxicol* 2013;29:397–405.
 46. Molina L, Bernal P, Udaondo Z, Segura A, Ramos JL. Complete genome sequence of a *Pseudomonas putida* clinical isolate, strain H8234. *Genome Announc* 2013;1:e00496-13.
 47. Cooksey DA. Genetics of bactericide resistance in plant pathogenic bacteria. *Annu Rev Phytopathol* 1990;28:201–219.
 48. Fernández-Olmos A, García-Castillo M, Alba JM, Morosini MI, Lamas A et al. Population structure and antimicrobial susceptibility of both nonpersistent and persistent *Pseudomonas aeruginosa* isolates recovered from cystic fibrosis patients. *J Clin Microbiol* 2013;51:2761–2765.
 49. Libisch B, Balogh B, Füzi M. Identification of two multidrug-resistant *Pseudomonas aeruginosa* clonal lineages with a countrywide distribution in Hungary. *Curr Microbiol* 2009;58:111–116.
 50. Slekovec C, Plantin J, Cholley P, Thouverez M, Talon D et al. Tracking down antibiotic-resistant *Pseudomonas aeruginosa* isolates in a wastewater network. *PLoS One* 2012;7:e49300.
 51. Martin K, Baddal B, Mustafa N, Perry C, Underwood A et al. Clusters of genetically similar isolates of *Pseudomonas aeruginosa* from multiple hospitals in the UK. *J Med Microbiol* 2013;62:988–1000.
 52. Bondy-Denomy J, Davidson AR. To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol* 2014;22:218–225.
 53. Pawluk A, Bondy-Denomy J, Cheung VH, Maxwell KL, Davidson AR. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio* 2014;5:e00896-14.
 54. Center for Genomic Epidemiology. 2015. *Pseudomonas aeruginosa* serotyper (PAst) [Internet]. Available from <https://cge.cbs.dtu.dk/services/PAst-1.0/>.
 55. Center for Genomic Epidemiology. 2017. CGE server [Internet]. Available from <https://cge.cbs.dtu.dk/services/MLST/>.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.

Annexe B

Article *Enterobacter cloacae*

Haenni, M., Saras, E., Ponsin, C., Dahmen, S., Petitjean, M., Hocquet, D., & Madec, J. Y. (2016). High prevalence of international ESBL CTX-M-15-producing *Enterobacter cloacae* ST114 clone in animals. *Journal of Antimicrobial Chemotherapy*, 71(6), 1497-1500.

High prevalence of international ESBL CTX-M-15-producing *Enterobacter cloacae* ST114 clone in animals

Marisa Haenni^{1*}, Estelle Saras¹, Cécile Ponsin¹, Safia Dahmen¹, Marie Petitjean²,
Didier Hocquet² and Jean-Yves Madec¹

¹Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES), Unité Antibiorésistance et Virulence Bactériennes, Lyon, France; ²Laboratoire d'Hygiène Hospitalière, Centre Hospitalier Universitaire, Laboratoire Chrono-environnement UMR CNRS 6249, Univ. Bourgogne Franche-Comté, F-25000 Besançon, France

*Corresponding author. Tel: +33-4-78696830; Fax: +33-4-78619145; E-mail: marisa.haenni@anses.fr

Received 3 September 2015; returned 22 October 2015; revised 19 December 2015; accepted 4 January 2016

Objectives: The objective of this study was to characterize ESBL-producing *Enterobacter cloacae* isolated from animals and to compare their clonal distribution with that of human-related isolates.

Methods: Among 635 clinical *E. cloacae* from horses, dogs and cats collected in France between 2010 and 2013, 36 were resistant to ceftiofur as determined by disc diffusion. ESBL genes were identified by sequencing. Plasmids carrying ESBL-encoding genes were characterized by PCR-based replicon typing, S1-PFGE and Southern blotting. IncHI2 plasmids were subtyped using the plasmid double-locus sequence typing scheme and multiplex amplification of the *hipA*, *smr0092* and *smr0183* genes. All *E. cloacae* were typed by PFGE and MLST. ST clustering was analysed by eBURST.

Results: All 36 ceftiofur-resistant *E. cloacae* produced an ESBL. Their PFGE patterns formed 23 clusters of high similarity and 13 STs and were isolated from epidemiologically unrelated animals (14 horses, 11 dogs and 11 cats) distributed throughout France. ST114, the most prevalent clone in humans, was over-represented in animals (16/36) compared with other human-related clones detected here. The *bla*_{CTX-M-15} gene was dominant (66.7%) and mostly carried on IncHI2 plasmids (ST1 subtype). ST114 isolates always produced CTX-M-15.

Conclusions: Most ESBL-producing *E. cloacae* from animals studied here (69.4%) belonged to potentially high-risk clones in humans, in particular ST114 (44.4%). These data raise questions and potential concerns about the transfer of *E. cloacae* between animals and humans.

Introduction

Enterobacter cloacae is one of the major pathogens causing nosocomial infections in humans. Antibiotic therapy is complicated by the intrinsic resistance of *E. cloacae* to first- and second-generation cephalosporins, penicillins and penicillin plus inhibitor combinations due to an inducible chromosomal AmpC-type cephalosporinase. In addition, *E. cloacae* can become resistant to extended-spectrum cephalosporins by mutational overproduction of AmpC and/or by production of acquired ESBLs.^{1,2}

Recently, a new MLST scheme was developed for *E. cloacae*³ and a first multinational study highlighted specific lineages of *E. cloacae* that frequently colonize the human intestine and thus are considered high-risk clones because of their potential clinical relevance. ST66, ST78, ST108 and ST114 were the most widespread *E. cloacae* STs, with ST114 being predominant. A second recent study on international *E. cloacae* human clones confirmed the dominance of ST114, especially in France.⁴ In both studies, STs could not be specifically associated with

particular β -lactamase content, even though ST114 isolates were often associated with the production of the ESBL enzyme CTX-M-15.

Data on *E. cloacae* resistant to third-generation cephalosporins in animals are scarce, probably because of the low number of animal infections due to this bacterial species.^{5,6} Here, we characterized a collection of ESBL-producing *E. cloacae* isolated from horses, dogs and cats and compared the population structure of *E. cloacae* isolated in animals with that of *E. cloacae* isolated in humans.

Materials and methods

Bacterial isolation and identification

Between 2010 and 2013, 635 clinical *E. cloacae* (213 from horses, 314 from dogs and 108 from cats) were isolated in the peripheral laboratories of the RESAPATH surveillance network of antimicrobial resistance in animals in France (www.resapath.anses.fr) and a routine antibiogram was performed on each of them. All non-duplicate ceftiofur-resistant

E. cloacae isolates were further studied at the French Agency for Food, Environmental and Occupational Health & Safety (ANSES) in Lyon. Identification was confirmed using MALDI-TOF MS (Microflex, Bruker) and API20E galleries (bioMérieux, Marcy-l'Étoile, France) according to the manufacturers' recommendations.

Antibiotic susceptibility testing

For all ceftiofur-resistant isolates, susceptibility to 32 antibiotics of veterinary and human interest was tested by disc diffusion according to the guidelines of the Antibiogram Committee of the French Society for

Microbiology (www.sfm-microbiologie.org). ESBL production was further confirmed by the double-disc synergy test on Mueller–Hinton agar plates without or with 200 mg/L cloxacillin (AES Chemunex, Bruz, France).⁷ *Escherichia coli* ATCC 25922 was used as quality control.

Identification of β -lactamase genes

The ESBL-encoding genes were detected by PCR.^{8,9} For the CTX-M-1 group, an additional PCR was performed using external primers (ISEcp1L1, 5'-CAGCTTTATGACTCG-3'; and P2D, 5'-CAGCGCTTTGCGGTCTAAG-3'). For the CTX-M-9 group, the external primers CTX-M-9A (5'-CTGATGAACCGG

Table 1. Characteristics and epidemiological data of 36 ESBL-producing *E. cloacae* isolated from animals in France between 2010 and 2013

Strain	Geographical origin	Year	Host	Isolation site	ESBL	Plasmid carrying the ESBL gene	Plasmid size (kb)	InchI2 pDLST	HI2 multiplex PCR			ST
									<i>hipA</i>	<i>smr0092</i>	<i>smr0183</i>	
32832	Val-d'Oise	2012	dog	bone	CTX-M-15	InchI2	290	1	pos	pos	pos	136
35785	Pyrénées-Atlantiques	2013	dog	respiratory tract	CTX-M-15	InchI2	320	1	pos	pos	pos	136
34385	Alpes-Maritimes	2013	cat	urinary tract	CTX-M-15	InchI2	320	1	pos	pos	pos	136
34141	Yvelines	2013	horse	skin	CTX-M-15	InchI2	320	NT ^b	pos	pos	pos	127
34871	Alpes-Maritimes	2013	horse	skin	CTX-M-15	InchI2	320	NT ^b	pos	pos	pos	372
32579	Gironde	2012	dog	unknown	CTX-M-15	NT	170					270
34040	Bas-Rhin	2013	dog	skin	CTX-M-14	NT	NT					102
32833	Seine-et-Marne	2012	dog	skin	CTX-M-3	NT	190					408
32894	Ardennes	2012	horse	skin	SHV-12	NT ^a	340					135
32159	Dordogne	2012	horse	unknown	CTX-M-1	NT	240					268
33444	Essone	2013	dog	respiratory tract	SHV-12	InchI2	290	1	pos	pos	pos	268
24232	Alpes-Maritimes	2010	cat	urinary tract	SHV-12	IncA/C	190					NT ^c
35787	Paris	2013	dog	unknown	CTX-M-15	InchI2	320	1	pos	pos	pos	100
32744	Gironde	2012	horse	urinary tract	CTX-M-1	IncP	40					107
32149	Dordogne	2012	horse	respiratory tract	SHV-12	InchI2	340	1	pos	pos	pos	145
32165	Corrèze	2012	horse	unknown	SHV-12	InchI2	340	1	pos	pos	pos	145
32745	Corrèze	2012	horse	respiratory tract	SHV-12	InchI2	340	1	pos	pos	pos	145
28684	Ain	2012	horse	unknown	CTX-M-15	InchI2	320					145
32143	Dordogne	2012	horse	respiratory tract	SHV-12	InchI2	290	1	pos	pos	pos	145
34042	Maine-et-Loire	2013	horse	reproductive tract	SHV-12	InchI2	320	1	pos	pos	pos	118
32743	Aube	2012	horse	skin	CTX-M-15	InchI2	290	1	pos	pos	pos	114
32746	Hauts-de-Seine	2012	horse	bone	CTX-M-15	InchI2	290	1	pos	pos	pos	114
34041	Côte-d'Or	2013	cat	urinary tract	CTX-M-15	NT ^a	320					114
29673	Alpes-Maritimes	2012	cat	skin	CTX-M-15	InchI2	290	1	neg	pos	pos	114
33445	Val-de-Marne	2013	cat	urinary tract	CTX-M-15	InchI2	320	1	pos	pos	pos	114
33447	Val-de-Marne	2013	cat	unknown	CTX-M-15	InchI2	320	1	pos	pos	pos	114
33739	Val-de-Marne	2013	cat	urinary tract	CTX-M-15	NT ^a	320					114
34142	Marne	2013	horse	skin	CTX-M-15	InchI2	320	1	pos	pos	pos	114
33716	Alpes-Maritimes	2012	dog	skin	CTX-M-15	InchI2	320	1	pos	pos	pos	114
35786	Drôme	2013	cat	urinary tract	CTX-M-15	InchI2	320	1	pos	pos	pos	114
35791	Val-de-Marne	2013	dog	bone	CTX-M-15	InchI2	320	1	pos	pos	pos	114
33446	Hauts-de-Seine	2013	cat	respiratory tract	CTX-M-15	InchI2	320	1	pos	pos	pos	114
35825	Hauts-de-Seine	2013	dog	skin	CTX-M-15	NT	NT					114
32905	Hauts-de-Seine	2012	cat	urinary tract	CTX-M-15	InchI2	290	1	pos	pos	pos	114
35790	Val-d'Oise	2013	cat	unknown	CTX-M-15	InchI2	320	1	pos	pos	pos	114
27309	Val-de-Marne	2011	dog	skin	CTX-M-15	InchI2	290	1	pos	pos	pos	114

NT, non-typeable; pos, positive; neg, negative.

^aAlthough the InchI2 replicon was detected by PCR, multiple hybridization experiments with the InchI2 probe were not conclusive.

^bDespite several attempts, the *smr00018* allele could not be amplified for isolates 34141 and 34871, whereas the *smr00199* amplifications matched with allele 1.

^cDespite several attempts, the *gyrB* allele could not be amplified.

ATTGAC-3') and CTX-M-9C (5'-AGCGCCCATATTGAGAG-3') were used. All *bla*_{CTX-M}-positive amplicons were sequenced.

Bacterial typing

PFGE was performed using the restriction enzyme XbaI.⁹ Electrophoresis was conducted in a CHEF DRIII system using 6 V/cm at 14°C for 17 h, with pulse times ramping from 1 to 25 s using an angle of 120°. PFGE results were interpreted according to international recommendations.¹⁰ MLST was carried out according to the protocol described on the *E. cloacae* MLST web site (<http://pubmlst.org/ecloacae/>).¹¹ The relatedness between STs was analysed by eBURST (<http://eburst.mlst.net/>).

Plasmid characterization

Plasmids were typed by PCR-based replicon typing (PBRT) using the PBRT kit from DIATHEVA (Fano, PU, Italy) according to the manufacturer's instructions.¹² Plasmid size was determined by S1-PFGE gels.⁹ Southern-blot hybridizations of plasmid DNA were performed on isolates using DIG-labelled probes (*bla*_{CTX-M}- or *bla*_{SHV}-specific probes and probes corresponding to the various replicon types found) according to the manufacturer's instructions (Roche Diagnostics, Germany). IncH12 plasmids were subtyped using the plasmid double-locus sequence typing scheme (pDLST; <http://pubmlst.org/plasmid/>) and multiplex amplification of the *hipA*, *smr0092* and *smr0183* genes.¹³

Results and discussion

The 36 (5.7%, 36/635) clinical *E. cloacae* isolates that were resistant to the extended-spectrum cephalosporin ceftiofur produced an ESBL. They were isolated from 14 horses, 11 dogs and 11 cats, leading to a prevalence within these animal species of 6.6% (14/213), 3.5% (11/314) and 10.2% (11/108), respectively.

This is in line with the prevalence of ESBL-producing *E. coli* of 3.7% previously reported in dogs in France.¹⁴ *E. cloacae* isolates were recovered from skin (*n*=11), respiratory tract (*n*=6), urinary tract (*n*=8), bone (*n*=3), reproductive tract (*n*=1) or non-documented infections (*n*=7) (Table 1). Bacterial isolates were collected from epidemiologically unrelated animals (i.e. from different owners and diverse locations). In addition to their resistance to extended-spectrum cephalosporins, they were mostly resistant to aminoglycosides with 100% (36/36) of strains resistant to streptomycin, 88.9% (32/36) to gentamicin and 80.6% (29/36) to kanamycin. They were also resistant to tetracyclines (88.9%, 32/36), sulphonamides (94.4%, 34/36), trimethoprim (97.2%, 35/36), quinolones (63.9%, 23/36) and fluoroquinolones (50.0%, 18/36). All of the isolates remained susceptible to carbapenems. Overall, these MDR patterns affected nearly all classes of antibiotics available for veterinary use.

The clonality of the isolates was first investigated by PFGE, which yielded 23 clusters of banding patterns with >85% similarity within each cluster (Figure S1, available as Supplementary data at JAC Online). Interestingly, isolates with the same PFGE pattern could be retrieved from different animals, either of the same or different species, indicating both local and nationwide spread of the *E. cloacae* ST114 clone, which has also spread in humans.

We further determined the ST of all isolates and found 13 different STs, among which one was not in the database (ST372). All STs were analysed by eBURST against the entire MLST database (as at 13 May 2015) (Figure 1). Most isolates belonging to the same PFGE cluster had the same ST. As expected, PFGE was more discriminatory than MLST, with several PFGE patterns clustering into one ST (Figure S1). In one case, the same pulsotype belonged to both ST127 and ST372, which were single-locus

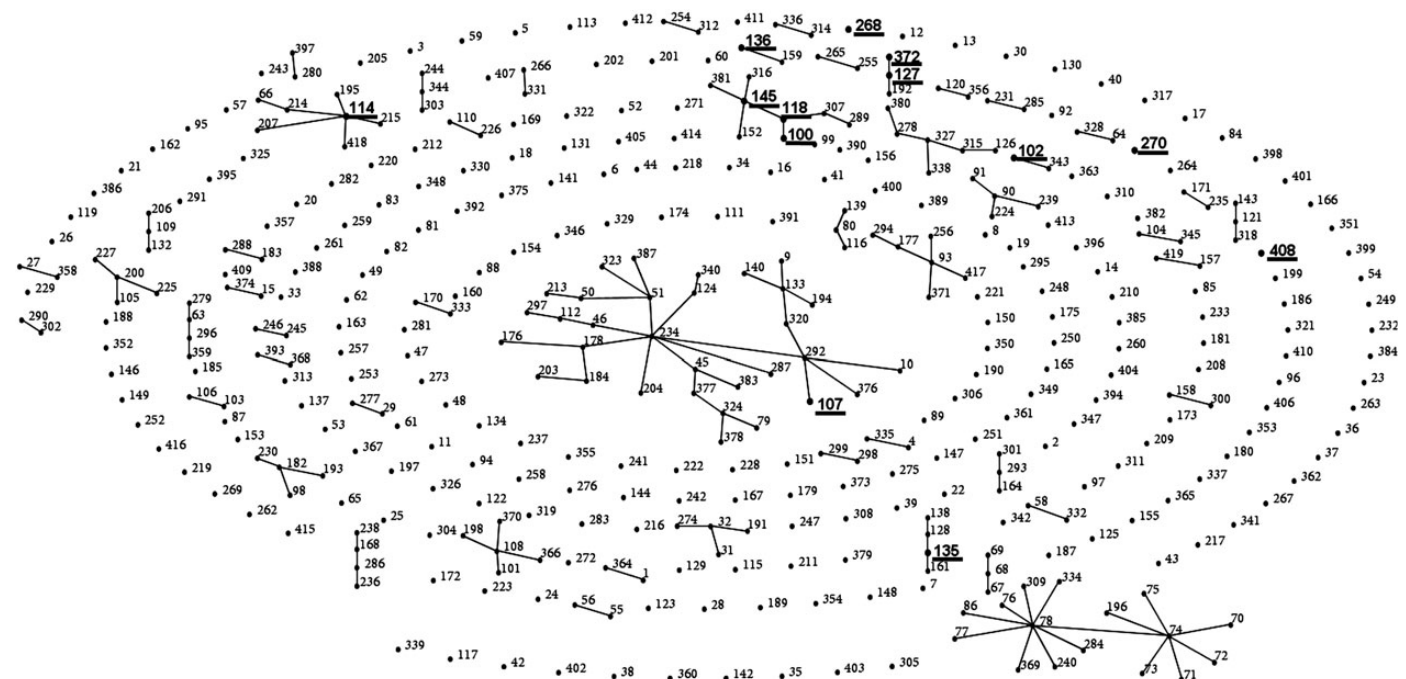


Figure 1. Population structure of the 36 ESBL-producing *E. cloacae* isolates, shown in the context of all 419 STs present in the global MLST database (<http://pubmlst.org/ecloacae/>, as at 13 May 2015). The scheme was constructed using eBURST analysis. Dots indicate the STs and STs identified in the study are in bold underlined type. SLVs (i.e. STs that share six MLST loci) are connected by a line.

variants (SLVs) at the *dnaA* allele. Strikingly, MLST revealed the over-representation (16/36, 44.4%) of ST114, mainly in cats and dogs (13/16). ST114 is a major international clone reported in humans in two recent studies.^{3,4} All 16 ST114 isolates shared highly similar PFGE patterns that were classified into a single cluster and a few of them even shared fully identical patterns. Other clones found in humans were also retrieved in the collection, such as ST135 ($n=1$), ST107 ($n=1$) and ST118 ($n=1$) and its two SLVs ST100 ($n=1$) and ST145 ($n=5$, in horses only). Larger-scale studies are now needed to investigate the possible transfer of *E. cloacae* clones from animals to humans or vice versa. Of note, all *E. cloacae* studied here produced ESBLs. Similarly, all human *E. cloacae* clones reported previously were recovered from media selecting for Enterobacteriaceae resistant to extended-spectrum cephalosporins. Consequently, further studies including susceptible isolates from both animals and humans would also help understand the clonal structure of the *E. cloacae* population.

