

## py Origine et évolution des bêta lactamases de classe D

**Auteur :** Lupo, Valérian

**Promoteur(s) :** Baurain, Denis

**Faculté :** Faculté des Sciences

**Diplôme :** Master en biochimie et biologie moléculaire et cellulaire, à finalité spécialisée en bioinformatique et modélisation

**Année académique :** 2017-2018

**URI/URL :** <http://hdl.handle.net/2268.2/5201>

---

### Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

---



---

# Origine et évolution des bêta-lactamases de classe D

---

Université de Liège

Faculté des Sciences  
Département des Sciences de la Vie  
Unité de Phylogénomique des Eucaryotes

*Promoteur* Denis Baurain

*Mémoire présenté par*

VALÉRIAN LUPO

*En vue de l'obtention du grade de*

Master en Biochimie et Biologie Moléculaire et Cellulaire, à finalité spécialisée en  
Bioinformatique et Modélisation

*Année académique* 2017-2018

# Remerciements

Tout d'abord, je souhaiterais remercier le Prof. Denis Baurain pour son encadrement tout au long de ce mémoire, de m'avoir appris toutes ces techniques de bioinformatique et donné goût à la programmation.

Je remercie tout particulièrement le Prof. Jean-Marie Frère pour avoir proposé ce sujet et également pour tous ses conseils.

Je tiens à remercier toutes les autres personnes sans qui la réalisation de ce mémoire n'aurait pas été possible.

Mes remerciements vont aux collègues de l'unité de Phylogénomique des Eucaryotes.

Enfin, j'aimerais remercier tous mes proches pour leur soutien tout au long de mon parcours universitaire.

## Résumé

La majorité des bactéries possèdent une paroi à l'extérieur de leur membrane cytoplasmique. Celle-ci leur confère une forme spécifique en résistant à la pression osmotique interne, tout en offrant une protection contre l'environnement extérieur. Le peptidoglycane (PG) est un des composants majeurs de la paroi bactérienne et la cible de nombreux antibiotiques à noyau bêta-lactame. En inhibant la biosynthèse du PG, les  $\beta$ -lactamines entraînent la lyse des cellules bactériennes. Pour se défendre, les bactéries ont développé différentes stratégies, dont la production d'enzymes spécifiques ( $\beta$ -lactamases) pour détruire les antibiotiques. Il existe quatre classes de  $\beta$ -lactamases : les classes A, C et D (ou OXA) regroupent les enzymes à sérine active, tandis que la classe B comprend les métallo- $\beta$ -lactamases.

A partir de séquences de  $\beta$ -lactamases de référence récoltées dans la base de données du NCBI, nous avons construit différents profils HMM afin d'identifier des séquences de  $\beta$ -lactamases putatives dans  $> 80\,000$  génomes complets de procaryotes. Ces recherches par homologie nous ont permis de récupérer des séquences putatives d'OXA dans des organismes où aucune OXA n'avait encore été décrite, tels que les *Cyanobacteria*, *Chlorobi* ou encore *Actinobacteria*. Les séquences d'OXA identifiées ont été alignées et soumises à des analyses phylogénétiques. Les arbres en résultant nous ont guidé dans la définition de 31 groupes. Nous avons observé que 30 groupes contenaient des OXAs putatives et un groupe des BlaR1 (un récepteur membranaire apparenté aux OXAs). Nous avons noté que 15 des 30 groupes étaient annotés par des séquences de référence. De cet arbre, nous avons conclu que la plupart des OXAs correspondent probablement au même gène orthologue du point de vue évolutif.

Enfin, une analyse des résidus du site actif a montré que les motifs SxxK, SxV, W et KTG (où x représente n'importe quel AA) étaient retrouvés dans la majorité des séquences de chaque groupe, ce qui témoigne d'une activité  $\beta$ -lactamase potentielle. De plus, le motif Y/FxN est très conservé à travers tous les groupes mais, situé en dehors de la poche du site actif, sa fonction dans l'enzyme est encore inconnue.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Antibiotiques . . . . .	1
1.2	Bêta-lactamines . . . . .	2
1.2.1	Pénicillines . . . . .	2
1.2.2	Céphalosporines . . . . .	3
1.2.3	Monobactames . . . . .	4
1.2.4	Carbapénèmes . . . . .	4
1.3	Paroi bactérienne . . . . .	4
1.4	Penicillin-binding proteins . . . . .	5
1.5	Mode d'action des $\beta$ -lactamines . . . . .	6
1.6	Résistance aux $\beta$ -lactamines . . . . .	6
1.7	Bêta-lactamases . . . . .	6
1.7.1	Classe A . . . . .	7
1.7.2	Classe B . . . . .	8
1.7.3	Classe C . . . . .	8
1.7.4	Classe D . . . . .	9
1.7.5	Distribution taxonomique des $\beta$ -lactamases . . . . .	10
<b>2</b>	<b>Objectifs</b>	<b>12</b>
<b>3</b>	<b>Matériels &amp; méthodes</b>	<b>13</b>
3.1	Environnement . . . . .	13
3.1.1	Hardware . . . . .	13
3.1.2	UNIX . . . . .	13
3.1.3	Perl . . . . .	13
3.2	Données et pré-traitement des données . . . . .	13
3.2.1	$\beta$ -lactamases de référence . . . . .	13
3.2.2	Génomes complets pour le data mining . . . . .	14
3.2.3	Sélection des $\beta$ -lactamases de référence . . . . .	14
3.2.4	Déréplication des $\beta$ -lactamases de référence . . . . .	14
3.3	Alignement multiple . . . . .	15

3.3.1	Alignement automatique . . . . .	15
3.3.2	Optimisation manuelle . . . . .	15
3.4	Recherche de $\beta$ -lactamases putatives . . . . .	15
3.4.1	Package HMMER . . . . .	15
3.4.2	Sélection des séquences putatives . . . . .	17
3.4.3	Annotation des OXAs putatives . . . . .	17
3.5	Inférences phylogénétiques . . . . .	17
3.6	Analyses complémentaires . . . . .	18
3.6.1	Distributions taxonomiques . . . . .	18
3.6.2	Conservation des résidus du site actif . . . . .	18
<b>4</b>	<b>Résultats</b>	<b>19</b>
4.1	Distribution taxonomique des génomes complets . . . . .	19
4.2	Recherche de $\beta$ -lactamases putatives . . . . .	19
4.3	Distribution taxonomique des OXAs putatives . . . . .	20
4.4	Annotation, alignement et phylogénie des OXAs putatives . . . . .	23
4.5	Annotation détaillée des OXAs putatives . . . . .	24
4.6	Phylogénie des OXAs putatives . . . . .	25
4.7	Conservation des résidus du site actif . . . . .	28
<b>5</b>	<b>Discussion</b>	<b>31</b>
5.1	Recherche de $\beta$ -lactamases putatives . . . . .	31
5.2	Distribution taxonomique des OXAs putatives . . . . .	31
5.3	Déréplication des OXAs putatives et de référence . . . . .	32
5.4	Phylogénie des OXAs putatives . . . . .	32
5.5	Classification des OXAs putatives . . . . .	33
5.6	Conservation des résidus du site actif . . . . .	33
5.7	Comparaison avec l'article de Keshri <i>et al.</i> (2018) . . . . .	33
<b>6</b>	<b>Conclusions</b>	<b>35</b>
<b>7</b>	<b>Annexes</b>	<b>I- 1</b>
7.1	Conservation des résidus du site actif . . . . .	I- 1
7.2	Scripts Perl . . . . .	II- 1



## Table des figures

1	Noyau $\beta$ -lactame . . . . .	2
2	Noyau 6-APA . . . . .	3
3	Différentes pénicillines . . . . .	3
4	Noyau 7-ACA . . . . .	4
5	Aztréonam . . . . .	4
6	Noyau carbapénème . . . . .	5
7	Structure du PG . . . . .	6
8	Réaction entre les $\beta$ -lactamases/PBPs et les $\beta$ -lactamines . . . . .	7
9	Structure 3D OXA . . . . .	9
10	Résidus conservés . . . . .	11
11	musted . . . . .	16
12	mustnet . . . . .	16
13	Taxonomie base de donnée . . . . .	20
14	Graphiques de ompa-pa.pl . . . . .	21
15	Taxonomie OXAs putatives . . . . .	22
16	Taxonomie OXAs putatives dérépliquées . . . . .	23
17	Arbre phylogénétique des OXAs . . . . .	26
18	Arbre phylogénétique des OXAs avec uniquement les séquences des 31 groupes	27
19	Comparaison des résidus conservés . . . . .	29
S.1	Résidus groupe 1 . . . . .	I- 1
S.2	Résidus groupe 2 . . . . .	I- 1
S.3	Résidus groupe 3 . . . . .	I- 1
S.4	Résidus groupe 4 . . . . .	I- 1
S.5	Résidus groupe 5 . . . . .	I- 1
S.6	Résidus groupe 6 . . . . .	I- 2
S.7	Résidus groupe 7 . . . . .	I- 2
S.8	Résidus groupe 8 . . . . .	I- 2
S.9	Résidus groupe 9 . . . . .	I- 2
S.10	Résidus groupe 10 . . . . .	I- 2
S.11	Résidus groupe 11 . . . . .	I- 2



S.12 Résidus groupe 12 . . . . .	I- 3
S.13 Résidus groupe 13 . . . . .	I- 3
S.14 Résidus groupe 14 . . . . .	I- 3
S.15 Résidus groupe 15 . . . . .	I- 3
S.16 Résidus groupe 16 . . . . .	I- 3
S.17 Résidus groupe 17 . . . . .	I- 3
S.18 Résidus groupe 18 . . . . .	I- 4
S.19 Résidus groupe 19 . . . . .	I- 4
S.20 Résidus groupe 20 . . . . .	I- 4
S.21 Résidus groupe 21 . . . . .	I- 4
S.22 Résidus groupe 22 . . . . .	I- 4
S.23 Résidus groupe 23 . . . . .	I- 4
S.24 Résidus groupe 24 . . . . .	I- 5
S.25 Résidus groupe 25 . . . . .	I- 5
S.26 Résidus groupe 26 . . . . .	I- 5
S.27 Résidus groupe 27 . . . . .	I- 5
S.28 Résidus groupe 28 . . . . .	I- 5
S.29 Résidus groupe 29 . . . . .	I- 5
S.30 Résidus groupe 30 . . . . .	II- 1
S.31 Résidus groupe 31 . . . . .	II- 1

## Abréviations

MRSA	<i>Staphylococcus aureus</i> résistant à la méthicilline
6-APA	Acide 6-aminopénicillanique
7-ACA	Acide 7-aminocéphalosporanique
PG	Peptidoglycane
GlcNAc	N-acétylglucosamine
MurNAc	Acide N-acétylmuramique
AA	Acide aminé
D-Ala	D-alanine
L-Ala	L-alanine
D-Glu	D-glutamate
méso-Dap	acide diaminopimélique
L-Lys	L-lysine
PBP	Penicillin-binding proteins
ESBL	$\beta$ -lactamase à spectre étendu
HMM	Modèles de Markov cachés (Hidden Markov Models)

# 1 Introduction

## 1.1 Antibiotiques

Le sol renferme une multitude de micro-organismes produisant naturellement des antibiotiques. Cette production leur permet d'éliminer d'autres espèces non productrices qui entreraient en compétition lorsque les ressources s'épuisent (Sengupta, Chattopadhyay, and Grossart 2013). Depuis des milliers d'années, l'homme exploite la production d'antibiotiques par ces micro-organismes pour lutter contre les infections. De nombreux récits relatent d'utilisation de moisissures à travers le monde, comme par exemple en Chine, vers l'an 3000 av. J.-C., où du soja moisi était appliqué sur des blessures infectées. Vers 1600 av. J.-C., en Grèce, une paysanne était réputée pour soigner des soldats en utilisant de la moisissure de fromage (Dougherty and Pucci 2014).

Plus récemment, vers 1890, Rudolf Emmerich et Oscar Löw avaient découvert que des bactéries vertes (*Pseudomonas aeruginosa*) isolées de bandages de patients inhibaient la croissance d'autres bactéries. De ces bactéries vertes a été extraite la Pyocyanase, le tout premier antibiotique utilisé pour traiter des infections chez l'homme. En 1897, Ernest Dushesne avait décrit que des levures inhibaient la croissance de bactéries (Gould 2016). Alexander Fleming a découvert la pénicilline en 1928 (Fleming 1929), mais elle n'a été purifiée qu'au début des années 1940 par Ernst Chain et Howard Florey (Gaynes 2017). La pénicilline a été largement utilisée à la fin de la seconde guerre mondiale pour soigner les soldats blessés. Depuis lors, l'utilisation d'antibiotiques a explosé et a permis de sauver de nombreuses vies (Ventola 2015).

De nos jours, de nombreux antibiotiques avec différents modes d'action sont présents sur le marché. Les bêta-lactamines et les glycopeptides permettent de cibler la paroi bactérienne. Les aminoglycosides, les tétracyclines, le chloramphenicol, les macrolides et les oxazolidinones inhibent la synthèse des protéines. Enfin, les sulfonamides inhibent le métabolisme des acides nucléiques (Etebu and Ariekpar 2016 ; Kapoor, Saigal, and Elongavan 2017).

A la suite de l'utilisation intensive des antibiotiques en médecine humaine et animale, de nombreuses souches bactériennes ont développé des résistances rendant leur élimination de plus en plus compliquée. Actuellement, on estime qu'environ 700 000 personnes meurent chaque année d'une infection causée par une bactérie résistante. D'ici 2050, ce chiffre pourrait s'accroître jusqu'à 10 millions (<http://amr-review.org/>, 19 mai 2016).

La plasticité génétique des bactéries leur a permis de mettre en place différents mécanismes de défense contre l'action des antibiotiques. Les deux mécanismes les plus répandus sont la destruction de l'antibiotique et la modification de sa cible (Munita et al. 2016 ; Lambert 2005). Les gènes de résistance peuvent se trouver sur le chromosome bactérien ou sur un plasmide. De par le phénomène de conjugaison, les plasmides contribuent à disséminer les gènes de résistance à travers différentes espèces (Nikaido 2009). Le cas le plus célèbre est le *Staphylococcus aureus* résistant à la

méthicilline (MRSA), bactérie résistante à pratiquement tous les agents anti-microbiens connus à ce jour, à l'exception des glycopeptides (dont la vancomycine) (Haaber, Penadés, and Ingmer 2017).

## 1.2 Bêta-lactamines

Les bêta-lactamines ( $\beta$ -lactamines) sont des antibiotiques bactéricides qui inhibent la biosynthèse de la paroi bactérienne. Ces molécules sont caractérisées par la présence d'un noyau bêta-lactame ( $\beta$ -lactame, figure 1) (Aoki and Okuhara 1980). Ils sont produits naturellement par divers micro-organismes eucaryotes, tels que des champignons filamenteux du genre *Penicillium*, mais également par des bactéries du genre *Streptomyces* (Tahlan and Jensen 2013). Sur base de  $\beta$ -lactamines naturelles, les industries pharmaceutiques créent des antibiotiques semi-synthétiques par modification chimique des chaînes latérales autour du noyau  $\beta$ -lactame (Elander 2003). Les pénicillines, les céphalosporines, les monobactames et les carbapénèmes composent les différentes familles de  $\beta$ -lactamines.

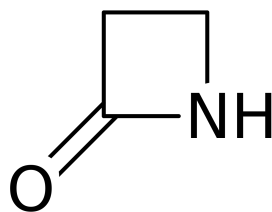


FIG. 1 : Noyau  $\beta$ -lactame.

### 1.2.1 Pénicillines

Les pénicillines regroupent toutes les  $\beta$ -lactamines possédant une structure similaire à la pénicilline G (première pénicilline utilisée à grande échelle), possédant un noyau d'acide 6-aminopénicillanique (6-APA, figure 2) (Ball, Gray, and Murdoch 1978). Il existe quatre générations de pénicillines, qui regroupent les antibiotiques selon leur spectre d'action. L'activité antibactérienne dépend ainsi des groupements présents autour du noyau 6-APA.

Les pénicillines de première génération incluent toutes les pénicillines naturelles (dont la pénicilline G). Elles possèdent une bonne activité contre les bactéries Gram-positives mais sont peu ou pas efficaces contre les bactéries Gram-négatives. En réaction à l'apparition de résistances contre ces pénicillines naturelles chez les bactéries Gram-positives, les industries pharmaceutiques ont développé les pénicillines de seconde génération. Ce sont des pénicillines semi-synthétiques, telles que l'oxacilline ou la méthicilline. Cependant, elles restent peu efficaces

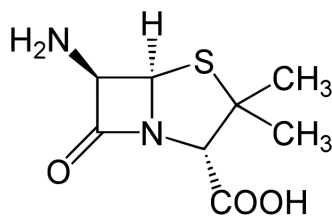


FIG. 2 : Noyau d'acide 6-aminopénicillanique.

contre les bactéries Gram-négatives. Les pénicillines de troisième et quatrième génération comprennent toutes les pénicillines à large spectre, comme l'ampicilline et l'amoxicilline, ou encore les carboxypénicillines, présentant toutes une meilleure activité contre les bactéries Gram-négatives (Lobanovska and Pilla 2017).

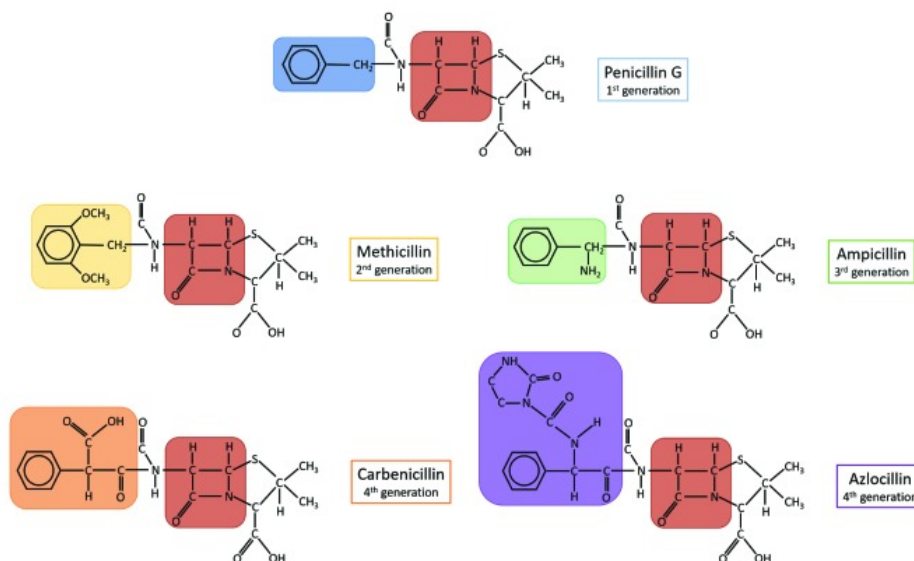


FIG. 3 : Structure de différentes générations de pénicillines. Le noyau  $\beta$ -lactame est encadré en rouge et les différentes chaînes latérales suivent un code couleur. On retrouve en bleu la pénicilline G (première génération), en jaune la méthicilline (deuxième génération), en vert l'ampicilline (troisième génération), en orange la carbénicilline (quatrième génération) et en mauve l'azlocilline (quatrième génération) (Lobanovska and Pilla 2017).

### 1.2.2 Céphalosporines

Les céphalosporines sont des antibiotiques structurellement similaires aux pénicillines. Elles ont la particularité de posséder un noyau d'acide 7-aminocéphalosporanique (7-ACA, figure 5) (Jago and Heatley 1961). En 1953, Newton et Abraham ont décrit la première céphalosporine, la céphalosporine C extraite d'*Acremonium chrysogenum* (Newton and Abraham 1955). Elle montrait un large spectre d'action contre les bactéries Gram-positives et négatives. De plus, elle était résistante aux pénicillinases produite par *Bacillus subtilis* et *S. aureus* (Bryskier 2000).

Tout comme les pénicillines, les céphalosporines sont classées en différentes générations (cinq) selon leur spectre d'action (Dougherty and Pucci 2014).

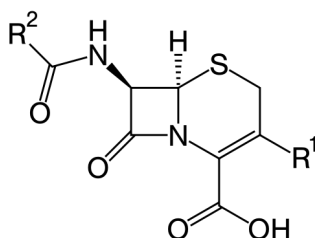


FIG. 4 : Noyau d'acide 7-aminocéphalosporanique.

### 1.2.3 Monobactames

Contrairement aux pénicillines et aux céphalosporines, les monobactames ne contiennent qu'un monocycle  $\beta$ -lactame (Fernandes, Amador, and Prudêncio 2013). L'aztréonam est l'unique monobactame utilisé dans le domaine hospitalier. Il est actif contre les bactéries Gram-négatives mais inefficace contre les bactéries Gram-positives et les bactéries Gram-positives et négatives à métabolisme anaérobique (Ramsey and MacGowan 2016).

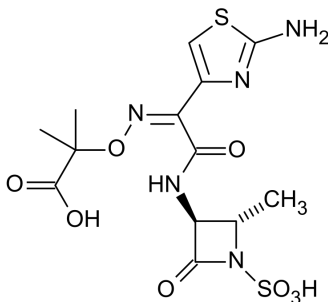


FIG. 5 : Structure de l'aztréonam.

### 1.2.4 Carbapénèmes

Les carbapénèmes ont une structure similaire aux pénicillines, mais possèdent une double liaison entre les carbone C-2 et C-3, tandis que l'atome de soufre en position 1 du cycle est remplacé par un carbone (figure 6). Ce sont les antibiotiques à noyau  $\beta$ -lactame présentant le spectre d'action le plus étendu. C'est pourquoi ils sont largement utilisés contre les bactéries multirésistantes (Papp-Wallace et al. 2011).

## 1.3 Paroi bactérienne

La plupart des bactéries possèdent une paroi à l'extérieur de leur membrane cytoplasmique. Une partie de cette paroi, le peptidoglycane (PG), confère aux bactéries une forme spécifique et

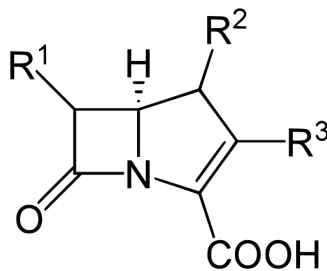


FIG. 6 : Structure du noyau carbapénème.

une résistance contre leur pression osmotique interne. Le PG est un hétéropolymère constitué de chaînes de glycanes reliées entre elles par de courts peptides. Les chaînes de glycanes sont une alternance de disaccharides composés de N-acétylglucosamine (GlcNAc) et d'acide N-acétylmuramique (MurNAc) liés en  $\beta$ -1  $\rightarrow$  4. Sur le groupement lactyle du MurNAc est lié un tétrapeptide constitué d'acides aminés (AA) en configurations L ou D. Le précurseur de synthèse de ce tétrapeptide est un pentapeptide comptant une D-alanine (D-Ala) supplémentaire.

La composition en acides aminés varie d'une espèce à l'autre. Chez les bactéries Gram-négatives, le tétrapeptide est composé de L-alanine (L-Ala), D-glutamate (D-Glu), acide diaminopimélique (méso-Dap) et D-Ala. Chez certaines bactéries Gram-positives, une L-lysine (L-Lys) peut remplacer le méso-Dap. Dans la majorité des cas, la liaison entre chaînes de glycanes se fait entre l'AA en position 4 et l'AA en position 3. Chez les bactéries Gram-négatives, cette liaison est directe, tandis que quelques peptides peuvent servir d'intermédiaires de liaison chez les bactéries Gram-positives (figure 7). La synthèse du PG commence dans le cytoplasme pour aboutir au lipide II membranaire. Ce dernier servira de précurseur du PG dans l'espace périplasmique (Vollmer, Blanot, and De Pedro 2008).

