

Université du Québec
Institut National de la Recherche Scientifique
Institut Armand-Frappier

**RECHERCHE DE NOUVELLES MOLECULES ANTIMICROBIENNES ACTIVES
CONTRE *STAPHYLOCCOCUS AUREUS* RÉSISTANT À LA MÉTHICILLINE
PAR MÉTAGÉNOMIQUE FONCTIONNELLE**

Par
Anissa Brahami

Thèse présentée pour l'obtention du grade de
Philosophiæ Doctor (Ph.D.) en biologie.

Jury d'évaluation

Président du jury et Examineur interne	Dr. Richard Villemur INRS – Institut Armand-Frappier
Examineur externe	Dr. François Malouin Université de Sherbrooke
Examineur externe	Dr. Luke Masson Université de Montréal-CNRC
Directeur de recherche	Dr. Annie Castonguay INRS – Institut Armand-Frappier
Co-directeur de recherche	Dr. Éric Déziel INRS – Institut Armand-Frappier

REMERCIEMENTS

Je dédie cette thèse à mes parents. Merci pour tous les sacrifices que vous avez faits pour moi. Merci pour vos encouragements et votre amour. Je sais que cela n'a pas été facile pour vous de me laisser partir poursuivre mes rêves et mes ambitions loin de vous. Merci Maman, Papa, Anis et Mélissa d'avoir accepté la place que le doctorat a eu dans ma vie.

Je tiens particulièrement à remercier mes directeurs de recherche, la Pr^e Annie Castonguay et le Pr Éric Déziel. Merci de m'avoir poussée à donner le meilleur de moi-même :

Annie, merci de m'avoir accepté dans ton laboratoire et de m'avoir donné l'opportunité de réaliser ce doctorat. Ce fut un grand plaisir de partager ces années de doctorat en ta compagnie. Tu m'as beaucoup appris et tu m'as permis d'évoluer en tant que scientifique, je t'en suis très reconnaissante. Merci d'avoir partagé ton expertise scientifique avec moi, et merci pour tout le temps que tu m'as accordée pour discuter des nombreuses problématiques de ce doctorat.

Éric, merci de m'avoir donné la liberté scientifique dont j'avais besoin sur ce projet, et de m'avoir appris à développer un esprit critique. Tu m'as toujours encouragé et tu as toujours cru en moi, je t'en suis très reconnaissante. Ton soutien autant que scientifique et moral a été pour moi un grand support. Merci pour ta grande disponibilité pour répondre à mes nombreuses questions et surtout à ton positivisme face à toutes les problématiques rencontrées durant ce doctorat.

Je tiens à remercier le professeur David Chatenet de m'avoir formé pour effectuer de la synthèse de peptides. Merci de m'avoir accordé ta confiance pour travailler en toute liberté dans ton laboratoire. Merci pour ta disponibilité, ton positivisme et ta bonne humeur contagieuse.

Je remercie aussi les membres de jury, pour avoir accepté de consacrer de leur temps pour lire, corriger, évaluer et de discuter de mon travail :

Pr François Malouin, Luke Masson, Richard Villemur, merci d'avoir répondu présent rapidement et merci d'avoir accepté de découvrir et d'évaluer ma thèse de doctorat.

Pr Philippe Constant, merci d'avoir participé à l'évaluation de tous les examens de mon doctorat. Merci pour les conversations enrichissantes que j'ai eues avec toi, merci pour ta présence, ta disponibilité et tes conseils.

Dr Vincent Burrus, merci d'avoir participé à l'évaluation de mon séminaire ainsi que ma thèse de doctorat.

Sincères et personnels remerciements à :

Sylvain Milot du laboratoire d'Éric Déziel pour avoir cru en moi et avoir été toujours derrière moi surtout durant la finale du concours ma thèse en 180 secondes 2017.

Marie-Christine Groleau du laboratoire d'Éric Déziel pour avoir répondu à toutes mes questions. Merci de m'avoir écoutée et encouragée durant tout le long de mon doctorat.

Myriam Letourneau du laboratoire de Nicolas Doucet pour ta disponibilité, ton soutien incessant et ton expertise. Tu m'as beaucoup appris.

Marlène Fortier du laboratoire de Yves St-Pierre pour ta disponibilité et ton partage de connaissance.