Among ESBL genes, *bla*_{CTX-M-15} was dominant ($n=24/36$, 66.7%) and all ST114 isolates produced CTX-M-15. Other ESBL genes were detected, such as *bla*_{SHV-12} ($n=8/36$, 22.2%), *bla*_{CTX-M-1} ($n=2/36$, 5.6%), *bla*_{CTX-M-3} ($n=1/36$, 2.8%) and *bla*_{CTX-M-14} ($n=1/36$, 2.8%) (Table 1). Most ESBL-encoding genes (26/36, 72.2%) were located on IncHI2 plasmids of 290–340 kb in size, as already reported in *E. cloacae*.^{1,15} One IncP, one IncA/C and eight non-typeable plasmids were also identified (Table 1). Here, all typeable IncHI2 plasmids belonged to the ST1 subtype, with the amplification of all three *hipA*, *smr0092* and *smr0183* genes (only one *hipA* gene could not be amplified despite several attempts).¹³ In animals, IncHI2 plasmids were also reported to spread *bla*_{CTX-M-2} in *Salmonella enterica* serovars Paratyphi B and Virchow in poultry and in *E. coli* in horses.^{16–18}

In conclusion, this study is the first clonal analysis of *E. cloacae* in animals. Remarkably, 69.4% (25/36) of the isolates belonged to the potentially high-risk clones reported in humans, with ST114 encompassing 44.4% of them. This collection of ESBL-producing *E. cloacae* isolated from animals also reveals the dominance of *bla*_{CTX-M-15} carried by IncHI2/ST1 plasmids. These data raise concerns about the transfer of *E. cloacae* between animals and humans.

Acknowledgements

We gratefully thank all of the peripheral laboratories of the RESAPATH network involved in the study. We also thank Tohru Miyoshi-Akiyama, curator of the *E. cloacae* MLST database, for assigning new allele numbers and new STs.

Funding

This work was supported by the French Agency for Food, Environmental and Occupational Health & Safety (ANSES).

Transparency declarations

None to declare.

Supplementary data

Figure S1 is available as Supplementary data at JAC Online (<http://jac.oxfordjournals.org/>).

References

- Nilsen E, Haldorsen BC, Sundsfjord A et al. Large IncHI2-plasmids encode extended-spectrum β -lactamases (ESBLs) in *Enterobacter* spp.: bloodstream isolates, and support ESBL-transfer to *Escherichia coli*. *Clin Microbiol Infect* 2013; **19**: 516–8.
- Miro E, Segura C, Navarro F et al. Spread of plasmids containing the *bla*_{VIM-1} and *bla*_{CTX-M} genes and the *qnr* determinant in *Enterobacter cloacae*, *Klebsiella pneumoniae* and *Klebsiella oxytoca* isolates. *J Antimicrob Chemother* 2010; **65**: 661–5.
- Izdebski R, Baraniak A, Herda M et al. MLST reveals potentially high-risk international clones of *Enterobacter cloacae*. *J Antimicrob Chemother* 2015; **70**: 48–56.
- Girlich D, Poirel L, Nordmann P. Clonal distribution of multidrug-resistant *Enterobacter cloacae*. *Diagn Microbiol Infect Dis* 2015; **81**: 264–8.
- Sidjabat HE, Hanson ND, Smith-Moland E et al. Identification of plasmid-mediated extended-spectrum and AmpC β -lactamases in *Enterobacter* spp. isolated from dogs. *J Med Microbiol* 2007; **56**: 426–34.
- Dierikx CM, van Duijkeren E, Schoormans AH et al. Occurrence and characteristics of extended-spectrum- β -lactamase- and AmpC-producing clinical isolates derived from companion animals and horses. *J Antimicrob Chemother* 2012; **67**: 1368–74.
- EUCAST. *EUCAST Guidelines for Detection of Resistance Mechanisms and Specific Resistances of Clinical and/or Epidemiological Importance*. 2013. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Resistance_mechanisms/EUCAST_detection_of_resistance_mechanisms_v1.0_20131211.pdf.
- Eckert C, Gautier V, Arlet G. DNA sequence analysis of the genetic environment of various *bla*_{CTX-M} genes. *J Antimicrob Chemother* 2006; **57**: 14–23.
- Dierikx C, van Essen-Zandbergen A, Veldman K et al. Increased detection of extended spectrum β -lactamase producing *Salmonella enterica* and *Escherichia coli* isolates from poultry. *Vet Microbiol* 2010; **145**: 273–8.
- Tenover FC, Arbeit RD, Goering RV et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 1995; **33**: 2233–9.
- Miyoshi-Akiyama T, Hayakawa K, Ohmagari N et al. Multilocus sequence typing (MLST) for characterization of *Enterobacter cloacae*. *PLoS One* 2013; **8**: e66358.
- Carattoli A, Bertini A, Villa L et al. Identification of plasmids by PCR-based replicon typing. *J Microbiol Methods* 2005; **63**: 219–28.
- Garcia-Fernandez A, Carattoli A. Plasmid double locus sequence typing for IncHI2 plasmids, a subtyping scheme for the characterization of IncHI2 plasmids carrying extended-spectrum β -lactamase and quinolone resistance genes. *J Antimicrob Chemother* 2010; **65**: 1155–61.
- Dahmen S, Haenni M, Chatre P et al. Characterization of *bla*_{CTX-M} IncFII plasmids and clones of *Escherichia coli* from pets in France. *J Antimicrob Chemother* 2013; **68**: 2797–801.
- Carattoli A. Resistance plasmid families in Enterobacteriaceae. *Antimicrob Agents Chemother* 2009; **53**: 2227–38.
- Garcia Fernandez A, Cloeckert A, Bertini A et al. Comparative analysis of IncHI2 plasmids carrying *bla*_{CTX-M-2} or *bla*_{CTX-M-9} from *Escherichia coli* and *Salmonella enterica* strains isolated from poultry and humans. *Antimicrob Agents Chemother* 2007; **51**: 4177–80.
- Smet A, Boyen F, Flahou B et al. Emergence of CTX-M-2-producing *Escherichia coli* in diseased horses: evidence of genetic exchanges of *bla*_{CTX-M-2} linked to ISCR1. *J Antimicrob Chemother* 2012; **67**: 1289–91.
- Doublet B, Praud K, Nguyen-Ho-Bao T et al. Extended-spectrum β -lactamase- and AmpC β -lactamase-producing β -tartrate-positive *Salmonella enterica* serovar Paratyphi B from broilers and human patients in Belgium, 2008–10. *J Antimicrob Chemother* 2014; **69**: 1257–64.

Annexe C

Article *P. aeruginosa* bruns

Hocquet, D., Petitjean, M., Rohmer, L., Valot, B., Kulasekara, H. D., Bedel, E., ... & Jacobs, M. A. (2016). Pyomelanin-producing *Pseudomonas aeruginosa* selected during chronic infections have a large chromosomal deletion which confers resistance to pyocins. *Environmental Microbiology*, 18(10), 3482-3493.

Pyomelanin-producing *Pseudomonas aeruginosa* selected during chronic infections have a large chromosomal deletion which confers resistance to pyocins

Didier Hocquet,^{1,2*} Marie Petitjean,^{1,2}
Laurence Rohmer,³ Benoît Valot,¹
Hemantha D. Kulasekara,³ Elodie Bedel,^{1,2}
Xavier Bertrand,^{1,2} Patrick Plésiat,⁴ Thilo Köhler,⁵
Alix Pantel,^{6,7} Michael A. Jacobs,³
Lucas R. Hoffman^{3,8} and Samuel I. Miller³

¹UMR CNRS 6249, Chrono-environnement, Université de Bourgogne-Franche-Comté, Besançon, France.

²Service d'Hygiène Hospitalière, Centre Hospitalier Régional Universitaire, Besançon, France.

³Department of Microbiology, University of Washington, Seattle, WA, USA.

⁴Service de Bactériologie, Centre Hospitalier Régional Universitaire, Besançon, France.

⁵Département de Génétique et de Microbiologie, Centre Médical Universitaire, Genève, Suisse.

⁶Service de Microbiologie, Centre Hospitalier Régional Universitaire, Nîmes, France.

⁷UMR INSERM U1047, Université de Montpellier, Nîmes, France.

⁸Department of Pediatrics, University of Washington, Seattle, WA, USA.

Summary

When bacterial lineages make the transition from free-living to permanent association with hosts, they can undergo massive gene losses, for which the selective forces within host tissues are unknown. We identified here melanogenic clinical isolates of *Pseudomonas aeruginosa* with large chromosomal deletions (66 to 270 kbp) and characterized them to investigate how they were selected. When compared with their wild-type parents, melanogenic mutants (i) exhibited a lower fitness in growth conditions found in human tissues, such as hyperosmolarity and pres-

ence of aminoglycoside antibiotics, (ii) narrowed their metabolic spectrum with a growth disadvantage with particular carbon sources, including aromatic amino acids and acyclic terpenes, suggesting a reduction of metabolic flexibility. Despite an impaired fitness in rich media, melanogenic mutants can inhibit their wild-type parents and compete with them in coculture. Surprisingly, melanogenic mutants became highly resistant to two intraspecific toxins, the S-pyocins AP41 and S1. Our results suggest that pyocins produced within a population of infecting *P. aeruginosa* may have selected for bacterial mutants that underwent massive gene losses and that were adapted to the life in diverse bacterial communities in the human host. Intraspecific interactions may therefore be an important factor driving the continuing evolution of pathogens during host infections.

Introduction

When bacterial lineages make the transition from free-living to permanent association with hosts, they often undergo a loss of genes (Moran, 2002). This long-term association with an unchanging environment requires little adaptability of the pathogen, rendering many bacterial genes useless and resulting in greater genetic drift and gene inactivation. Commonly, gene loss and genome reduction improve the ability of pathogens to survive in the host but also restrict the range of hosts available to the bacterium (Bentley and Parkhill, 2004; Ochman and Davalos, 2006).

Pseudomonas aeruginosa is an extremely common opportunistic pathogen. Its large genome (~6.5 Mbp) accounts for its metabolic versatility and affinity for various environmental niches, but also for its infectious capability towards a large set of hosts (Aujoulat *et al.*, 2012). *P. aeruginosa* is a leading cause of hospital-acquired infections. It also causes a wide range of acute and chronic infections in predisposing situations such as chronic wounds, burn wounds and chronic obstructive pulmonary disease, the

Received 8 January, 2016; revised 4 April, 2016; accepted 5 April, 2016. *For correspondence. E-mail dhocquet@chu-besancon.fr; Tel. + 33 3 70 63 21 34; Fax + 33 3 70 63 21 27.

latter of which is particularly common among cystic fibrosis patients (Driscoll *et al.*, 2007). Chronic infections with *P. aeruginosa* are associated with the high-density bacterial assemblages, such as bacterial biofilms, that favour the emergence of variants mostly arising through homologous recombination, recombinatorial DNA repair, and DNA mismatch repair deficiency (Oliver *et al.*, 2000; Stoodley *et al.*, 2002; Boles *et al.*, 2004). The population dynamics within these populations can be affected by bacterial competition through the production of pyocins that kill susceptible sub-populations (Waite and Curtis, 2009).

The evolution and diversification of *P. aeruginosa* has been extensively studied in the respiratory chronic infection in individuals with cystic fibrosis (CF), during which this pathogen commonly loses genetic information either by gene inactivation or by genome reduction (Ernst *et al.*, 2003; Marvig *et al.*, 2015a,b). For example, mutations of the genes *mexZ*, *mucA* and *lasR* occur very frequently in lineages adapted to CF patients and confer selective advantages in host environments including antibiotic resistance, mucoidy and growth on abundant nitrogen sources, respectively (Marvig *et al.*, 2015a). During in-host evolution, the chromosome of *P. aeruginosa* can also undergo large deletions of hundreds of genes (Ernst *et al.*, 2003). However, the selective forces for such massive genetic losses in bacteria during chronic infections remain unknown.

In this work, we aimed at exploring the features of chronic infection bacterial isolates with large chromosomal deletions and at identifying the selective forces that could have driven their selection in the host. We focused on pyomelanin-producing mutants in which others have shown large chromosomal deletions (Ernst *et al.*, 2003; Le *et al.*, 2014). Overproduction of the pigment pyomelanin results in an easily recognizable red-brownish pigmented colony. Pyomelanin-producing mutants are regularly retrieved from the cultures of urine, sputum, bile or wounds, and up to 13% of the chronically infected CF patients harbour pyomelanin-producing mutants (Ogunnariwo and Hamilton-Miller, 1975; Mayer-Hamblett *et al.*, 2014). Their pigmentation is due to homogentisic acid, which accumulates due to the inactivation of the gene *hmgA* (*pa2009* in *P. aeruginosa* strain PAO1), which encodes homogentisate-1,2-dioxygenase (Rodríguez-Rojas *et al.*, 2009). Although the inactivation of *hmgA* may occur in clinical strains by large chromosomal deletion that included *hmgA* (Ernst *et al.*, 2003; Le *et al.*, 2014), other authors have described single point mutations in this gene (Murugan *et al.*, 2014; Tielen *et al.*, 2014). First, we compared the genomes and phenotypes of a collection of clinical pyomelanin-producing mutants with those of their wild-type parents. We found that large chromosomal reduction was the main genetic event leading to pyomelanin production. Second, we identified consequent alterations in the fitness, metabolism, production of virulence factors, and resistance to antibiotics

and toxins. These results identify the likely selective pressure that drove the emergence of these mutants: intraspecific competition through production of pyocins.

Results and discussion

The genome modifications that occur in *P. aeruginosa* during chronic infections often include large genomic deletions (>50 kb). These deletions, which tend to occur at specific hotspots distributed throughout the genome, often do not have easily identified phenotypic consequences (Ernst *et al.*, 2003; Spencer *et al.*, 2003; Rau *et al.*, 2012). In contrast, deletions including *hmgA* lead to an easily recognizable brown colony phenotype. Therefore, pyomelanin-producing mutants are a convenient model to study the causes and consequences of large deletions that occur during chronic infections.

Pyomelanin-producing mutants contain large chromosomal deletions

We first compared the chromosome genetic sequences of one laboratory and four clinical brown mutants of *P. aeruginosa* with those of their genetically related precursor isolates (Table 1). This comparison revealed large deletions in each case, ranging in size from 66 kbp (isolate B_{pm}) to 270 kbp (isolate A_{pm}) and representing from 1.05% to 4.32% of the genome. As expected, each deletion eliminated the *hmgA* gene (Fig. 1A and 1B). Forty-six genes in the vicinity of *hmgA* (from *pa1974* to *pa2019*, representing 54.6 kbp) were also deleted in all the pyomelanin-producing isolates (Fig. 1C). No other mutations were found in the genome of the brown mutants compared with their wild-type precursors, including in the gene *lasR*.

We confirmed that deletion of the region *pa1974*–*pa2018*, which includes *hmgA*, occurs commonly during chronic infections by identifying this change in an independent additional collection of 5 other pyomelanin-producing isolates compared with their wild-type parents (Supporting Information table S1). Only one additional brown isolate (3447) had a nucleotide substitution leading to a Leu95Pro change in HmgA when compared with its parent. Because single point mutations in *hmgA* are fairly uncommon in clinical isolates (i.e., in 1 out of 10 isolates tested), it is unlikely that the *in vivo* selection of these mutants is only due to properties conferred by *hmgA* deletion itself. Rather, the brown color of the bacterial colonies due to the *hmgA* mutation likely represents a surrogate marker of a large chromosomal deletion encompassing this gene that was selected for other reasons.

Comparative sequence analysis of wild-type isolates PAO1, A_{wt}, B_{wt}, C_{wt} and D_{wt} revealed direct repeats (from 8 bp to 12 bp) at the boundaries of the genomic region missing from the brown clinical isolates, among which

Table 1. Characteristics of the pyomelanin-producing isolates of *P. aeruginosa* and their precursor wild-type isolates.

Pair of isolates	Sequence type ^a	Source	Isolate ^b	Production of pyomelanin	Isolation location	Type of infection or colonization	Age of the infection/colonization (days) ^c	Age of the patient (years)	Antibiotic treatment ^d	Reference
PAO1/PAO1 $\Delta hmgA$	ST549	Laboratory	PAO1 PAO1 $\Delta hmgA$	No Yes	– ^e	–	–	–	–	(Jacobs <i>et al.</i> , 2003)
PAO1/CIP	ST549	Laboratory	PAO1 CIP	No Yes	–	–	–	–	Ciprofloxacin	This work
A	ST253	Clinical	A _{wt} A _{pm}	No Yes	Sinus	Sinusitis	10	72	Ceftriaxone, fosfomycin, amikacin	This work
B	ST298	Clinical	B _{wt} B _{pm}	No Yes	Blood culture	Urinary sepsis	22	56	Cotrimoxazole	This work
C	ST253	Clinical	C _{wt} C _{pm}	No Yes	Tracheal aspirate	Tracheal colonization	36	49	Co-amoxiclav	This work
D	ST974	Clinical	D _{wt} D _{pm}	No Yes	Sputum	Pulmonary infection in cystic-fibrosis patient	51	0.8	Colimycin	This work

a. As determined using the MLST scheme described by Curran *et al.* (Curran *et al.*, 2004).

b. WT, wild-type; PM, pyomelanin-producing isolate.

c. Duration as first *P. aeruginosa* isolation.

d. Antibiotic treatment received by the patients before the emergence of pyomelanin-producing isolates.

e. Not applicable.

these repeats occurred in only a single copy at the deletion sites. We confirmed the presence of the direct repeats using PCR and sequencing in the pairs PAO1/CIP, A, and D (Fig. 1B). These data were consistent with a recombination event between these direct repeats, facilitating the excision of the large chromosomal sequences.

Pyomelanin-producing mutants exhibit narrowed metabolic substrate utilization profiles

Screening for phenotypic differences between pyomelanin-producing mutants and their wild-type parents was performed using the Biolog method. This analysis indicated relative growth defects among pyomelanin-producing clinical mutants in high (9.5) pH, in 6% NaCl, and when aromatic amino acids provided the sole sources of either nitrogen or carbon. We first experimentally confirmed that one of these metabolic differences was exhibited by all pyomelanin-producing mutants: a growth defect when tyrosine was the sole C-source (Fig. 2A). As seen with the PAO1 *hmgA* mutant, this difference was expected due to the fact that HmgA is involved in the catabolic pathway for tyrosine.

Other metabolic genes were missing from the chromosomes of all tested pyomelanin-producing clinical isolates (Fig. 1C) including the *liu* genes, necessary for the catabolism of the amino acid leucine and that of acyclic terpenes (e.g., citronellol, geraniol). We confirmed that the brown mutants did not grow with the geraniol as the sole C-source, in contrast with their isogenic parents (Fig. 2B). In

addition, deletions of the metabolic genes *erbR*, *eraSR*, *exaA* and *pqqC* have been shown to impair growth on ethanol, and the deletion of genes *dhcRAB* are necessary for the use of carnitine (Wargo and Hogan, 2009; Mern *et al.*, 2010; Beaudoin *et al.*, 2012). Taken together, these data suggest a reduction of metabolic versatility, and therefore, of niche range, of the pyomelanin-producing mutants.

Pyomelanin-producing mutants are more susceptible to harsh environments

The gene *mexY* was missing on the chromosome of all the melanogenic-producing mutant isolates. The deletion of this protein transporter, involved in the intrinsic resistance to aminoglycosides, resulted in a twofold to fourfold higher susceptibility of the brown mutants to tobramycin, an aminoglycoside antibiotic used widely to treat *P. aeruginosa* infection (Fig. 3A; Ramos Aires *et al.*, 1999; Vogne *et al.*, 2004). In contrast, we did not observe any clear trend in the susceptibility of brown mutants to β -lactams, to ciprofloxacin, and to colistin (Supporting Information table 2).