#### 1.4 Penicillin-binding proteins

Les penicillin-binding proteins (PBPs) sont des enzymes-clés de la synthèse du PG. Elles permettent l'élongation et le pontage des chaînes de glycanes. Leur activité transpeptidase assure chez les bactéries Gram-négatives le pontage entre le groupement carboxyle du D-Ala en position 4 et le groupement amine de la chaîne latérale du méso-Dap en position 3 (Vollmer, Blanot, and De Pedro 2008). Lors de cette réaction, le D-Ala en position 5 est éliminé et l'énergie de la liaison D-Ala-D-Ala est utilisée pour former la nouvelle liaison (Frère and Joris 1985).

Les PBPs peuvent être divisées en trois groupes : la classe A présente une activité transglycosylase et une activité transpeptidase, la classe B une activité transpeptidase seulement et la classe C des activités transpeptidase et carboxypeptidase (Goffin and Ghuysen 1998 ; Sauvage et al. 2008). Les bactéries peuvent posséder de nombreuses PBPs, comme par exemple *E.coli* qui en dispose de 12 différentes, dont trois de la classe A et deux de la classe B (Georgopapadakou and Liu 1980 ; Sauvage et al. 2008).

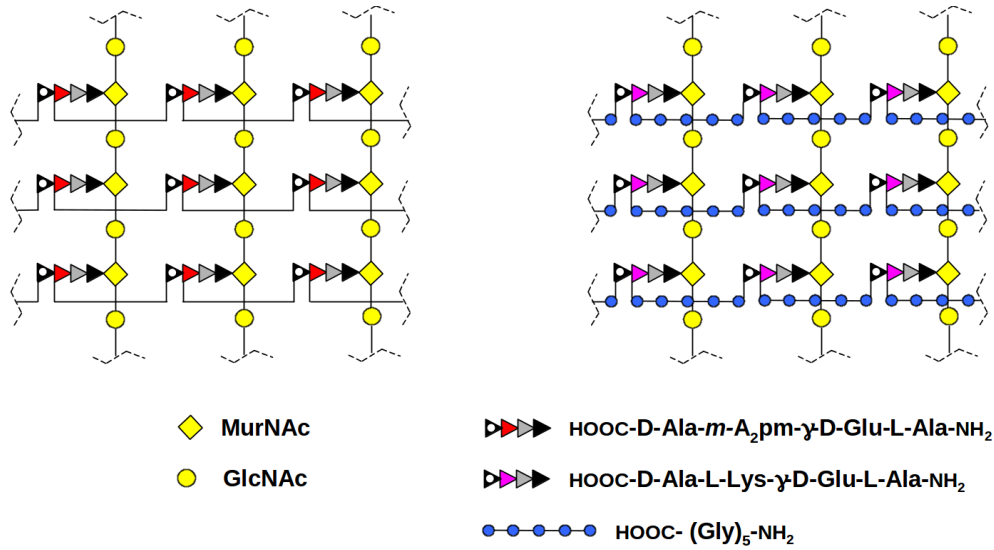


FIG. 7 : Structure du PG chez *Escherichia coli* (à gauche) et *Staphylococcus aureus* (à droite). Les chaînes de glycanes sont représentées en jaune et le tétrapeptide par des triangles suivant un code couleur : L-Ala en noir, D-Glu en gris, méso-Dap en rouge, L-Lys en mauve et D-Ala en noir avec point blanc. Le pontage est direct chez *E. coli*, alors que cinq glycines (ronds bleus) font le lien chez *S. aureus* (figure de B. Joris).

### 1.5 Mode d'action des $\beta$ -lactamines

Les  $\beta$ -lactamines sont structuellement similaires au substrat naturel des PBPs, le D-Ala-D-Ala présent sur le pentapeptide. De ce fait, ces antibiotiques agissent comme substrat suicide en bloquant le site actif de l'enzyme. Lors de l'étape d'acylation, l'antibiotique est lié de façon covalente au résidu sérine du site actif. L'enzyme étant incapable d'hydrolyser cette liaison (figure 8), l'antibiotique inhibe l'activité transpeptidase du PBP (Bush and Bradford 2016), ce qui arrête le pontage entre chaîne de glycanes et provoque la lyse de la bactérie (Kohanski, Dwyer, and Collins 2010).

### 1.6 Résistance aux $\beta$ -lactamines

Pour contrer l'action des  $\beta$ -lactamines, les bactéries ont développé deux stratégies de défense principales. La première est la synthèse de PBPs résistantes, tandis que la seconde est la production d'enzymes capables de neutraliser les antibiotiques (Frère 1995). De plus, les bactéries Gram-négatives réduisent la concentration de l'antibiotique en diminuant sa vitesse de diffusion et en augmentant son efflux au niveau de la membrane externe (Munita et al. 2016).

### 1.7 Bêta-lactamases

Les bêta-lactamases ( $\beta$ -lactamases) sont des enzymes capables d'hydrolyser la liaison amide du noyau  $\beta$ -lactame (figure 8), cette réaction rend l'antibiotique inactif. C'est le mécanisme de résistance le plus courant chez les bactéries Gram-négatives (Poole 2004). Il existe quatre classes



de  $\beta$ -lactamases : classe A, classe B, classe C et classe D. Cette classification est basée sur la séquence en AA (Ambler 1980). Les classes A, C et D regroupent des  $\beta$ -lactamases à sérine active, tandis que la classe B comprend des métallob- $\beta$ -lactamases (Babic, Hujer, and Bonomo 2006). Des similarités de séquences laissent penser que les  $\beta$ -lactamases à sérine active et les PBPs partagent une origine commune (Massova and Mobashery 1998).

Des mutations ponctuelles des séquences de  $\beta$ -lactamases permettent de modifier leur spectre d'action. Ainsi, les  $\beta$ -lactamases à spectre étendu (ESBLs) sont des  $\beta$ -lactamases conférant non seulement une résistance contre toutes les pénicillines, mais également contre les céphalosporines de première, deuxième et troisième générations. Toutefois, certaines ESBLs sont inhibées par l'acide clavulanique. La plupart des ESBLs font partie de la classe A, mais on en retrouve également dans la classe D (Paterson and Bonomo 2005; Rawat and Nair 2010). Les carbapénémases, quant à elles, sont des  $\beta$ -lactamases à très large spectre d'action capables d'hydrolyser en plus (ou uniquement) les carbapénèmes (Queenan and Bush 2007). Elles sont présentes dans les classes A, B et D. A ce jour, plus de 4000  $\beta$ -lactamases ont été répertoriées (Naas et al. 2017).

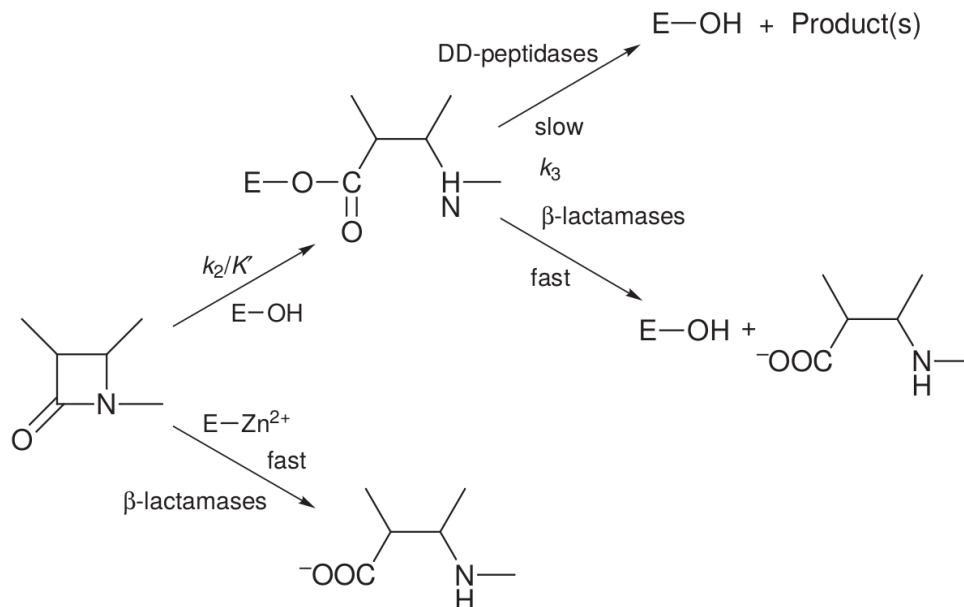


FIG. 8 : Réaction entre les  $\beta$ -lactamases/PBPs et les  $\beta$ -lactamines.  $E-OH$  représente les enzymes à sérine active (PBPs et classes A,C et D) et  $E-Zn^{2+}$  les métallob- $\beta$ -lactamases.  $E-OH$  et les  $\beta$ -lactamines forment un intermédiaire covalent (acyl-enzyme). Les  $\beta$ -lactamases peuvent hydrolyser le substrat, tandis que les PBPs hydrolysent peu ou pas du tout le substrat et restent sous forme d'acyl-enzyme inactif.  $E-Zn^{2+}$  ne passe pas par un intermédiaire covalent et hydrolyse directement le noyau  $\beta$ -lactame (Matagne et al. 1999).

### 1.7.1 Classe A

Historiquement, la classe A était caractérisée comme possédant une activité pénicillinase supérieure à son activité céphalosporinase. Cette classe est la mieux caractérisée et la plus

diversifiée en terme de séquences décrites. Les sous-familles les plus répandues sont les TEM, SHV, CTX-M et KPC (Fisher, Meroueh, and Mobashery 2005; Philippon, Slama, and Labia 2016). La sous-famille TEM regroupe des  $\beta$ -lactamases dont les gènes de résistance sont portés par un plasmide. En 1963, TEM-1 a été la première  $\beta$ -lactamase encodée sur un plasmide à être identifiée. Depuis, de nombreux variants de TEM-1 ont été décrits. Tous sont des ESBLs et diffèrent de TEM-1 par une à cinq substitutions (Salverda, Visser, and Barlow 2010). Comme TEM, SHV est une sous-famille de classe A portée sur un plasmide. Dans les années 1970, SHV-1 a été la première à être identifiée chez *E.coli*. Il existe également de nombreux variants ESBLs et non ESBLs de SHV-1. En 1997, on a découvert qu'un gène de type SHV-1 était encodé par le chromosome de *Klebsiella pneumoniae* et cela a été interprété comme la situation ancestrale (Liakopoulos, Mevius, and Ceccarelli 2016). Les enzymes CTX-M sont toutes des ESBLs portées sur un plasmide. Depuis le début des années 2000, on a observé une importante dispersion de ces enzymes, ce qui fait des CTX-M les ESBLs les plus répandues au monde de nos jours (Cantón, González-Alba, and Galán 2012). Les KPC sont une sous-famille de classe A à activité carbapénémase encodée sur un plasmide (Queenan and Bush 2007).

### 1.7.2 Classe B

Les métallo- $\beta$ -lactamases de classe B (MBLs) sont structurellement très divergentes des autres classes. De plus, elles ont comme particularité d'inclure un ou deux ions zinc au niveau de leur site catalytique. Cette caractéristique fait que les MBLs sont inhibées par des agents chélatants tels que l'EDTA. Les MBLs présentent un large spectre d'action et sont capables d'inhiber tous les antibiotiques à noyau bêta-lactame, à l'exception des monobactames (Palzkill 2013). Sur base de leurs séquences, on peut diviser les MBLs en trois sous-classes : B1, B2 et B3. La sous-classe B1 est celle qui contient le plus de représentants (Galleni, Lamotte-brasseur, and Rossolini 2001). Par exemple, IMP, NDM et VIM sont trois types de  $\beta$ -lactamases B1 portées sur un plasmide que l'on retrouve principalement chez *Pseudomonas* spp., mais également chez des *Enterobacteriaceae* telles que *Klebsiella pneumoniae* et *Citrobacter freundii* (Laraki et al. 1999; Lauretti et al. 1999).

### 1.7.3 Classe C

Les  $\beta$ -lactamases de classe C (ou AmpC) sont à l'origine des enzymes encodées dans le chromosome de plusieurs espèces d'*Enterobacteriaceae*. Ces gènes ont la particularité d'être inductibles et confèrent une résistance contre les céphalosporines et les monobactames (Philippon, Arlet, and Jacoby 2002; G. A. Jacoby 2009). Il a été démontré que des bactéries encodant une AmpC à faible activité contrebalancent ce défaut en la surproduisant massivement (Lakaye et al. 1999). Ces dernières années, de nombreux plasmides portant ce type de  $\beta$ -lactamases ont été retrouvés dans des espèces bactériennes n'encodant pas des AmpC sur leur chromosome (*Salmonella* spp. par exemple) (Pozzi et al. 2018).

#### 1.7.4 Classe D

Historiquement, la classe D ou OXA  $\beta$ -lactamases était différenciée des autres  $\beta$ -lactamases à sérine active par sa capacité à hydrolyser l'oxacilline. Les premières OXAs identifiées étaient toutes portées par un plasmide. On a ensuite découvert que les chromosomes d'*Acinetobacter* spp. encodent des OXAs (B. A. Evans and Amyes 2014) à large spectre d'action pouvant hydrolyser les carbapénèmes (Brown and Amyes 2006). On retrouve les OXAs dans de nombreuses espèces Gram-négatives pathogènes. Cependant, une étude récente a montré que des OXAs putatives étaient présentes dans le génome de bactéries Gram-positives des familles des *Bacillaceae*, *Clostridiaceae* et *Eubacteriaceae*. Il a également été démontré que certaines espèces de *Bacillus* possèdent des  $\beta$ -lactamases actives (Toth et al. 2016).

La comparaison de structures 3D a démontré que les OXAs sont apparentées au domaine C-terminal du récepteur membranaire BlaR1 (figure 9) (Leonard, Bonomo, and Powers 2013) présent chez *Staphylococcus aureus*, *Bacillus licheniformis* et *Clostridium botulinum*. Localisé dans le périplasme, le domaine C-terminal de BlaR1 permet la transduction du signal de détection de  $\beta$ -lactamines vers le cytoplasme, ce qui déclenche la synthèse de  $\beta$ -lactamases (Llarrull et al. 2011 ; Berzigotti et al. 2012 ; Mazuet et al. 2016).

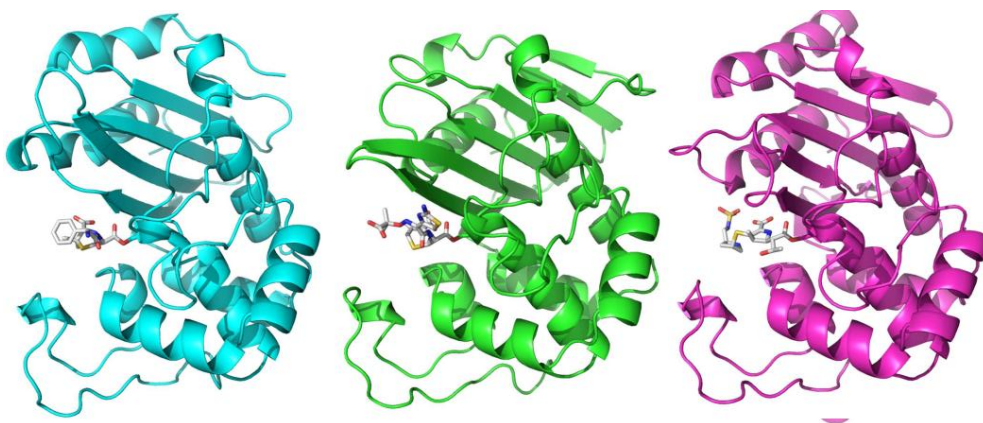


FIG. 9 : Comparaison des structures 3D : OXA-10 (ESBL) en cyan et formant un acyl-enzyme avec l'ampicilline ; OXA-24 (carbapénémase) en vert et formant un acyl-enzyme avec un carbapénème ; domaine C-terminal de BlaR1 en magenta et formant un acyl-enzyme avec une céphalosporine (Leonard, Bonomo, and Powers 2013).

Au niveau de la poche du site actif des OXAs se retrouvent plusieurs résidus conservés, tels que le triplet Sérine-Alanine-Valine (SAV) situé sur la boucle entre les hélices  $\alpha$  4 et  $\alpha$  5 et, en face, le triplet Lysine-Thréonine-Glycine (KTG) du feuillet- $\beta$  5. A l'entrée du site actif, on retrouve la boucle  $\omega$  qui contient l'hélice  $\alpha$  7. Sur cette hélice se trouve un Tryptophane (W) conservé chez toutes les OXAs connues. Enfin, dans la partie N-terminale de l'hélice  $\alpha$  3 se trouve le motif Sérine-Thréonine-Phénylalanine-Lysine (STFK) (figure 10). La sérine sur l'hélice  $\alpha$  3 est la sérine active. Une particularité du site actif des OXAs est que la lysine sur l'hélice  $\alpha$  3 peut être carboxylée (Kerff et al., n.d. ; Leonard, Bonomo, and Powers 2013).

### 1.7.5 Distribution taxonomique des $\beta$ -lactamases

Tout au long de cette introduction nous avons subdivisé les bactéries sous les termes traditionnels de *Gram-positives* et *Gram-négatives* pour être en adéquation avec la littérature consacrée aux antibiotiques. Cependant, cette classification n'est pas correcte d'un point de vue phylogénétique et ne reflète qu'une partie de la diversité des enveloppes bactériennes. En effet, le terme Gram-positives désigne les bactéries possédant une épaisse couche de PG qui retient le colorant cristal violet. Ces bactéries apparaissent donc en violet lors de la coloration de Gram. Quant au terme Gram-négatives, il s'applique aux bactéries possédant une couche de PG plus fine et une membrane externe asymétrique qui ne retient pas le cristal violet. Dès lors, ces dernières apparaissent fuchsia lors de la coloration de Gram. Néanmoins, certaines bactéries ne possédant ni PG ni membrane externe présentent une coloration caractéristique des Gram-négatives. Il est donc plus correct (même si pas parfait) d'utiliser les termes *monodermes* et *didermes* (Léonard 2014). Durant tout le reste de ce travail, nous raisonnerons pour notre part en termes d'embranchement (ou phylum) bactérien. Actuellement, les bactéries sont réparties dans 76 phyla, dont 17 uniquement représentés par des métagénomomes (Parks et al. 2017). (La métagénomique est une méthode d'analyse de génomes provenant d'échantillons de l'environnement. Elle permet l'étude génomique de micro-organismes incultivables en laboratoire.)

Lorsqu'on se réfère à la base de données "Beta-Lactamase DataBase" (Naas et al. 2017) (utilisée plus loin dans ce travail), les enzymes de classe A sont majoritairement présentes au sein des *Proteobacteria*, mais on en retrouve également chez les *Bacteroidetes*, les *Actinobacteria* et les *Firmicutes*. En ce qui concerne la classe B, la distribution taxonomique varie énormément en fonction des trois sous-classes décrites. Les enzymes de la sous-classe B1 sont majoritairement présentes chez les *Firmicutes*, alors que le reste se retrouve chez les *Proteobacteria* et les *Bacteroidetes*, mais également chez les *Spirochaetes*. La sous-classe B2 n'est présente que chez les *Proteobacteria*, tandis que la sous-classe B3 est retrouvée chez les *Proteobacteria* et les *Bacteroidetes*. On retrouve la classe C exclusivement chez les *Proteobacteria*, à l'exception d'une enzyme décrite chez *Mycobacterium smegmatis* (*Actinobacteria*). Quant aux OXAs (classe D), elles sont largement présentes chez les *Proteobacteria*, alors que quelques OXAs ont été récemment décrites chez des *Firmicutes* (voir 1.7.4). Quelques OXAs sont également retrouvées chez les *Spirochaetes*, les *Bacteroidetes*, les *Fusobacteria* et dans des métagénomomes océaniques.