Je remercie mes collègues de l'IAF : Sophie, Arvin, Servane, German, Ahmad, Fabrice, Fadi, Ghizlane, Sarah, Pauline, Kanak, May, Snizhana, Charles, Audrey-Anne du laboratoire Déziel pour les bons moments partagés et les discussions enrichissantes, Golar, Mehdi, Ahmed, Hala, Van, Robin du laboratoire Castonguay pour vos encouragements, Mustapha, Hassan, Étienne du laboratoire Chatenet pour vos enseignements et votre amabilité.

Je remercie également mes amis: Carlos, Mathilde, Marie-Noel, Marie-Anne, Aurore, Antoine, Stéphanie, Imene, Morgane, Kamel, Sabrina, Sofiane, pour votre soutien incessant, vos encouragements et votre support. Merci pour tous ces bons moments partagés et merci d'avoir été présent pendant les moments les moins joyeux.

Je tiens à témoigner ma profonde reconnaissance à ma deuxième famille de Montréal : Hadjira, Tsoufik, Katia, Malika et Lynda. Merci pour votre générosité incroyable, votre support et votre amour. Vous faites maintenant partie de ma vie. Merci d'avoir pris soin de moi. Spécial remerciement à Dalila et Mourad.

Mes remerciements ne seraient pas complets si je ne les adressais pas à l'homme que j'aime : Wassim. Ce doctorat nous a tenu loin l'un de l'autre pendant 4 ans avant que tu puisses me rejoindre. Ce fut long, mais nous avons tenu bon. Merci d'avoir accepté mes ambitions, merci de m'avoir encouragé de l'autre côté de l'océan, merci pour ton amour, ton support et ton dévouement indéfectible. Je remercie également mes beaux-parents Sekoura et Rachid pour leurs encouragements et leur compréhension.

Merci à tous d'avoir contribué à ce que je suis devenue

PRÉAMBULE

Cette thèse est présentée sous forme de thèse par article.

Le premier chapitre est une revue de littérature rédigée sous forme de deux parties; la première fait état du problème de résistance aux antibiotiques et du besoin crucial de rechercher de nouvelles molécules antibiotiques, et la deuxième partie décrit les généralités sur la métagénomique fonctionnelle et ses applications pour la recherche d'antibiotiques. Une partie de cette introduction ainsi que la majorité des études métagénomiques fonctionnelles axées sur la découverte de nouveaux antibiotiques ont été rédigées sous forme de revue en anglais qui sera soumis prochainement, ajouté dans l'annexe A.

Le deuxième chapitre correspond au 1^{er} article. Il s'agit d'un article qui traite du développement de méthodes de construction de banques d'ADN métagénomiques et de criblage fonctionnel basés sur l'ADN génomique de la bactérie *Burkholderia thailandensis*. Les étapes de construction de banques d'ADN métagénomiques ont été mises au point afin de maximiser l'efficacité de transformation, et une nouvelle méthode de criblage fonctionnel plus sensible a été développée.

Le troisième chapitre est l'étude métagénomiques effectuées à partir d'ADN prélevé à partir des biofilms d'hôpitaux.

Le quatrième chapitre correspond au 2^{ème} article. Il s'agit d'une étude qui a permis de découvrir un nouveau peptide antimicrobien ayant un effet sur *Staphylococcus aureus* résistant à la méthicilline (SARM) à partir d'ADN échantillonnée à partir de conduits d'égout d'un hôpital par métagénomique fonctionnelle.

Le cinquième chapitre est une discussion générale de la thèse où les méthodes utilisées sont critiquées et les résultats obtenus sont en perspectives par rapport à la littérature existante sur la métagénomique fonctionnelle.

Le sixième chapitre fait état des conclusions obtenues suite à ce projet ainsi que les perspectives proposées pour la suite de cette étude. Des perspectives à plus long terme sur les études métagénomiques fonctionnelles ont aussi été proposées.

RÉSUMÉ

La résistance aux antibiotiques représente un défi important pour la santé publique mondiale. Il y a donc urgence de découvrir de nouveaux antibiotiques efficaces contre ces bactéries multirésistantes.

Etant donné que les microorganismes du sol sont à l'origine de la plupart des antibiotiques connus et utilisés mais que la grande majorité des microorganismes demeure incultivable au laboratoire, la métagénomique fonctionnelle permet un accès à ce potentiel génétique qui représente une source prometteuse de nouveaux antibiotiques.