We experimentally confirmed the higher susceptibility to basic pH (pH9.5) of three out of four pyomelanin-producing clinical mutants. This feature was independent of the deletion of *hmgA* (Supporting Information fig. S1a). Although PAO1 and its two melanogenic derivatives behaved similarly in a hyperosmotic media containing 2 M of NaCl (Fig. 3B), all the clinical mutants producing pyomelanin were more susceptible to this osmotic shock than were their isogenic wild-type parents (Fig. 3C). This finding was

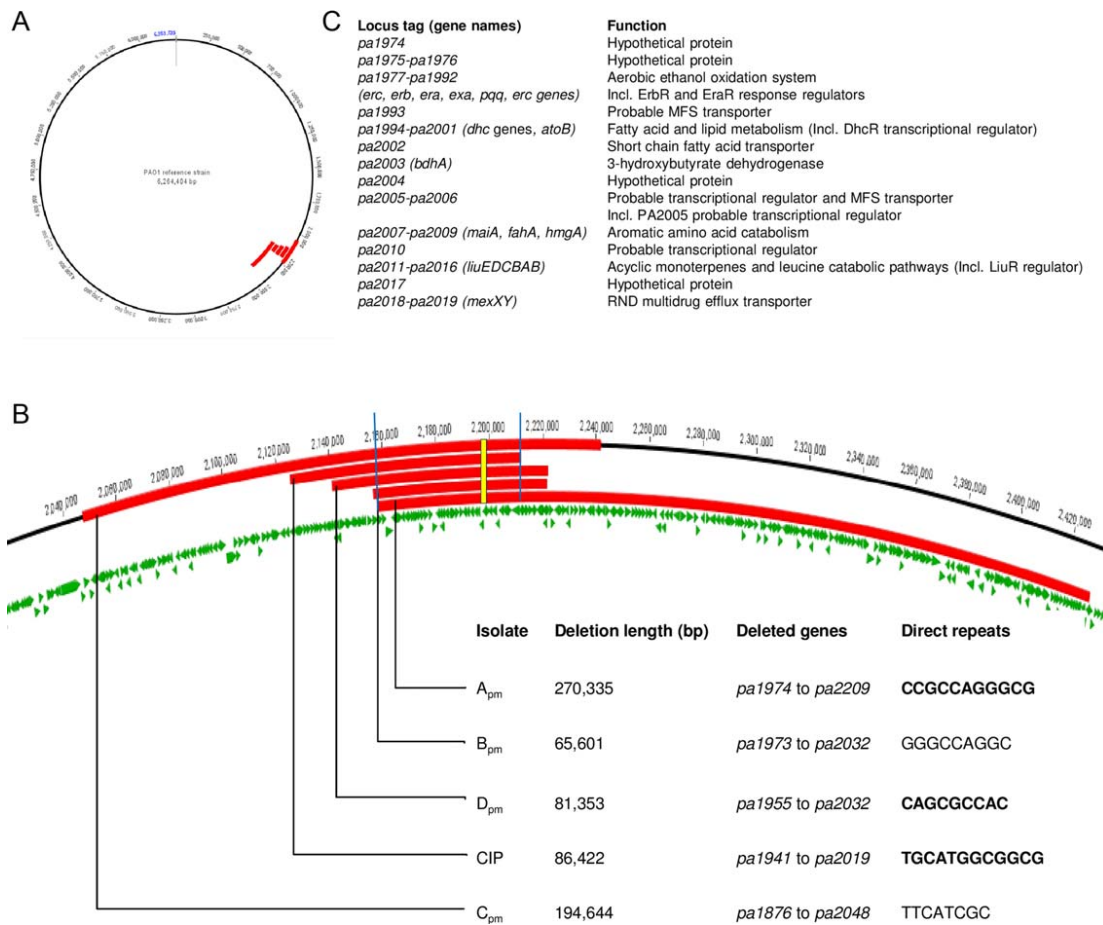


Fig. 1. Large chromosomal deletions in pyomelanin-producing *P. aeruginosa* clinical isolates.

A. Location (in red) relative to sequenced *P. aeruginosa* strain PAO1 of chromosomal deletions in five brown *P. aeruginosa* mutants compared with wild-type precursors.

B. Close-up 1A, indicating the deleted region for each pyomelanin-producing mutant in this study. The blue bars surround the genes deleted in all the pyomelanin-producing mutants. The yellow rectangle indicates *hmgA* gene. The length of the deletion and the deleted genes are indicated. The direct repeats putatively involved in the deletion are detailed (PCR-verified repeats are bold-typed).

C. Name and function of the genes deleted in all the pyomelanin-producing mutants.

particularly unexpected for the isolate D_{pm} retrieved from the lung of a CF-patient, as the osmolarity of CF airway secretions are thought to be high in osmolarity (Henderson *et al.*, 2014).

It has been shown that the production of pyomelanin after the inactivation of *hmgA* increases the resistance of *P. aeruginosa* to oxidative stress (Rodriguez-Rojas *et al.*, 2009). Here, although we confirmed the higher resistance of an engineered *hmgA* mutant to challenge with 50 mM of H₂O₂ for 45 min compared with its parent PAO1, only one clinical mutant (A_{pm}) out of the four that underwent a large deletion was more resistant to H₂O₂ than its isogenic counterparts (Supporting Information fig. S1b). This finding makes it unlikely that pyomelanin-producing mutants are selected in the host based on a higher resistance to reactive oxygen species. Therefore, loss of other genes in the vicinity of *hmgA* presumably counterbalanced the protec-

tive effect of pyomelanin. We found that all the pyomelanin-producing clinical mutants were more susceptible to hyperosmolarity and the presence of aminoglycosides, and that three out of four were more susceptible to high pH. These stress conditions are found in the human tissues. However, these findings must be interpreted with caution. For example, oxidative stress *in vivo* presumably occurs primarily due to superoxide anions produced by macrophages and monocytes in response to engulfed bacteria or bacterial products and that help to kill phagocytosed bacteria. However, exposure to these products may not occur *in vivo* as they do in *in vitro* cultures due either to biofilm formation, physicochemical variability and/or gradients, or other unknown mechanisms (Costeron *et al.*, 1999). Therefore, the relationships between *in vitro* and *in vivo* susceptibilities to specific conditions are difficult to predict.

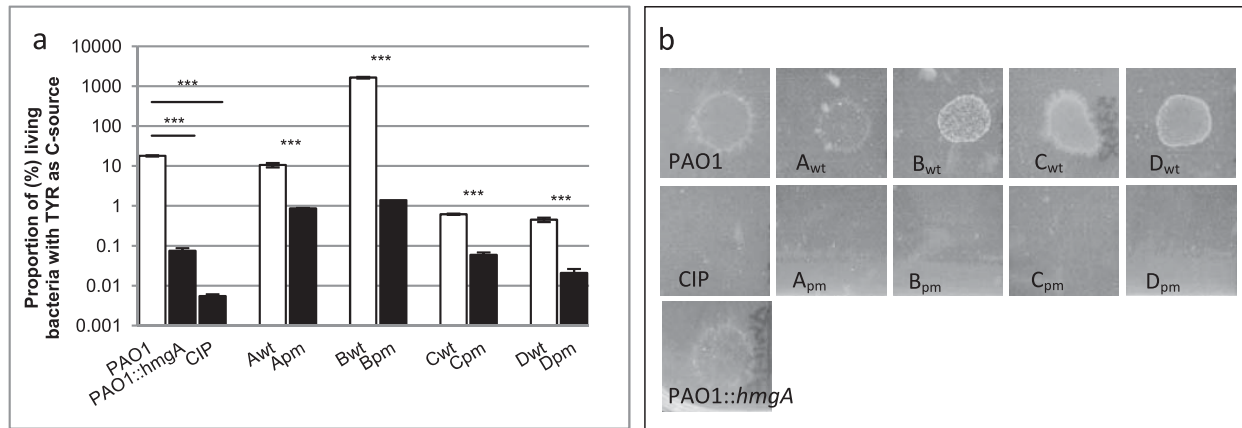


Fig. 2. Metabolic changes in pyomelanin-producing mutants of *P. aeruginosa*.

A. Proportion of living bacteria after 24 h of incubation in minimal media M63 supplemented with 0.05% of tyrosine as the sole C-source as described before (see Supplementary Information). Ratio of means \pm SEM from at least 3 independent experiments; two-sided Student's *t*-tests; NS, $P > 0.05$; ***, $P < 0.001$.

B. Bacterial growth on M63 plates supplemented with 0.1% geraniol as the only C-source and incubated for 48 h at 37°C followed by 10 days at 30°C.

Pyomelanin-producing mutants produced less virulence factors in vitro

We compared the *in vitro* production of virulence factors (elastase, rhamnolipids and pyocyanin), biofilm formation, haemolytic activities, and swarming, swimming and twitching motilities among these isolates and strains. Production of rhamnolipids, swimming, swarming and twitching motilities were not affected in the pyomelanin-producing mutants (data not shown). In contrast, the production levels of pyocyanin, elastase and the haemolytic activity were significantly reduced in three, one and three clinical mutants, respectively (Supporting Information fig. S2). The haemolytic activity is presumably attributed to PlcH, which gene (*pa0844*) is present in the genome of all the mutants and parental strains. In contrast, the production of none of the virulence factors tested was affected in the PAO1 *hmgA* mutant (Supporting Information fig. S2). The gene *galU* (*pa2023*) encodes a UDP-glucose pyrophosphorylase that catalyzes an essential step for the synthesis of a complete lipopolysaccharide (LPS) outer core. *GalU* mutants produce truncated (rough) LPS molecules and display attenuated virulence in part due to higher serum sensitivity (Dean and Goldberg, 2002; Priebe *et al.*, 2004). As determined by the O-serotyping by specific antisera, the wild-type isolates PAO1, A_{wt}, B_{wt} and C_{wt} harboured the O:5, O:16, O:11, and O:16 antigens, respectively. Consistent with the deletion of the *galU* gene from their genomes (Fig. 1C), the pyomelanin-producing clinical isolates A_{pm}, B_{pm}, and C_{pm} autoaggregated, presumably due to the lack of the parental O-antigen (Dean and Goldberg, 2002). We measured the kinetics of biofilm formation of all the isolates and found that three out of the four pyomelanin-producing clinical isolates (B_{pm}, C_{pm}, D_{pm}) sig-

nificantly produced less early biofilm than their wild-type parents (Fig. 4). This impairment was not clearly related to the size and the gene content of the chromosomal deletions underwent by the brown mutants.

Fitness and coculture

The persistence of the pyomelanin-producing mutants in the source patients led us to suspect that the chromosome deletions observed in the pyomelanin-producing mutants conferred a growth advantage *in vivo* (Mayer-Hamblett *et al.*, 2014). Therefore, we searched for *in vitro* conditions under which these mutants had a growth advantage. Surprisingly, we found a lower relative fitness of pyomelanin-producing strains under both aerobic and anaerobic conditions in rich liquid media in comparison with their wild-type parents (Fig. 5A and 5B). We then hypothesized that the brown mutants had an advantage over their wild-type parents in biofilm conditions (Costerton *et al.*, 1999). To test this hypothesis, we cocultured each pair of isolates in growth conditions that favor the formation of a biofilm (Waite and Curtis, 2009). While the CIP laboratory mutant outgrew its wild-type parent PAO1 after 24 h of biofilm growth, the isolates C_{pm} and D_{pm} stayed in similar proportion with their wild-type parent after 24 h of coculture despite their lower fitness in anaerobiosis (Fig. 5C). In contrast, the PAO1-derived *hmgA* mutant and the clinical isolates A_{pm} and B_{pm} were outcompeted by their wild-type parents after 24 h of coculture. To determine how some pyomelanin-producing mutants could compete with their parents in biofilms, we tested the ability each strain's supernatant to inhibit growth of its isogenic counterpart. Others have shown that a general stress response known

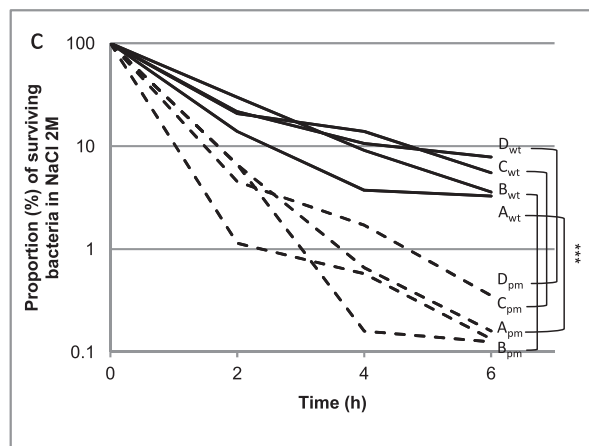
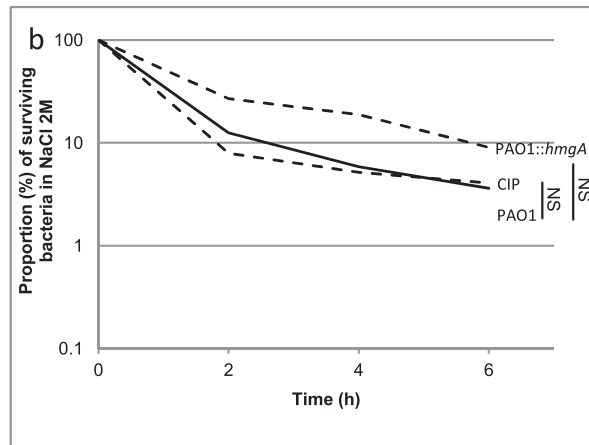
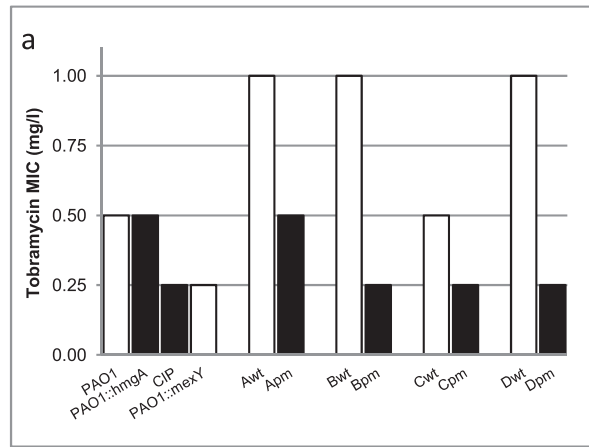


Fig. 3. Survival of the pyomelanin-producing mutants of *P. aeruginosa* in harsh environments. A. The minimal inhibitory concentrations of tobramycin were determined by Etest. B and C. The resistance to osmotic shock was determined in Luria-Bertani broth supplemented with 2M NaCl. Ratio of means \pm SEM from at least 3 independent experiments; two-sided Student's *t*-tests; NS, $P > 0.05$; **** $P < 0.001$.

as the SOS response is triggered in biofilms (Beloin *et al.*, 2004; Waite and Curtis, 2009). To stimulate the SOS response and to mimic this biofilm stress, each strain was

Table 2. Susceptibility to S-pycocins of the pyomelanin-producing isolates of *P. aeruginosa* and their wild-type parents.

Isolate	Lysis by S-type pycocins	
	AP41	S1
PAO1	+ ^a	-
PAO1::hmgA	+	-
CIP	-	-
A _{wt}	+	+
A _{prm}	-	-
B _{wt}	+	+
B _{prm}	-	-
C _{wt}	+	-
C _{prm}	-	-
D _{wt}	-	+
D _{prm}	-	-

a. +, lysis; -, no lysis.

cultured in this experiment in media containing 1 mg/l of the SOS-inducing mitomycin C (Hocquet *et al.*, 2012). Consistent with the biofilm competition results, the supernatants of the three pyomelanin-producing mutants that effectively competed with their wild-type parents (CIP, C_{prm} and D_{prm}) also inhibited growth of those parent strains on agar media (Fig. 5D). The role of the 19 genes (*pa1955-pa1973*; Fig. 1b) that are deleted only in CIP, C_{prm} and D_{prm} while present in the other brown mutants still remains to be explored.

Pyomelanin-producing mutants are resistant to S-pycocins

Bacteria are known to diversify both *in vivo* during chronic infections and *in vitro* during growth in biofilms, and biofilms are thought to underlie many chronic infections (Boles *et al.*, 2004; Marvig *et al.*, 2015b). It has been

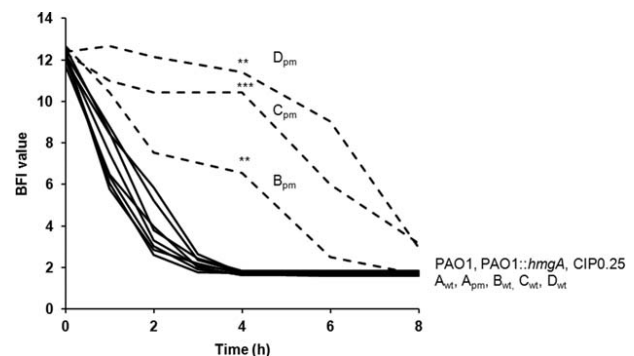


Fig. 4. Kinetics of the early phase of biofilm formation of pyomelanin-producing mutants of *P. aeruginosa* and their wild-type parents. Measures were done by the BioFilm Ring Test® (BioFilm Control). Means of BFI (BioFilm Indice) for four replicates from two independent experiments. Standard deviations were not displayed for ease of reading. Statistical differences at 4 h were calculated with two-sided Student's *t*-tests; ** $P < 0.01$; *** $P < 0.001$.

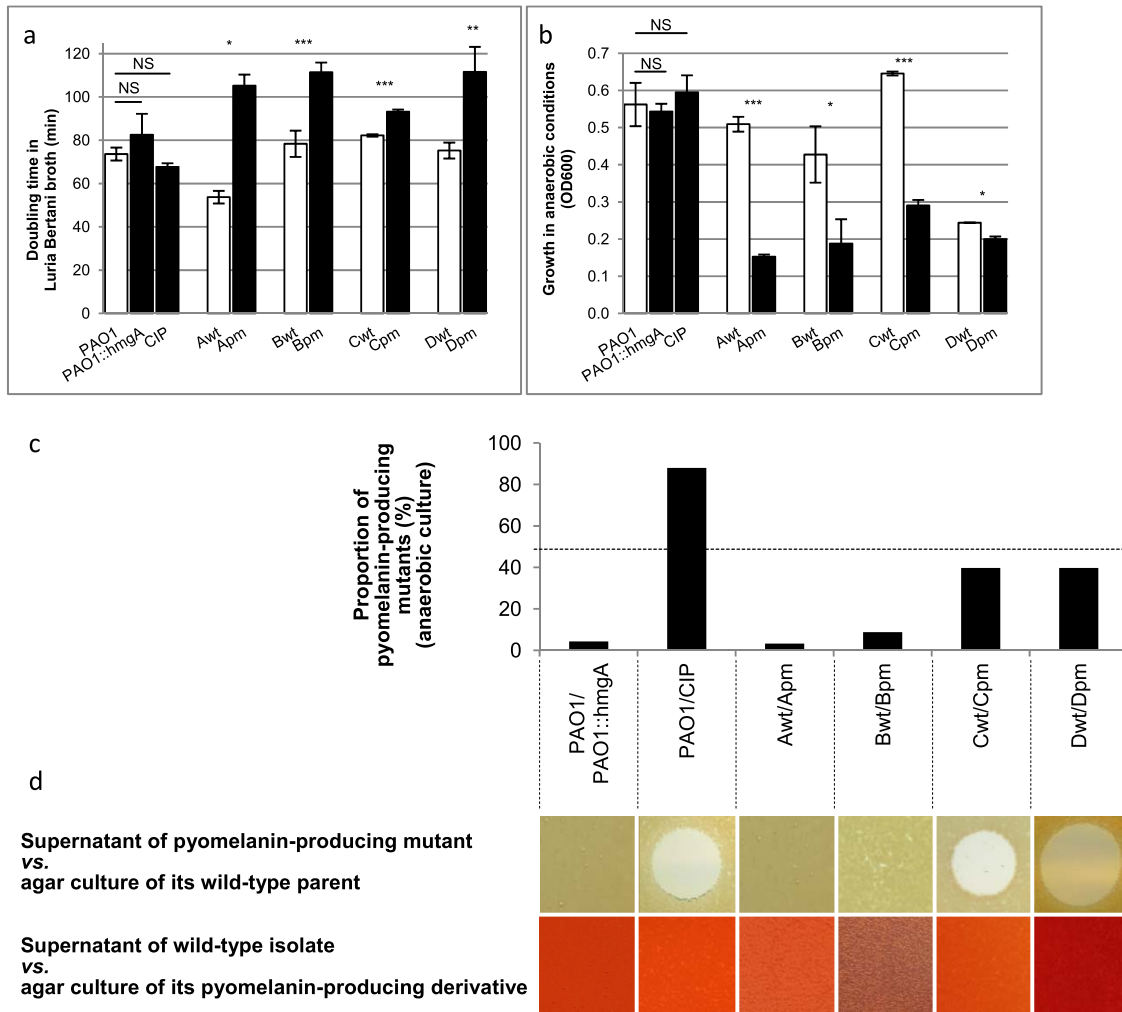


Fig. 5. Brown mutants can compete with their wild-type parents in coculture despite lower fitness in monoculture. Means \pm SEM from at least 3 independent experiments; two-sided Student's *t*-tests; NS, $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

A. Growth in Luria-Bertani broth under aerobic conditions at 37°C under shaking at 150 r.p.m.

B. Growth in Luria-Bertani broth supplemented with 0.4 mM KNO₃ under anaerobic conditions without shaking.

C. Relative proportion of pyomelanin-producing mutants after 24 h of coculture with their isogenic parents in biofilm under anaerobic conditions as described elsewhere (Waite and Curtis, 2009).