<b>Dimeric</b>	
I OXA-10	..... MKTFAAYV I I ACLS..... STALAGS <span style="border: 1px solid cyan;">I T E N T S</span> <span style="border: 1px solid cyan;">W N K E F S A</span> EAVN <span style="border: 1px solid cyan;">G V F V L C K S</span> 46
I OXA-13	..... MKTFAAYV I T ACLS..... STALASS <span style="border: 1px solid cyan;">I T E N T S</span> <span style="border: 1px solid cyan;">W N K E F S A</span> EAVN <span style="border: 1px solid cyan;">G V F V L C K S</span> 46
I OXA-48	-- MRVLALS AVFLVAS I I G..... MPAVAKE <span style="border: 1px solid cyan;">W Q E N K S</span> <span style="border: 1px solid cyan;">W N A H F T E</span> H K S Q <span style="border: 1px solid cyan;">G V V V L W N E</span> 49
II OXA-2	--- MAIRIFAILFSIFSLAT..... FAHAQEG <span style="border: 1px solid cyan;">T L E R S D</span> <span style="border: 1px solid cyan;">W R K F F S E F</span> QAK <span style="border: 1px solid cyan;">G T I V V A D E</span> 49
II OXA-46	--- MAIRFFTILLSTFFLT S..... FVYAQEH <span style="border: 1px solid cyan;">V V I R S D</span> <span style="border: 1px solid cyan;">W K K F F S D L</span> QAE <span style="border: 1px solid cyan;">G A I V I A D E</span> 49
<b>Monomeric</b>	
III OXA-1	--- MKNTI H I N F A I F L I I A N..... I I Y S S A S A S T D I <span style="border: 1px solid cyan;">S T V A S P L F</span> E G T E <span style="border: 1px solid cyan;">G C F L L Y D</span> 48
VI OXA-24	MKKFILPIFSI SILVSLSACSSI KTKSEDNFHI <span style="border: 1px solid cyan;">S S Q Q H E K A I K S Y F D E A</span> Q T Q <span style="border: 1px solid cyan;">G V I I I K E G</span> 60
<b>Dimeric</b>	
I OXA-10	S.. SKS <span style="border: 1px solid cyan;">C A T N</span> <span style="border: 1px solid cyan;">D L A R A S</span> <span style="border: 1px solid cyan;">K E Y L</span> <span style="border: 1px solid cyan;">P A S T F K I</span> <span style="border: 1px solid cyan;">P N A I I G L E T</span> G V I K N E H Q V <span style="border: 1px solid cyan;">F K W D G K P R A M</span> <span style="border: 1px solid cyan;">M K Q W E R</span> 104
I OXA-13	S.. SKS <span style="border: 1px solid cyan;">C A T N</span> <span style="border: 1px solid cyan;">N L A R A S</span> <span style="border: 1px solid cyan;">K E Y L</span> <span style="border: 1px solid cyan;">P A S T F K I</span> <span style="border: 1px solid cyan;">P S A I I G L E T</span> G V I K N E H Q V <span style="border: 1px solid cyan;">F K W D G K P R A M</span> <span style="border: 1px solid cyan;">M K Q W E R</span> 104
I OXA-48	N.. KQ <span style="border: 1px solid cyan;">Q G F T N</span> <span style="border: 1px solid cyan;">N L K R A N</span> <span style="border: 1px solid cyan;">Q A F L</span> <span style="border: 1px solid cyan;">P A S T F K I</span> <span style="border: 1px solid cyan;">P N S L I A L D L</span> G V V K D E H Q V <span style="border: 1px solid cyan;">F K W D G Q T R D I</span> <span style="border: 1px solid cyan;">A T W N R</span> 107
II OXA-2	RQADRAMLVF <span style="border: 1px solid cyan;">D P V R S K</span> <span style="border: 1px solid cyan;">K R Y S</span> <span style="border: 1px solid cyan;">P A S T F K I</span> <span style="border: 1px solid cyan;">P H T L F A L D A</span> G A V R D E F Q I <span style="border: 1px solid cyan;">F R W D G V N R G F A G H N Q</span> 108
II OXA-46	RQAKH <span style="border: 1px solid cyan;">T L S V F</span> <span style="border: 1px solid cyan;">D Q E R A A</span> <span style="border: 1px solid cyan;">K R Y S</span> <span style="border: 1px solid cyan;">P A S T F K I</span> <span style="border: 1px solid cyan;">P H T L F A L D A</span> D A V R D E F Q V <span style="border: 1px solid cyan;">F R W D G V N R S F A G H N Q</span> 108
<b>Monomeric</b>	
III OXA-1	VSTNA <span style="border: 1px solid cyan;">E I A Q F</span> <span style="border: 1px solid cyan;">N K A K C A</span> T Q M A <span style="border: 1px solid cyan;">P D S T F K I</span> <span style="border: 1px solid cyan;">A L S L M A F D A</span> E.. I I D Q K T I <span style="border: 1px solid cyan;">F K W D K T P K G</span> <span style="border: 1px solid cyan;">M E I W N S</span> 107
VI OXA-24	K.. <span style="border: 1px solid cyan;">N L S T Y G N</span> <span style="border: 1px solid cyan;">A L A R A N K</span> E Y V <span style="border: 1px solid cyan;">P A S T F K I</span> <span style="border: 1px solid cyan;">M L N A L I G L E N</span> .. H K A T T N E I <span style="border: 1px solid cyan;">F K W D G K K R T</span> <span style="border: 1px solid cyan;">Y P M W E K</span> 107
<b>Dimeric</b>	
I OXA-10	D L T <span style="border: 1px solid cyan;">L R G A I Q V</span> <span style="border: 1px solid cyan;">S A V P V F Q Q I A R E V</span> G E V R M Q K Y L K K F S Y G N Q N I S G G..... I D <span style="border: 1px solid cyan;">K F W L E G Q</span> 158
I OXA-13	D L S <span style="border: 1px solid cyan;">L R G A I Q V</span> <span style="border: 1px solid cyan;">S A V P V F Q Q I A R E V</span> G E V R M Q K Y L K K F S Y G N Q N I S G G..... I D <span style="border: 1px solid cyan;">K F W L E D Q</span> 158
I OXA-48	D H N L I T A M K Y <span style="border: 1px solid cyan;">S V V P V Y Q E F A R Q I</span> G E A R M S K M L H A F D Y G N E D I S G N..... V D <span style="border: 1px solid cyan;">S F W L D G G</span> 161
II OXA-2	D Q D <span style="border: 1px solid cyan;">L R S A M R N</span> <span style="border: 1px solid cyan;">S T V W V Y E L F A K E I</span> G D D K A R R Y L K K I D Y G N A D P S T S..... N G D Y <span style="border: 1px solid cyan;">W I E G S</span> 163
II OXA-46	D Q D <span style="border: 1px solid cyan;">L R S A M R N</span> <span style="border: 1px solid cyan;">S T V W V Y E L F A K D I</span> G E D K A R R Y L K K I D Y G N V D P S T I..... K G D Y <span style="border: 1px solid cyan;">W I D G N</span> 163
<b>Monomeric</b>	
III OXA-1	N H T <span style="border: 1px solid cyan;">P K T W M Q F</span> <span style="border: 1px solid cyan;">S V V W V S Q E I T Q K I</span> G L N K I K N Y L K D F D Y G N Q D F S G D K E R N <span style="border: 1px solid cyan;">N G L T E A W L E S S</span> 167
VI OXA-24	D M T <span style="border: 1px solid cyan;">L G E A M A L</span> <span style="border: 1px solid cyan;">S A V P V Y Q E L A R R T</span> G I E L M Q K E V K R V N F G N T N I G T Q..... V D <span style="border: 1px solid cyan;">N F W L V G P</span> 171
<b>Dimeric</b>	
I OXA-10	L R I <span style="border: 1px solid cyan;">S A V N Q V E F L E S L Y L N</span> <span style="border: 1px solid cyan;">K L S A</span> <span style="border: 1px solid cyan;">S K E N Q L I V K E A L V</span> <span style="border: 1px solid cyan;">T E A A P E</span> .. <span style="border: 1px solid cyan;">Y L V H S K T G F</span> <span style="border: 1px solid cyan;">S G V G T E S N P</span> 217
I OXA-13	L R I <span style="border: 1px solid cyan;">S A V N Q V E F L E S L F</span> <span style="border: 1px solid cyan;">L N K L S A</span> <span style="border: 1px solid cyan;">S K E N Q L I V K E A L V</span> <span style="border: 1px solid cyan;">T E A A P E</span> .. <span style="border: 1px solid cyan;">Y L V H S K T G F</span> <span style="border: 1px solid cyan;">S G V G T E S N P</span> 217
I OXA-48	I R I <span style="border: 1px solid cyan;">S A T E Q I S F L R K L Y H N</span> <span style="border: 1px solid cyan;">K L H V</span> <span style="border: 1px solid cyan;">S E R S Q R I V K Q A M L</span> <span style="border: 1px solid cyan;">T E A</span> N G D.. <span style="border: 1px solid cyan;">Y I I R A K T G Y S</span> .. - - T R I E P 217
II OXA-2	L A I <span style="border: 1px solid cyan;">S A Q E Q I A F L R K L Y R N</span> <span style="border: 1px solid cyan;">E L P F</span> <span style="border: 1px solid cyan;">R V E H Q R L V K D L M I</span> <span style="border: 1px solid cyan;">V E A G R N</span> .. <span style="border: 1px solid cyan;">W I L R A K T G W E G</span> ..... 215
II OXA-46	L K I <span style="border: 1px solid cyan;">S A H E Q I L F L R K L Y R N</span> <span style="border: 1px solid cyan;">Q L P F</span> <span style="border: 1px solid cyan;">K V E H Q R L V K D L M I</span> <span style="border: 1px solid cyan;">T E A</span> G R S.. <span style="border: 1px solid cyan;">W I L R A K T G W E G</span> ..... 215
<b>Monomeric</b>	
III OXA-1	L K I <span style="border: 1px solid cyan;">S P E E Q I Q F L R K I I N H</span> <span style="border: 1px solid cyan;">N L P V</span> <span style="border: 1px solid cyan;">K N S A I E N T I E N M</span> <span style="border: 1px solid cyan;">Y L Q D</span> L E N S T <span style="border: 1px solid cyan;">K L Y G K T G A G F</span> <span style="border: 1px solid cyan;">T A N R</span> .. T L 227
VI OXA-24	L K I <span style="border: 1px solid cyan;">T P V Q E V N F A D D L A H N R L P F</span> <span style="border: 1px solid cyan;">K L E T Q</span> .. E V K K M L <span style="border: 1px solid cyan;">L I K E</span> V N G S.. <span style="border: 1px solid cyan;">K I Y A K S G W G</span> .. - - M G.. V T P 227
<b>Dimeric</b>	
I OXA-10	<span style="border: 1px solid cyan;">G V A W W V G W V E K E T</span> <span style="border: 1px solid cyan;">E.. V Y F F A F N M D I D N</span> .. <span style="border: 1px solid cyan;">E S K L P L R K S I P T K I M E S E</span> <span style="border: 1px solid cyan;">G I I G G</span> ..... 266
I OXA-13	<span style="border: 1px solid cyan;">G V A W W V G W V E K G T</span> <span style="border: 1px solid cyan;">E.. V Y F F A F N M D I D N</span> .. <span style="border: 1px solid cyan;">E N K L P L R K S I P T K I M A S E</span> <span style="border: 1px solid cyan;">G I I G G</span> ..... 266
I OXA-48	<span style="border: 1px solid cyan;">K I G W W V G W V E L D D N</span> .. <span style="border: 1px solid cyan;">V W F F A F N M D M P T</span> .. <span style="border: 1px solid cyan;">S D G L G L R Q A I T K E V L K Q E</span> <span style="border: 1px solid cyan;">K I I P</span> ..... 265
II OXA-2	<span style="border: 1px solid cyan;">R M G W W V G W V E W P T</span> <span style="border: 1px solid cyan;">G.. S V F F A L N I D T</span> <span style="border: 1px solid cyan;">P N R M D D L F K R E A I V R A I L R S I</span> <span style="border: 1px solid cyan;">E A L P P N P A V N S D A A</span> 274
II OXA-46	<span style="border: 1px solid cyan;">R F G W W V G W I E W P T</span> <span style="border: 1px solid cyan;">G.. P V F F A L N I D T</span> <span style="border: 1px solid cyan;">P N R T D D L F K R E A I A R A I L R S I</span> <span style="border: 1px solid cyan;">D A L P P N</span> ..... 266
<b>Monomeric</b>	
III OXA-1	<span style="border: 1px solid cyan;">Q N G W F E G F I I</span> <span style="border: 1px solid cyan;">S K S G H</span> <span style="border: 1px solid cyan;">K Y V F V S A L T G</span> <span style="border: 1px solid cyan;">N L G S N L</span> <span style="border: 1px solid cyan;">T S S I K A K K N A I T</span> .. - - I L N T L N L..... 276
VI OXA-24	<span style="border: 1px solid cyan;">Q V G W L T G W V E</span> <span style="border: 1px solid cyan;">Q A N G K</span> <span style="border: 1px solid cyan;">K I P F S L N L E M K E</span> .. G M <span style="border: 1px solid cyan;">S G S I R N E I T Y K S L</span> <span style="border: 1px solid cyan;">E N L G I I</span> ..... 274

FIG. 10 : Alignement multiple de quelques séquences d'OXAs. En cyan sont représentés les hélices- $\alpha$  et en magenta les feuillets- $\beta$ . Les résidus conservés au niveau de la poche du site actif sont encadrés en jaune (figure de F. Kerff).

## 2 Objectifs

Lorsqu'une nouvelle  $\beta$ -lactamase appartenant à la classe D est découverte, un nom arbitraire lui est attribué (OXA + numéro de série) (Srivastava et al. 2014). On lui associe en outre un nom de sous-famille sur base de sa seule similarité avec une autre séquence déjà décrite (généralement la plus proche dans les banques de données). Cependant, ces annotations ne sont pas correctes d'un point de vu évolutif. De ce fait, nous voudrions mettre en place une annotation des OXAs qui soit en accord avec la phylogénie de la classe.

Pour ce faire, nous allons procéder en plusieurs étapes. Premièrement, nous allons établir la distribution des OXAs au travers des procaryotes afin de déterminer quels phyla possèdent ce type de  $\beta$ -lactamase. Ensuite, à partir des séquences putatives d'OXA récoltées à travers les différents phyla, nous allons réaliser un arbre phylogénétique. Cet arbre nous permettra de définir différentes sous-familles (ou groupes) au sein de la classe D. Nous pourrons alors caractériser les groupes définis d'un point de vue taxonomique et de celui de la conservation de leur site actif. Enfin, nous tenterons de nommer ces groupes en fonction de séquences d'OXA de référence.

## 3 Matériels & méthodes

### 3.1 Environnement

#### 3.1.1 Hardware

Deux ordinateurs ont été utilisés lors de ce mémoire. Le premier est le cluster de calcul *durandal* opéré par l'Unité de Phylogénomique des Eucaryotes. Il s'agit d'un système Flex d'IBM/Lenovo contenant un gros noeud de calcul (x440) et neuf plus petits (x240). Il totalise 196 cores physiques (392 en Hyper-Threading), 2880 cores CUDA (GPU), 2 TB de RAM et 112 TB d'espace de stockage. *durandal* utilise un système d'exploitation Linux CentOS 6.6.

Mon ordinateur personnel à également été utilisé. Il s'agit d'un Packard Bell EasyNote LS44-HR à un processeur Processeur Intel® Core™ i3-2310M à 2 cores physiques, 4 Go de RAM et 500 Go d'espace de stockage.

#### 3.1.2 UNIX

Créé en 1969 par Kenneth Thompson, UNIX est un système d'exploitation constituant la base de **Linux** et qui inclut notamment un interpréteur en ligne de commandes (shell). Le shell sert d'interface homme-machine et permet à l'homme de réaliser différentes opérations qui seront interprétée par le Shell.

#### 3.1.3 Perl

Perl est un langage de programmation crée par le linguiste Larry Wall en 1987 permettant de traiter de l'information textuelle. Il est dérivé du langage C et s'inspire également d'autres langages comme sed, awk ou encore le shell d'UNIX. Tous les scripts utilisés lors de ce travail ont été rédigés en Perl et les plus importants d'entre eux sont repris en annexe.

### 3.2 Données et pré-traitement des données

#### 3.2.1 $\beta$ -lactamases de référence

Lors de ce projet, deux jeux de données de référence ont été exploités. Le premier provenait du projet **pathogens detection** (<ftp://ftp.ncbi.nlm.nih.gov/pathogen/>) développé par le National Center for Biotechnology Information (NCBI) et a été téléchargé le 1er décembre 2017. Il contenait des séquences protéiques représentant les quatre classes de  $\beta$ -lactamases provenant du NCBI et du Lahey Hospital & Medical Center (un hôpital rattaché à l'université de Tufts). Au total, 2 711 séquences étaient réparties entre les deux fichiers dont seulement 1 617 uniques. Ces séquences ont été utilisées pour générer les profils HMMs (voir 3.4.1). Le second jeu de données, téléchargé

le 30 avril 2018, provenait de la Beta-Lactamase Data Base (BLDB) (<http://www.bldb.eu/>). Il regroupait 613 séquences de  $\beta$ -lactamases de classe D qui ont été utilisées pour l’annotation finale de l’arbre phylogénétique (voir 3.4.3).

### 3.2.2 Génomes complets pour le data mining

Une base de données (nommée **prok\_batch\_db**) incluant 80 490 génomes procaryotes complets a été utilisée pour rechercher *in silico* des séquences putatives de  $\beta$ -lactamases (voir 3.4.1).

### 3.2.3 Sélection des $\beta$ -lactamases de référence

**get\_accession\_num.pl** est un script Perl permettant de partitionner le fichier FASTA du NCBI contenant les quatre classes de  $\beta$ -lactamases. Pour ce faire, il est nécessaire de fournir au programme un fichier tabulaire contenant trois colonnes : nom du gène, numéro d’accession, description du gène (ex. : blaOXA-100, CAJ77817.2, OXA-51 family carbapenem-hydrolyzing class D beta-lactamase OXA-100). Le script va ‘lire’ la troisième colonne du fichier tabulaire, repérer la classe de  $\beta$ -lactamase spécifiée avec l’option **--class** et utiliser le numéro d’accession pour extraire la séquence correspondante du fichier FASTA.

Après utilisation, les 1617 séquences de référence ont été réparties dans 4 fichiers en fonction de leur classe. Ainsi, 640 séquences appartenaient à la classe A, 136 à la classe B, 371 à la classe C et 470 à la classe D. L’option **--forann** a également été utilisée pour générer des noms pour les séquences de la classe D qui seront utilisés par le programme **annotate.pl** (voir 3.4.3).

Pour sa part, le script **ann-classd.pl** permet de générer des noms pour les séquences de la BLDB en fournissant au programme un fichier tabulaire correspondant à la page web <http://bldb.eu/BLDB.php?prot=D>. Ces noms seront également utilisés par le programme **annotate.pl**.

### 3.2.4 Déréplication des $\beta$ -lactamases de référence

Développé par le Dr. Weizhong Li, **cd-hit** version 4.6 est un programme permettant de réduire la taille d’une base de données en se basant sur la déréplication de séquences. C’est un algorithme de type glouton<sup>1</sup> qui procède en deux étapes. Premièrement, les séquences sont classées de la plus longue à la plus courte. La séquence la plus longue est utilisée comme représentante du premier ‘cluster’. Ensuite, la séquence suivante est intégrée dans le cluster si elle est au moins similaire à un seuil d’identité prédéterminé (95%, 98% ou 100%); sinon elle forme un nouveau cluster. Ces étapes sont répétées jusqu’à ce que toutes les séquences fassent partie d’un cluster (éventuellement ne contenant qu’une seule séquence) (Fu et al. 2012).

---

<sup>1</sup>Un algorithme glouton est une heuristique qui procède étape par étape en privilégiant les optima locaux afin de s’approcher de l’optimum global.



Dans la suite de ce mémoire, les séquences représentant un cluster en particulier sont nommées **représentantes** et les autres séquences du cluster **membres**.

### 3.3 Alignement multiple

#### 3.3.1 Alignement automatique

L'alignement multiple de séquences est réalisé par le programme **MAFFT** (Multiple Alignment using Fast Fourier Transform) version v7.273. La transformée de Fourier rapide sert à l'identification des régions homologues au sein de l'alignement (Kato et al. 2002). Ensuite, une matrice de similarité et un système de score sont utilisés pour l'introduction de gaps dans les séquences alignées.

Les fonctions `--add` et `--keeplength` ont également été utilisées pour l'ajout des séquences au sein d'alignements déjà existants tout en évitant l'introduction de nouveaux gaps dans l'alignement enrichi (Kato and Frith 2012).

#### 3.3.2 Optimisation manuelle

**MUST-C** est un package regroupant divers programmes écrits en langage C développés par le Prof. H. Philippe du CNRS de Moulis et améliorés par le Prof. D. Baurain.

**musted** est un programme interactif permettant de modifier manuellement un alignement multiple au format ALI. Durant ce mémoire, nous avons procédé à de nombreux 'rounds' d'optimisation de l'alignement des OXAs. Nous avons utilisé la version 5.19a du programme.

**mustnet** est un programme exploitant notamment **Gblocks** (Castresana 2000) et qui permet de sélectionner manuellement ou automatiquement des colonnes à analyser dans un alignement multiple contenant des régions moins bien alignées (figure 12). La version 5.04 de **mustnet** a été utilisée.

### 3.4 Recherche de $\beta$ -lactamases putatives

#### 3.4.1 Package HMMER

**HMMER** est un package d'analyse de séquences développé par le Dr. Sean Roberts Eddy. Il regroupe un ensemble de programmes qui reposent sur les modèles statistiques de Markov cachés (Hidden Markov Models : HMM). A partir d'un profil, il permet la recherche de séquences homologues dans une base de données. Comparé à d'autres outils de recherche, tel que **BLAST**, **HMMER** est à la fois plus sensible et plus spécifique car il modélise des alignements multiples et non des alignements par paire (Eddy, Wheeler, and Development Team 2015). Lors de ce travail, la version 3.1b2 du package a été utilisée.

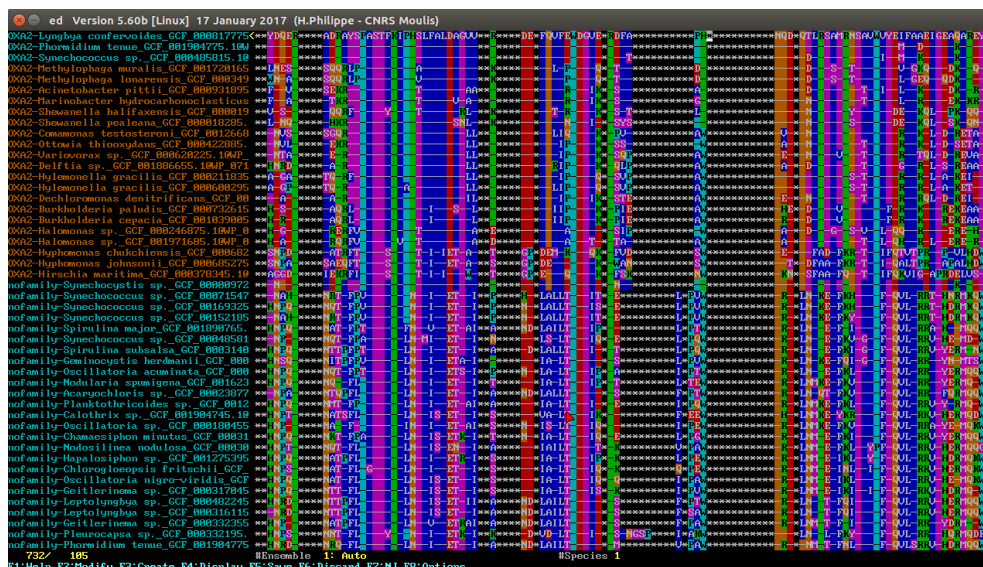


FIG. 11 : Interface du programme musted. A gauche, on retrouve le nom des séquences avec un code couleur dépendant de la taxonomie. A droite se trouve l'alignement multiple, les étoiles blanches représentant les gaps, alors que les AAs sont représentés par un code couleur fonction de leurs propriétés physico-chimiques.

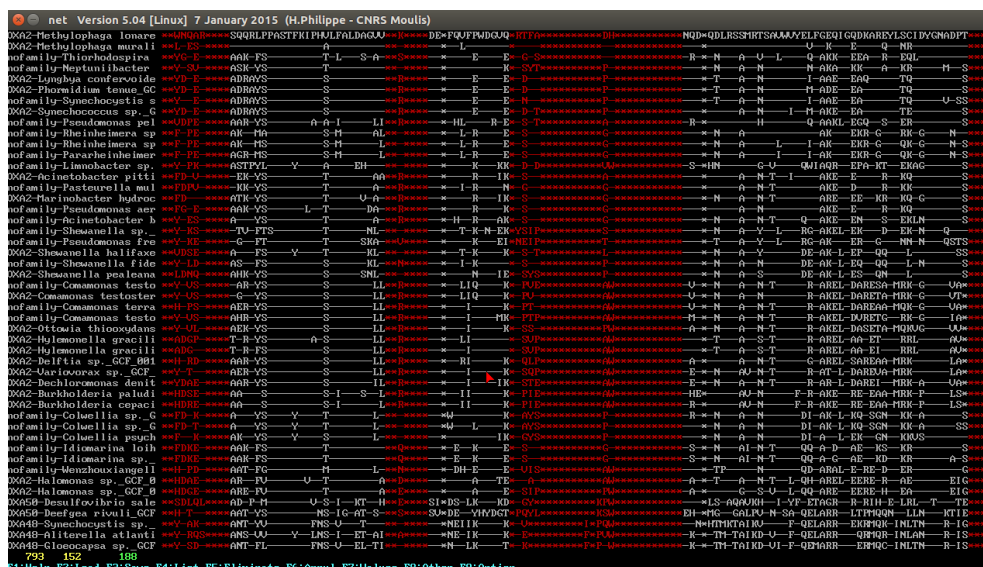


FIG. 12 : Interface du programme mustnet. On retrouve à gauche le nom des séquences et à droite l'alignement multiple, dans lequel les gaps sont représentés par des étoiles et les colonnes sélectionnées sont montrées en blanc.

Le programme **hmmbuild** a été utilisé afin construire des profils HMM à partir des fichiers d'alignement multiple. Un profil HMM est un modèle statistique d'un alignement multiple permettant de modéliser quelles positions de l'alignement sont conservées, les acides aminés attendus à ces positions et les régions où des insertions ou des délétions sont acceptables.

La recherche des séquences putatives a été effectuée par le programme **hmmsearch**. Le profil créé à partir de **hmmbuild** était confronté à la base de données de génomes complets **prok\_batch\_db**.

On obtenait en sortie un fichier avec la liste des séquences correspondant au profil, chacune avec une e-value associée.

### 3.4.2 Sélection des séquences putatives

`ompa-pa.pl` est un outil graphique interactif développé par le Prof. D. Baurain qui permet d'extraire des séquences d'une base de données sur base de recherches d'homologie. Pour ce faire, il suffit de fournir au programme le fichier de sortie produit par `hmmsearch` et le chemin vers la base de données (ex. `prok_batch_db`). Ensuite, `ompa-pa.pl` génère un graphique qui représente les séquences obtenues en fonction de leur longueur et du logarithme de leur e-value vis-à-vis du profil HMM. On peut alors sélectionner à la souris la zone du graphique contenant les séquences que l'on souhaite conserver.