D. Inhibitory activity of 30 μ l of the supernatant of an overnight culture of pyomelanin-producing mutants in MH broth containing 1 mg/l of the SOS response inducer mitomycin C against a culture on MH agar of their wild-type parent (up). Inhibitory activity of 30 μ l of the supernatant of a 12-h culture of the wild-type parent in MH broth containing 1 mg/l of mitomycin C against a culture on MH agar of their pyomelanin-producing derivative (down). As a control, 30 μ l of MH broth containing 1 mg/l of mitomycin C incubated overnight had no inhibitory activity against any of the strains and isolates shown on agar culture (data not shown). The 2 variant colonies growing in the inhibition spot of the C_wt were found – after an overnight subculture on MH agar – as susceptible to the C_pm supernatant as their parent (data not shown).

shown that the dynamics of variants that arise in biofilms can be shaped by the production of pyocins, which are produced at increased levels during anaerobic biofilm growth (Waite and Curtis, 2009). *P. aeruginosa* is capable of producing at least three different pyocins (of S-, R- and F-types) that kill other strains from the same species. More specifically, S-pyocins kill bacteria through DNase, tRNase, or pore-forming activity (Michel-Briand and Baysse, 2002). Interestingly, Holloway *et al.* described in 1973 two mutants

tolerant to the S-pyocin AP41 that 'produced copious brown pigment in the media', suggesting a potential relationship between pyocin resistance and brown mutant selection (Holloway *et al.*, 1973). Therefore, we tested the inhibitory effect of the S-pyocins AP41, S1, S2, S3, and the R-pyocins R1, R2, R3, R5 on our collection and found that, unlike their wild-type parents, the mutants CIP, A_pm, B_pm and C_pm were all resistant to AP41, and the mutants A_pm, B_pm and D_pm were resistant to S1 (Table 2; Additional

file 4: Supporting Information). Therefore, each brown clinical mutant had acquired new resistance to at least one tested S-pyocin. These resistances were not conferred by deletion of *hmgA*, since the PAO1 *hmgA* mutant remained as susceptible to AP41 as its parent strain was. The resistance to some pyocins has been attributed to modification of the lipopolysaccharides which production depends (?) on *galU* gene, deleted in all the brown mutants (Kohler *et al.*, 2010). We then tested the resistance to the S-pyocins AP41 and S1 of the PAO1*galU* mutant (El Garch *et al.*, 2007) and found that it was as susceptible to the S-pyocins tested as its wild-type parent PAO1. No change in the susceptibility to pyocins S2 and S3 and to the R-pyocins R1, R2, R3, R5 occurred in pyomelanin-producing isolates (data not shown). Unlike the killing mechanism of pyocins S2 and S3, which depends on the FpvA ferrityoverdine receptors, the resistance mechanisms to pyocins S1 and AP41 are still unknown (Denayer *et al.*, 2007). Our data confirm that the resistance to pyocins mostly occurs via mechanisms independent from specific immunity genes (Ghoul *et al.*, 2015). The resistances to S1 and AP41 pyocins could be due to the loss of a membrane receptor that allows the entry of these toxins into the bacterial cell. As we did not find any obvious receptor-encoding genes within the chromosomal region missing from the brown mutants, we speculated the presence within this region of a gene encoding a regulator which deletion abolishes or downregulates the production of a pyocin receptor. These results could indicate that the relative resistance to AP41 and/or S1-pyocins of the pyomelanin-producing mutants plays a key role in the survival of these adapted mutants in patients; in support of this concept, nearly all (98%) the clinical isolates of *P. aeruginosa* isolated from CF-patients have been observed to produce at least one S-pyocin (Ghoul *et al.*, 2015). However, we are aware that the pyocins that we did not test (S-pyocins S4, S5, S6, R-pyocin R4 and F-pyocins) could also play a role in bacterial competition.

Emergence of pyomelanin-producing mutants

The four patients from whom pyomelanin-producing mutants were isolated differed in terms of the anatomic location of their infections, which included sinusitis, endocarditis, pulmonary infection and endotracheal colonization, all of which have been proposed to involve bacterial biofilms (Costerton *et al.*, 1999). In addition, the source patients were treated with unrelated antibiotics (Table 1). Elsewhere, Cabot *et al.* have shown that meropenem, another antibiotic, can select pyomelanin-producing mutants (Cabot *et al.*, 2016). Biofilms are structured multicellular communities of microorganisms associated with surfaces that are believed to play a role in at least 65% of all human infections, in particular device-related infections,

body surface infections, and other chronic infections (Costerton *et al.*, 1999). Therefore, the brown mutants tested here could all have occurred in biofilms *in vivo*.

The resistance of pyomelanin-producing mutants to S-pyocins was the only unifying characteristic that could have played a selective role. We then hypothesized that pyocin resistance was selected during biofilm growth in these mutants. Pyocin-resistant mutants have been readily selected *in vitro* by the S-pyocin AP41, and Boles *et al.* further showed that these mutants mostly emerge in biofilm, depending on RecA for recombination (Holloway *et al.*, 1973; Boles *et al.*, 2004). To test whether pyomelanin-producing mutants could have been specifically selected from bacterial biofilms of wild-type isolates, we grew PAO1, A_{wt} , B_{wt} and C_{wt} both in planktonic and biofilm cultures before selection on MH agar plates supplemented with AP41 and quantified the proportion of pyocin AP41-resistant mutants that produced pyomelanin. D_{wt} was deliberately not tested since already resistant to AP41. While no pyomelanin-producing mutants were retrieved from planktonic cultures, such mutants were easily selected from biofilms, as previously shown (Boles *et al.*, 2004). The average frequency of emergence of AP41-resistant mutants was 1.9×10^{-3} from biofilms, with pyomelanin-producing mutants representing a large part of pyocin-resistant mutants (from 31% to 64%, depending on the strain tested, PAO1, A_{wt} , B_{wt} , or C_{wt}), with an average frequency of pyomelanin-producing mutants of 1.1×10^{-3} . The emergence of pyomelanin-producing mutants depends on the presence of the DNA recombinase RecA, which is activated in biofilm, while independent from the error-prone DNA polymerase DinB (Boles *et al.*, 2004). Given our observation of direct repeats at the boundaries of the chromosomal deletions (Fig. 1B) and the relative rarity of point mutation in *hmgA* in our pyomelanin-producing mutant collection, these results indicate that recombination is the primary mechanism of selection of these mutants.

We found here that all the tested pyomelanin-producing clinical isolates contained large chromosomal deletions (≥ 65.6 kbp). The diversity of the genetic background of the strains (Table 1; Supporting Information fig. S3) and the size change of the chromosomal deletion in mutants (Fig. 1B) certainly account for the variability of their phenotypes. However, all the brown mutants tested displayed a lower fitness, a narrowed substrate utilization profiles and an increased *in vitro* susceptibility to hyperosmolarity and antibiotics. While these isolates also exhibited reduced growth in monoculture compared with their wild-type parents, the mutants could inhibit their parents in coculture. Importantly, the pyomelanin-producing mutants all displayed increased resistance to S-pyocins, likely favoring their survival in mixed population of *P. aeruginosa* in patients. Our results suggest that the production of pyomelanin resulting from the deletion of *hmgA* is a phenotypic marker of a large

chromosomal deletion in *P. aeruginosa* mutants adapted to the life in diverse bacterial communities in the human host. These pyomelanin-producing mutants lost much of the metabolic flexibility exhibited by free-living bacteria and produced less virulence factors *in vitro*.

The pathogen chromosome reductions that occur in many chronic infections often result in the loss of unnecessary regulatory elements. The relatively static environment of host tissues eliminates the extreme environmental fluctuations encountered by free-living bacteria (Moran, 2002). Interactions between the host and the bacteria can drive the microbe evolution (Marvig *et al.*, 2015a). We found that pyomelanin-producing mutants emerged in diverse clinical situations and from various genetic backgrounds. This suggests that the selection of pyomelanin-producing mutants is unlikely due to host interactions or antibiotic treatments. Here, we propose that the bacterial adaptation to the host can be shaped by pyocin-mediated interbacterial interactions that are facilitated by growth in biofilms.

Experimental procedures

Bacterial isolates

We used six pyomelanin-producing mutants of *P. aeruginosa* (suffixed with 'pm') and their non-pyomelanin-producing parents (suffixed with wt, for wild-type). Each clinical pair (A, B, C and D) was isolated the same day in the same specimen from the same patient. The genetic similarity of *P. aeruginosa* clinical isolates was investigated by pulsed field gel electrophoresis (PFGE; CHEF-DR III; Bio-Rad, Hercules, California) with the use of *DraI* enzyme, as described elsewhere (Supporting Information fig. S3; Talon *et al.*, 1996). We also used two derivatives from the reference strain PAO1: the mutant PAO1::hmgA with *hmgA* gene inactivated with the transposon ISphoA/hah (Jacobs *et al.*, 2003) and the mutant CIP fortuitously isolated from PAO1 culture on agar containing 0.25 mg/l of ciprofloxacin. Table 1 details the characteristics of the clinical isolates and the patient source.

Genome comparison

The reads from the chromosomes of the pyomelanin-producing isolates and their wild-type parents were generated with Illumina GAIIx sequencer (Illumina, San Diego, CA). The reads were mapped onto the genomic sequence of the reference strain *P. aeruginosa* PAO1 using BWA and Samtools (Stover *et al.*, 2000; Li and Durbin, 2009; Li *et al.*, 2009). For each pair, we looked for regions of the PAO1 genome that were covered by reads from the wild-type parent but not by reads from its pyomelanin-producing derivative. Only regions with a minimum coverage of 10 and a minimum quality score of 15 were considered. The position of the deleted region was validated by the graphical view of reads distribution. To confirm the position of the deletions in pyomelanin-producing isolates, we designed four primers for each pair of strains to perform three PCR reactions (Supporting Information table S3). A first PCR with primers 1 and 2 overlapped the 3'-edge of the deletion in the chromosome of the wild-type parent. A

second PCR with primers 3 and 4 overlapped the 5'-edge of the deletion. A third PCR with primers 1 and 4 overlapped the 'scar' of the deletion in the chromosome of the pyomelanin-producing isolates. Direct repeats potentially involved in the deletion were identified with Replib (http://genometools.org). We predicted the sequence type (ST) of the isolates from their whole genome sequence data with the MLST 1.7 tool (Larsen *et al.*, 2012).

Confirmation of deletion by PCR

To confirm the chromosomal deletion in other pyomelanin-producing clinical mutants, we PCR amplified internal fragments of the genes *pa1974*, *pa1989*, *pa2009* and *pa2019* in an independent collection of six pyomelanin-producing clinical mutants and in their wild-type parents (Supporting Information table S1).

Phenotype determination

We screened the metabolic changes in the pyomelanin-producing mutants with the Phenotype MicroArray analysis (Biolog, Hayward, CA) in duplicate according to the manufacturer's instructions. The bacterial growth of the pyomelanin-producing mutants was assessed in 950 conditions (i.e., with various carbon, nitrogen, phosphorus, sulfur and peptide nitrogen sources, various nutrient supplements, with various osmolytes) and compared with that of their isogenic parent. The conditions are available at http://www.biolog.com/pdf/pm_lit/PM1-PM10.pdf. The supplemental materials detail the conditions under which we assessed the bacterial fitness, the changes in utilization of aromatic amino acids and terpene compounds, the survival in harsh environment, and the production of virulence factors. We evaluated the kinetics of early biofilm formation of all the isolates of the collection using the Biofilm Ring Test (BioFilm Control, Saint Beaulieu, France) as previously described (Olivares *et al.*, 2016).

Coculture

Cocultures of pyomelanin-producing mutants with their wild-type parents were performed in anaerobic biofilms which favor the production of pyocins (Waite and Curtis, 2009). Briefly, 10^5 CFU of an overnight culture of the pyomelanin-producing mutants and 10^5 CFU of their wild-type parents were mixed and grown on 0.45- μ m nitrocellulose filters (Millipore) placed on reduced-strength LB agar (20%) containing 1% KNO_3 to permit anaerobic respiration. Dishes were incubated 24 h at 37°C in an anaerobic cabinet (MACS – MG 500 Work Station; DW Scientific). The proportion of each cocultured isolate was assessed by plating an appropriate dilution of the culture on MHA incubated overnight at 37°C. Randomly-chosen colonies ($n > 100$) were streaked on MHA incubated overnight at 37°C before observation of the colony pigmentation.

Pyocin resistance determination

E. coli DH5 α strains harboring plasmids pYMSS11, pYMPS1, pPYS3-3pY or pYK211 (encoding for the pyocins S1, S2, S3 or AP41, respectively) were cultured 6 h at 37°C in LB broth

containing 100 mg/l of ampicillin and then centrifuged (Sano and Kageyama, 1993; Sano *et al.*, 1993; Duport *et al.*, 1995). The pellet was resuspended in PBS buffer and sonicated on ice. The lysate was cleared by centrifugation 20 min, 1200 g at 4°C. The supernatant was sterilized through a 0.22 µm pore size filter and stored at –80°C upon utilization. MHA plates were inoculated with a lawn of *P. aeruginosa* isolates, according to CLSI disk diffusion method, before dropping 30 µl of the pyocin solution and overnight incubation at 37°C.

Emergence of pyomelanin-producing mutants

To assess the conditions under which the pyomelanin-producing mutants emerge, we cultured the wild-type isolates PAO1, A_{wt}, B_{wt}, C_{wt} and D_{wt} in 5 ml of MH broth at 37°C under shaking either 18 h or 5 days with daily media changes. Dilutions of the resulting cultures were then plated on MHA plates containing 1.5% of the pyocin AP41 and incubated 24 h at 37°C. Fifty randomly-chosen AP41-resistant mutants were streaked on MHA for the observation of colony pigmentation.

Statistical analysis

Student's t-tests were used to determine statistical significance for comparisons of survival in tyrosine and in harsh environment, of biofilm and growth parameters, and of production of virulence factors. Graphical examination supports the assumption of normality and homogeneous variation across experiments. The chosen significance threshold was 0.05 for all tests.

Acknowledgments

We thank Beth Ramage, Marie-Pierre Blanc, and Marion Manzoni for their technical support and also Yumico Sano for the gift of pYMPS1 and pYK211. We thank Emmanuelle Dé, Thierry Jouenne and Christine Baysse for helpful discussion.

Conflict of interests

The authors have declared that no competing interests exist.

Ethics statement

Approval and written informed consent from all subjects or their legally authorized representatives were obtained before study initiation. The study was approved by the ethical committee 'Comité d'Etude Clinique' of the Besançon hospital, Besançon, France.

References

Aujoulat, F., Roger, F., Bourdier, A., Lotthe, A., Lamy, B., Marchandin, H., and Jumas-Bilak, E. (2012) From environment to man: genome evolution and adaptation of human opportunistic bacterial pathogens. *Genes (Basel)* **3**: 191–232.
 Beaudoin, T., Zhang, L., Hinz, A.J., Parr, C.J., and Mah, T.-F. (2012) The biofilm-specific antibiotic resistance gene *ndvB* is important for expression of ethanol oxidation genes in

Pseudomonas aeruginosa biofilms. *J Bacteriol* **194**: 3128–3136.
 Beloin, C., Valle, J., Latour-Lambert, P., Faure, P., Kzreminski, M., Balestrino, D. *et al.* (2004) Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Mol Microbiol* **51**: 659–674.
 Bentley, S.D., and Parkhill, J. (2004) Comparative genomic structure of prokaryotes. *Annu Rev Genet* **38**: 771–791.
 Boles, B.R., Thoendel, M., and Singh, P.K. (2004) Self-generated diversity produces “insurance effects” in biofilm communities. *Proc Natl Acad Sci USA* **101**: 16630–16635.
 Cabot, G., Zamorano, L., Moyà, B., Juan, C., Navas, A., Blázquez, J., and Oliver, A. (2016) Evolution of *Pseudomonas aeruginosa* antimicrobial resistance and fitness under low and high mutation rates. *Antimicrobial Agents and Chemotherapy* **60**: 1767–1778.
 Costerton, J.W., Stewart, P.S., and Greenberg, E.P. (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284**: 1318–1322.
 Dean, C.R., and Goldberg, J.B. (2002) *Pseudomonas aeruginosa galU* is required for a complete lipopolysaccharide core and repairs a secondary mutation in a PA103 (serogroup O11) *wbpM* mutant. *FEMS Microbiol Lett* **210**: 277–283.
 Denayer, S., Matthijs, S., and Cornelis, P. (2007) Pyocin S2 (Sa) kills *Pseudomonas aeruginosa* strains via the FpvA Type I ferripyoverdine receptor. *J Bacteriol* **189**: 7663–7668.
 Driscoll, J., Brody, S., and Kollef, M. (2007) The epidemiology, pathogenesis and treatment of *Pseudomonas aeruginosa* infections. *Drugs* **67**: 351–368.
 Duport, C., Baysse, C., and Michel-Briand, Y. (1995) Molecular characterization of pyocin S3, a novel S-type pyocin from *Pseudomonas aeruginosa*. *J Biol Chem* **270**: 8920–8927.
 El Garch, F., Jeannot, K., Hocquet, D., Llanes-Barakat, C., and Plésiat, P. (2007) Cumulative effects of several nonenzymatic mechanisms on the resistance of *Pseudomonas aeruginosa* to aminoglycosides. *Antimicrob Agents Chemother* **51**: 1016–1021.
 Ernst, R.K., D'Argenio, D.A., Ichikawa, J.K., Banger, M.G., Selgrade, S., Burns, J.L. *et al.* (2003) Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis. *Environ Microbiol* **5**: 1341–1349.
 Ghoul, M., West, S.A., Johansen, H.K., Molin, S., Harrison, O.B., Maiden, M.C.J., Jelsbak, L., Bruce, J.B., Griffin, A.S. (2015) Bacteriocin-mediated competition in cystic fibrosis lung infections. *Proc. R. Soc. B* **282**: 20150972. <http://dx.doi.org/10.1098/rspb.2015.0972>
 Henderson, A.G., Ehre, C., Button, B., Abdullah, L.H., Cai, L.-H., Leigh, M.W. *et al.* (2014) Cystic fibrosis airway secretions exhibit mucin hyperconcentration and increased osmotic pressure. *J Clin Invest* **124**: 3047–3060.
 Hocquet, D., Llanes, C., Thouverez, M., Kulasekara, H.D., Bertrand, X., Plésiat, P. *et al.* (2012) Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS Pathog* **8**: e1002778.
 Holloway, B.W., Rossiter, H., Burgess, D., and Dodge, J. (1973) Aeruginocin tolerant mutants of *Pseudomonas aeruginosa*. *Genet Res* **22**: 239–253.
 Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S. *et al.* (2003) Comprehensive

- transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **100**: 14339–14344.
- Kohler, T., Donner, V., and van Delden, C. (2010) Lipopolysaccharide as shield and receptor for R-pyocin-mediated killing in *Pseudomonas aeruginosa*. *J Bacteriol* **192**: 1921–1928.
- Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L. et al. (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* **50**: 1355–1361.
- Le, S., Yao, X., Lu, S., Tan, Y., Rao, X., Li, M. et al. (2014) Chromosomal DNA deletion confers phage resistance to *Pseudomonas aeruginosa*. *Sci Rep* **4**: 4738.
- Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Marvig, R.L., Sommer, L.M., Molin, S., and Johansen, H.K. (2015a) Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat Genet* **47**: 57–64.
- Marvig, R.L., Dolce, D., Sommer, L.M., Petersen, B., Ciofu, O., Campana, S. et al. (2015b) Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients. *BMC Microbiol* **15**: 1–13.
- Mayer-Hamblett, N., Rosenfeld, M., Gibson, R.L., Ramsey, B.W., Kulasekara, H.D., Retsch-Bogart, G.Z. et al. (2014) *Pseudomonas aeruginosa* in vitro phenotypes distinguish cystic fibrosis infection stages and outcomes. *Am J Respir Crit Care Med* **190**: 289–297.
- Mern, D.S., Ha, S.-W., Khodaverdi, V., Gliese, N., and Görisch, H. (2010) A complex regulatory network controls aerobic ethanol oxidation in *Pseudomonas aeruginosa*: indication of four levels of sensor kinases and response regulators. *Microbiology* **156**: 1505–1516.
- Michel-Briand, Y., and Baysse, C. (2002) The pyocins of *Pseudomonas aeruginosa*. *Biochimie* **84**: 499–510.
- Moran, N.A. (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583–586.
- Murugan, N., Malathi, J., Umashankar, V., and Madhavan, H.N.R. (2014) Comparative genomic analysis of multidrug-resistant *Pseudomonas aeruginosa* clinical Isolates VRFPA06 and VRFPA08 with VRFPA07. *Genome Announc* **2**: e00140–00114.
- Ochman, H., and Davalos, L.M. (2006) The nature and dynamics of bacterial genomes. *Science* **311**: 1730–1733.
- Ogunnariwo, J., and Hamilton-Miller, J.M.T. (1975) Brown- and red-pigmented *Pseudomonas aeruginosa*: differentiation between melanin and pyorubrin. *J Med Microbiol* **8**: 199–203.
- Olivares, E., Badel-Berchoux, S., Provot, C., Jaulhac, B., Prévost, G., Bernardi, T., and Jehl, F. (2016) The BioFilm Ring Test: a rapid method for routine analysis of *Pseudomonas aeruginosa* biofilm formation kinetics. *J Clin Microbiol* **54**: 657–661.
- Oliver, A., Canton, R., Campo, P., Baquero, F., and Blazquez, J. (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**: 1251–1254.
- Priebe, G.P., Dean, C.R., Zaidi, T., Meluleni, G.J., Coleman, F.T., Coutinho, Y.S. et al. (2004) The *galU* Gene of *Pseudomonas aeruginosa* is required for corneal infection and efficient systemic spread following pneumonia but not for infection confined to the lung. *Infect Immun* **72**: 4224–4232.
- Ramos A.J., Köhler, T., Nikaido, H., and Plésiat, P. (1999) Involvement of an efflux system in the natural resistance of *Pseudomonas aeruginosa* to aminoglycosides. *Antimicrob Agents Chemother* **43**: 2624–2628.
- Rau, M.H., Marvig, R.L., Ehrlich, G.D., Molin, S., and Jelsbak, L. (2012) Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environ Microbiol* **14**: 2200–2211.
- Rodriguez-Rojas, A., Mena, A., Martin, S., Borrell, N., Oliver, A., and Blazquez, J. (2009) Inactivation of the *hmgA* gene of *Pseudomonas aeruginosa* leads to pyomelanin hyperproduction, stress resistance and increased persistence in chronic lung infection. *Microbiology* **155**: 1050–1057.
- Sano, Y., and Kageyama, M. (1993) A novel transposon-like structure carries the genes for pyocin AP41, a *Pseudomonas aeruginosa* bacteriocin with a DNase domain homology to E2 group colicins. *Mol Gen Genet* **237**: 161–170.
- Sano, Y., Matsui, H., Kobayashi, M., and Kageyama, M. (1993) Molecular structures and functions of pyocins S1 and S2 in *Pseudomonas aeruginosa*. *J Bacteriol* **175**: 2907–2916.
- Spencer, D.H., Kas, A., Smith, E.E., Raymond, C.K., Sims, E.H., Hastings, M. et al. (2003) Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. *J Bacteriol* **185**: 1316–1325.
- Stoodley, P., Sauer, K., Davies, D.G., and Costerton, J.W. (2002) Biofilms as complex differentiated communities. *Annu Rev Microbiol* **56**: 187–209.
- Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J. et al. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959–964.
- Talon, D., Cailleaux, V., Thouverez, M., and Michel-Briand, Y. (1996) Discriminatory power and usefulness of pulsed-field gel electrophoresis in epidemiological studies of *Pseudomonas aeruginosa*. *J Hosp Infect* **32**: 135–145.
- Tielen, P., Wibberg, D., Blom, J., Rosin, N., Meyer, A.-K., Bunk, B. et al. (2014) Genome sequence of the small-colony variant *Pseudomonas aeruginosa* MH27, isolated from a chronic urethral catheter infection. *Genome Announc* **2**: e01174–13.
- Vogne, C., Aires, J.R., Bailly, C., Hocquet, D., and Plésiat, P. (2004) Role of the multidrug efflux system MexXY in the emergence of moderate resistance to aminoglycosides among *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis. *Antimicrob Agents Chemother* **48**: 1676–1680.
- Waite, R.D., and Curtis, M.A. (2009) *Pseudomonas aeruginosa* PAO1 pyocin production affects population dynamics within mixed-culture biofilms. *J Bacteriol* **191**: 1349–1354.
- Wargo, M.J., and Hogan, D.A. (2009) Identification of genes required for *Pseudomonas aeruginosa* carnitine catabolism. *Microbiology* **155**: 2411–2419.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Survival of the pyomelanin-producing mutants of *P. aeruginosa* in high pH and under oxidative stress. **a**, proportion of surviving bacteria after 24h of incubation at 37°C under shaking in LB broth pH9.5. Ratio of means \pm SEM from at least 3 independent experiments; two-sided Student's *t*-tests; NS, $P > 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **b**, proportion of surviving bacteria after 45 min of exposure to 50 mM H₂O₂ (Rodriguez-Rojas *et al.*, 2009). Ratio of means \pm SEM from at least 3 independent experiments; two-sided Student's *t*-tests; NS, $P > 0.05$; *, $P < 0.05$, **, $P < 0.01$; ***, $P < 0.001$.

Figure S2. Production of virulence factors by pyomelanin-producing *P. aeruginosa*. Means or ratio of means \pm SEM from at least 3 independent experiments; two-sided Student's *t*-tests; NS, $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **a**, elastase activity was quantitated with Elastin Congo Red assay. **b**, proportion of sheep red blood cells that were hemolyzed after incubation with the supernatant

of an overnight bacterial culture. **c**, the production of pyocyanin was quantified as previously described (see Suppl. materials).

Figure S3. Pulsed-field gel electrophoresis profiles of *DraI*-digested DNA from *P. aeruginosa* isolates. **a**, first series with A, B, C, and D pairs. **b**, complementary series. The brackets indicate two isogenic isolates isolated from the same patient - a wild-type isolate and its pyomelanin-producing mutant.

Table S1. PCR detection of the genes *pa1974*, *pa1989*, *pa2009* and *pa2018* among a collection of pyomelanin-producing clinical isolates of *P. aeruginosa* and in their isogenic parents. + and - signs indicate a positive and a negative PCR reaction with specific primers, respectively.

Table S2. Susceptibility of the pyomelanin-producing mutants and their isogenic counterparts to the major anti-pseudomonal compounds.

Table S3. Primers used in the study.

Annexe D

Article *Escherichia coli* ST131

Nicolas-Chanoine, M. H., Petitjean, M., Mora, A., Mayer, N., Lavigne, J. P., Boulet, O., ... & Hocquet, D. (2017). The ST131 *Escherichia coli* H 22 subclone from human intestinal microbiota : Comparison of genomic and phenotypic traits with those of the globally successful H 30 subclone. *BMC Microbiology*, 17(1), 71.

RESEARCH ARTICLE

Open Access



The ST131 *Escherichia coli* H22 subclone from human intestinal microbiota: Comparison of genomic and phenotypic traits with those of the globally successful H30 subclone

Marie-Hélène Nicolas-Chanoine^{1,2,3*}, Marie Petitjean^{4,5}, Azucena Mora⁶, Noémie Mayer¹, Jean-Philippe Lavigne^{7,8}, Olivier Boulet⁹, Véronique Leflon-Guibout¹, Jorge Blanco⁶ and Didier Hocquet^{4,5}

Abstract

Background: In 2006, we found healthy subjects carrying ST131 *Escherichia coli* in their intestinal microbiota consisting of two populations: a subdominant population of fluoroquinolone-resistant *E. coli* belonging to subclone H30 (H30-R or subclade C1), the current worldwide dominant ST131 subclone, and a dominant *E. coli* population composed of antibiotic-susceptible *E. coli* belonging to subclone H22 (clade B), the precursor of subclone H30. We sequenced the whole genome of fecal H22 strain S250, compared it to the genomes of ExPEC ST131 H30-Rx strain JJ1886 and commensal ST131 H41 strain SE15, sought the H22-H30 genomic differences in our fecal strains and assessed their phenotypic consequences.

Results: We detected 173 genes found in the Virulence Factor Database, of which 148 were shared by the three ST131 genomes, whereas some were genome-specific, notably those allowing determination of virotype (D for S250 and C for JJ1886). We found three sequences of the FimH site involved in adhesion: two in S250 and SE15 close and identical, respectively, to that previously reported to confer strong intestinal adhesion, and one in JJ1886, corresponding to that commonly present in uropathogenic *E. coli*. Among the genes involved in sugar metabolism, one encoding a gluconate kinase lacked in S250 and JJ1886. Although this gene was also absent in both our fecal H22 and H30-R strains, H22 strains showed a higher capacity to grow in minimal medium with gluconate. Among the genes involved in gluconate metabolism, only the *ghrB* gene differed between S250/H22 and JJ1886/H30-R strains, resulting in different gluconate reductases. Of the genes involved in biofilm formation, two were absent in the three genomes and one, *fimB*, in the JJ1886 genome. Our fecal H30-R strains lacking intact *fimB* displayed delayed biofilm formation relative to our fecal H22 strains. The H22 strains differed by subclade B type and plasmid content, whereas the H30-R strains were identical.

Conclusions: Phenotypic analysis of our fecal strains based on observed genomic differences between S250 and JJ1886 strains suggests the presence of traits related to bacterial commensalism in our H22 strains and traits commonly found in uropathogenic *E. coli* in our H30-R strains.

Keywords: *E. coli* ST131, H22 genome, Sugar metabolism, Mannose-binding FimH region, Biofilm, Subclades B, Plasmid replicons

* Correspondence: mhn.chanoine@aphp.fr

¹Service de Microbiologie, Hôpital Beaujon, AP-HP, Clichy, France

²Faculté de Médecine Paris Diderot, Paris, France

Full list of author information is available at the end of the article



Background

Phylogenetic group B2, sequence type (ST) 131 *Escherichia coli* has been a worldwide dominant human extra-intestinal pathogenic *E. coli* (ExPEC) since the beginning of the 2000s, and is among those resistant to fluoroquinolones and/or producing the extended-spectrum β -lactamase (ESBL) CTX-M-15 [1]. Its dominance was shown to be driven by the expansion of a subclone harboring type 1 fimbriae-encoding *fimH* allele 30 (subclone H30), comprised of mostly strains resistant to fluoroquinolones (H30-R) [2]. Within this subclone, a subgroup called H30-Rx, mostly comprising strains resistant to fluoroquinolones and producing CTX-M-15, has quickly emerged and disseminated [2]. The evolutionary history of clone ST131 revealed that before the emergence and dissemination of subclone H30 (also called clade C), the ST131 population consisted of mostly two subclones called H22 (clade B) and H41 (clade A), with subclone H22 comprised of mostly fluoroquinolone-susceptible isolates [3–5]. It also revealed that subclone H22 was the precursor of subclone H30 and that separate F-type plasmids have shaped the evolution of subclone H30 [2, 3, 5, 6]. The most recent ST131 *E. coli* phylogenetic reconstruction carried out by Ben Zakour et al. using 3779 non-recombinant single nucleotide polymorphisms (SNP) found in the high quality genomes of 172 clade B and C strains, identified five B subclades, for which the independent evolutionary trajectories are characterized by successive insertions and a recombination from ancestral H22 (subclade B1) [7]. Subclade B2 is characterized by a Flag-2 locus insertion, subclade B3 by a Flag-2 locus insertion and *par-C2* recombination, subclade B4 by Flag-2 locus and GI-PheV insertions, and subclade B5 by Flag-2 locus and Phi3 insertions. Other insertions, notably that of GI-PheV, recombination events, and mutations in *gyrA* and *parC* occurred within strains of subclade B5, resulting in their evolution to clade C and subclades C1 (H30-R) and C2 (H30-Rx). Epidemiologically, the 172 isolates consisted of mostly those from North America obtained between 1948 and 2011, with most collected between 2000 and 2010, irrespective of the clade and subclade types. Since the first description of the *fimH* lineage in 2013 [3], the *fimH* type has been found in several epidemiological studies published on ExPEC or fecal ST131 isolates [8–17]. However, these have mostly concerned fluoroquinolone-resistant and /or ESBL-producing isolates. In 2006, we retrieved non-ESBL-producing ST131 *E. coli* isolates, susceptible or resistant to fluoroquinolones, from the intestinal microbiota of 7% of healthy subjects living in the Paris area [18]. The fluoroquinolone-resistant isolates accounted for a subdominant *E. coli* population in four independent healthy subjects, whereas those susceptible to fluoroquinolones accounted for the dominant *E. coli* population in

three other independent subjects. The latter finding strongly suggests that ST131 *E. coli* is part of the normal intestinal microbiota of humans. In the present study, we aimed to determine whether these isolates belong to different H subclones, based on their fluoroquinolone-susceptibility pattern. The dominant fluoroquinolone-susceptible ST131 *E. coli* populations belonged to subclone H22. Assembled and annotated genomes of H22 strains were not available when we performed the present study. Thus, we first sequenced the whole genome of one of our commensal H22 strains (S250) and compared it to the genome of two reference ST131 strains: the ExPEC H30-Rx strain JJ1886 and the commensal H41 strain SE15. This genome comparison focused on virulence factor (VF)-encoding genes and those encoding processes or structures (sugar metabolism, biofilm formation, and the FimH mannose-binding site) known to be involved in the adaptation of the bacteria to different environments including human intestine. The four fluoroquinolone-resistant fecal ST131 isolates were H30-R strains. Thus, we compared them to the three fecal H22 strains considering the genomic differences identified between H22 strain S250 and H30-Rx strain JJ1886, and analyzed the phenotypic impact of some of these differences.

Methods

Bacterial strains

The seven fecal ST131 isolates (strains 02, 39, 183, 187, 196, 208, and S250) obtained from the intestinal microbiota of seven healthy subjects living in the Paris area in 2006 were included in the study. They display serotype O25:H4, do not produce ESBL, and are either susceptible (196, 208 and S250) or resistant (02, 39, 183 and 187) to fluoroquinolones [18]. The fluoroquinolone-resistant strains were retrieved by plating the feces of healthy subjects on nalidixic acid-containing plates. The fluoroquinolone-susceptible strains accounted for the dominant *E. coli* population of three subjects. We previously assessed the virulence potential of strain S250 in the *Caenorhabditis elegans* and zebrafish models and analyzed its genome by optical mapping. This strain had a level of virulence similar to that of the multidrug resistant ST131 isolates, with which it shared 86% genome similarity [19]. We also included the *E. coli* K-12 MG1655 reference strain in the study as a control in the experiments assessing the use of gluconate as the sole source of carbon.

fimH type

fimH typing was based on the internal 489-nucleotide (nt) sequence of the *fimH* gene as previously described [20].

Antibiotic susceptibility

Antibiotic susceptibility was determined by the agar diffusion method and interpreted following the 2015 EUCAST recommendations (www.eucast.org). The following antibiotics were tested: amoxicillin, amoxicillin + clavulanic acid, ciprofloxacin, gentamicin, amikacin, cotrimoxazole and fosfomycin.

Molecular analysis of resistance mechanisms

The genes encoding resistance to amoxicillin (TEM and SHV enzymes) were identified by PCR and sequencing methods as previously described [21]. The *qnr* genes encoding plasmid-mediated resistance to fluoroquinolones were tested and the quinolone resistance determining region (QRDR) of the genes *gyrA*, *gyrB*, *parC* and *parE* amplified and sequenced using methods previously described [21–23]. The QRDRs of our strains were compared with those of the fluoroquinolone-susceptible reference strain *E. coli* K-12 MG1655 [GenBank: CP014225.1] and the allelic *gyrA* and *parC* profiles were compared with those previously described for ST131 isolates [3].

Sequencing and analysis of the whole genome of H22 strain S250

The complete genomic sequence was determined for H22 strain S250. Total DNA was extracted using the Qiagen Blood & Cell Culture DNA Mini Kit (Qiagen, Courtaboeuf, France). Libraries were constructed using Nextera technology and sequenced on an Illumina HiSeq-2000 using a 2 × 100 nucleotides (nt) paired-end strategy. All reads were processed to remove low quality or artefactual nucleotides, using sequentially sickle (www.github.com/najoshi/sickle), AlienTrimmer [24] and fqDuplicate (ftp.pasteur.fr/pub/gensoft/projects/fqtools). Read pairs were assembled using *clc_assembler* from the CLC Genomics Workbench analysis package (www.clcbio.com/products/clc-genomics-workbench). All contigs of ≥500 nt were reordered and reoriented, using the genomic sequence of strain *E. coli* K-12 MG1655 as a reference, with Mauve Contig Mover [25]. The reordered contigs were analyzed. The genome of H22 strain S250 was compared to that of two reference ST131 strains, JJ1886 (H30-Rx) [GenBank: CP006784.1] and SE15 (H41) [GenBank: AP009378.1], focusing on VF-encoding genes and genes involved in sugar metabolism, biofilm formation, and methylation. We downloaded VFs (www.mgc.ac.cn/VFs/) available from the Virulence Factor Database (VFDB) and searched the three genomes for their presence using Prodigal v2.6.1 [26] and clustered them at 90% identity using CD-hit v4.6 [27]. We extracted the sequences of the genes involved in biofilm formation in strain *E. coli* K12 BW25113 [28] from its genome [GenBank: CP009273.1]

and clustered them at 90% identity with CD-hit. We assessed the percentage identity between the genes from *E. coli* K12 BW25113 and the three studied genomes. Using the NCBI basic local alignment search tool (BLAST), we blasted genes from *E. coli* K12 MG1655 against the S250, JJ1886, and SE15 genomes to verify the possible absence of any genes involved in biofilm formation. We then searched the genomes of ST131 strains S250, JJ1886, and SE15 for genes involved in sugar metabolism as defined by Maltby et al. [29] in commensal strains *E. coli* HS [GenBank: CP000802.1] and *E. coli* Nissle 1917 [GenBank: CP007799.1] using BLAST. Moreover, we also searched for all the genes involved in gluconate metabolism (main pathway: *gntR*, *gntT*, *gntU*, *gntP*, *gntK*; secondary pathway: *idnT*, *idnDOTR*; gene *idnK* (*gntV*) which plays a role in the two pathways; the Entner-Doudoroff pathway: *edd* and *eda*; additional genes involved in other pathways: *kdgT*, *kdgK*, *gnd*, *tkrA*, *ghrB*, *kduD*, *dkgA* and *dkgB*) [30], as well as the 100 base pairs (bps) upstream of the start codon of each, in the genome of strains S250 and JJ1886. The nucleotide sequence of the 100 bp-upstream regions and the sequences of the deduced proteins for each gene of the strains S250 and JJ1886 were compared. We also searched for genes encoding methyltransferases in ST131 strain EC958 [GenBank: HG941718.1] [31] in the genome of ST131 strains S250, JJ1886, and SE15 using BLAST.

Subclade B and clade C typing

According to the work of Ben Zakour et al. [7], we determined the subclade B type of H22 strain S250 and the clade C type of strain JJ1886 by direct blasting of the genes of the Flag-2 locus (Flag-2 locus from strain *E. coli* 042 [EMBL: CR 753847]), Phi3 (from strain EC958), and GI-PheV (from strain JJ1886) against the genome of strains S250 and JJ1886. We then determined the type of subclade B displayed by strain S250 and the two other fecal H22 strains by PCR using primers specific for the Flag-2 locus, Phi3, and GI-PheV (Additional file 1: Table S1) and our H30-R fecal strains as a positive control.