### 3.4.3 Annotation des OXAs putatives

Les séquences putatives d'OXA récupérées après utilisation de `ompa-pa.pl` sont identifiées par un numéro d'accèsion incluant un code GCA. A partir de ces codes GCA et de la taxonomie du NCBI, un nom d'organisme (genre, espèce) est attribué aux séquences en utilisant le script Perl `fetch-tax.pl` écrit par le Prof. D. Baurain.

Un tag a été ajouté à ces séquences grâce au script Perl `annotate.pl` qui permet d'annoter des séquences inconnues à partir de séquences de référence. Cette annotation est basée sur la similarité de séquence. Il suffit de fournir au programme deux fichiers au format FASTA, un avec des séquences inconnues et le second avec les séquences de référence. Ensuite, le programme réalise un BLAST des séquences inconnues contre les séquences connues et utilise le nom de chaque premier 'hit' (le cas échéant) pour annoter les séquences inconnues. Une valeur seuil de e-value et/ou d'identité peut être spécifiée pour contrôler la sensibilité de l'annotation.

Les bases de données du NCBI et BLDB ont été utilisées pour annoter les séquences putatives de  $\beta$ -lactamases de classe D. Avant annotation, trois séquences de BlaR1 (*Clostridium difficile* (CDT53463.1), *Staphylococcus aureus* (P18357) et *Bacillus licheniformis* (P12287)) et trois séquences de PBPs (PBP1a (P71707), PBP1b (P02919-2) et PBP1c (P76577)) ont été ajoutées aux séquences de référence d'OXA.

Après déréduplication, chaque séquence **représentante** porte un tag avec le nombre de séquences **membres** qu'elle représente. Cette annotation est réalisée par le script Perl `create-mapper.pl`.

## 3.5 Inférences phylogénétiques

`ali2phyliip.pl` est un script Perl développé par le prof. D. Baurain qui permet de filtrer un alignement multiple en ne gardant que les régions homologues et de convertir le fichier

d'alignement au format PHYLIP nécessaire pour l'utilisation de RAxML. Les options `--max`, `--min` et `--gb-mask` ont été utilisées pour filtrer l'alignement multiple. `--max` permet d'éliminer les colonnes de l'alignement contenant trop de gaps, tandis que `--min` permet d'éliminer les séquences trop courtes.

L'analyse phylogénétique est réalisée par le programme RAxML (Randomized Axelerated Maximum Likelihood) qui est particulièrement adapté pour des alignements de grande taille. RAxML utilise le maximum de vraisemblance (maximum likelihood) comme critère d'optimalité lors de la recherche du meilleur arbre. A partir d'un fichier d'alignement, il permet de produire un arbre phylogénétique avec de bons scores de vraisemblance en un temps raisonnable (Stamatakis 2014). Le modèle statistique retenu lors de ce mémoire était PROTGAMMALGF. Nous avons utilisé la version 8.1.17 de RAxML.

Le paramétrage de RAxML a été réalisée à l'aide du script Perl `phy2raxml.pl` développé par le Prof. D. Baurain. En fournissant un fichier d'alignement au format PHYLIP, le programme génère un job à lancer sur *durandal*.

Les valeurs de bootstrap ont été optimisées à l'aide de BOOSTER (BOOtstrap Support by TransfER, <https://booster.pasteur.fr/new/>) en fournissant au programme les fichiers `bipartition` et `bootstrap` générés par RAxML (Lemoine et al. 2018).

## 3.6 Analyses complémentaires

### 3.6.1 Distributions taxonomiques

Les analyses taxonomiques ont été réalisées avec le script `fetch-tax.pl` qui génère une taxonomie complète pour chaque identifiant de séquence.

Le script `OTUsamples2krona.sh` a été utilisé pour générer les graphiques circulaires de distribution taxonomique.

### 3.6.2 Conservation des résidus du site actif

Les logos de séquences représentant la poche du site actif ont été générés par le programme `weblogo` version 3.6.

## 4 Résultats

### 4.1 Distribution taxonomique des génomes complets

Notre base de données **prok\_batch\_db** est composé à 99% de génomes de bactéries et 1% d'archées (figure 13). Les génomes des bactéries sont répartis entre deux grands groupes les *Terrabacteria\_group* (48%) et *Proteobacteria* (44%). On retrouve dans le supraphylum *Terrabacteria\_group* deux phyla majeurs, les *Firmicutes* et les *Actinobacteria* (d'où proviennent la majorité des monodermes), le reste des génomes appartenant en partie à des *Cyanobacteria* ou des *Tenericutes* (bactéries sans paroi). Les autres génomes bactériens de la base de données appartiennent au *FCB\_group* (comprenant les *Bacteroidetes* et les *Chlorobi*), aux *Spirochaetes*, *PVC\_group*, *Fusobacteria* et à d'autres phyla minoritaires en termes de nombres de génomes (non montrés sur la figure 13). Ces phyla sont : *Thermotogae*, *Acidobacteria*, *Aquificae*, *Synergistetes*, *Nitrospirae*, *Thermodesulfobacteria*, *Deferribacteres*, *Balneolaeota*, *Elusimicrobia*, *Chrysiogenetes*, *Dictyoglomi*, *Caldiserica*, *Nitrospinae/Tectomicrobia* et *Calditrichaeota*.

Il est à noter que cet échantillonnage est biaisé en faveur de certains groupes de bactéries (voire de certains genres ou espèces, telles que *Escherichia coli*), mais ce biais n'est pas très facile à corriger à l'échelle des génomes complets. En effet, représenter certains groupes par certaines espèces ou souches seulement aurait pu nous faire manquer des  $\beta$ -lactamases dans la suite du travail.

### 4.2 Recherche de $\beta$ -lactamases putatives

Nous avons commencé ce projet en exploitant la base de données de  $\beta$ -lactamases du NCBI. Après séparation des séquences de référence en 4 fichiers, nous avons procédé à une déréplication des  $\beta$ -lactamases à un seuil de 98%. Les 640 séquences de la classe A ont été regroupées en 65 clusters, les 135 séquences de la classe B en 40 clusters, les 371 séquences de la classe C en 74 clusters et les 470 séquences de la classe D (OXA) en 160 clusters. Nous avons également généré un fichier contenant les 1481 séquences des trois classes de  $\beta$ -lactamases à sérine active (A+C+D) et réalisé une déréplication à 98%.

Nous avons ensuite généré des alignements multiples de ces cinq fichiers, puis construit cinq profils HMM nous permettant de rechercher des séquences homologues dans la base de données **prok\_batch\_db**. Le passage par *ompa-pa.pl* nous a permis de sélectionner 35 987 séquences putatives de  $\beta$ -lactamases de classe A, 2176 séquences de classe B, 17 288 séquences de classe C et 24 916 séquences de classe D. Le profil regroupant les  $\beta$ -lactamases de trois classes A+C+D nous a permis pour sa part de récupérer 95 761 séquences (figure 14).

Nous avons alors établi la redondance entre les différents fichiers de séquences putatives et observé que trois séquences étaient partagées par les fichiers des classes C et D. Le fichier contenant les



FIG. 13 : Graphique krona de la diversité taxonomique de la base de données prok\_batch\_db regroupant 80 490 génomes procaryotes.

séquences homologues aux trois classes de  $\beta$ -lactamases a également été comparé aux fichiers individuels des classes A, C et D. 537 séquences de la classe A, 104 séquences de la classe C et 3 séquences de la classe D étaient manquantes dans le fichier regroupant les trois classes. Cependant, le profil combiné nous a permis d'identifier 18 217 séquences de plus que les profils individuels. Aucune séquence putative de classe B n'a été retrouvée en dehors du fichier obtenu avec le profil HMM de la classe B.

### 4.3 Distribution taxonomique des OXAs putatives

Le profil de la classe D (OXA) nous a permis d'identifier des séquences putatives d'OXA mais également des séquences putatives de BlaR1. Pour analyser la distribution taxonomique des organismes possédant des OXAs (figure 15), nous avons d'abord écarté ceux ne possédant que des BlaR1 putatives (7733 séquences). Nous pouvons observer qu'aucune séquence archéenne, ni provenant de phyla bactériens dit "minoritaires" (à l'exception des *Balneolaeota*), n'a été

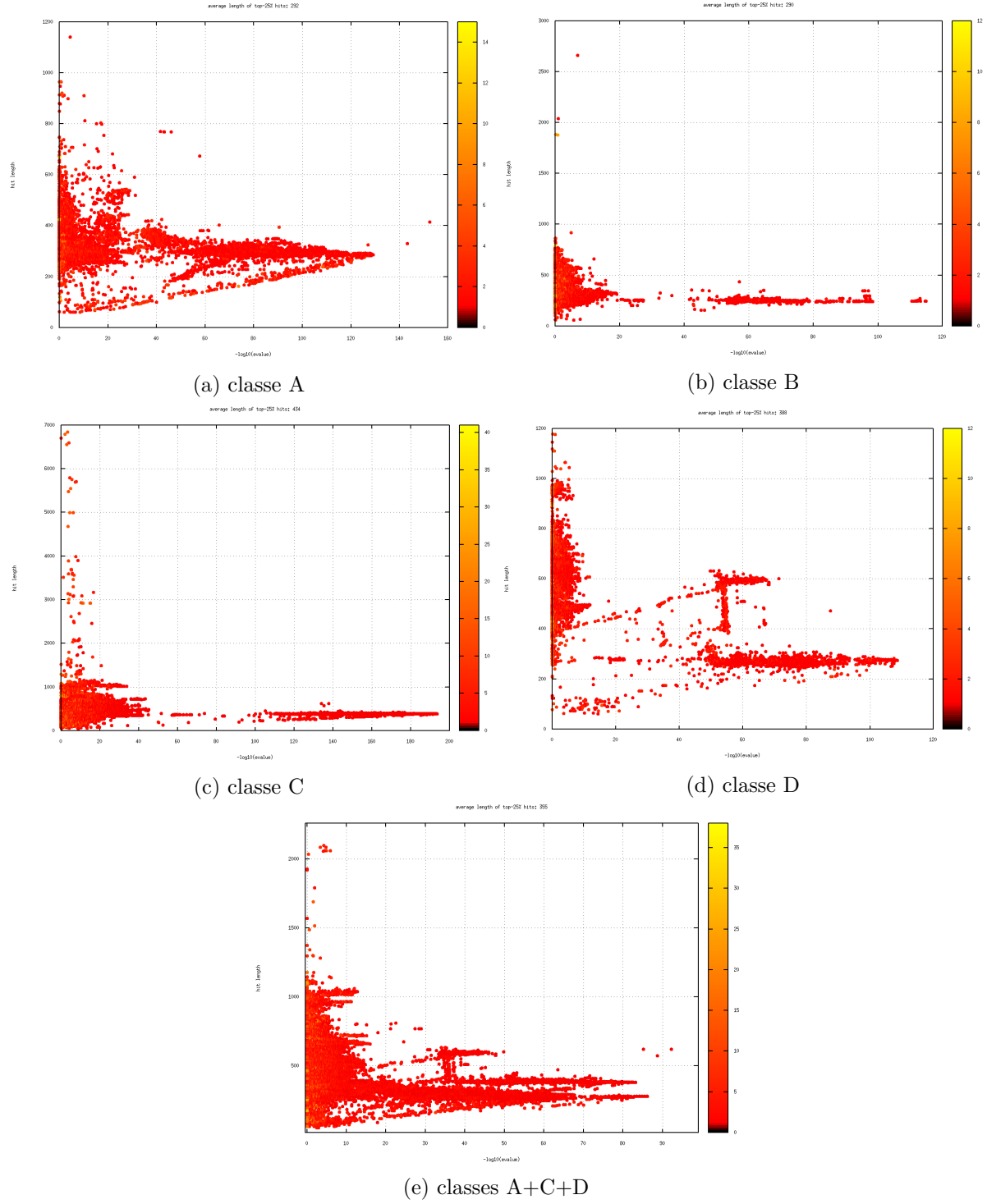


FIG. 14 : Graphiques de ompa-pa.pl obtenus à partir des cinq profils HMM : classe A, classe B, classe C, classe D et classes A+C+D. Les axes X et Y correspondent respectivement au logarithme négatif de la e-value et à la longueur des séquences identifiées par HMMER (représentées par des points rouges).

identifiée par le profil de la classe D. En revanche, des OXAs putatives ont été retrouvées au sein de phyla jamais décrits comme possédant des OXAs. Pour la plupart de ces phyla, une OXA putative a été détectée dans plusieurs organismes différents. Il est donc vraisemblable que ces résultats soient fiables. Ces phyla sont les *Cyanobacteria* et les *Actinobacteria* (du superphylum *Terrabacteria\_group*) les *Chlorobi* (du superphylum *FCB\_group*), les *Chlamydiae*, les *Planctomycetes* et *Verrucomicrobia* (du superphylum *PVC\_group*) et les *Balneolaeota* (genre *Aliifodinibius*). Les OXAs putatives sont majoritairement présentes au sein du phylum des *Proteobacteria* (74% des OXAs putatives) et plus particulièrement dans la classe des *Gammaproteobacteria* où le genre *Acinetobacter* possède à lui seul 22% des OXAs putatives. Cela dit, les *Gammaproteobacteria* représentent 29% de nos 80 490 génomes, illustrant la difficulté de tirer des conclusions taxonomiques quantitatives sans un échantillon équilibré.

FIG. 15 : Graphique krona de la diversité taxonomique des séquences putatives identifiées avec le profil HMM de la classe D. Ce graphique exclut les organismes possédant uniquement BlaR1.



des OXAs putatives dérèpliquées et les séquences des *Spirochaetes* plus que 1% (contre 7%). Ces résultats laissent supposer que ces séquences étaient très redondantes.

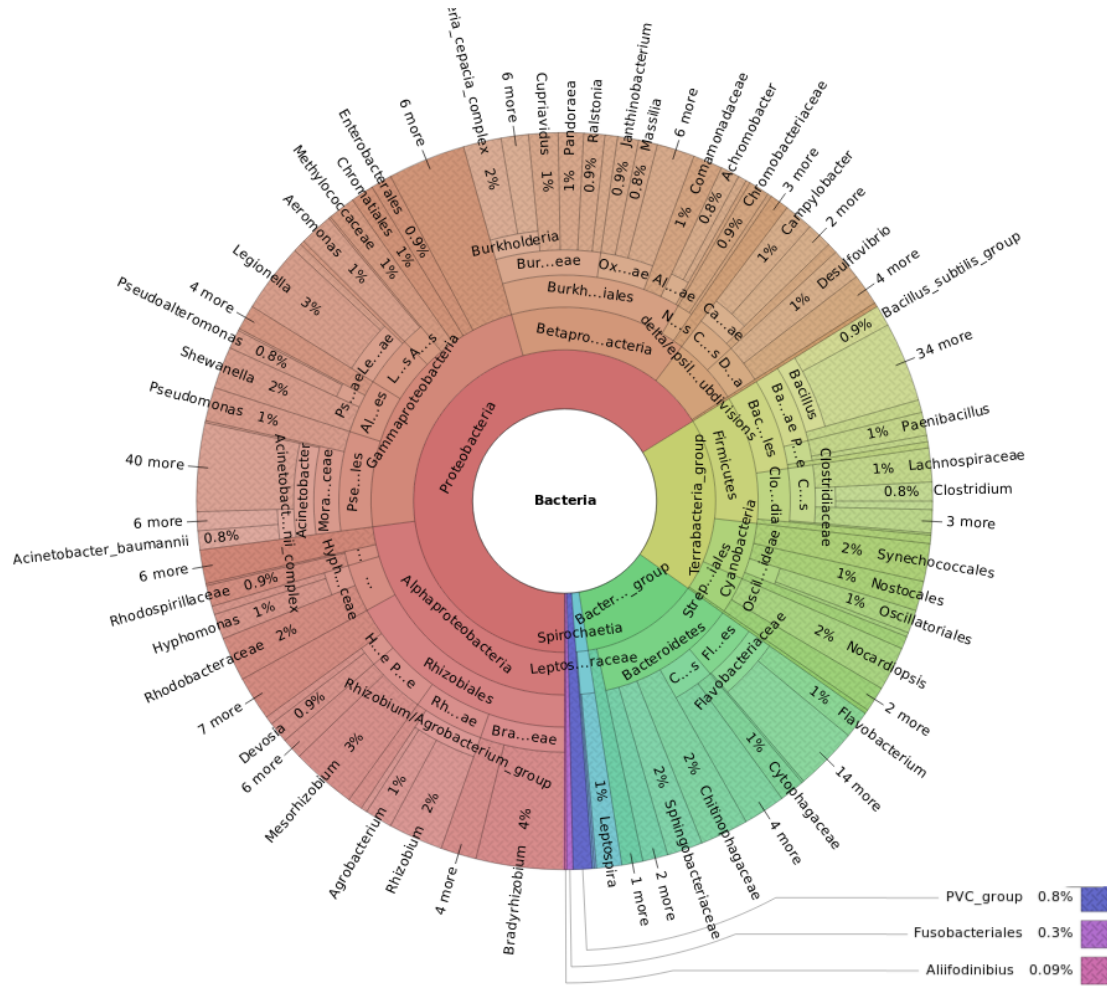


FIG. 16 : Graphique krona de la diversité taxonomique des séquences putatives d'OXA dérèpliquées à 95%.

#### 4.4 Annotation, alignement et phylogénie des OXAs putatives

L'ajout d'un tag aux séquences **représentantes** à une e-value seuil de  $1e^{-20}$  a été réalisé en utilisant les séquences de la classe D du NCBI (dérèpliquées à 100%) comme séquences de référence. Après annotation, 1406 séquences sur 1413 portent un tag et de nombreuses séquences portent le même tag. Plus de 200 séquences sont annotées BlaR1. Cependant, aucune séquence n'est annotée par une PBP. Cinq séquences non-annotées appartiennent à des *Leptospira* spp. et deux à des *Lysobacter* spp.

Plusieurs arbres phylogénétiques préliminaires successifs nous ont servi de guides pour optimiser à la main (avec **musted**) l'alignement automatique des 1413 séquences représentantes. La partie N-terminale des séquences annotées BlaR1 a été supprimée pour réduire la longueur de l'alignement. Les séquences écartées par application du filtre `--min` d'**ali2philyp.pl** et celles trop divergentes

ont été ré-introduites a posteriori à l'alignement optimisé manuellement à l'aide de **mafft**.

Avec **mustnet**, nous avons finalement extrait 188 colonnes bien conservées hors de l'alignement optimisé. Un filtre **ali2phylip.pl** a également été appliqué sur cet alignement final, mais aucune séquence n'a plus été écartée. C'est ce fichier qui nous a permis de calculer l'arbre phylogénétique présenté dans ce travail avec **RAxML** et le modèle **PROTGAMMALGF**.

## 4.5 Annotation détaillée des OXAs putatives

A côté des séquences de référence du NCBI, nous avons aussi testé celles de la banque de données BLDB (à nouveau dérépliquées à 100%). Comme ci-dessus, ces dernières nous ont permis d'ajouter un tag aux séquences putatives de la classe D à une e-value seuil de  $1e^{-20}$ . Aucune séquence n'est annotée par des PBPs et les sept séquences non-annotées avec la base de données du NCBI ne sont pas non plus annotées avec la base de données BLDB. Ensuite, nous avons comparé les annotations obtenues avec les deux bases de données de référence. Sur les 1406 séquences annotées, seules 300 étaient annotées différemment. En fin de compte, nous avons décidé d'utiliser les séquences BLDB pour annoter notre arbre car la banque BLDB contenait plus de séquences et des informations plus détaillées pour chaque séquence d'OXA.

Comme plusieurs séquences portaient un même tag, nous avons décidé d'ajouter un seuil d'identité sévère (en plus du seuil d'e-value) lors de l'annotation. Lorsque ce seuil d'identité est fixé à 95%, uniquement 105 séquences l'arbre sont annotées par 102 séquences de la base de données BLDB. Si on annoté les séquences représentantes et les séquences membres des clusters, 12 187 séquences sont annotées par 279 séquences de la base de données.

Afin d'examiner la raison pour laquelle les 337 séquences restantes de la base de données BLDB n'annotaient aucune des séquences putatives, nous avons utilisé l'option **--hitlist** du programme **annotate.pl**. Cette option permet d'obtenir une liste de tous les hits BLAST de chaque séquence putative à des seuils d'e-value et d'identité prédéterminés. Après utilisation, nos 12 187 séquences ont été annotées par 595 séquences de la base de données. Ces 306 séquences supplémentaires n'annotaient auparavant aucune séquence putative car elles n'étaient jamais un premier hit. Vu que plusieurs séquences de référence peuvent annoter la même séquence putative, la base de données BLDB contient massivement beaucoup de séquences redondantes.

En conséquence, l'annotation finale a été réalisée en utilisant la base de données BLDB dérépliquée à 95% avec une e-value seuil de  $1e^{-20}$  et une identité seuil de 95%. 105 séquences représentantes sont annotées par 98 séquences de la base de données. Cependant, hormis les 3 séquences de PBPs, 18 séquences BLDB sont absentes de l'arbre : OXA-18, OXA-45, OXA-46, OXA-53, OXA-63, OXA-158, OXA-159, OXA-182, OXA-191, OXA-198, OXA-209, OXA-361, OXA-372, OXA-527, OXA-548, OXA-549, OXA-550 et OXA-568.

Pour déterminer la raison de ces absences, nous avons réalisé un quick BLASTp de ces 18

séquences contre la banque **nr** sur le site du NCBI. Nous avons observé que ces séquences ne sont pas retrouvées dans l'arbre car les génomes dont proviennent ces séquences de référence ne sont pas présents dans notre base de données **prok\_batch\_db**. De plus, le seuil d'identité de 95% ne leur permet pas d'annoter une séquence de nos. Par exemple, les deux séquences OXA-158 et OXA-159 proviennent de deux souches de *Pandoraea* sp., Va8523 et HD7676 respectivement. Pour ces deux souches, seules deux séquences protéiques, dont la séquences d'OXA, sont disponibles sur le site du NCBI (et donc aucun génome complet). Ces séquences ne pouvaient donc pas être présentes dans notre base de données car elle ne contient que des génomes complets. OXA-158 et OXA-159 sont similaires à 99%, lors du BLASTp, ces deux OXAs se retrouvent elles-mêmes, mais la troisième séquence récupérée par BLASTp est OXA-156, qui ne leur est similaire qu'à 88%. Etant assez divergentes, ces OXAs ne peuvent donc annoter aucune des séquences présentes dans notre arbre lorsque le seuil d'identité est fixé à 95%. Ces deux explications sont applicables aux 16 autres séquences.

## 4.6 Phylogénie des OXAs putatives

Au sein de l'arbre phylogénétique final, nous avons défini différents groupes tenant compte de plusieurs critères. Ainsi, le groupe doit être, si possible, monophylétique et contenir au moins 5 séquences (en tenant compte des représentants et des membres). De plus, il doit idéalement être soutenu par une branche ayant une valeur de support bootstrap élevé. Durant ce projet, une nouvelle méthode d'optimisation de bootstrap est parue (Lemoine et al. 2018). Cette méthode a été appliquée à notre arbre, ce qui a permis de consolider les nos valeurs de bootstrap et la définition de nos groupes d'OXAs. Suivant ces critères, 31 groupes ont été définis. Cependant, 19 séquences représentantes n'ont pu être intégrées à aucun groupe (figure 17).

Pour vérifier la robustesse des groupes définis, nous avons calculé un arbre phylogénétique excluant les 19 séquences difficiles à positionner et utilisant les mêmes options que l'arbre précédent. Nous avons observé que les groupes 1- 3, 8-15, 17-26, 28-29 et 31 restent identiques, tandis que les autres groupes éclatent. Ainsi, les groupe 16 et 30 deviennent paraphylétiques, les groupes 4, 6, 7, 27 polyphylétiques et le groupe 5 perd une séquence. Les résultats obtenus plus bas pour ces sept groupes seront donc à confirmer (figure 18).

On retrouve des séquences appartenant à des *Proteobacteria* et des *Firmicutes* dans 27 groupes. Les quatres groupes restants sont composés de séquences appartenant à des *Bacteroidetes* (groupes 17, 18 et 19) et des *Leptospira* (groupe 25). 15 groupes sont composés à 100% de *Proteobacteria* et deux groupes à 100% de *Firmicutes*. Le groupe 24 contient les trois séquences de BlaR1 et la majorité des séquences d'OXA brutes (10 534 séquences). Sur base de la séquences en AA, toute les séquences de ce groupe semblent être des BlaR1 putatifs. Les OXAs de la base de donnée BLDB sont réparties en 15 groupes, dont le groupe 3 qui en contient 341 (55% de la base de données) (tableau 1).

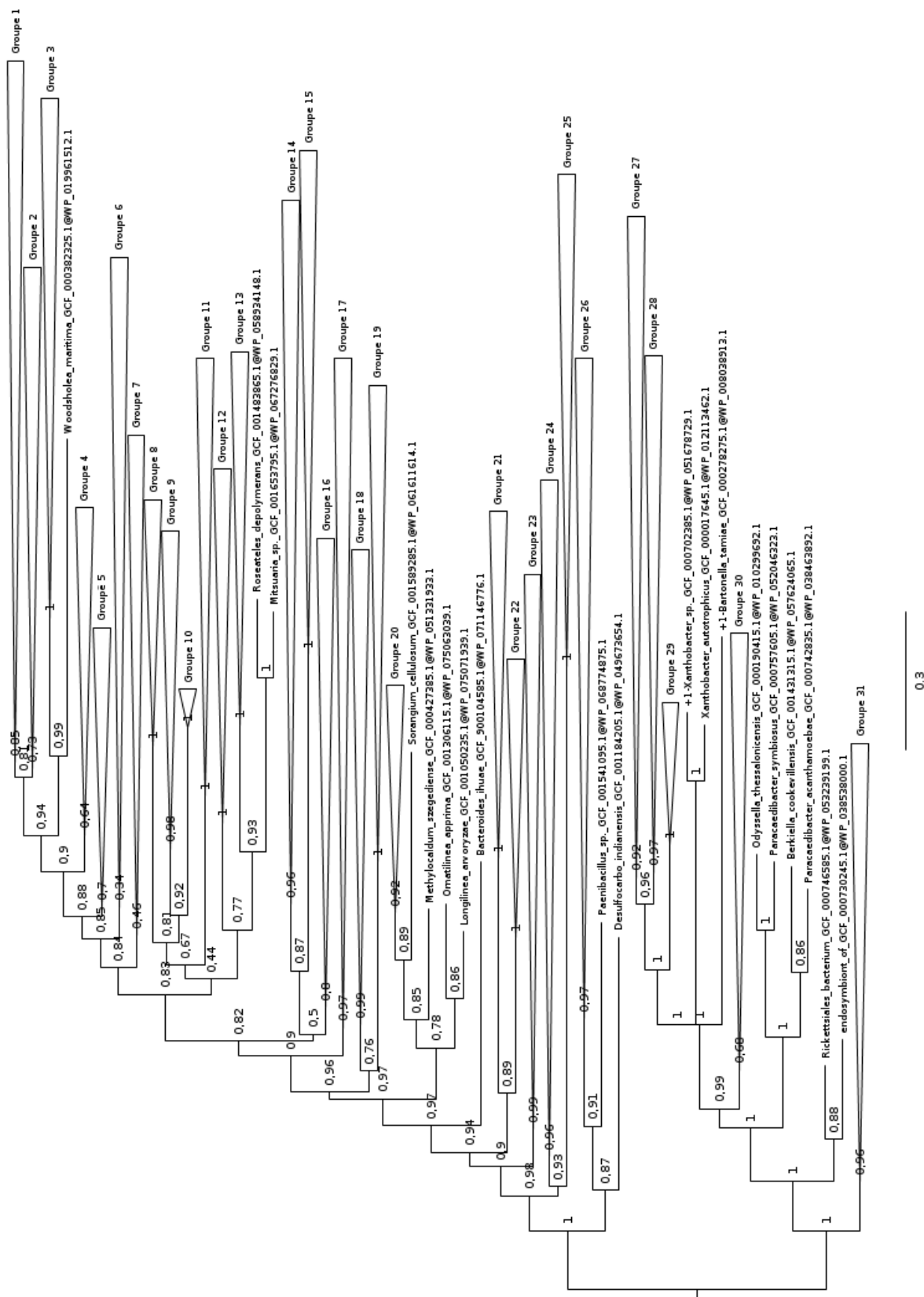


FIG. 17 : Arbre phylogénétique des OXAs putatives. Les groupes définis sont représentés par des triangles blancs et les séquences ne faisant pas partie d'un groupe apparaissent en noir. Ces séquences sont représentées par un nom de genre et d'espèce, suivi d'un numéro d'accèsion unique. Un tag précède une séquence si elle est représentante d'un cluster.

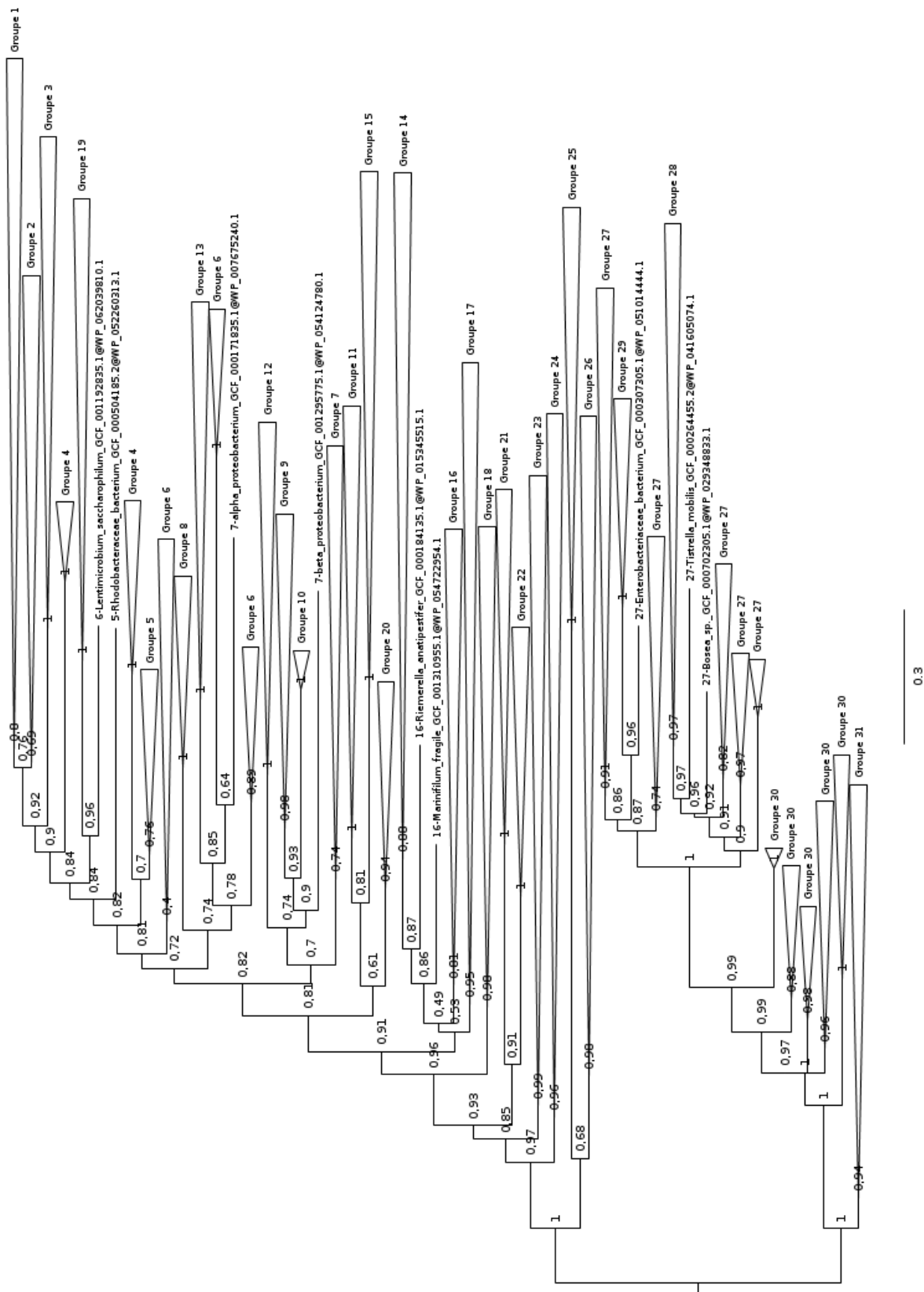


FIG. 18 : Arbre phylogénétique des OXAs putatives calculé en excluant les 19 séquences difficiles à positionner. Les groupes sont représentés par des triangles blancs. Les séquences des groupes qui éclatent sont représentées par un nom de genre et d'espèce, suivi d'un numéro d'accèsion unique. Le tag qui précède ces séquences indique le numéro du groupe auquel elles appartiennent.

En considérant les 24 916 OXAs et BlaR1 putatifs, nous avons observé que 77,8% des organismes possèdent une seule OXA/BlaR1, tandis que 21,8% en possèdent deux et 0,4% en possèdent trois ou quatre. Nous avons ensuite analysé les 4505 organismes possédant au moins deux séquences OXA et/ou BlaR1. Nous avons ainsi noté que 1449 organismes possèdent plusieurs OXAs appartenant à un même des 31 groupes, principalement dans les groupes 3 et 30 (respectivement 1208 et 161 organismes), tandis que 1847 organismes du groupe 24 présentent plusieurs BlaR1 putatifs. Par contre, les 1209 organismes restants possèdent des séquences appartenant à au moins deux groupes. Ainsi, 269 organismes ont plusieurs OXAs putatives et 940 organismes ont simultanément un BlaR1 et une OXA putative (groupes 21-24 : 904 organismes ; groupes 23-24 : 35 organismes ; groupes 17-24 : un organisme). Le groupe 21 ne contient aucune séquence de référence de la base de données BLDB et est composé à 99% de *Clostridiales*. Le groupe 23 est composé à 97% de séquences provenant de *Bacillales* et on y retrouve neuf séquences de référence BLBD. Ces neuf séquences correspondent aux OXAs retrouvées chez les bactéries monodermes du phylum des *Firmicutes* (Toth et al. 2016). Quant au groupe 17, il est composé à 96% de *Bacteroidetes* et ne contient aucune séquence de référence. Enfin, 11 groupes contiennent des organismes ne possédant qu'une seule OXA putative.

#### 4.7 Conservation des résidus du site actif

Pour chacun des groupes définis précédemment, nous avons analysé la conservation des résidus au niveau du site actif. Nous avons observé que dans tous les groupes, on retrouve des séquences présentant les motifs SxxK en positions 68 à 70, SxV en positions 115 à 117, W en position 153 et KTG en positions 204 à 206 (x représente n'importe quel AA ; la numérotation correspond à celle d'OXA-10). Ces motifs sont les plus importants au niveau du site actif et témoignent d'une activité  $\beta$ -lactamase potentielle. Cependant, on retrouve des séquences, notamment dans le groupe 26, où la sérine active du motif SxxK est mutée en alanine (A). Le groupe 25 semble être le plus divergent et possède des motifs SxxH, SxH et AxG, quoique le tryptophane (W) soit toujours conservé. Nous avons également observé que le motif Y/FxN en positions 141 à 143, absent de la poche du site actif, est conservé dans tous les groupes à l'exception du groupe 25, où seule la phénylalanine (F) est conservée (voir annexe).

Les carbapénèmes sont des antibiotiques utilisés en dernier recours lorsqu'un patient est infecté par une bactérie produisant des  $\beta$ -lactamases. De ce fait, les  $\beta$ -lactamases à activité carbapénémase sont de véritables fléaux en milieu hospitalier. Les carbapénémases de la classe A posent peu de problème car elles sont inhibées par des inhibiteurs de  $\beta$ -lactamases (comme l'acide clavulanique), mais ce n'est pas le cas des OXAs à activité carbapénémase. De plus, il est difficile de caractériser la mutation qui induit cette activité. Pour essayer de déterminer quel(s) résidu(s) pourrai(en)t être impliqué(s) dans cette activité carbapénémase, nous avons comparé les séquences de quatre groupes : les groupes 3 et 4, dont toutes les séquences de référence

possèdent une activité carbapénémase, et les groupes 1 et 12, dont les séquences de référence ne possèdent pas d'activité carbapénémase (figure 19). Hélas, nous n'avons observé aucune différence frappante entre les deux paires de séquences niveau des résidus du site actif.

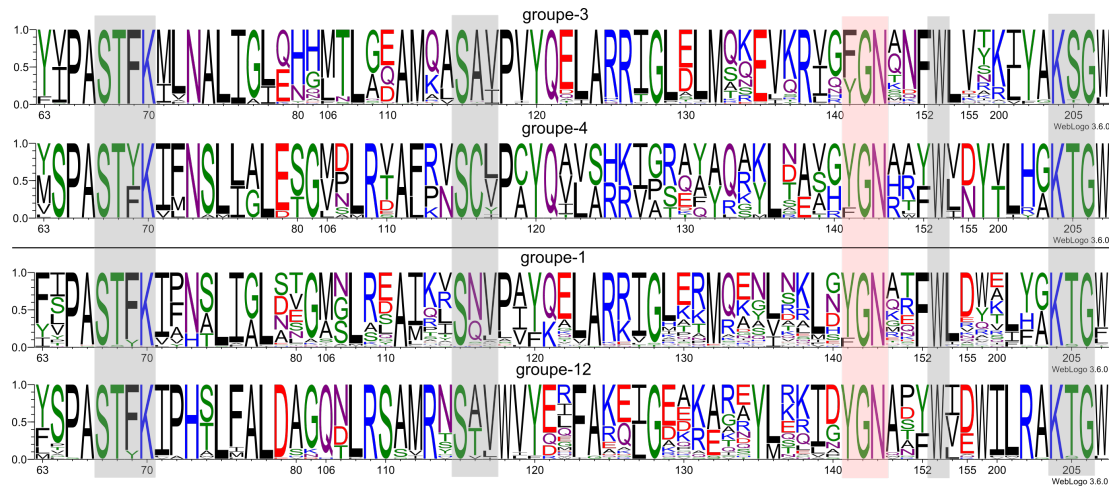


FIG. 19 : Comparaison de séquences entre deux groupes (3 et 4) possédant des séquences d'OXA de référence avec une activité carbapénémase et deux groupes (1 et 12) ne possédant pas de séquences d'OXA avec une activité carbapénémase. Les résidus conservés du site actif sont encadrés en gris et les résidus conservés hors du site actif sont encadrés en rose.

TAB. 1 : Résumé du nombre de séquences d'OXA représentantes et du nombre de séquences total par groupe, ainsi que de la taxonomie des organismes auxquels appartiennent ces séquences. La colonne **BLDB** donne le nombre de séquences de référence de la base de données BLDB que l'on retrouve dans chaque groupe. L'étoile rouge désigne les groupes où l'on retrouve des organismes possédant plusieurs séquences putatives d'OXA au sein d'un même groupe, tandis que l'étoile verte pointe les groupes où l'on retrouve des organismes possédant des séquences putatives réparties en au moins deux groupes. Le point (a) représente les groupes devenus paraphylétiques après calcul de l'arbre contenant uniquement les 31 groupes, le point (b) représente les groupes devenus polyphylétiques et le point (c) les groupes ayant perdu une séquence.

Groupe	Séquences représentantes		Séquences totales		BLDB
	nombre	taxonomie	nombre	taxonomie	
1 <sup>**</sup>	70	<i>Proteobacteria</i> : 87% <i>Verrucomicrobia</i> : 5% <i>Cyanobacteria</i> : 3% <i>Fusobacteria</i> : 3% <i>Spirochaetes</i> : 2%	3628	<i>Proteobacteria</i> : 98% <i>Verrucomicrobia</i> : 0,5% <i>Cyanobacteria</i> : 0,5% <i>Fusobacteria</i> : 0,5% <i>Spirochaetes</i> : 0,5%	40
2 <sup>*</sup>	48	<i>Actinobacteria</i> : 70% <i>Proteobacteria</i> : 30%	74	<i>Actinobacteria</i> : 59% <i>Proteobacteria</i> : 41%	5
3 <sup>**</sup>	66	<i>Proteobacteria</i> : 100%	3472	<i>Proteobacteria</i> : 100%	341
4(b)	21	<i>Proteobacteria</i> : 100%	43	<i>Proteobacteria</i> : 100%	8
5(c)	6	<i>Proteobacteria</i> : 100%	7	<i>Proteobacteria</i> : 100%	
6(b)	9	<i>Proteobacteria</i> : 89% <i>Bacteroidetes</i> : 11%	9	<i>Proteobacteria</i> : 89% <i>Bacteroidetes</i> : 11%	
7(b)	6	<i>Proteobacteria</i> : 100%	6	<i>Proteobacteria</i> : 100%	
8	55	<i>Proteobacteria</i> : 100%	141	<i>Proteobacteria</i> : 100%	
9 <sup>*</sup>	16	<i>Proteobacteria</i> : 100%	142	<i>Proteobacteria</i> : 100%	95
10 <sup>*</sup>	3	<i>Proteobacteria</i> : 100%	121	<i>Proteobacteria</i> : 100%	27
11 <sup>*</sup>	17	<i>Proteobacteria</i> : 100%	30	<i>Proteobacteria</i> : 100%	
12 <sup>*</sup>	43	<i>Proteobacteria</i> : 90% <i>Cyanobacteria</i> : 10%	217	<i>Proteobacteria</i> : 94% <i>Cyanobacteria</i> : 6%	22
13	16	<i>Proteobacteria</i> : 100%	18	<i>Proteobacteria</i> : 100%	
14 <sup>**</sup>	58	<i>Proteobacteria</i> : 92% <i>Chlorobi</i> : 6% <i>Firmicute</i> : 2%	98	<i>Proteobacteria</i> : 95% <i>Chlorobi</i> : 4% <i>Firmicute</i> : 1%	5
15	8	<i>Proteobacteria</i> : 100%	8	<i>Proteobacteria</i> : 100%	
16 <sup>*(a)</sup>	56	<i>Cyanobacteria</i> : 84% <i>Bacteroidetes</i> : 7% <i>Proteobacteria</i> : 5% <i>Spirochaetes</i> : 2% <i>Verrucomicrobia</i> : 2%	85	<i>Cyanobacteria</i> : 77% <i>Bacteroidetes</i> : 14% <i>Proteobacteria</i> : 4% <i>Spirochaetes</i> : 3% <i>Verrucomicrobia</i> : 1%	
17 <sup>**</sup>	65	<i>Bacteroidetes</i> : 96% <i>Fusobacteria</i> : 2% <i>Balneolaeota</i> : 2%	82	<i>Bacteroidetes</i> : 96% <i>Fusobacteria</i> : 2% <i>Balneolaeota</i> : 2%	
18 <sup>*</sup>	52	<i>Bacteroidetes</i> : 100%	90	<i>Bacteroidetes</i> : 100%	1
19 <sup>*</sup>	23	<i>Bacteroidetes</i> : 100%	25	<i>Bacteroidetes</i> : 100%	
20	10	<i>Proteobacteria</i> : 100%	11	<i>Proteobacteria</i> : 100%	
21 <sup>*</sup>	38	<i>Firmicutes</i> : 100%	956	<i>Firmicutes</i> : 100%	
22	9	<i>Chlorobi</i> : 78% <i>Proteobacteria</i> : 11% <i>Verrucomicrobia</i> : 11%	9	<i>Chlorobi</i> : 78% <i>Proteobacteria</i> : 11% <i>Verrucomicrobia</i> : 11%	
23 <sup>**</sup>	83	<i>Firmicutes</i> : 100%	374	<i>Firmicutes</i> : 100%	9
24 <sup>**</sup>	223	<i>Firmicutes</i> : 99% <i>Proteobacteria</i> : 0,5% <i>Fusobacteria</i> : 0,5%	10534	<i>Firmicutes</i> : 99,91% <i>Proteobacteria</i> : 0,07% <i>Fusobacteria</i> : 0,02%	3
25	12	<i>Leptospira</i> : 100%	416	<i>Leptospira</i> : 100%	
26	12	<i>Proteobacteria</i> : 100%	22	<i>Proteobacteria</i> : 100%	
27 <sup>**</sup> (b)	195	<i>Proteobacteria</i> : 100%	1768	<i>Proteobacteria</i> : 100%	24
28 <sup>**</sup>	98	<i>Proteobacteria</i> : 98% <i>PVC_group</i> : 2%	719	<i>Proteobacteria</i> : 99% <i>PVC_group</i> : 1%	1
29 <sup>*</sup>	16	<i>Proteobacteria</i> : 100%	164	<i>Proteobacteria</i> : 100%	5
30 <sup>**</sup> (a)	40	<i>Proteobacteria</i> : 93% <i>Chlamydiae</i> : 7%	761	<i>Proteobacteria</i> : 93% <i>Chlamydiae</i> : 7%	1
31 <sup>*</sup>	20	<i>Proteobacteria</i> : 100%	865	<i>Proteobacteria</i> : 100%	8



## 5 Discussion

### 5.1 Recherche de $\beta$ -lactamases putatives

Lors de ce travail, nous avons construit différents profils HMM en fonction des différentes classes de  $\beta$ -lactamases. Ces profils nous ont permis d'identifier des milliers de séquences homologues putatives. Nous avons dû réduire la quantité de séquences en sélectionnant manuellement une population intéressante sur les graphiques produits par `ompa-pa.pl` (figure 14). Nos critères de sélection étaient : un nuage bien défini, des longueurs de séquences similaires aux séquences de référence et une e-value faible. La sélection des séquences pour les classe B et C a été plutôt simple (figures 14b et 14c) car les nuages de points correspondants étaient assez circonscrits. Par contre, elle était moins évidente pour les profils A, D et A+C+D (figures 14a, 14d et 14e). Pour éviter de perdre des séquences potentiellement homologues, nous avons décidé de sélectionner des séquences en dehors des nuages principaux en prenant le risque de récupérer des enzymes qui n'étaient pas des  $\beta$ -lactamases.

Nous avons observé que trois séquences sont identifiées à la fois par les profils C et D. Elles semblent donc présenter des caractéristiques des deux classes. Les séquences portent les identifiants WP\_056397197.1, WP\_071849985.1 et WP\_073220251.1. Sur la base de données du NCBI, la première séquence est identifiée comme classe D appartenant à la souche Root418 de *Massilia* sp., la seconde comme classe C de la souche HH105 de *Duganella* sp. et la troisième comme classe C de la souche CF038 de *Massilia* sp. Ces organismes sont proches et appartiennent à la famille des *Oxalobacteraceae* (*Betaproteobacteria*). Toutefois, ces trois séquences contiennent 600 AA, ce qui est près du double d'une OXA classique. Par ailleurs, le NCBI inclut une séquence métagénomique (WP\_063839877.1) annotée "bifunctional class C beta-lactamase/class D beta-lactamase fusion protein" et qui compte 609 AA. Dès lors, l'explication la plus simple est que nos trois protéines sont des fusions d'une enzyme de classe C et d'une OXA de classe D. Et en effet, une recherche BLASTp sur le site du NCBI à partir de la séquence environnementale retrouve bien ces trois protéines à côté de quatre autres séquences de *Massilia* sp. et une de *Janthinobacterium* sp.