Plasmid content

The plasmid content of H22 strain S250 was determined using the PlasmidFinder system [32] and the FIL, FIA, and FIB formula of the detected IncF plasmid by PCR-based replicon typing (<http://pubmlst.org/plasmid/>). The latter method was also applied to the two remaining fecal H22 strains and the four fecal H30-R strains.

ExPEC status and virotype

According to the study of Johnson and Adam, ExPEC status is defined by the presence of ≥2 VF genes among the following genes: *pap*, *sfa/focDE*, *afa/draBC*, *iutA*,

and *kpsMT II* [33]. As the latter gene is not included in the VFDB, we searched for it by blasting the *kpsMT II* genes [GenBank: X53819] against the S250, JJ1886, and SE15 genomes. Based on the study of Blanco et al., the major virotypes of *E. coli* ST131 are defined using four genes as follows: virotype A = *afaFM955459*⁺ *iroN*⁻ *ibeA*⁻ *sat*^{+/-}, virotype B = *afaFM955459*⁻ *iroN*⁺ *ibeA*⁻ *sat*^{+/-}, virotype C = *afaFM955459*⁻ *iroN*⁻ *ibeA*⁻ *sat*⁺, and virotype D = *afaFM955459*^{+/-} *iroN*^{+/-} *ibeA*⁺ *sat*^{+/-} [34]. As *afaFM955459* is not included in the VFDB, we searched for it by blasting the *afaFM955459* operon [EMBL: FM955459] against the S250, JJ1886, and SE15 genomes [35]. Then, classic multiplex PCR was used to search for genes encoding ExPEC-associated VFs (Additional file 2: Table S2) [34] in the seven fecal strains to confirm the results provided by the direct genome analysis of strain S250 and to determine the VF profile, virotype, and ExPEC status of the six remaining fecal strains.

FimH structure

According to the study of Sokurenko et al., the amino acid variations observed within the sequence of the adhesin, FimH, of type 1 fimbriae, result in different levels of binding to mono-mannose (M₁) structures, whereas they have no impact on the normal high level of binding to tri-mannose (M₃) structures [36]. By measuring the ratio of M₁/M₃ binding in different experimental models, they defined low M₁-binding (M₁/M₃ < 0.1) and high M₁-binding (M₁/M₃ > 0.90) FimH phenotypes and showed that they are related to specific FimH sequences [37]. Therefore, the deduced protein sequence of FimH of H22 strain S250 was compared with that of H30-Rx ST131 strain JJ1886 [GenBank: AGY86963.1], H41 ST131 strain SE15 [GenBank: BAI57801.1], phylogroup A *E. coli* strain MG1655 [GenBank: AMC97175.1], and two other phylogroup B2 strains: ST73 strain CFT073 [GenBank: AAN83822.1] and ST131 uk_P46212 [GenBank: ALT52319.1]. The deduced protein sequence of FimH was then characterized for the six remaining fecal strains by using primers for which the sequences have been previously published [20].

Amplification and sequencing of the *idnK* (*gntV*) and *ghrB* genes of the seven fecal strains and growth with gluconate as the sole carbon source

The *idnK* (*gntV*) gene, encoding a thermosensitive D-gluconate kinase, and the *ghrB* gene, encoding a gluconate reductase, were amplified with primers indicated in Additional file 1: Table S1. Additional primers were used to sequence the *ghrB* gene (Additional file 1: Table S1). We evaluated the ability of the seven strains to grow in the presence of gluconate as the sole carbon source, as previously described [38]. Experiments were conducted three times independently and all incubations were

performed overnight at 37 °C with shaking (150 rpm). Briefly, bacteria were first cultured in Mueller Hinton broth before washing twice in minimal media M63 [15 mM (NH₄)₂SO₄; 100 mM KH₂PO₄; 0.002 mM FeSO₄ (7H₂O)]. We inoculated 5 ml of M63 supplemented with 0.2% glucose (Sigma-Aldrich, France) with 10 µl of the washed bacteria. One ml of this culture was washed twice in M63. We transferred 10 µl of this washed culture into 5 ml of M63 with 0.2% gluconate (Sigma-Aldrich, Saint-Quentin Fallavier, France). The cultures were then adjusted to 0.002 at OD₆₀₀ in 30 ml fresh M63 with 0.2% gluconate and incubated at 37 °C with shaking (150 rpm). We estimated the bacterial growth after 48 h by measuring the OD₆₀₀ [39]. Tukey's test was used for intergroup comparisons and R software for statistical analyses. *P* values < 0.01 were considered to be statistically significant.

Amplification and sequencing of the *fimB* gene and kinetics of biofilm formation in the seven fecal strains

The *fimB* gene was amplified and sequenced with the primers indicated in Additional file 1: Table S1. The kinetics of early biofilm formation was assessed using the BioFilm Ring Test* (BioFilm Control, Saint Beauzire, France), as described [40]. Briefly, standardized bacterial cultures were incubated at 37 °C in a 96-well microtiter plate in the presence of magnetic beads. At various time points, the plates were placed onto a magnetic test block and put in the reader. The images of each well before and after magnetic attraction were analyzed using BioFilm Control software that gives a BioFilm Index (BFI). The BFI was converted into the proportion of immobilized beads relative to a reference condition (% RBI) using the formula: % RBI = $\sqrt{[(1-(BFI_{\text{assay}}-BFI_{\text{min}}))/(BFI_{\text{control}}-BFI_{\text{min}})]} \times 100$, where BFI_{assay} is the BFI of the tested strain, BFI_{control} is the BFI of the control, corresponding to the maximum BFI, and BFI_{min} is the minimal observed BFI when all the beads are blocked. The more RBI approaches a value of 1, the more the biofilm is fully formed (beads are immobilized). Three experiments were performed in duplicated per strain and per incubation time. The kinetics of biofilm formation were compared using a two-way ANOVA followed by Dunnett's multiple comparisons test.

Results

fimH type, antibiotic susceptibility, molecular analysis of resistance mechanisms, and allelic profiles of the *gyrA* and *parC* genes

The three fluoroquinolone-susceptible strains (S250, 208 and 196) were of the *fimH22* type. All but one were susceptible to all of the antibiotics tested (Table 1). H22 strain 196 was resistant to both amoxicillin and cotrimoxazole. The four fluoroquinolone-resistant strains

Table 1 Characterization of the seven fecal strains of *E. coli* ST131

Strain/ <i>FimH</i> type	Susceptibility								β-lactamase	Allele type ^a (amino acid substitution)			Flag-2	Phi3	GI-PheV	Subclade	IncF plasmid replicon
	AMX	AMC	CIP	GEN	AMK	SXT	FOS			<i>gyrA</i>	<i>parC</i>	<i>parE</i>					
S250/H22	S	S	S	S	S	S	S	–	1a	1	NA (I529L)	–	–	–	B1	F89A-B62	
208/H22	S	S	S	S	S	S	S	–	1a	1	NA (I529L)	+	–	–	B2	–	
196/H22	R	S	S	S	S	R	S	TEM-1	1a	1	NA (I529L)	+	–	+	B4	F24:A-B6	
187/H30	R	S	R	S	S	S	S	TEM-1	1AB (S83 L/D87N)	1aAB (S80I/E84V)	NA (I529L)	+	+	+	C1	F1:A2:B20	
183/H30	R	S	R	S	S	S	S	TEM-1	1AB (S83 L/D87N)	1aAB (S80I/E84V)	NA (I529L)	+	+	+	C1	F1:A2:B20	
39/H30	R	S	R	S	S	S	S	TEM-1	1AB (S83 L/D87N)	1aAB (S80I/E84V)	NA (I529L)	+	+	+	C1	F1:A2:B20	
02/H30	R	S	R	S	S	S	S	TEM-1	1AB (S83 L/D87N)	1aAB (S80I/E84V)	NA (I529L)	+	+	+	C1	F1:A2:B20	

^aaccording to reference [3], AMX amoxicillin, AMC amoxicillin + clavulanic acid, CIP ciprofloxacin, GEN gentamicin, AMK amikacin, SXT cotrimoxazole, FOS fosfomycin, R resistant, S susceptible, –: absence, +: presence, NA not available

(187, 183, 39 and 02), which were all resistant to amoxicillin but susceptible to the other antibiotics tested, were of the *fimH30* type (*H30-R*) (Table 1). QRDR nucleotide sequence analysis showed that the three *H22* strains displayed *gyrA1a* and *parC1* alleles, whereas the four *H30-R* strains displayed the *gyrA1AB* allele, encoding amino acid substitutions S83 L and D87N, and the *parC1aAB* allele, encoding amino acid substitutions S80I and E84V (Table 1). We found the substitution I529L in ParE in the seven fecal strains (Table 1). None of these strains harbored plasmid-mediated *qnr* genes. *H22* 196 and the four *H30-R* strains that were resistant to amoxicillin harbored a TEM-1-encoding gene (Table 1).

Genomic and phenotypic characterization

We assembled the whole genome sequence of the fecal *H22* strain S250 into 50 contigs and analyzed and compared it with those of the ExPEC *H30-Rx* strain JJ1886 and the commensal *H41* strain SE15.

Subclade B and C type

We were unable to find the genes composing the Flag-2 locus in the genome of *H22* strain S250 using BLAST. We only found a fragment of approximately 1700 bp which was very similar to the end of the first gene, *IfhA*, and another of 769 bp similar to the end to the last gene, *lafU*, of the Flag-2 locus (data not shown). We found neither Phi3 nor GI-PheV in the genome of *H22* strain S250, whereas we found them, as well as the Flag-2 locus, in the genome of JJ1886. The use of specific primers allowed us to confirm the absence of these genetic elements in *H22* strain 250 and to detect the Flag-2 locus in *H22* strains 196 and 208, as well as GI-PheV in

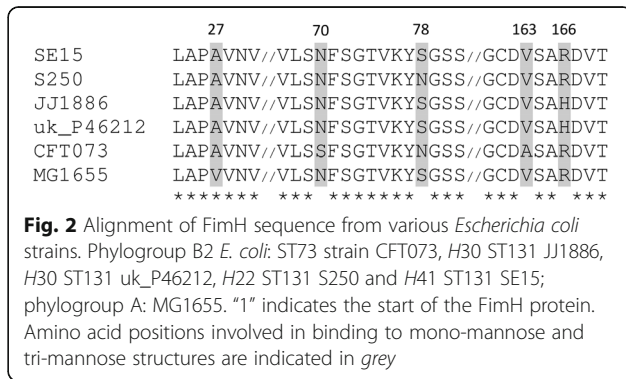
H22 strain 196. These three genetic elements were amplified from our four *H-30R* strains (Table 1).

VF-encoding genes

Using a gene identity level of $\geq 90\%$, a total of 173 genes among the 2520 *Escherichia* sp. VF-encoding genes of the VFDB was identified in the genomes of the three ST131 strains studied. *H22* strain S250, *H30-Rx* strain JJ1886, and *H41* strain SE15 had 160, 159, and 152 VF-encoding genes, respectively. The three ST131 genomes shared 148 virulence genes (Fig. 1). Nine genes were found specifically in *H22* strain S250: the five *iroBCDEN* genes that encode proteins related to a catecholate siderophore, the three *pixCDH* genes encoding Pix pilus adhesion, and the *ibeA* gene involved in invasion of brain endothelium (Fig. 1). Ten genes were specifically found in *H30-Rx* strain JJ1886: the four *iucABCD* and *iutA* genes encoding proteins involved in the binding and transport of iron, the *pap1* and *papX* genes encoding *pap* operon regulatory proteins, the *sat* gene encoding a toxin, the *flu* gene encoding autotransporter protein Ag43a, and the *iha* gene encoding the adhesion-siderophore receptor. Three genes, including the *fimB* gene, were present in strains *H22* S250 and *H41* SE15, but not *H30-Rx* strain JJ1886 (Fig. 1). The ECP_2810 gene, encoding a Val-Gly Repeats-related protein, was present in *H41* strain SE15 and *H30-Rx* strain JJ1886.

Virotype and ExPEC status

The VF genes identified in the VFDB would suggest virotype D (*ibeA*⁺, *iroN*⁺, and *sat*⁻) for strain S250 and virotype C (*sat*⁺, *ibeA*⁻, and *iroN*⁻) for strain JJ1886. These virotypes were confirmed by the absence of the



the study of Sokurenko et al. in strain SE15. We found the substitutions A27V and V163A, that are each related to a ten-fold increase in the M_1/M_3 ratio in the Sokurenko et al., study in strain MG1655 and strain CFT073, respectively. We also found the substitution R166H related to a M_1/M_3 ratio of 0.33 in the same study in H30-Rx strains JJ1886 and uk_P46212. H22 strain S250 displayed a sequence not published in the Sokurenko et al. study as residue N was present at both positions 70 and 78. However, we found none of the substitutions known to induce an increase in the M_1/M_3 ratio in the FimH sequence of strain S250. We found the same feature in the two other fecal H22 strains, whereas the FimH sequence of our four H30-R strains was identical to that of H30-Rx strains JJ1886 and uk_P46212 (data not shown).

Genes involved in sugar metabolism and growth with gluconate as the sole carbon source

All genes involved in sugar metabolism in commensal *E. coli* strains HS and Nissle were detected in the genome of H41 strain SE15, whereas one, the D-gluconate kinase-encoding *idnK* (*gntV*) gene, was not detected in either strain H30-Rx JJ1886 or H22 S250 (Table 3). Its absence in our six remaining strains was revealed by the PCR-sequencing assay. Therefore, we assessed the ability of our strains to grow with gluconate as the sole carbon source in the absence of the *idnK* (*gntV*) gene and the presence of the *gntK* gene, which both transform D-gluconate to 6P-gluconate. After 48 h of growth, the biomass of our H30-R strains was significantly lower than that for our H22 strains (Tukey's test $p < 0.01$) (Fig. 3). This unexpected difference led us to compare the genomes of strains H22 S250 and H30-Rx JJ1886 focusing on all the genes, other than the *idnK* (*gntV*) gene, involved in gluconate metabolism, as well as the 100 bp upstream of the start codon of the operons or genes. All were present in the genome of strains S250 and JJ1886. The 100 bp upstream of the start codon and the sequence of the deduced proteins showed 100% identity (data not shown) between these two strains, except for the coding region of the *ghrB* gene,

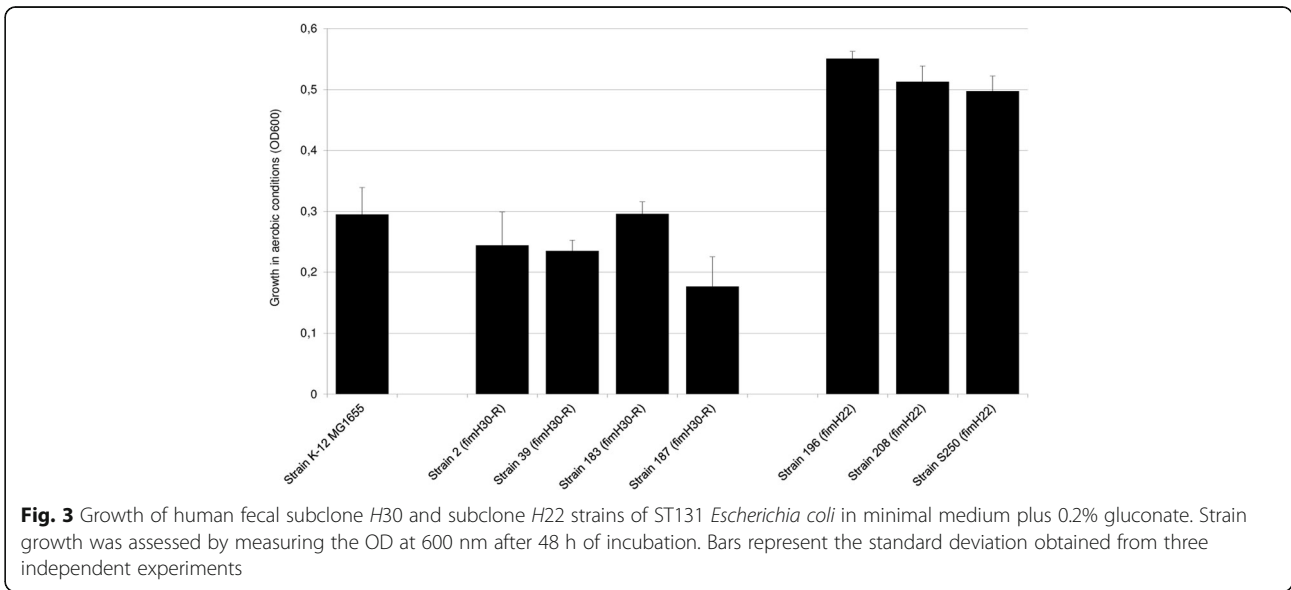
Table 3 Distribution of the genes involved in sugar metabolism in commensal *Escherichia coli* strains HS and Nissle [29] in the genomes of representative H30, H22, and H41 subclones of *Escherichia coli* ST131

Commensal strain/gene	Catabolized sugar	H30 JJ1886	H22 S250	H41 SE15
<i>E. coli</i> HS				
<i>araA</i>	Arabinose	+	+	+
<i>araB</i>	Arabinose	+	+	+
<i>araD</i>	Arabinose	+	+	+
<i>fucK</i>	Fucose	+	+	+
<i>galk</i>	Galactose	+	+	+
<i>gntK</i>	Gluconate	+	+	+
<i>idnK</i>	Gluconate	-	-	+
<i>lacZ</i>	Lactose	+	+	+
<i>manA</i>	Mannose	+	+	+
<i>nagE</i>	N-acetylglucosamine	+	+	+
<i>nanA</i>	N-acetylneuraminate	+	+	+
<i>nanT1</i>	N-acetylneuraminate	+	+	+
<i>rbsK</i>	Ribose	+	+	+
<i>uxaC</i>	Glucuronate	+	+	+
<i>E. coli</i> Nissle 1917				
<i>agaA</i>	N-acetylglactosamine	+	+	+
<i>agaE</i>	N-acetylglactosamine	+	+	+
<i>agaF</i>	N-acetylglactosamine	+	+	+
<i>agaW</i>	N-acetylglactosamine	+	+	+

which showed a T968C substitution, resulting in amino acid substitution V323A in strain JJ1886. This substitution was confirmed in our four fecal H30-R strains, whereas V323 was identified in the two remaining H22 fecal strains.

Biofilm genotype and phenotype

The expression of a wide panel of genes can influence biofilm formation [28]. We specifically searched for these genes in the genome of strains JJ1886, S250, and SE15 (Additional file 3: Table S3). Two genes of the biofilm gene panel, *fliC* and *fliD*, were missing from the three tested genomes. We aligned the *fliC* and *fliD* genes from *E. coli* strain MG1655 onto the three genomes (JJ1886, S250 and SE15) to verify their absence and found deletions within *fliC* and the absence of *fliD* in the three strains. An intact version of the *fimB* gene was present in strains H22 S250 and H41 SE15, but it was disrupted in H30-Rx JJ1886 due to the insertion of IS3-like, as previously described in various H30 strains [34, 41]. The difference in the structure of the *fimB* gene in strains H30-Rx JJ1886 and H22 250 led us to sequence the *fimB* gene of the remaining fecal strains. We found that the *fimB* nucleotide sequence in the four H30-R strains was identical to that of H30-Rx strain JJ1886,



whereas that of the two remaining *H22* strains was identical to that of strain S250. Niba et al. previously reported that the deletion of *fliC* and/or *fliD* induces a substantial decrease in biofilm formation, whereas deletion of *fimB* results in its near complete absence (Additional file 3: Table S3). We thus tested whether this *fimB* polymorphism affects biofilm formation. Kinetic measurements of early biofilm formation showed that the three *H22* strains formed biofilms significantly earlier ($p < 0.0001$) than the four *H30*-R strains (Fig. 4). Of note, *H30*-R strain 39, which was the single fecal strain with an ExPEC status, showed significantly ($p < 0.0001$)

slower biofilm production (incomplete biofilm formation after 24 h) than the other *H30*-R strains.

Genes encoding methyltransferases

We found all but one of the ten methyltransferases, previously identified in the genome of *H30* ST131 strain EC958 [31], in the genome of *H30*-Rx strain JJ1886 (Table 4). We found six in the genome of *H41* strain SE15 and only three in the genome of *H22* strain S250 (Table 4).