### 5.2 Distribution taxonomique des OXAs putatives

Lorsque nous avons analysé la taxonomie de notre base de données de génomes complets (`prok_batch_db`), nous avons observé qu'elle était composée majoritairement de génomes provenant des *Firmicutes* et des *Proteobacteria*. Par conséquent, la majorité des séquences identifiées à l'aide du profil de la classe D appartiennent principalement à ces deux phyla (les séquences des *Firmicutes* sont majoritairement des BlaR1). Comme la recherche de séquences putatives dépend des génomes séquencés et présents dans notre base de données, nous n'avons probablement capturé qu'une infime partie de la diversité des OXAs qui existent réellement

dans la nature. L'étape suivante serait de rechercher des OXAs putatives dans les riches données métagénomiques.

### 5.3 Déréplication des OXAs putatives et de référence

`cd-hit` nous a permis de réduire le nombre de séquences putatives d'OXA afin de traiter plus facilement les résultats. Cependant, le programme ne tient compte que de la similarité entre séquences lors de la déréplication. Certains de nos clusters contiennent dès lors des séquences provenant d'organismes différents mais phylogénétiquement proches. Néanmoins, il se pourrait que le programme rassemble des séquences de phyla différents au sein d'un même cluster, ce qui fausserait les interprétations taxonomiques de l'arbre phylogénétique de la famille. Dans de futures expériences, il faudrait plutôt utiliser un programme capable de dérépliquer les séquences tout en tenant compte de la taxonomie des organismes les contenant. Ce type de programme n'existe pas encore, mais des doctorants du laboratoire y travaillent.

Lors de l'annotation des séquences putatives, nous avons adopté une stratégie qui reprend toutes les séquences d'OXA de référence qui correspondent à une séquence inconnue (à une identité seuil d'au moins 95%) afin de déterminer quelles séquences de la base de données BLDB étaient présentes dans notre arbre phylogénétique. Nous avons observé que de nombreuses séquences putatives étaient annotées par les mêmes séquences de référence, mais aussi que plusieurs séquences de référence pouvaient annoter la même séquence putative. Ceci indique qu'au-delà de la redondance notée dans les séquences putatives, il y a également énormément de redondance dans les séquences de référence. Cela plaide pour une approche dérépliquative plus agressive des séquences de référence, mais tenant compte de leur taxonomie, comme pour les séquences putatives (voir section précédente).

### 5.4 Phylogénie des OXAs putatives

Lorsque nous avons inféré un second arbre phylogénétique avec uniquement les séquences contenues dans les 31 groupes définis, nous avons observé que sept groupes n'étaient pas stables. Il faudrait donc inférer de nouveaux arbres avec d'autres modèles d'évolution des séquences afin de vérifier si ces groupes n'éclatent plus dans l'arbre et si les 31 groupes restent stables. Sinon, la définition de nos groupes devra être revue pour mieux correspondre à la réalité.

Durant nos analyses, nous avons remarqué qu'environ 80% des organismes ne possèdent qu'une seule OXA ou un BlaR1, laissant supposer que la totalité des OXAs ne sont que des orthologues d'un seul gène ancestral. Cependant, dans certains cas, des organismes portent plusieurs OXAs et/ou BlaR1. Environ 25% des organismes possédant un BlaR1 (groupe 24) possèdent également soit une OXA soit un autre BlaR1. Quand les organismes ont deux OXAs, celles-ci peuvent appartenir au même groupe (duplications assez récentes) ou se retrouver dans deux groupes

différents (duplications plus anciennes avec pertes secondaires ou transfert de gène). Dès lors, il faudrait examiner ces duplications afin d'évaluer si elles ont eu lieu avant ou après spéciation des organismes en question. Il serait également intéressant d'étudier les transferts des gènes d'OXA.

## 5.5 Classification des OXAs putatives

Nous avons comme objectif de donner de nouveaux noms aux groupes d'OXA définis sur notre arbre. Nous avons observé que l'on retrouve des séquences de référence dans 16 de ces groupes, dont un qui ne contient que des séquences similaires à BlaR1. Néanmoins, comment étendre l'annotation des séquences de référence aux autres séquences putatives ? Est-ce que cette annotation dépend de la taxonomie ? Quelle annotation donner aux groupes ne contenant aucune séquence de référence ? Par exemple, le groupe 3 est composé de 66 séquences représentantes, dont 61 appartiennent à des organismes du genre *Acinetobacter* (*Gammaproteobacteria*). Les autres séquences du groupe appartiennent également à des organismes du phylum des *Proteobacteria* : trois *Gammaproteobacteria*, une *Alphaproteobacteria* et une *Betaproteobacteria*. Les OXAs de ce groupe pourraient donc être annotées "OXA d'*Acinetobacter*", dont certaines potentiellement transférées vers d'autres *Proteobacteria*. On le voit, il n'est pas aisé de nommer ces groupes au-delà d'une annotation purement taxonomique d'un gène majoritairement orthologue chez la plupart des bactéries où il est retrouvé. C'est pourquoi, à ce stade, nous n'avons transformé aucun de nos groupes en sous-famille.

## 5.6 Conservation des résidus du site actif

Lorsque nous avons analysé les résidus du site actif conservés dans chaque groupe, nous avons remarqué que les motifs caractéristiques des  $\beta$ -lactamases étaient bien présents. Dans certains cas, on observe la mutation de la sérine active en alanine (ex. groupe 26). Les enzymes possédant cette mutation sont potentiellement inactives. Ces résultats indiquent que les bactéries possédant ces OXAs putatives sont des réservoirs potentiels de  $\beta$ -lactamases de classe D encore inconnues qui pourraient être transférées vers des bactéries pathogènes pour l'homme. Nous avons également observé que le motif Y/FxN, hors du site actif, est conservé à travers tous les groupes. Au vu de cette conservation, ce motif doit avoir une fonction encore à découvrir au sein de l'enzyme.

## 5.7 Comparaison avec l'article de Keshri *et al.* (2018)

Durant la rédaction de ce mémoire est paru un article utilisant les mêmes approches que nous mais sur les quatre classes de  $\beta$ -lactamases (Keshri et al. 2018). Les auteurs ont recherché des séquences putatives dans quatre bases de données : environmental protein sequence database, the human microbiome project reference genome, the human microbiome project metagenomic database et the NCBI nonredundant (nr) database. Les séquences de référence de  $\beta$ -lactamases

ont pour leur part été obtenues à partir de la base de données ARG-ANNOT (contenant moins de séquences de référence que les bases de données BLDB et NCBI). Un profil HMM a été produit à partir de 210 séquences de la classe D. Ce profil leur a permis d'identifier 4127 séquences putatives (non-dérépliquées). Ils ont également identifié deux motifs conservés : SxxK et KxG.

Les résultats obtenus lors de notre projet sont différents de ceux discutés dans cet article. En effet, nous avons identifié environ 20 000 séquences putatives de plus en dépit du fait que notre profil a été généré avec moins de séquences (160 séquences obtenues en déréplicant les 470 séquences de la base de données du NCBI). Dans leur article, ils n'ont pas défini de nouvelles sous-familles et aucune caractérisation taxonomique n'a été faite. Néanmoins, nous avons obtenu des résultats similaires concernant les motifs conservés dans les séquences, et ce même si nous avons montré que d'autres motifs sont également conservés : SxV, W en position 153 et Y/FxN.

## 6 Conclusions

Au cours de ce mémoire, nous avons montré que des OXAs putatives sont présentes dans divers nouveaux phyla bactériens (ex. *Cyanobacteria* ou *Chlorobi*), mais sont absentes chez les *Thermotogae*, *Acidobacteria*, *Aquificae*, *Synergistetes*, *Nitrospirae*, *Thermodesulfobacteria*, *Deferribacteres*, *Elusimicrobia*, *Chrysiogenetes*, *Dictyoglomi*, *Caldiserica*, *Nitrospinae/Tectomicrobia*, *Calditrichaeota*. Les archées en semblent également dépourvues.

Nous avons observé qu'environ 80% des organismes ne possèdent qu'une seule OXA. La majorité des OXAs putatives présentent des motifs d'acides aminés typiques des  $\beta$ -lactamases : SxxK, SxV, W et KTG (où x représente n'importe quel AA). Nous avons également constaté que le motif Y/FxN est très conservé, mais sa fonction dans l'enzyme reste à élucider. Malgré l'étude des motifs, nous n'avons pas réussi à associer l'activité carbapénémase à un motif de séquence particulier.

L'inférence d'un arbre phylogénétique nous a permis de définir 31 groupes au sein des OXAs, dont sept semblent être moins stables. 15 groupes contiennent des OXAs connues et caractérisées, tandis qu'un groupe rassemble les récepteurs membranaire BlaR1 et 15 autres groupes ne contiennent aucune séquence connue. Nous avons caractérisé ces groupes d'un point de vue taxonomique et noté que 27 groupes sont surtout composées de séquences provenant de *Proteobacteria* et/ou de *Firmicutes*. Cependant, nous n'avons pas encore attribué de noms de sous-familles aux groupes car notre impression est que la protéine OXA est majoritairement encodée par un gène orthologue simple à travers les bactéries où elle est présente.

## 7 Annexes

### 7.1 Conservation des résidus du site actif

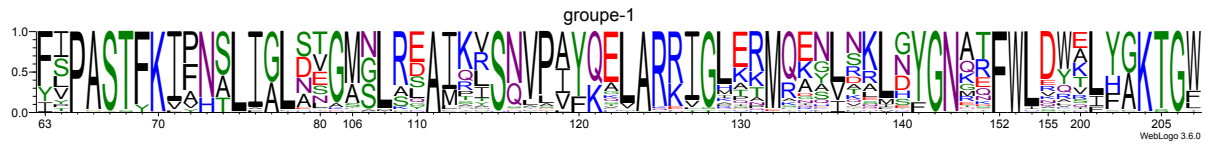


FIG. S.1 : Résidus du site actif du groupe 1 (la numérotation correspond à celle d'OXA-10)

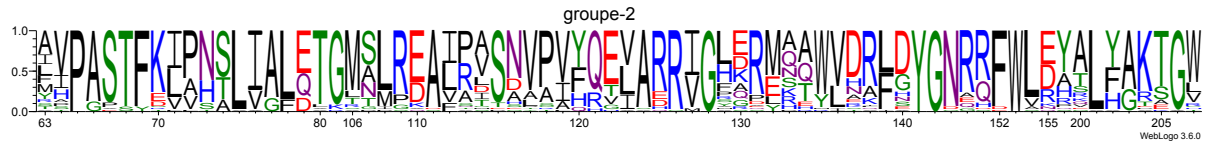


FIG. S.2 : Résidus du site actif du groupe 2 (la numérotation correspond à celle d'OXA-10)

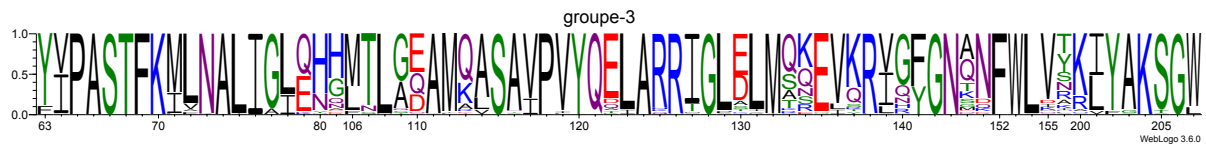


FIG. S.3 : Résidus du site actif du groupe 3 (la numérotation correspond à celle d'OXA-10)

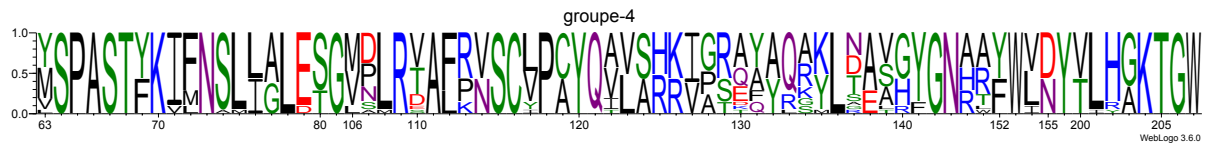


FIG. S.4 : Résidus du site actif du groupe 4 (la numérotation correspond à celle d'OXA-10)

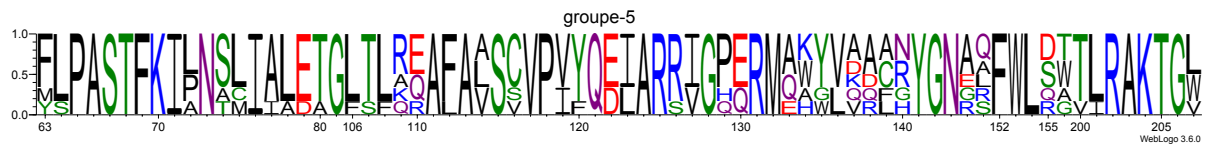


FIG. S.5 : Résidus du site actif du groupe 5 (la numérotation correspond à celle d'OXA-10)

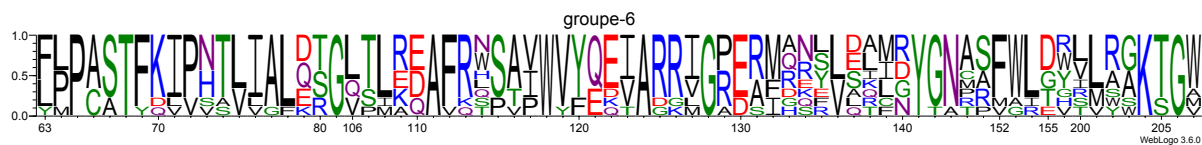


FIG. S.6 : Résidus du site actif du groupe 6 (la numérotation correspond à celle d'OXA-10)

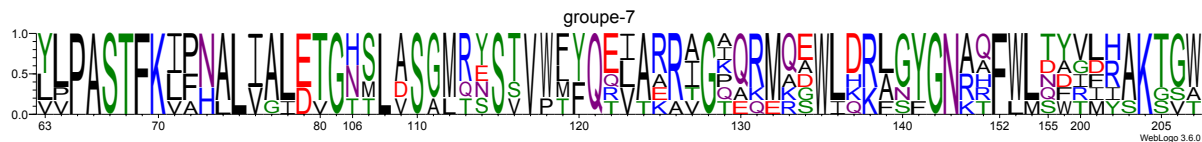


FIG. S.7 : Résidus du site actif du groupe 7 (la numérotation correspond à celle d'OXA-10)

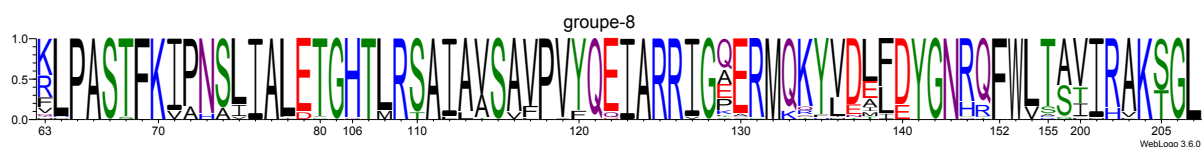


FIG. S.8 : Résidus du site actif du groupe 8 (la numérotation correspond à celle d'OXA-10)

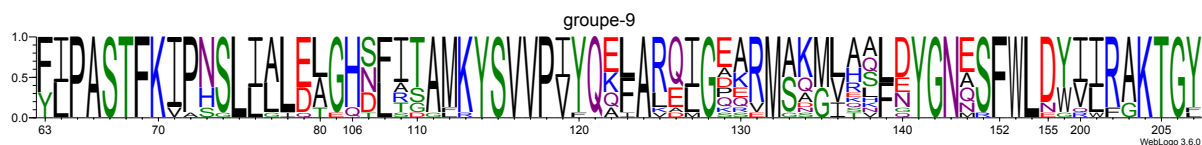


FIG. S.9 : Résidus du site actif du groupe 9 (la numérotation correspond à celle d'OXA-10)

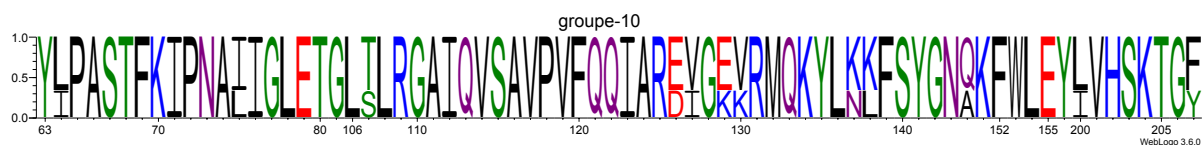


FIG. S.10 : Résidus du site actif du groupe 10 (la numérotation correspond à celle d'OXA-10)

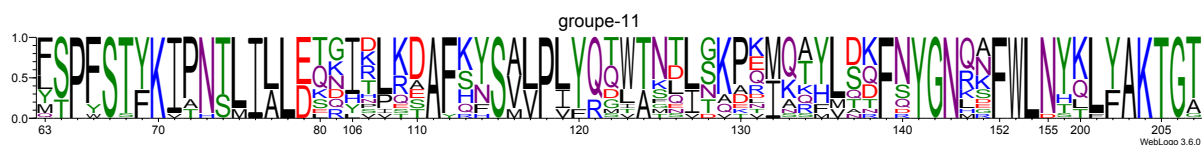


FIG. S.11 : Résidus du site actif du groupe 11 (la numérotation correspond à celle d'OXA-10)

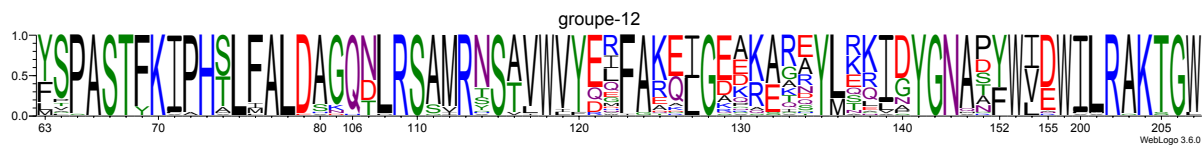


FIG. S.12 : Résidus du site actif du groupe 12 (la numérotation correspond à celle d'OXA-10)

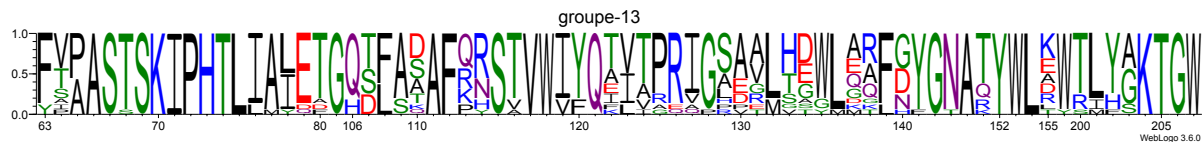


FIG. S.13 : Résidus du site actif du groupe 13 (la numérotation correspond à celle d'OXA-10)

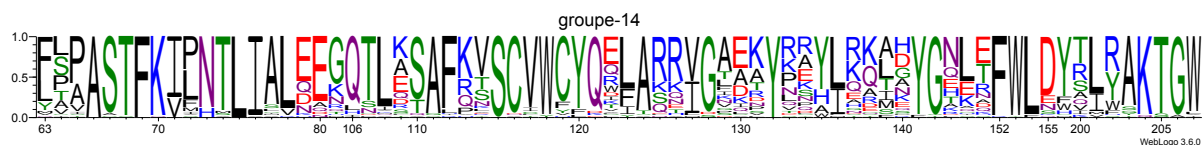


FIG. S.14 : Résidus du site actif du groupe 14 (la numérotation correspond à celle d'OXA-10)

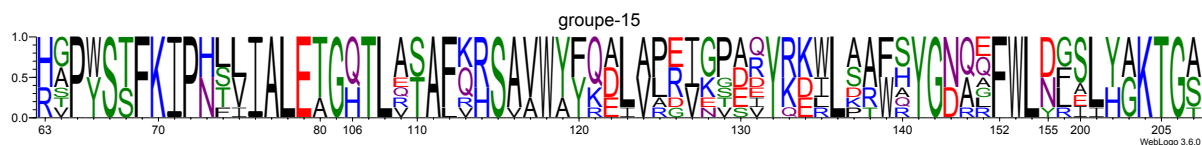


FIG. S.15 : Résidus du site actif du groupe 15 (la numérotation correspond à celle d'OXA-10)

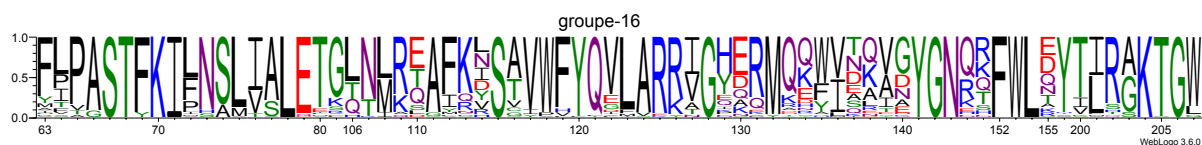


FIG. S.16 : Résidus du site actif du groupe 16 (la numérotation correspond à celle d'OXA-10)

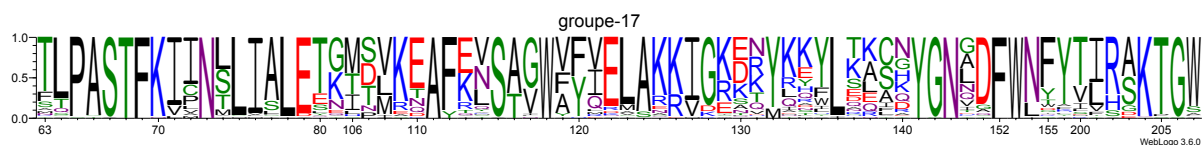


FIG. S.17 : Résidus du site actif du groupe 17 (la numérotation correspond à celle d'OXA-10)



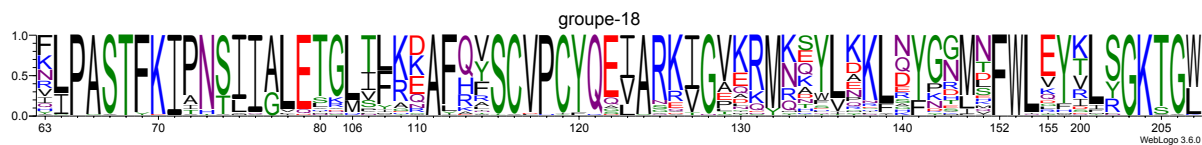


FIG. S.18 : Résidus du site actif du groupe 18 (la numérotation correspond à celle d'OXA-10)

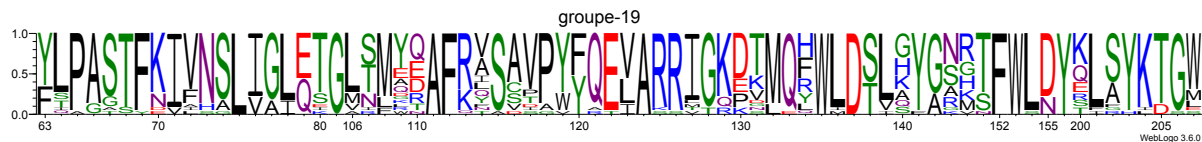


FIG. S.19 : Résidus du site actif du groupe 19 (la numérotation correspond à celle d'OXA-10)

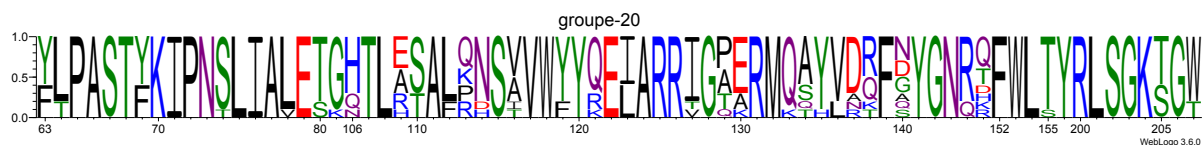


FIG. S.20 : Résidus du site actif du groupe 20 (la numérotation correspond à celle d'OXA-10)