Plasmid content

We detected an IncFII plasmid in contig 4 and an IncFIB element in contig 41 of the S250 genome using PlasmidFinder. Using NCBI BLAST and the plasmid MSLT typing system, we identified the replicon alleles

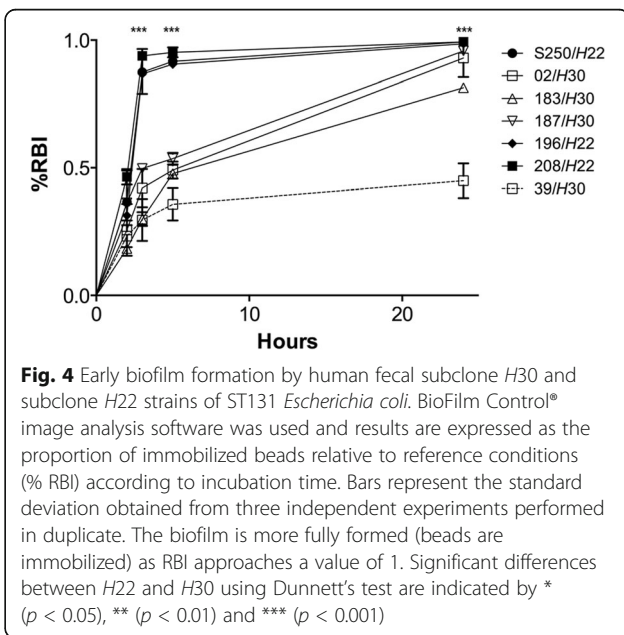


Table 4 Distribution of the 10 methyltransferases previously described in *H30* strain EC958, [31] in the genomes of representative *H30*, *H22*, and *H41* subclones of *Escherichia coli* ST131

Methyltransferase (<i>H30</i> EC958)	<i>H30</i> JJ1886	<i>H22</i> S250	<i>H41</i> SE15
M.EcoMIV	+	-	-
M.EcoMV	+	-	-
M.EcoMDcm	+	+	+
EcoMIII	+	-	+
M.EcoMVI	+	+	+
M.EcoMDam	+	+	+
M.EcoMVII	-	-	-
M1.EcoMI	+	-	+
M2.EcoMI	+	-	+
M.EcoMII	+	-	-
Total number	9	3	6

+: presence, -: absence

F89 and B62 in the whole genome of strain S250. We confirmed this result for strain S250 by PCR-sequencing, and also obtained the following results for the six remaining fecal strains: F1:A2:B20 for the four *H30-R* strains, no IncF plasmid for *H22* strain 208, and F24:B6 for *H22* strain 196 (Table 1).

Discussion

CTX-M-15 producing, fluoroquinolone-resistant ST131 *E. coli*, which has been shown to be a worldwide human ExPEC since the beginning of the 2000s, has also been shown to colonize the human digestive tract [1]. In 2006, we found healthy subjects with a subdominant intestinal population of fluoroquinolone-resistant ST131 *E. coli*, as well as those with a dominant intestinal population of ST131 *E. coli* susceptible to fluoroquinolones [18]. The latter finding suggests that ST131 *E. coli* may be a human intestinal commensal. Here, we first showed that our three antibiotic-susceptible fecal ST131 *E. coli* strains displayed *fimH22*, whereas the four resistant to fluoroquinolones displayed *fimH30* (*H30-R*). We thus centered our study on the comparison between these two groups of strains focusing on processes and structures involved in environmental adaptation, considering that subclone *H22* is the precursor of subclone *H30*, which emerged at the end of 1990s [3, 7]. We first sequenced the whole genome of one of our fecal *H22* strains, strain S250, as there were no assembled and annotated genomes of *H22* strains when we started our work. In contrast, there were sequenced whole genomes of representative multidrug resistant *H30* ExPECs, such as *H30-Rx* strain JJ1886 and a commensal ST131 *E. coli*, strain SE15, belonging to another ST131 subclone, characterized by the allele *fimH41*. Comparison of the genomes of *H22* strain S250, *H30-Rx* strain JJ1886, and *H41* strain SE15 showed high similarity in terms of the number and types of VF genes from the VFDB. This may explain the similar level of virulence that we found previously between *H22* strain S250 and an *H30-Rx* strain in the *C. elegans* model [19, 42]. Nevertheless, strains S250 and JJ1886 displayed two different virotypes: D for strain S250 and C for strain JJ1886. Moreover, strain S250 did not display the VF genes required for ExPEC status, whereas strain JJ1886 did [43]. There was also high similarity between the genes involved in biofilm formation in the three genomes [28], except for the *fimB* gene, which was disrupted in the *H30-Rx* strain JJ1886 and intact in both *H22* strain S250 and *H41* strain SE15. As the intact *fimB* gene was also absent from our fecal *H30-R* strains and present in our remaining fecal *H22* strains, we analyzed the kinetics of early biofilm formation in our seven fecal strains. Biofilm formation was significantly delayed in the *H30-R* strains. This is the first study to compare the formation of

biofilms by *H22* and non-ESBL-producing *H30-R* strains [4, 42, 44–46]. Further studies are required to verify the involvement of the *fimB* gene in the two different biofilm phenotypes. Comparison of the genes of the three genomes involved in the metabolism of various sugars showed the absence of one of the two genes encoding gluconate kinases in strains S250 and JJ1886. Although this gene was absent in both our fecal *H22* and *H30-R* strains, we found that *H22* strains had a higher capacity to grow in medium with gluconate as the sole carbon source. Thus, gluconate consumption may confer an advantage to *H22* strains for residing in the human intestine, as gluconate is a component of intestinal mucus [47]. Indeed, it has been shown that *E. coli* laboratory mutants with impaired growth on gluconate are less able to colonize the large intestine of mice [38]. Comparison of all genes, and their promoter regions, involved in the different pathways of gluconate metabolism showed one difference between strains S250 and JJ1886, namely a mutation leading to an amino acid substitution in the 2-keto-D-gluconate reductase GhrB in strain JJ1886. Although this genetic difference was found between the fecal *H22* and *H30-R* strains, further studies are required to clarify the role of this mutation for the difference in the growth of the two strains when gluconate is the sole carbon source. Another difference we identified between strain S250 and strain JJ1886 concerned the sequence of the type 1 fimbriae FimH adhesin, shown by Sokurenko et al. to be involved in variations of the M_1/M_3 ratio associated with tissue tropism and the shift of bacterial adaptation from a commensal to pathological habitat [36, 37]. The FimH sequence of strain JJ1886, which was also present in our *H30-R* fecal strains, corresponded to that identified by Sokurenko et al. in strains with a M_1/M_3 ratio = 0.33. This sequence and M_1/M_3 ratio are commonly observed in uropathogenic strains [36, 37]. The FimH sequence identified in strain S250 and the other fecal *H22* strains was not found in the strains studied by Sokurenko et al. and was only found at a low frequency in those studied by Chen et al. [48]. This sequence showed none of the amino acid substitutions involved in the increase of the M_1/M_3 ratio. Thus it is likely to be associated with the lowest M_1/M_3 ratio shown by Sokurenko et al., which was associated with strong adhesion to intestinal cells [36, 37]. Altogether, although the *H22* and *H30-R* strains investigated in this work were all isolated from the digestive tract of healthy subjects, only *H22* strains displayed properties previously shown to be associated with human intestinal commensalism, including competition for nutrients in the intestine (gluconate use) and a high intestinal adhesiveness [49].

We sought to confirm the genetic differences identified here between the lineages *H22* and *H30* with the

whole genome sequences of ST131 isolates recently deposited on NCBI databases (i.e. 28 *H22* strains, 9 *H30* strains, see Additional file 4: Table S4). As expected, we retrieved all but one of these genetic differences in this wider *H22* and *H30* strain collection. Indeed, the gluconate reductase-encoding *ghrB* gene in two *H22* clinical isolates (strain JJ1897 and strain GN02448 from the bio-project PRJNA290784: Additional file 4: Table S4) displayed the same sequence as in all *H30* strains and not that found in the other *H22* strains. Of note, these two strains possessed Phi3 and the *par-C1a* variant, the two evolutionary markers present in the *H22* “intermediate” strain group B0/C0 described by Ben Zakour et al. [7]. The intermediate status was based on the SNP pattern of *H22* strains which progressively acquired clade C-defining point mutations.

The last comparison made between the S250, JJ1886 and SE15 genomes concerned the genes encoding the 10 methyltransferases previously identified in the genome of *H30-Rx* ST131 strain EC958 [31]. We found nine of these genes in strain JJ1886, six in strain SE15, and only three in strain S250. Forde et al. showed that six of the 10 genes detected in strain EC958 are components of defined mobile genetic elements (MGE) that we showed to be absent from the genome of *H22* strain S250. Indeed, we showed that this strain belongs to subclade B1, meaning to the ancestor of subclone *H22* from which evolved the four other B subclades and subclone *H30* with its subclades C1 and C2. Subclone *H30* arose by acquiring, among others, MGEs near which the methyltransferase-encoding genes are located in *H30-Rx* strain EC958. The three genes identified in strain S250 encode orphan methyltransferases known to be involved in regulation and not methyltransferases involved in restriction modification systems known to inhibit the uptake of non-self DNA and restrict horizontal gene transfer.

We also showed that strain S250 had a *gyrA1* allele. This allele was not found in the subclade B1 strains studied by Ben Zakour et al., whereas it was found in some of their strains belonging to subclades B4 and B5 [7]. Concerning our two other fecal *H22* strains that harbored the *gyrA1a* allele, we showed that one belonged to subclade B2 and the other to subclade B4. Plasmid content also varied within our *H22* strains: no IncF plasmid in one strain and IncF plasmids with different replicons, F24:B6 and F89:B62, in the two remaining *H22* strains, with allele B62 identified here (strain S250) for the first time. Johnson et al. also found plasmids with FII and FIB replicons in most of their *H22* strains. However, the two FII and FIB allelic patterns identified by Johnson et al. were different from ours [6]. In contrast, our *H30-R* strains were all of the same subclade with the same plasmid content: subclade C1 and the IncF plasmid with the replicon F1:A2:B20-. This replicon was previously identified in France in CTX-

M-27-producing, fluoroquinolone-resistant *H30* strains isolated from stool of children attending different day-care centers in 2012 [14]. It was also very recently shown by Johnson et al. to be the dominant IncF plasmid replicon type among the fluoroquinolone-resistant *H30* strains (called *H30R1*) from different sources whereas replicon F2:A1 was shown to be dominant in those *H30* strains producing CTX-M-15 (*H30-Rx*) [6].

Conclusions

This study shows, for the first time, that the O25b:H4 ST131 *E. coli* isolates that were carried in the intestinal microbiota of healthy subjects living in the Paris area in 2006 consisted of strains belonging to both the ancient dominant subclone *H22* and the current dominant subclone *H30*. The genotypic and phenotypic differences identified in this study between these two subclones affect their ability to adapt to the environment. The fecal *H22* strains displayed traits increasing their ability to colonize human gut, whereas the fecal *H30-R* strains displayed traits allowing extra-intestinal adaptation. Thus, resistance to antibiotics may not be the sole advantage of subclone *H30-R* over its precursor subclone *H22* in causing worldwide infections.

Additional files

Additional file 1: Table S1. List of primers designed and used in this study (DOCX 24 kb)

Additional file 2: Table S2. *Escherichia coli* virulence factor encoding genes tested by the PCR method. (DOCX 25 kb)

Additional file 3: Table S3. Distribution and identity percentage of the genes previously shown to be involved in biofilm formation [28] in representatives of different ST131 *H* subclones. (DOCX 41 kb)

Additional file 4: Table S4. *H22* and *H30* strains with a whole genome sequence available in NCBI databases either under a circular DNA form or a fragmented DNA form. (DOCX 16 kb)

Abbreviations

BFI: Biofilm index; BLAST: Basic Local Alignment Search Tool; bp: Base pair; ESBL: Extended-spectrum β -lactamase; ExPEC: Extra-intestinal pathogenic *E. coli*; M₁: Mono-mannose; M₃: Tri-mannose; MGE: Mobile genetic elements; NCBI: National Center for Biotechnology Information; nt: nucleotide; QRDR: Quinolone-resistance determining region; SNP: Single-nucleotide polymorphism; ST: Sequence type; VF: Virulence factor; VFDB: Virulence factor database; WGS: Whole genome shotgun

Acknowledgments

We thank Bill Hempel for his contribution in the writing of the MS in English.

Funding

This work was partially supported by grants AGL2013-47852-R (Ministerio de Economía y Competitividad, Gobierno de España), CN2012/303 and EM2014/001 (Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia and The European Regional Development Fund, ERDF), and RCT-2012-02 (Direction Générale de l'Offre de Soins et Institut National de la Santé et de la Recherche Médicale, Paris, France).

Availability of data and materials

The whole genome shotgun (WGS) project of *H22* strain S250 is available in the [DDBJ/EMBL/GenBank] repositories under the accession LXQL00000000

[https://www.ncbi.nlm.nih.gov/assembly/GCF_001648715.1/]. Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline. The version described in this paper is version LXQL01000000.

Authors' contributions

MHNC conceived the study, supervised the research, and drafted the manuscript. MP performed the genomic analyses. AM, NM, JPL, OB, and VLB performed the molecular and bacterial experiments. JB participated in the design of the molecular experiments and analyzed the data. DH supervised and participated in the genomic analyses and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Service de Microbiologie, Hôpital Beaujon, AP-HP, Clichy, France. ²Faculté de Médecine Paris Diderot, Paris, France. ³INSERM UMR 1137, Université Paris 7, Paris, France. ⁴Laboratoire d'Hygiène Hospitalière, CHRU Besançon, Besançon, France. ⁵UMR 6249 Chrono-Environnement, Besançon, France. ⁶Laboratorio de Referencia de E. coli (LREC), Facultad de Veterinaria, Universidad de Santiago de Compostela, Lugo, Spain. ⁷Institut National de la Santé et de la Recherche Médicale, U1047, Université de Montpellier, UFR de Médecine, Nîmes, France. ⁸Service de Microbiologie, CHU Carémieu, Nîmes, France. ⁹Laboratoire d'Analyse Médicale du Centre, Coulommiers, France.

Received: 19 November 2016 Accepted: 14 March 2017

Published online: 27 March 2017

References

- Nicolas-Chanoine MH, Bertrand X, Madec JY. *Escherichia coli* ST131, an intriguing clonal group. *Clin Microbiol Rev.* 2014;27:543–74.
- Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chattopadhyay S, et al. The epidemic of extended-spectrum- β -lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio.* 2013;4:1–11.
- Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, Riddell K, Rogers P, Qin X, et al. Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *J Infect Dis.* 2013;207:919–28.
- Olesen B, Frimodt-Moller J, Leihof RF, Struve C, Johnston B, Hansen DS, Scheutz F, Krogfelt KA, Kuskowski MA, et al. Temporal trends in antimicrobial resistance and virulence-associated traits within the *Escherichia coli* sequence type 131 clonal group and its H30 and H30-Rx subclones, 1968 to 2012. *Antimicrob Agents Chemother.* 2014;58:6886–95.
- Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, et al. Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *MBio.* 2016;7:e02162.
- Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, et al. Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. *mSphere.* 2016;1:1–15.
- Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, Schembri MA, Beatson SA. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *MBio.* 2016;7:e00347–16.
- Can F, Kurt-Azap O, Ispir P, Nurtop E, Seref C, Loclar I, Aktas ON, Orhan YC, Ergonul O. The clinical impact of ST131 H30-Rx subclone in urinary tract infections due to multidrug-resistant *Escherichia coli*. *J Glob Antimicrob Resist.* 2016;4:49–52.
- Rodrigues C, Machado E, Fernandes S, Peixe L, Novais A. An update on faecal carriage of ESBL-producing *Enterobacteriaceae* by Portuguese healthy humans: detection of the H30 subclone of B2-ST131 *Escherichia coli* producing CTX-M-27. *J Antimicrob Chemother.* 2016;71:1120–2.
- Ranjan A, Shaik S, Hussain A, Nandanwar N, Semmler T, Jadhav S, Wieler LH, Ahmed N. Genomic and functional portrait of a highly virulent, CTX-M-15-producing H30-Rx subclone of *Escherichia coli* sequence type 131. *Antimicrob Agents Chemother.* 2015;59:6087–95.
- Johnson JR, Johnston B, Kuskowski MA, Sokurenko EV, Tchesnokova V. Intensity and mechanisms of fluoroquinolone resistance within the H30 and H30Rx subclones of *Escherichia coli* sequence type 131 compared with other fluoroquinolone-resistant *E. coli*. *Antimicrob Agents Chemother.* 2015;59:4471–80.
- Dahbi G, Mora A, Mamani R, Lopez C, Alonso MP, Marzoa J, Blanco M, Herrera A, Viso S, et al. Molecular epidemiology and virulence of *Escherichia coli* O16:H5-ST131: comparison with H30 and H30-Rx subclones of O25b:H4-ST131. *Int J Med Microbiol.* 2014;304:1247–57.
- Rogers BA, Ingram PR, Runnegar N, Pitman MC, Freeman JT, Athan E, Havers S, Sidjabat HE, Gunning E, et al. Sequence type 131 *fimH30* and *fimH4* subclones amongst *Escherichia coli* isolates in Australia and New Zealand. *Int J Antimicrob Agents.* 2015;45:351–8.
- Blanc V, Leflon-Guibout V, Blanco J, Haenni M, Madec JY, Rafignon G, Bruno P, Mora A, Lopez C, et al. Prevalence of day-care centre children (France) with faecal CTX-M-producing *Escherichia coli* comprising O25b:H4 and O16:H5 ST131 strains. *J Antimicrob Chemother.* 2014;69:1231–7.
- Zhong YM, Liu WE, Liang XH, Li YM, Jian ZJ, Hawkey PM. Emergence and spread of O16-ST131 and O25b-ST131 clones among faecal CTX-M-producing *Escherichia coli* in healthy individuals in Hunan Province. *China J Antimicrob Chemother.* 2015;70:2223–7.
- Gurnee EA, Ndao IM, Johnson JR, Johnston BD, Gonzalez MD, Burnham CA, Hall-Moore CM, McGhee JE, Mellmann A, et al. Gut colonization of healthy children and their mothers with pathogenic ciprofloxacin-resistant *Escherichia coli*. *J Infect Dis.* 2015;212:1862–8.
- Mohamed M, Clabots C, Porter SB, Thurax P, Johnson JR. Isolation and characterization of *Escherichia coli* sequence type 131 (ST131) and other antimicrobial-resistant gram-negative bacilli from clinical stool samples from veterans. *Antimicrob Agents Chemother.* 2016;60:4638–45.
- Leflon-Guibout V, Blanco J, Amaqdouf K, Mora A, Guize L, Nicolas-Chanoine MH. Absence of CTX-M enzymes but high prevalence of clones, including clone ST131, among fecal *Escherichia coli* isolates from healthy subjects living in the area of Paris. *France J Clin Microbiol.* 2008;46:3900–5.
- Lavigne JP, Vergunst AC, Goret L, Sotto A, Combescurre C, Blanco J, O'Callaghan D, Nicolas-Chanoine MH. Virulence potential and genomic mapping of the worldwide clone *Escherichia coli* ST131. *PLoS One.* 2012;7:e34294.
- Weissman SJ, Johnson JR, Tchesnokova V, Billig M, Dykhuizen D, Riddell K, Rogers P, Qin X, Butler-Wu S, et al. High-resolution two-locus clonal typing of extraintestinal pathogenic *Escherichia coli*. *Appl Environ Microbiol.* 2012;78:1353–60.
- Leflon-Guibout V, Jurand C, Bonacorsi S, Espinasse F, Guelfi MC, Duportail F, Heym B, Bingen E, Nicolas-Chanoine MH. Emergence and spread of three clonally related virulent isolates of CTX-M-15-producing *Escherichia coli* with variable resistance to aminoglycosides and tetracycline in a French geriatric hospital. *Antimicrob Agents Chemother.* 2004;48:3736–42.
- Guillard T, de Champs C, Moret H, Bertrand X, Scheftel JM, Cambau E. High-resolution melting analysis for rapid characterization of *qnr* alleles in clinical isolates and detection of two novel alleles, *qnrB25* and *qnrB42*. *J Antimicrob Chemother.* 2012;67:2635–9.
- Paltansing S, Kraakman ME, Ras JM, Wessels E, Bernards AT. Characterization of fluoroquinolone and cephalosporin resistance mechanisms in *Enterobacteriaceae* isolated in a Dutch teaching hospital reveals the presence of an *Escherichia coli* ST131 clone with a specific mutation in *parE*. *J Antimicrob Chemother.* 2013;68:40–5.
- Criscuolo A, Brisse S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics.* 2013;102:500–6.
- Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the mauve aligner. *Bioinformatics.* 2009;25:2071–3.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 2010;11:119.
- Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics.* 2001;17:282–3.