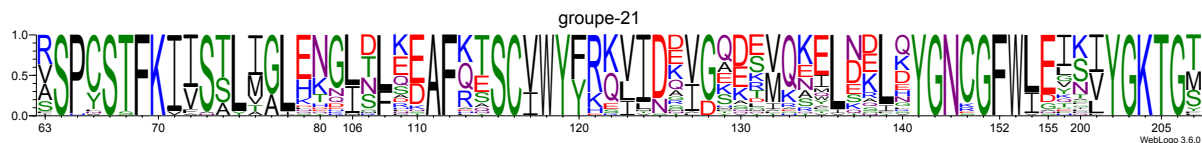


FIG. S.21 : Résidus du site actif du groupe 21 (la numérotation correspond à celle d'OXA-10)

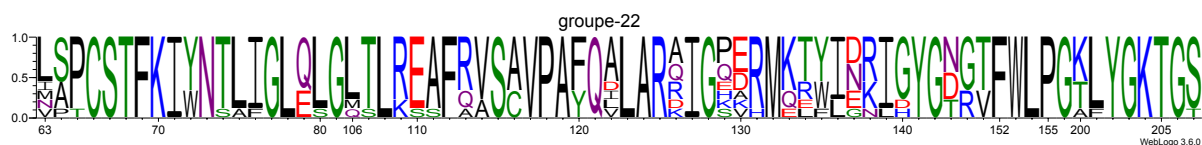


FIG. S.22 : Résidus du site actif du groupe 22 (la numérotation correspond à celle d'OXA-10)

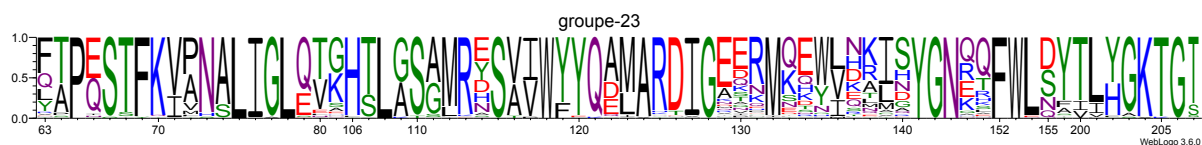


FIG. S.23 : Résidus du site actif du groupe 23 (la numérotation correspond à celle d'OXA-10)

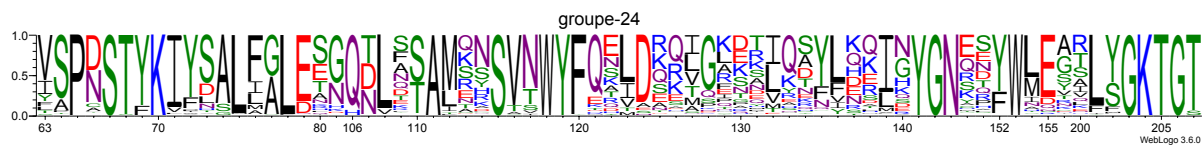


FIG. S.24 : Résidus du site actif du groupe 24 (la numérotation correspond à celle d'OXA-10)

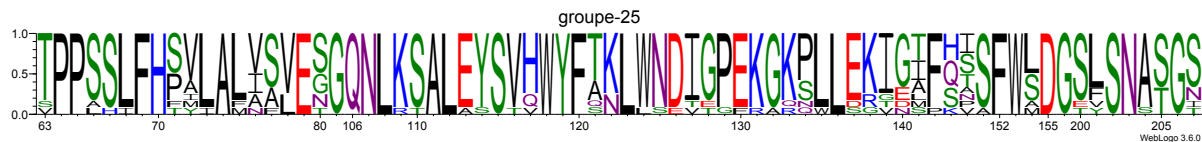


FIG. S.25 : Résidus du site actif du groupe 25 (la numérotation correspond à celle d'OXA-10)

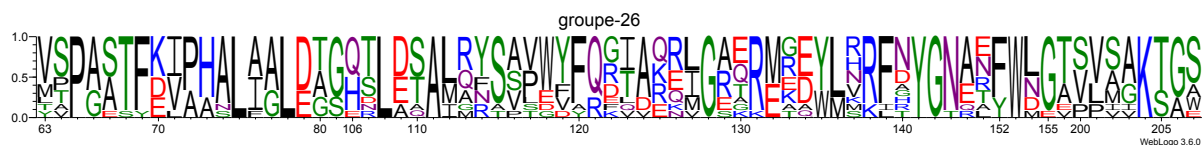


FIG. S.26 : Résidus du site actif du groupe 26 (la numérotation correspond à celle d'OXA-10)

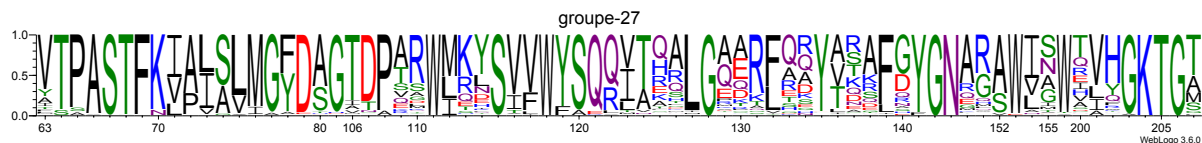


FIG. S.27 : Résidus du site actif du groupe 27 (la numérotation correspond à celle d'OXA-10)



FIG. S.28 : Résidus du site actif du groupe 28 (la numérotation correspond à celle d'OXA-10)

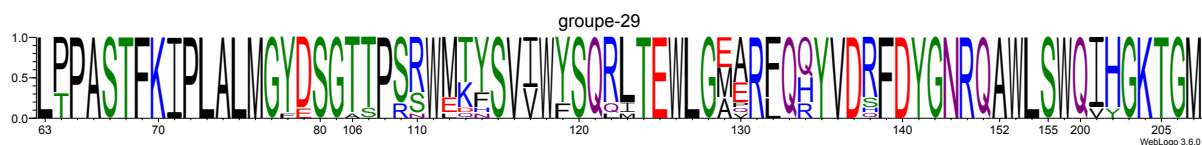


FIG. S.29 : Résidus du site actif du groupe 29 (la numérotation correspond à celle d'OXA-10)

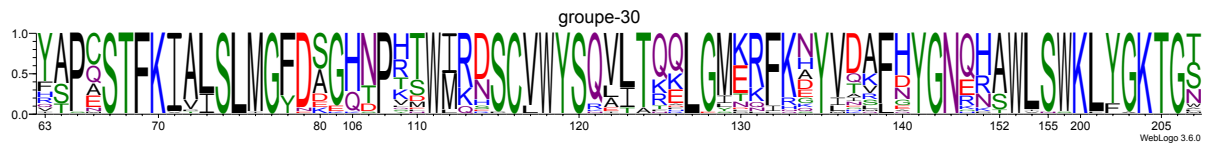


FIG. S.30 : Résidus du site actif du groupe 30 (la numérotation correspond à celle d'OXA-10)

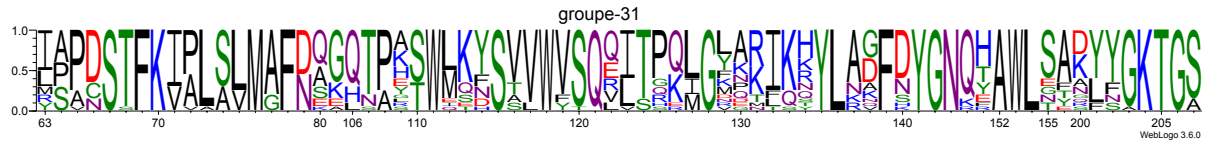


FIG. S.31 : Résidus du site actif du groupe 31 (la numérotation correspond à celle d'OXA-10)

## 7.2 Scripts Perl

Programme get\_accession\_num.pl

```

1  #!/usr/bin/env perl
2
3  use Modern ::Perl '2011';
4  use autodie;
5
6  use Getopt ::Euclid;
7  use Smart ::Comments '###';
8
9  use Tie ::IxHash;
10 use Const ::Fast;
11
12 my $seq_for = read_fasta( $ARGV{'<infile>'} );
13 my $classes_for = get_classes( $ARGV{'--tab'} );
14 ### $classes_for
15 my $class = $ARGV{'--class'};
16 $class = lc $class;
17
18 ### You choose : 'class ' . $class
19
20 # Create an outfile
21 my $outfile = $ARGV{'--out'};
22
23 open my $out, '>', $outfile;
24
25 SEQ :
26 while (my ($id, $prot) = each %$seq_for) {
27     for my $bla ( %{ $classes_for->{$class} } ){
28         if ($id =~ m/\b$bla\b/xms){
29             if ($ARGV{'--forann'}){
30                 my $bl = $classes_for->{$class}{$bla}{gene};
31                 my $fam = $classes_for->{$class}{$bla}{family} // 'nofamily';
32 #                 my ($bl, $fam) = @{ $classes_for->{$class}{$bla} }{ qw(gene family) };
33                 $id = $class . '|' . $bl . '|' . $fam . '|' . $bla . "\t" . $id;
34             }
35             say {$out} '>' . $id;
36             say {$out} _wrap_seq($prot);
37         next SEQ;
38     }
39 }
40 }
41
42 close $out;
43
44 ### Writing output file : $outfile;

```

```

45
46 sub read_fasta {
47     my $infile = shift;
48
49     open my $in, '<', $infile;
50
51     my $seq_id;
52     my $seq;
53     tie my %seq_for, 'Tie ::IxHash';
54
55     LINE :
56     while (my $line = <$in>){
57         chomp $line;
58
59         next LINE if $line =~ m/^\s*$/xms;      # skip empty line
60         next LINE if $line =~ m/^\s*#/xms;      # skip comment line
61
62         if ($line =~ m/^\>/xms){
63             if ($seq) {
64                 $seq_for{$seq_id}= $seq;
65                 $seq = q();
66             }
67             $seq_id = substr($line, 1);
68             next LINE;
69         }
70         $seq .= $line;
71     }
72
73     $seq_for{$seq_id}= $seq if $seq;
74
75     return \%seq_for;
76 }
77
78 sub get_classes {
79     my $infile = shift;
80
81     open my $in, '<', $infile;
82
83     my %classes_for;
84
85     LINE :
86     while (my $line = <$in>) {
87         chomp $line;
88
89         next LINE if $line =~ m/^\s*$/xms;
90         next LINE if $line =~ m/^\s*#/xms;
91
92         my ($bla, $id, $name) = split /\t/, $line;
93
94         if ($line =~ m/class\s[a,A]\s/xmsg){
95             my $class = 'a';
96             ($bla) = $bla =~ m/^bla(.+)\$/xms? $1 : $bla =~ m/^(.+)$/xms? $1 : "unnamed";
97             $bla =~ tr/-//d;
98             $classes_for{$class}{$id}{gene} = $bla;
99         }
100
101         if ($line =~ m/class\s[c,C]\s/xmsg){
102             my $class = 'c';
103             ($bla) = $bla =~ m/^bla(.+)\$/xms? $1 : $bla =~ m/^(.+)$/xms? $1 : "unnamed";
104             $bla =~ tr/-//d;
105             $classes_for{$class}{$id}{gene} = $bla;
106         }
107
108         if ($line =~ m/class\s[b,B]\s/xmsg || $line =~ m/metallo/xmsg){
109             my $class = 'b';
110             ($bla) = $bla =~ m/^bla(.+)\$/xms? $1 : $bla =~ m/^(.+)$/xms? $1 : "unnamed";
111             $bla =~ tr/-//d;
112             $classes_for{$class}{$id}{gene} = $bla;
113         }
114
115         if ($line =~ m/class\s[d,D]\s/xmsg || $line =~ m/OXA\-\d+/xmsg){
116             my $class = 'd';
117             ($bla) = $bla =~ m/^bla(.+)\$/xms? $1 : $bla =~ m/^(.+)$/xms? $1 : "unnamed";
118             $bla =~ tr/-//d;

```

```

119         my ($fam) = $name =~ m/^(OXA.+)\sfamily|-like/xms? $1
120             : $name =~ m/(\S+.spectrum)/xms           ? "ESBL"
121             : $name =~ m/(carbapenem)\S+/xms          ? $1
122             :                                         "nofamily"
123     ;
124     $fam =~ tr/-//d;
125     $classes_for{$class}{$id} = {
126         gene    => $bla,
127         family => $fam,
128     };
129 }
130 }
131
132 return \%classes_for;
133 }
134
135 sub _wrap_seq {
136     my $seq = shift;
137     const my $CHUNK_LEN => 70;
138
139     my $seq_len = length $seq;
140
141     # wrap seq in chunks of length $CHUNK_LEN
142     my $wrapped_seq;
143     for (my $i = 0; $i < $seq_len; $i += $CHUNK_LEN) {
144         $wrapped_seq .= substr($seq, $i, $CHUNK_LEN) . "\n";
145     }
146
147     return $wrapped_seq;
148 }
149
150 =head1 NAME
151
152 Extract protein sequence by extract id from a tabular file.
153
154 =head1 VERSION
155
156 This documentation refers to get_accession_num version 0.0.1
157
158 =head1 USAGE
159
160     get_accession_num.pl <infile> --tab <infile>... --class <name> [options]
161
162 =head1 REQUIRED ARGUMENTS
163
164 =over
165
166 =item <infile>
167
168 Path to input FASTA file.
169
170 =for Euclid :
171     infile.type : readable
172
173 =item --tab [=] <infile>
174
175 Path to a table database.
176
177 =for Euclid :
178     infile.type : readable
179
180 The database is a tab-delimited 3-column flat file structured as follows :
181
182     1. allele name
183     2. protein accession number
184     3. curated gene product name
185
186 Example :
187
188     ACC-1   ACD12694.1  cephalosporin-hydrolyzing class C beta-lactamase ACC-1
189     ACC-1a  AAF86699.1  cephalosporin-hydrolyzing class C beta-lactamase ACC-1a
190
191 =item --class [=] <name>
192

```

```

193 The name of the beta-lactamase class you want to extract (a, b, c or d).
194
195 =item --out <outfile>
196
197 Outfile FASTA name [default : no].
198
199 =for Euclid :
200     outfile.type : writable
201
202 =back
203
204 =head1 OPTIONS
205
206 =item --forann
207
208 Make file adapted for annotation : name of class | name of proteine | name of sub-family [default : no].
209
210 =over
211
212 =item --version
213
214 =item --usage
215
216 =item --help
217
218 =item --man
219
220 Print the usual program information
221
222 =back
223
224 =head1 AUTHOR
225
226 Val rian Lupo (valerian.lupo@student.ulg.ac.be)
227
228 =head1 BUGS
229
230 There are undoubtedly serious bugs lurking somewhere in this code.
231 Bug reports and other feedback are most welcome.
232
233 =head1 COPYRIGHT
234
235 Copyright (c) 2013, Val rian Lupo. All Rights Reserved.
236 This program is free software. It may be used, redistributed
237 and/or modified under the terms of the Perl Artistic License
238 (see http://www.perl.com/perl/misc/Artistic.html)

```

Programme ann-classd.pl

```

1
2  #!/usr/bin/env perl
3
4  use Modern ::Perl '2011';
5  use autodie;
6
7  use Getopt ::Euclid;
8  use Smart ::Comments '###';
9
10 use Tie ::IxHash;
11 use Const ::Fast;
12
13 my $seq_for = read_fasta( $ARGV{'<infile>'} );
14 my $ann_for = annotate( $ARGV{'--tab'} );
15 my $outfile = $ARGV{'--out'};
16
17 open my $out, '>', $outfile;
18
19 SEQ :
20 while (my ($id, $prot) = each %$seq_for) {      ### Elapsed time |===[%]
21     for my $bla ( %{$ann_for->{D}} ){
22         if ($id =~ m/\b$bla\b/xms){
23             my ($acquire, $fam, $gene, $res) = @{$ann_for->{D}}{$bla} ){ qw(acquire family gene resistance) };
24             say {$out} '>' . $gene . '|' . $fam . '|' . $res . '|' . $acquire . '|' . $bla . "\t" . $id;

```

```

25         say {$out} _wrap_seq($prot);
26     next SEQ;
27 }
28 }
29 }
30
31 ### writing outfile : $outfile
32
33 sub read_fasta {
34     my $infile = shift;
35
36     open my $in, '<', $infile;
37
38     my $seq_id;
39     my $seq;
40     tie my %seq_for, 'Tie ::IxHash';
41
42     LINE :
43     while (my $line = <$in>){
44         chomp $line;
45
46         next LINE if $line =~ m/^\s*$/xms;      # skip empty line
47         next LINE if $line =~ m/^\#/xms;        # skip comment line
48
49         if ($line =~ m/^\>/xms){
50             if ($seq) {
51                 $seq_for{$seq_id} = $seq;
52                 $seq = q();
53             }
54             $seq_id = substr($line, 1);
55             next LINE;
56         }
57         $seq .= $line;
58     }
59
60     $seq_for{$seq_id} = $seq if $seq;
61
62     return \%seq_for;
63 }
64
65 sub annotate {
66     my $infile = shift;
67
68     open my $in, '<', $infile;
69
70     my %ann_for;
71
72     LINE :
73     while (my $line = <$in>){
74
75         next LINE if $line =~ m/^\s*$/xms;
76
77         my ($class, $name, undef, $fam, $peptid, $bankid,
78             undef, undef, undef, undef, undef, $resis, $acquire) = split /\t/, $line;
79         $name =~ tr/-*/d;
80         ($fam) = $fam =~ m/^(w+|\w+|-)like/xms? $1
81             :
82             'nofamily'
83         ;
84         $fam =~ tr/-*/d;
85         ($resis) = $resis =~ m/Carbapenemase/xms? 'carba'
86             : $resis =~ m/ESBL/xms ? 'ESBL'
87             : $resis =~ m/[N,n]arrow/xms ? 'narrow'
88             :
89             'no'
90         ;
91         ($acquire) = $acquire =~ m/^(w+)/xms? $1
92             :
93             'no'
94         ;
95         $ann_for{$class}{$peptid} = {
96             gene => $name,
97             family => $fam,
98             resistance => $resis,
99             acquire => $acquire,
100         };
101         $ann_for{$class}{$bankid} = {

```

```

99         gene => $name,
100         family => $fam,
101         resistance => $resis,
102         acquire => $acquire,
103     };
104
105 }
106
107 return \%ann_for;
108 }
109
110 sub _wrap_seq {
111     my $seq = shift;
112     const my $CHUNK_LEN => 70;
113
114     my $seq_len = length $seq;
115
116     # wrap seq in chunks of length $CHUNK_LEN
117     my $wrapped_seq;
118     for (my $i = 0; $i < $seq_len; $i += $CHUNK_LEN) {
119         $wrapped_seq .= substr($seq, $i, $CHUNK_LEN) . "\n";
120     }
121
122     return $wrapped_seq;
123 }
124
125 =head1 NAME
126
127 Annotate class D beta-lactamase from the Beta-Lactamase Data Base.
128
129 =head1 VERSION
130
131 This documentation refers to get_accession_num version 0.0.1
132
133 =head1 USAGE
134
135     get_accession_num.pl <infile> --tab <infile>...
136
137 =head1 REQUIRED ARGUMENTS
138
139 =over
140
141 =item <infile>
142
143 Path to input FASTA file.
144
145 =for Euclid :
146     infile.type : readable
147
148 =item --tab [=] <infile>
149
150 Path to the table information from Beta-Lactamase Data Base.
151
152 =for Euclid :
153     infile.type : readable
154
155 =item --out <outfile>
156
157 Outfile FASTA name [default : no].
158
159 =for Euclid :
160     outfile.type : writable
161
162 =back
163
164 =head1 OPTIONS
165
166 =over
167
168 =item --version
169
170 =item --usage
171
172 =item --help

```



```

173
174 =item --man
175
176 Print the usual program information
177
178 =back
179
180 =head1 AUTHOR
181
182 Valérian Lupo (valerian.lupo@student.ulg.ac.be)
183
184 =head1 BUGS
185
186 There are undoubtedly serious bugs lurking somewhere in this code.
187 Bug reports and other feedback are most welcome.
188
189 =head1 COPYRIGHT
190
191 Copyright (c) 2013, Valérian Lupo. All Rights Reserved.
192 This program is free software. It may be used, redistributed
193 and/or modified under the terms of the Perl Artistic License
194 (see http://www.perl.com/perl/misc/Artistic.html)

```

Programme annotate.pl

```

1
2 #!/usr/bin/env perl
3
4 use Modern ::Perl '2011';
5 use autodie;
6
7 use Getopt ::Euclid;
8 use Smart ::Comments '###';
9 use Template;
10
11 use File ::Basename;
12 use Path ::Class 'file';
13
14 use IPC ::System ::Simple qw(system);
15 use List ::AllUtils 'any';
16
17 use Forem ::FastaFile qw( :io prefix_ids );
18 use Forem ::BlastTable 'get_parser';
19
20 ##### Arguments : %ARGV
21
22 ### Reading infile : $ARGV{'<infile>'}
23 my ($seq_file, $intype, $infile, $inid_for) = process_file($ARGV{'<infile>'});
24
25 ### Reading infile : $ARGV{'--ref-file'}
26 my ( undef, $reftype, $refile, $refid_for) = process_file($ARGV{'--ref-file'});
27
28 # determine blast program based on type combination
29 my %pgm_for = (
30     'prot :prot' => 'blastp',
31     'nucl :prot' => 'blastx',
32     'prot :nucl' => 'tblastn',
33     'nucl :nucl' => 'tblastx',
34 );
35 my $pgm = $pgm_for{ "$intype :$reftype" };
36
37 # define command template
38 my $report = "$infile.blast.fmt7";
39 my $template = <<'EOT';
40 makeblastdb -in [% refile %] -dbtype [% reftype %]
41 [% pgm %] -query [% infile %] -db [% refile %] -evalvalue [% E %] -outfmt 7 -out [% report %]
42 EOT
43
44 # build command
45 my %vars = (
46     refile => $refile,
47     reftype => $reftype,
48     pgm => $pgm,

```

```

49     infile => $infile,
50     E => $ARGV{'--evaluate'},
51     report => $report,
52 );
53 my $command;
54 my $tt = Template->new;
55 $tt->process(\$template, \%vars, \$command);
56 #### $command
57
58 ### Performing BLAST...
59 system($command);
60
61 ### Parsing BLAST report...
62 my $parser = get_parser($report);
63 my %ann_for;
64 my %hit_id_for;
65 my $curr_id = q{};
66
67 HSP :
68 while ( my $hsp_ref = $parser->() ) {
69     my ($qid, $hid, $evalue, $identity) = @{$hsp_ref}{qw(query_id hit_id evalue identity)};
70     next HSP if $evalue > $ARGV{'--evaluate'};      # skip weak hits
71     next HSP if $identity < $ARGV{'--identity'};     # skip weak identity
72
73     unless ( $ARGV{'--hitlist'} ) {
74         next HSP if $qid eq $curr_id;                # skip non-first hits
75         $curr_id = $qid;
76     }
77
78     # capture annotation bit in ref seq id using regex
79     my ($annotation) = $refid_for->{$hid} =~ $ARGV{'--ref-regex'};
80     $ann_for{ $inid_for->{$qid} } = $annotation;
81     # take all the hit for a given e-value and identity
82     push @{$hit_id_for{ $inid_for->{$qid} }}, $annotation;
83 }
84 #### Annotations : %ann_for
85
86 unless ( $ARGV{'--hitlist'} ){
87     say '# ' . join "\t#", qw(tag id);
88     for my $id (sort keys %ann_for) {
89         say join "\t", $ann_for{$id}, $id;
90     }
91 }
92
93 if ( $ARGV{'--hitlist'} ){
94     say '# ' . join "\t#", qw(tag ids);
95     for my $id (sort keys %hit_id_for){
96         say $id . "\t" . join "\n\t", @{$hit_id_for{$id}};
97     }
98 }
99
100 if ( $ARGV{'--write-ann-file'} ) {
101     my ($basename, $dir, $ext) = fileparse($ARGV{'<infile>'}, qr{\.[^.]*}xms);
102     my $outfile = file($dir, $basename . '_ann' . $ext);
103     ### Writing annotated file : $outfile->stringify
104     write_fasta( $outfile, prefix_ids($seq_for, \%ann_for) );
105 }
106
107 sub process_file {
108     my $infile = shift;
109
110     my $seq_for = read_fasta($infile);
111     my $type = ( any { m/[EFILPQ]/i } values %$seq_for )? 'prot' : 'nucl';
112     my ($tmpfile, $id_for) = write_tmp_fasta( {
113         map { $_ => $seq_for->{$_} =~ tr/--//dr } keys %$seq_for
114     } ); # degap seqs on the fly
115
116     return ($seq_for, $type, $tmpfile, $id_for);
117 }
118
119 =head1 NAME
120
121 annotate - Annotate a sequence file by similarity to reference sequences
122