28. Niba ET, Naka Y, Nagase M, Mori H, Kitakawa M. A genome-wide approach to identify the genes involved in biofilm formation in *E. coli*. *DNA Res.* 2007;14:237–46.
29. Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T. Nutritional basis for colonization resistance by human commensal *Escherichia coli* strains HS and Nissle 1917 against *E. coli* O157:H7 in the mouse intestine. *PLoS One.* 2013;8:e53957.
30. Gomez KM, Rodriguez A, Rodriguez Y, Ramirez AH, Isturiz T. The subsidiary GntII system for gluconate metabolism in *Escherichia coli*: alternative induction of the *gntV* gene. *Biol Res.* 2011;44:269–75.
31. Forde BM, Phan MD, Gawthorne JA, Ashcroft MM, Stanton-Cook M, Sarkar S, Peters KM, Chan KG, Chong TM, et al. Lineage-specific methyltransferases define the methylome of the globally disseminated *Escherichia coli* ST131 clone. *MBio.* 2015;6:e01602–15.
32. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother.* 2014;58:3895–903.
33. Johnson JR, Stell AL. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis.* 2000;181:261–72.
34. Blanco J, Mora A, Mamani R, Lopez C, Blanco M, Dahbi G, Herrera A, Marzoa J, Fernandez V, et al. Four main virotypes among extended-spectrum- β -lactamase-producing isolates of *Escherichia coli* O25b:H4-B2-ST131: bacterial, epidemiological, and clinical characteristics. *J Clin Microbiol.* 2013;51:3358–67.
35. Blanco M, Alonso MP, Nicolas-Chanoine MH, Dahbi G, Mora A, Blanco JE, Lopez C, Cortes P, Llagostera M, et al. Molecular epidemiology of *Escherichia coli* producing extended-spectrum β -lactamases in Lugo (Spain): dissemination of clone O25b:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother.* 2009;63:1135–41.
36. Sokurenko EV, Courtney HS, Maslow J, Siitonen A, Hasty DL. Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in *fimH* genes. *J Bacteriol.* 1995;177:3680–6.
37. Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, Struve C, Schembri MA, Hasty DL. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A.* 1998;95:8922–6.
38. Leatham MP, Stevenson SJ, Gauger EJ, Krogfelt KA, Lins JJ, Haddock TL, Autieri SM, Conway T, Cohen PS. Mouse intestine selects nonmotile *fliHDC* mutants of *Escherichia coli* MG1655 with increased colonizing ability and better utilization of carbon sources. *Infect Immun.* 2005;73:8039–49.
39. D'Argenio DA, Wu M, Hoffman LR, Kulasekara HD, Deziel E, Smith EE, Nguyen H, Ernst RK, Larson Freeman TJ, et al. Growth phenotypes of *Pseudomonas aeruginosa lasR* mutants adapted to the airways of cystic fibrosis patients. *Mol Microbiol.* 2007;64:512–33.
40. Chavant P, Gaillard-Martinie B, Talon R, Hebraud M, Bernardi T. A new device for rapid evaluation of biofilm formation potential by bacteria. *J Microbiol Methods.* 2007;68:605–12.
41. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, et al. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One.* 2011;6:e26578.
42. Pantel A, Dunyach-Remy C, Ngba Essebe C, Mesureur J, Sotto A, Pages JM, Nicolas-Chanoine MH, Lavigne JP. Modulation of membrane influx and efflux in *Escherichia coli* sequence type 131 has an impact on bacterial motility, biofilm formation, and virulence in a *Caenorhabditis elegans* model. *Antimicrob Agents Chemother.* 2016;60:2901–11.
43. Owens RCJ, Johnson JR, Stogsdill P, Yarmus L, Lolans K, Quinn J. Community transmission in the United States of a CTX-M-15-producing sequence type ST131 *Escherichia coli* strain resulting in death. *J Clin Microbiol.* 2011;49:3406–8.
44. Hussain A, Ranjan A, Nandanwar N, Babbar A, Jadhav S, Ahmed N. Genotypic and phenotypic profiles of *Escherichia coli* isolates belonging to clinical sequence type 131 (ST131), clinical non-ST131, and fecal non-ST131 lineages from India. *Antimicrob Agents Chemother.* 2014;58:7240–9.
45. Morgand M, Vimont S, Bleibtreu A, Boyd A, Thien HV, Zahar JR, Denamur E, Arlet G. Extended-spectrum β -lactamase-producing *Escherichia coli* infections in children: are community-acquired strains different from nosocomial strains? *Int J Med Microbiol.* 2014;304:970–6.
46. Sarkar S, Vagenas D, Schembri MA, Totsika M. Biofilm formation by multidrug resistant *Escherichia coli* ST131 is dependent on type 1 fimbriae and assay conditions. *Pathog Dis.* 2016;74(3):ftw013
47. Peekhaus N, Conway T. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J Bacteriol.* 1998;180:3495–502.
48. Chen SL, Hung CS, Pinkner JS, Walker JN, Cusumano CK, Li Z, Bouckaert J, Gordon JI, Hultgren SJ. Positive selection identifies an *in vivo* role for FimH during urinary tract infection in addition to mannose binding. *Proc Natl Acad Sci U S A.* 2009;106:22439–44.
49. Conway T, Cohen PS. Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiol Spectr.* 2015;3:1–15.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

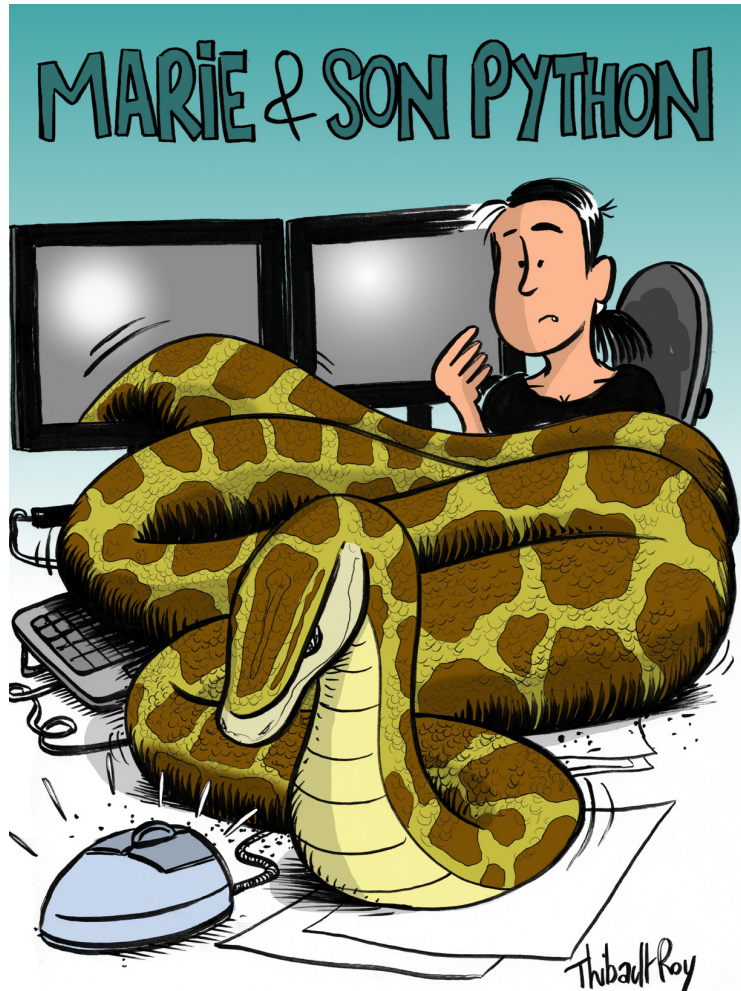


Annexe E

Bande dessinée de vulgarisation scientifique

Réalisée par le Service Science, arts et culture de l'Université de Franche-Comté et la Mission Culture Scientifique de l'Université de Bourgogne.

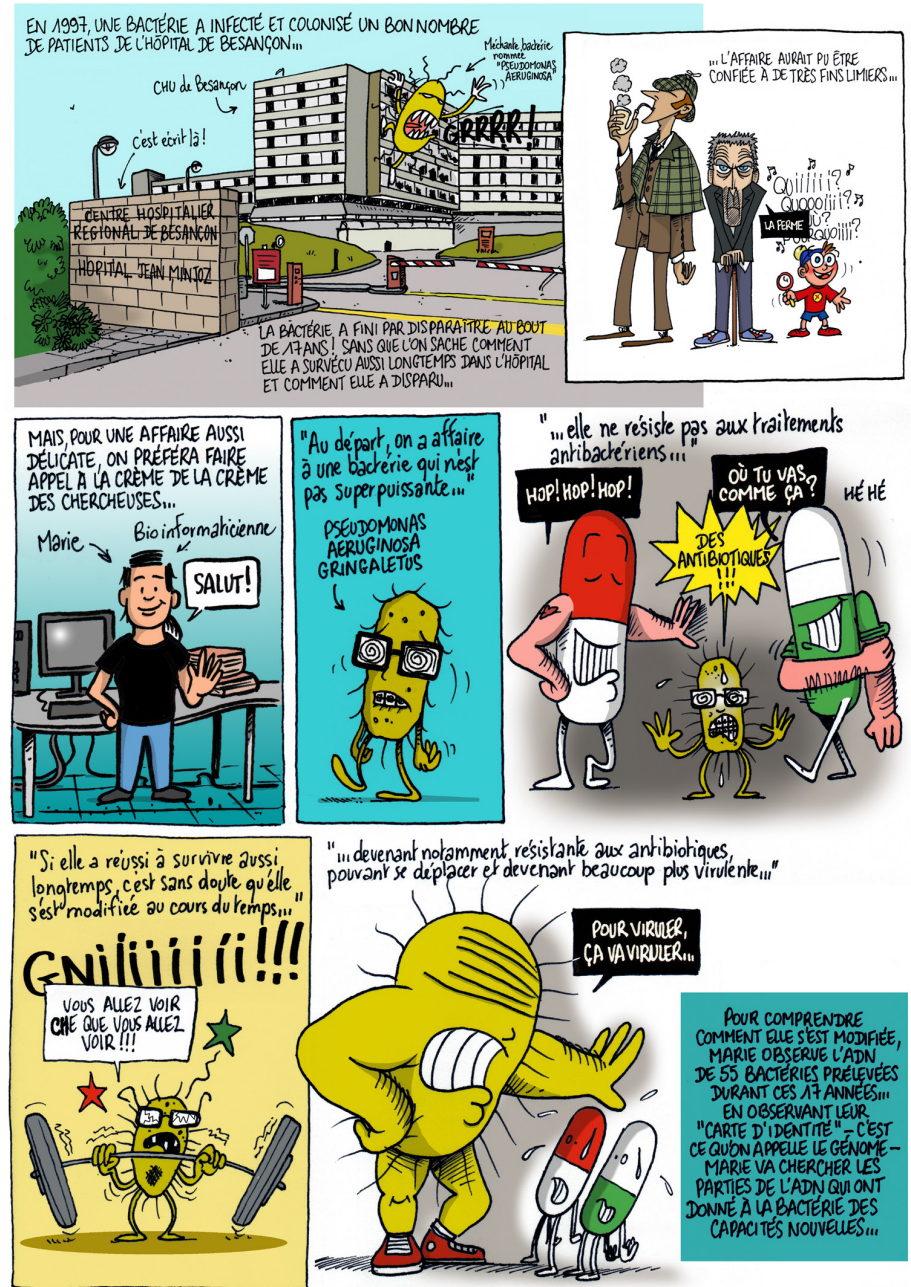
Et si la SCIENCE



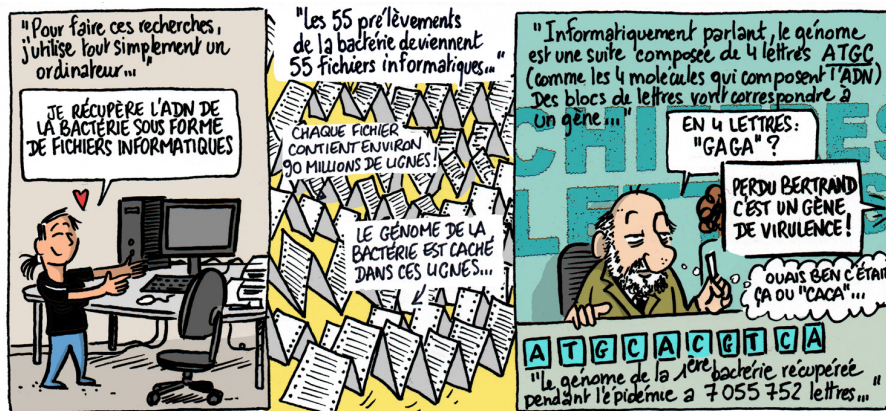
MARIE & SON PYTHON

était un OBJET

OU LA RECHERCHE EN DESSINS



Marie PETITJEAN est jeune chercheuse en bio-informatique au laboratoire d'Hygiène Hospitalière du Centre Hospitalier Régional Universitaire de Besançon.



"Je vais reconstituer les génomes pour voir les modifications au cours du temps et je vais chercher des gènes spécifiques..."

"Je vais utiliser des bases de données présentes sur le net pour chercher les gènes en question..."

"Pour cela, j'utilise un des nombreux langages informatiques. LE PYTHON!"

COMME LES GÈNES DE RÉSISTANCE AUX ANTIBIOTIQUES, PAR EXEMPLE!

CERTAINES RECHERCHES SE FONT AUTOMATIQUEMENT MAIS POUR D'AUTRES, JE DOIS ÉCRIRE UN MORCEAU DE CODE...

JE LUI DEMANDE DE FAIRE UNE TÂCHE PRÉCISE COMME LIRE LETTRE PAR LETTRE UN GÉNOME...



"L'écriture de code en Python n'est pas compliquée en soi... C'est généralement au lancement que ça plante. Il faut être précis! Et prévoir une partie "débugage"..."

"Ensuite, je vais faire des tests en laboratoire, pour vérifier ce que j'ai trouvé, grâce à mon ordinateur..."

LIS LES GÉNOMES, COMPARE LES, DÈS QUE TU TROUVES DES SÉQUENCES QUI SE RESSEMBLENT, DIS MOI OÙ ELLES SE TROUVENT DANS LA BASE DE DONNÉES. ENSUITE NOTE TOUT ÇA DANS UN FICHIER QUE TU APPELLERAS "RÉSULTATS" @

"Ce qui est bien c'est que je peux lancer de gros calculs le soir avant de partir, je sais que le lendemain j'aurai mes résultats affichés..."



Ⓜ En Python dans le texte



Battle, Daphné, Erik, Guillaume, Maïmouna, Marie, Pierre et Sindy sont de jeunes chercheurs du Réseau des Experimentarium. Leurs questionnements, leurs tâtonnements, les instruments qu'ils apprennent à manipuler constituent les coulisses de la recherche.

Thibault Roy, Jo-Ann Campion et Marie-Laure Baudement ont recueilli leurs témoignages qui vous font découvrir le quotidien de leur travail. En résultent ces dessins de Thibault Roy, fruits de discussions passionnantes et d'heures passées devant la table à dessin.

Ce projet s'inscrit dans le cadre du Programme d'Investissement d'Avenir du Réseau des Experimentarium, qui organise des rencontres avec des jeunes chercheurs. Il a été réalisé par la Mission Culture Scientifique de l'université de Bourgogne en collaboration avec le Service Sciences, arts, culture de l'université de Franche-Comté.

Retrouvez les autres planches sur le site du Réseau des Experimentarium, qui favorise les rencontres avec des jeunes chercheurs : www.experimentarium.fr

CONTACTS

Service Sciences, arts, culture de l'université de Franche-Comté
1 rue Goudimel
25030 Besançon cedex
jeremy.querenet@univ-fcomte.fr

Mission Culture Scientifique université de Bourgogne
Sciences Gabriel
6, boulevard Gabriel - 21 000 Dijon
03 80 39 90 99
culture-scientifique@u-bourgogne.fr

Dessin : © Thibault Roy
Coordination de la publication : Marie-Laure Baudement et Jo-Ann Campion
Conception graphique et mise en page : Jérôme Berthier et Mathilde Bosson

Tous droits réservés, reproduction interdite

Titre : Évolution génotypique et phénotypique d'une souche épidémique de *Pseudomonas aeruginosa* au cours des 11 ans de sa diffusion hospitalière.

Mots clés : *P. aeruginosa*, bactérie, bioinformatique, épidémie, gram négatif

Résumé : *P. aeruginosa* est une bactérie pathogène de l'homme, responsable d'infections nosocomiales chez les patients immunodéprimés. Bien que son évolution au sein d'un patient soit bien décrite, son évolution génomique globale au cours de sa propagation dans un hôpital est très mal connue. Le clone à haut-risque ST395 multirésistant aux antibiotiques a diffusé dans le Centre Hospitalier Régional Universitaire de Besançon entre 1997 et 2008 en infectant ou colonisant plus de 300 patients. Une approche WGS a été utilisée afin d'identifier l'origine de l'épidémie, les caractéristiques ayant pu contribuer à son installation à l'hôpital ainsi que les modifications accompagnant la fin de l'épidémie. Les génomes de 54 isolats représentatifs de l'épidémie ont été séquencés. L'arbre phylogénétique a mis en évidence deux clusters distincts indiquant la présence de deux épidémies parallèles. La datation d'un ancêtre commun en 1979, date de début de la construction de l'hôpital, indiquerait une contamination précoce du réseau d'eau de l'hôpital. Cette hypothèse est soutenue par la présence d'un îlot génomique spécifique de ST395 portant les gènes codant 6 transporteurs du cuivre et associée à une résistance phénotypique à ce métal constituant les tuyaux du réseaux de distribution d'eau potable. Les isolats tardifs présentaient des signatures génomiques d'adaptation à l'infection chronique (altération du lipopolysaccharide et de la porine OprD – objectivées phénotypiquement, et extinction de la surproduction de la pompe d'efflux MexAB-OprM – contrôlée par RT-qPCR) suite à des mutations indépendantes. Certaines de ces mutations ont été associées à une perte de *fitness* bactérien *in vitro*. Nous émettons l'hypothèse que l'émergence indépendante d'isolats adaptés à l'infection chronique, et ainsi l'accumulation de culs-de-sac épidémiologiques, a participé à l'épuisement de l'épidémie hospitalière de *P. aeruginosa* ST395.

Title : Genotypic and phenotypic evolution of a *Pseudomonas aeruginosa* ST395 strain during 11-year in hospital spread.

Keywords : *P. aeruginosa*, bacteria, bioinformatic, outbreak, gram negative

Abstract : *P. aeruginosa* is an opportunistic pathogen responsible of hospital-acquired infections in immunocompromised patients. Although in-host evolution of *P. aeruginosa* is well documented, little is known about this pathogen evolution during its spread on a hospital scale. The high-risk multidrug resistant clone ST395 spread among more than 300 patients in the University Hospital of Besançon between 1997 and 2008. We used a WGS approach to identify the origin of the outbreak, the features that could have helped its implantation in our hospital and those associated with the end of the epidemics. The genomes of 54 representative isolates were fully sequenced. The phylogenetic tree indicated two distinct clusters corresponding to two parallel outbreaks. The ancestor of the ST395 clone possibly contaminated our hospital water network during its construction in 1979. This hypothesis is supported by the fact that the ST395 strain had a specific genomic island carrying 6 copper transporter genes implicated in copper resistance, correlated with the resistance to this metal which water supply network is made of. The late isolates displayed independent genomic signatures of chronic adaptation in patients (altered LPS and porin OprD, and extinction of MexAB-oprM efflux pump overproduction). Some of these mutations were associated with a decreased *in vitro* fitness. We hypothesize that the independent emergence of isolates adapted to chronic infection, and thus the accumulation of epidemiological dead-ends, participated to the end of the hospital outbreak of *P. aeruginosa* ST395.