```

```

123 =head1 VERSION
124
125 This documentation refers to annotate version 0.0.1
126
127 head1 USAGE
128
129     annotate.pl <infile> --ref-file <infile> --ref-regex <regex> [options]
130
131 =head1 REQUIRED ARGUMENTS
132
133 =over
134
135 =item <infile>
136
137 Path to input FASTA file.
138
139 =for Euclid :
140     infile.type : readable
141
142 =item --ref-file [=] <infile>
143
144 Path to reference FASTA file.
145
146 =for Euclid :
147     infile.type : readable
148
149 =item --ref-regex [=] <regex>
150
151 Regular expression for capturing the reference seq id part that has to be used
152 for annotating infile seq ids [default : none].
153
154 =for Euclid :
155     regex.type : string
156
157 =back
158
159 =head1 OPTIONS
160
161 =over
162
163 =item --evaluate [=] <float>
164
165 E-value threshold for annotating a sequence [default : float.default].
166
167 =for Euclid :
168     float.type : number
169     float.default : 1e-10
170
171 =item --identity [=] <number>
172
173 Identity threshold for annotating a sequence [default : number.default].
174
175 =for Euclid :
176     number.type : number
177     number.default : 0
178
179 =item --write-ann-file
180
181 Write an annotated version (with prefixed ids) of the infile [default : no];
182
183 =item --hitlist
184
185 Print a id-hit list - at a given e-value and identity - instead of annotation report [default : no];
186
187 =item --version
188
189 =item --usage
190
191 =item --help
192
193 =item --man
194
195 Print the usual program information
196

```

Programme create\_mapper.pl

```

1
2  #!/usr/bin/env perl
3
4  use Modern ::Perl '2011';
5  use autodie;
6
7  use Smart ::Comments '###';
8
9  use Bio ::MUST ::Core;
10 use aliased 'Bio ::MUST ::Core ::SeqId';
11 use Bio ::MUST ::Drivers ::CdHit;
12
13 use Path ::Class 'file';
14 use File ::Basename;
15
16 my $class = 'Bio ::MUST ::Drivers ::CdHit';
17 my $cdh = $class->new( seqs =>file( '../find_fin_ann/bla_d98.fasta'), cdhit_args => { -c => 0.95 } );
18
19 my $bldb = shift;
20 my $phydm = shift;
21
22 my %repr_for;
23 my %org_for;
24 my $ann_for = create_idm($bldb);
25 my $idm_for = create_idm($phydm);
26
27 # create hash from cd-hit report
28 for my $repr ( $cdh->all_cluster_names ){
29
30     my $seq_id = SeqId->new( full_id => $repr );
31     my $gca = $seq_id->gca;
32     my $acc = $seq_id->accession;
33     my $fullgca = $gca . "@" . $acc;
34
35     my @members = map { $_->full_id } @{ $cdh->seq_ids_for($repr) };
36     my $count = @members;
37
38 }
39
40 # create new idm
41 IDM :
42 for my $id (keys %repr_for){
43     my $count = $repr_for{$id};
44     my $ann = $ann_for->{$id}{ann} // 0;
45     my $seq = $idm_for->{$id}{ann} // 0;
46
47     if ($ann eq 0 || $seq eq 0){
48         next IDM
49     };
50
51     if ($ann && $count > 0){
52         $ann = "+$count|$ann";
53     }
54
55     unless ($ann){
56         $ann = $idm_for->{$id}{fullname};
57
58         if ($count > 0){
59             $ann = "+$count-$ann";
60         }
61     }
62     say join "\t", $ann, $seq
63 }
64
65 # create hash from annotate idm and reverse phylip idm
66 sub create_idm {
67     my $infile = shift;
68     my %idm_for;
69     open my $in, '<', $infile;

```

```

70
71 LINE :
72 while (my $line = <$in>){
73     chomp $line;
74     next LINE if $line =~ m/^\#/xms;
75     next LINE if $line =~ m/^\s*/xms;
76
77     my ($ann, $id) = split /\t/, $line;
78     #~ ($id) = $id =~ m/.[A-Z]{3}\_d+\.\d?\@ [A-Z]{2}\_d+\.\d?)/xms;
79
80     my $seq_id = SeqId->new( full_id => $id );
81     my $gca = $seq_id->gca;
82     my $acc = $seq_id->accession;
83     my $fullgca = $gca . "@" . $acc;
84     my $taxon = $seq_id->org;
85     $taxon =~ tr/ _/ /;
86
87     $idm_for{$fullgca} = {
88         ann      => $ann,
89         fullname => $taxon . "_" . $fullgca,
90     };
91 }
92 return \%idm_for;
93 }

```

# Bibliographie

Ambler, R. P. 1980. “The Structure of  $\beta$ -Lactamases.” *Philosophical Transactions of the Royal Society of London B : Biological Sciences* 289 (1036). The Royal Society : 321–31. doi :10.1098/rstb.1980.0049.

Aoki, H, and M Okuhara. 1980. “Natural Beta-Lactam Antibiotics.” *Annual Review of Microbiology* 34 (1) : 159–81. doi :10.1146/annurev.mi.34.100180.001111.

Babic, Maja, Andrea M. Hujer, and Robert A. Bonomo. 2006. “What’s New in Antibiotic Resistance ? Focus on Beta-Lactamases.” *Drug Resistance Updates* 9 (3) : 142–56. doi :10.1016/j.drug.2006.05.005.

Ball, A P, J A Gray, and J McM. Murdoch. 1978. “The Natural Penicillins — Benzylpenicillin (Penicillin G) and Phenoxymethylpenicillin (Penicillin V).” In *Antibacterial Drugs Today*, 6–18. Dordrecht : Springer Netherlands. doi :10.1007/978-94-011-8004-7\_3.

Berzigotti, Stéphanie, Kamal Benlafya, Jérémy Sépulchre, Ana Amoroso, and Bernard Joris. 2012. “Bacillus Licheniformis BlaR1 L3 Loop Is a Zinc Metalloprotease Activated by Self-Proteolysis.” *PLoS ONE* 7 (5) : 1–10. doi :10.1371/journal.pone.0036400.

Brown, Susan, and Sebastian Amyes. 2006. “OXA  $\beta$ -Lactamases in Acinetobacter : The Story so Far.” *Journal of Antimicrobial Chemotherapy* 57 (1) : 1–3. doi :10.1093/jac/dki425.

Bryskier, a. 2000. “Cephems : Fifty Years of Continuous Research.” *The Journal of Antibiotics* 53 (10) : 1028–37. doi :10.7164/antibiotics.53.1028.

Bush, Karen, and Patricia A Bradford. 2016. “ $\beta$ -Lactams and  $\beta$ -Lactamase Inhibitors : An Overview.” *Cold Spring Harbor Perspectives in Medicine* 6 (8). doi :10.1101/cshperspect.a025247.

Cantón, Rafael, José María González-Alba, and Juan Carlos Galán. 2012. “CTX-M Enzymes : Origin and Diffusion.” *Frontiers in Microbiology* 3 (APR). doi :10.3389/fmicb.2012.00110.

Castresana, J. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution* 17 (4) : 540–52.

Dougherty, Thomas J., and Michael J. Pucci. 2014. *Antibiotic Discovery and Development*. doi :10.1007/978-1-4614-1400-1.

Eddy, Sean R., Travis J. Wheeler, and Development Team. 2015. “HMMER User Guide,” no. February : 120. <http://eddylab.org/software/hmmer3/3.1b2/Userguide.pdf>.

Elander, R. P. 2003. “Industrial Production of  $\beta$ -Lactam Antibiotics.” *Applied Microbiology and Biotechnology* 61 (5-6) : 385–92. doi :10.1007/s00253-003-1274-y.

Etebu, Ebimiewei, and Ibemologi Ariekpar. 2016. “Antibiotics : Classification and Mechanisms

of Action with Emphasis on Molecular Perspectives.” *International Journal of Applied Microbiology and Biotechnology Research* 4 : 90–101.

Evans, Benjamin A., and Sebastian G B Amyes. 2014. “OXA  $\beta$ -Lactamases.” *Clinical Microbiology Reviews* 27 (2) : 241–63. doi :10.1128/CMR.00117-13.

Fernandes, Rúben, Paula Amador, and Cristina Prudêncio. 2013. “ $\beta$ -Lactams : Chemical Structure, Mode of Action and Mechanisms of Resistance.” *Reviews in Medical Microbiology* 24 (1) : 7–17. doi :10.1097/MRM.0b013e3283587727.

Fisher, Jed F., Samy O. Meroueh, and Shahriar Mobashery. 2005. “Bacterial Resistance to  $\beta$ -Lactam Antibiotics : Compelling Opportunism, Compelling Opportunity.” *Chemical Reviews* 105 (2) : 395–424. doi :10.1021/cr030102i.

Fleming, A. 1929. “On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenae.” *British Journal of Experimental Pathology* 10 : 226–36.

Frère, Jean Marie. 1995. “Beta-Lactamases and Bacterial Resistance to Antibiotics.” *Molecular Microbiology* 16 : 385–95. doi :10.1111/j.1365-2958.1995.tb02404.x.

Frère, Jean Marie, and Bernard Joris. 1985. “Penicillin-Sensitive Enzymes in Peptidoglycan Biosynthesis.” *CRC Critical Reviews in Microbiology* 11 (4) : 299–396.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. “CD-HIT : Accelerated for Clustering the Next-Generation Sequencing Data.” *Bioinformatics* 28 (23) : 3150–2. doi :10.1093/bioinformatics/bts565.

Galleni, Moreno, Josette Lamotte-brasseur, and Gian Maria Rossolini. 2001. “GUEST COMMENTARY Standard Numbering Scheme for Class B  $\beta$ -Lactamases.” *Society* 45 (3) : 660–63. doi :10.1128/AAC.45.3.660.

Gaynes, Robert. 2017. “The Discovery of Penicillin—new Insights After More Than 75 Years of Clinical Use.” *Emerging Infectious Diseases* 23 (5) : 849–53. doi :10.3201/eid2305.161556.

Georgopapadakou, N. H, and F.Y Liu. 1980. “Penicillin-Binding Proteins in Bacteria.” *Annals of Internal Medicine* 96 (1) : 502–4. doi :10.1128/AAC.18.1.148.

Goffin, Colette, and Jean-Marie Ghuysen. 1998. “Multimodular Penicillin-Binding Proteins : An Enigmatic Family of Orthologs and Paralogs.” *Microbiol. Mol. Biol. Rev.* 62 (4) : 1079–93. <http://mmbr.asm.org/content/62/4/1079.short>.

Gould, Kate. 2016. “Antibiotics : From Prehistory to the Present Day.” *Journal of Antimicrobial*

Haaber, Jakob, José R Penadés, and Hanne Ingmer. 2017. “Transfer of Antibiotic Resistance in *Staphylococcus Aureus*.” Elsevier Ltd. doi :10.1016/j.tim.2017.05.011.

Jacoby, George A. 2009. “AmpC B-Lactamases.” *Clinical Microbiology Reviews* 22 (1) : 161–82. doi :10.1128/CMR.00036-08.

Jago, M, and N G Heatley. 1961. “Some Biological Properties of Cephalosporin C and a Derivative.” *British Journal of Pharmacology* 16 : 170–79.

Kapoor, Garima, Saurabh Saigal, and Ashok Elongavan. 2017. “Action and Resistance Mechanisms of Antibiotics : A Guide for Clinicians.” *Journal of Anaesthesiology Clinical Pharmacology* 33 (3) : 300–305. doi :10.4103/joacp.JOACP.

Katoh, Kazutaka, and Martin C. Frith. 2012. “Adding Unaligned Sequences into an Existing Alignment Using MAFFT and LAST.” *Bioinformatics* 28 (23) : 3144–6. doi :10.1093/bioinformatics/bts578.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. “MAFFT : A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14) : 3059–66. doi :10.1093/nar/gkf436.

Kerff, Frédéric, Eric Sauvage, Lionel Vercheval, and Paulette Charlier. n.d. “Structures of Class D Beta-Lactamase.”

Keshri, Vivek, Arup Panda, Anthony Levasseur, Jean-Marc Rolain, Pierre Pontarotti, and Didier Raoult. 2018. “Phylogenomic Analysis of  $\beta$ -Lactamase in Archaea and Bacteria Enables the Identification of Putative New Members.” *Genome Biology and Evolution* 10 (4) : 1106–14. doi :10.1093/gbe/evy028.

Kohanski, Michael A., Daniel J. Dwyer, and James J. Collins. 2010. “How Antibiotics Kill Bacteria : From Targets to Networks.” *Nature Reviews Microbiology* 8 (6) : 423–35. doi :10.1038/nrmicro2333.

Lakaye, Bernard, Alain Dubus, Sophie Lepage, S. Gros Lambert, and Jean Marie Frère. 1999. “When Drug Inactivation Renders the Target Irrelevant to Antibiotic Resistance : A Case Story with  $\beta$ -Lactams.” *Molecular Microbiology* 31 (1) : 89–101. doi :10.1046/j.1365-2958.1999.01150.x.

Lambert, Peter A. 2005. “Bacterial Resistance to Antibiotics : Modified Target Sites.” *Advanced Drug Delivery Reviews* 57 (10) : 1471–85. doi :10.1016/j.addr.2005.04.003.

Laraki, Nezha, Moreno Galleni, Iris Thamm, Maria Letizia Riccio, Gianfranco Amicosante, Jean Marie Frère, and Gian Maria Rossolini. 1999. “Structure of In31, a Bla(IMP)-Containing



Pseudomonas Aeruginosa Integron Phyletically Related to In5, Which Carries an Unusual Array of Gene Cassettes.” *Antimicrobial Agents and Chemotherapy* 43 (4) : 890–901.

Lauretti, L, M L Riccio, A Mazzariol, G Cornaglia, G Amicosante, R Fontana, and G M Rossolini. 1999. “Cloning and Characterization of BlaVIM, a New Integron-Borne Metallo-Beta-Lactamase Gene from a Pseudomonas Aeruginosa Clinical Isolate.” *Antimicrobial Agents and Chemotherapy* 43 (7) : 1584–90.

Lemoine, F., J. B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, and O. Gascuel. 2018. “Renewing Felsenstein’s Phylogenetic Bootstrap in the Era of Big Data.” *Nature* 556 (7702) : 452–56. doi :10.1038/s41586-018-0043-0.

Leonard, David A., Robert A. Bonomo, and Rachel A. Powers. 2013. “Class D  $\beta$ -Lactamases : A Reappraisal After Five Decades.” *Accounts of Chemical Research* 46 (11) : 2407–15. doi :10.1021/ar300327a.

Léonard, Raphaël. 2014. “Evolution Du Cluster Dcw et Phylogénie Des Bactéries.” Thèse de doctorat, Université de Liège.

Liakopoulos, Apostolos, Dik Mevius, and Daniela Ceccarelli. 2016. “A Review of SHV Extended-Spectrum  $\beta$ -Lactamases : Neglected yet Ubiquitous.” *Frontiers in Microbiology* 7 (SEP). doi :10.3389/fmicb.2016.01374.

Llarrull, Leticia I., Marta Toth, Matthew M. Champion, and Shahriar Mobashery. 2011. “Activation of BlaR1 Protein of Methicillin-Resistant Staphylococcus Aureus, Its Proteolytic Processing, and Recovery from Induction of Resistance.” *Journal of Biological Chemistry* 286 (44) : 38148–58. doi :10.1074/jbc.M111.288985.

Lobanovska, Mariya, and Giulia Pilla. 2017. “Penicillin’s Discovery and Antibiotic Resistance : Lessons for the Future ?” *Yale Journal of Biology and Medicine* 90 (1) : 135–45.

Massova, Irina, and Shahriar Mobashery. 1998. “Kinship and Diversification of Bacterial Penicillin-Binding Proteins and  $\beta$ -Lactamases.” *Antimicrobial Agents and Chemotherapy* 42 (1) : 1–17.

Mazuet, C., E. J. Yoon, S. Boyer, S. Pignier, T. Blanc, I. Doehring, D. Meziane-Cherif, et al. 2016. “A Penicillin- and Metronidazole-Resistant Clostridium Botulinum Strain Responsible for an Infant Botulism Case.” *Clinical Microbiology and Infection* 22 (7). Elsevier : 644.e7–644.e12. doi :10.1016/j.cmi.2016.04.011.

Munita, Jose M, Cesar A Arias, Antimicrobial Resistance Unit, and Alemana De Santiago. 2016. “Mechanisms of Antibiotic Resistance.” *Microbiology Spectrum* 4 (2) : 1–37. doi :10.1128/microbiolspec.VMBF-

Naas, Thierry, Saoussen Oueslati, Rémy A. Bonnin, Maria Laura Dabos, Agustin Zavala, Laurent Dortet, Pascal Retailleau, and Bogdan I. Iorga. 2017. "Beta-Lactamase Database (BLDB)—structure and Function." *Journal of Enzyme Inhibition and Medicinal Chemistry* 32 (1). Informa UK Ltd. : 917–19. doi :10.1080/14756366.2017.1344235.

Newton, G G F, and E P Abraham. 1955. "Cephalosporin C, a New Antibiotic Containing Sulphur and D- $\alpha$ -Aminoadipic Acid [1]." *Nature* 175 : 548.

Nikaido, Hiroshi. 2009. "Multidrug Resistance in Bacteria." *Annu Rev Biochem.*, no. 2 : 119–46. doi :10.1146/annurev.biochem.78.082907.145923.Multidrug.

Palzkill, Timothy. 2013. "Metallo- $\beta$ -Lactamase Structure and Function." *Annals of the New-York Academy of Sciences*, 91–104. doi :10.1111/j.1749-6632.2012.06796.x.Metallo-.

Papp-Wallace, Krisztina M., Andrea Endimiani, Magdalena A. Taracila, and Robert A. Bonomo. 2011. "Carbapenems : Past, Present, and Future." *Antimicrobial Agents and Chemotherapy* 55 (11) : 4943–60. doi :10.1128/AAC.00296-11.

Parks, Donovan H., Christian Rinke, Maria Chuvochina, Pierre Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz, and Gene W. Tyson. 2017. "Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life." *Nature Microbiology* 2 (11). Springer US : 1533–42. doi :10.1038/s41564-017-0012-7.

Paterson, David L, and Robert A Bonomo. 2005. "Extended-Spectrum Beta-Lactamases : A Clinical Update." *Clinical Microbiology Reviews* 18 (4) : 657–86. doi :10.1128/CMR.18.4.657.

Philippon, Alain, Guillaume Arlet, and George A Jacoby. 2002. "Plasmid-Determined AmpC-Type-Beta-Lactamases." *Antimicrobial Agents and Chemotherapy* 46 (1) : 1–11. doi :10.1128/AAC.46.1.1.

Philippon, Alain, Patrick Slama, and Roger Labia. 2016. "A Structure-Based Classification of Class A Beta-Lactamases , a Broadly." *Clinical Microbiology Reviews* 29 (1) : 29–57. doi :10.1128/CMR.00019-15.

Poole, K. 2004. "Resistance to  $\beta$ -Lactam Antibiotics." *Cellular and Molecular Life Sciences* 61 (17) : 2200–2223. doi :10.1007/s00018-004-4060-9.

Pozzi, Cecilia, Di Pisa, Filomena De Luca, and Manuela Benvenuti. 2018. "Atomic-Resolution Structure of a Class C  $\beta$ -Lactamase and Its Complex with Avibactam." *ChemMedChem*, 1–11. doi :10.1002/cmdc.201800213.

Queenan, Anne Marie, and Karen Bush. 2007. "Carbapenemases : The Versatile  $\beta$ -Lactamases."

*Clinical Microbiology Reviews* 20 (3) : 440–58. doi :10.1128/CMR.00001-07.

Ramsey, Christopher, and Alasdair P. MacGowan. 2016. “A Review of the Pharmacokinetics and Pharmacodynamics of Aztreonam.” *Journal of Antimicrobial Chemotherapy* 71 (10) : 2704–12. doi :10.1093/jac/dkw231.

Rawat, Deepti, and Deepthi Nair. 2010. “Extended-Spectrum  $\beta$ -Lactamases in Gram Negative Bacteria.” *Journal of Global Infectious Diseases* 2 (3) : 263. doi :10.4103/0974-777X.68531.

Salverda, Merijn L M, J. Arjan G M de Visser, and Miriam Barlow. 2010. “Natural Evolution of TEM-1  $\beta$ -Lactamase : Experimental Reconstruction and Clinical Relevance.” *FEMS Microbiology Reviews* 34 (6) : 1015–36. doi :10.1111/j.1574-6976.2010.00222.x.

Sauvage, Eric, Frédéric Kerff, Mohammed Terrak, Juan A. Ayala, and Paulette Charlier. 2008. “The Penicillin-Binding Proteins : Structure and Role in Peptidoglycan Biosynthesis.” *FEMS Microbiology Reviews* 32 (2) : 234–58. doi :10.1111/j.1574-6976.2008.00105.x.

Sengupta, Saswati, Madhab K. Chattopadhyay, and Hans Peter Grossart. 2013. “The Multifaceted Roles of Antibiotics and Antibiotic Resistance in Nature.” *Frontiers in Microbiology* 4 (MAR) : 1–13. doi :10.3389/fmicb.2013.00047.

Srivastava, Abhishikha, Neelja Singhal, Manisha Goel, Jugsharan Singh Viridi, and Manish Kumar. 2014. “CBMAR : A Comprehensive  $\beta$ -Lactamase Molecular Annotation Resource.” *Database* 2014 : 1–8. doi :10.1093/database/bau111.

Stamatakis, Alexandros. 2014. “RAxML Version 8 : A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics* 30 (9) : 1312–3. doi :10.1093/bioinformatics/btu033.

Tahlan, Kapil, and Susan E. Jensen. 2013. “Origins of the  $\beta$ -Lactam Rings in Natural Products.” *Journal of Antibiotics* 66 (7). Nature Publishing Group : 401–10. doi :10.1038/ja.2013.24.

Toth, Marta, Nuno Tiago Antunes, Nichole K. Stewart, Hilary Frase, Monolekha Bhattacharya, Clyde Smith, and Sergei Vakukenko. 2016. “Class D  $\beta$ -Lactamases Do Exist in Gram-Positive Bacteria.” *Nature Chemical Biology* 12 (1) : 39–46. doi :10.1038/nchembio.1950.

Ventola, C Lee. 2015. “The Antibiotic Resistance Crisis : Part 1 : Causes and Threats.” *P & T : A Peer-Reviewed Journal for Formulary Management (2015)* 40 (4) : 277–83. doi :Article.

Vollmer, Waldemar, Didier Blanot, and Miguel A. De Pedro. 2008. “Peptidoglycan Structure and Architecture.” *FEMS Microbiology Reviews* 32 (2) : 149–67. doi :10.1111/j.1574-6976.2007.00094.x.