L'hypothèse exploitée est qu'un environnement enrichi en BMR sélectionnera positivement par antibiose la croissance de bactéries capables de compétitionner grâce à la production d'antibiotiques efficaces contre ces BMR. De ce fait, dans un premier temps, la construction de banques d'ADN génomiques en utilisant l'hôte d'expression *Escherichia coli* DH10B a été effectuée. Nous avons démontré comment la déphosphorylation de l'ADN du vecteur / déphosphorylation de l'ADNe, l'inactivation de la ligase après ligation, la dialyse du produit de ligation et le rapport ADN vecteur / ADN génomique influencent l'efficacité de transformation. De plus, nous avons décrit l'utilisation d'un aérographe pour le criblage fonctionnel de cette banque d'ADN génomique construite en vaporisant la bactérie cible *Staphylococcus aureus* souche Newman. Cette approche de criblage fonctionnel permettant de détecter la plus petite zone d'inhibition produite par les clones transformants a été nommée *bacteriospray*. Une fine et uniforme couche de bactéries cible est déposée, ce qui permet le criblage de 5 à 10 fois plus de clones que la méthode de double couche d'agar conventionnelle. Pour valider cette méthode, l'ADN génomique de *Burkholderia thailandensis* E264 a servi à la construction d'une banque d'ADN où quatre des 70,000 clones criblés inhibaient la croissance de *S. aureus*. L'analyse des inserts de ces clones actifs a révélé que ces régions génomiques n'avaient jamais été rapportées responsables de la production de molécules antimicrobiennes.

Par la suite, l'ADN de biofilms de différents hôpitaux a été utilisé pour la construction et l'expression hétérologue de l'ADN chez *Escherichia coli* DH10B à l'aide du vecteur pBeloBAC11 suivant les méthodes optimisées. Le criblage fonctionnel par bacteriospray de 150,000 clones contre *Staphylococcus aureus* résistant à la méticilline (SARM) a permis d'identifier 10 clones actifs. Certains clones avaient des morphologies particulières en étoiles ou formaient de petites colonies, mais un clone en particulier présentait un défaut de croissance en milieu liquide et en milieu solide. Le résultat BLASTx de la séquence indique 100% de similarité avec la sous unité de la cobaltochélatase CobN de *Mycobacterium* sp. Une analyse des cadres de lecture ouverts avec le codon ATG comme codon de départ a révélé deux différents peptides qui ont été synthétisés sur support solide afin de tester leur activité. Une prédiction structurale, une analyse par dichroïsme circulaire ainsi qu'une analyse par résonance magnétique nucléaire de ces deux peptides ont montré que le peptide-1 (dérivé de ORF1) adopte une conformation en α -hélice, une structure commune aux peptides antibactériens, tandis que le peptide-2 (dérivé de ORF2) adopte une structure aléatoire. Un résultat inattendu obtenu était que le peptide-2 était faiblement actif contre SARM alors que le peptide-1 ne présentait aucune activité. Des études de structure-activité sont proposées en perspective pour améliorer l'activité biologique du peptide-2.

Les résultats obtenus au cours de cette thèse nous ont en premier lieu permis d'améliorer notre compréhension des facteurs influençant l'expression hétérologue d'ADN exogène, particulièrement en métagénomique fonctionnelle. En deuxième lieu, le développement de la nouvelle technique de criblage décrite dans cette thèse pourra contribuer à accélérer la découverte de nouveaux antibiotiques en permettant le criblage de volumineuses bibliothèques métagénomiques d'ADNe. Finalement, l'identification d'un peptide antimicrobien en exploitant l'ADNe d'environnements riches en microorganismes multirésistants tels les biofilms hospitaliers encouragera la communauté scientifique à explorer davantage de telles sources d'ADNe pour la découverte d'antibiotiques capables d'éradiquer les bactéries multirésistantes.

ABSTRACT

Multidrug-resistant infections represent a serious human threat and there is therefore an urgent need for the discovery of new effective antibiotics.

Since soil microorganisms produce most of the antibiotics currently known and used in human medicine, but that most microorganisms cannot be cultured using standard laboratory conditions, functional metagenomics is a key method for the identification of new metabolites as it allows on to the access to a promising source of genetic material.

The hypothesis exploited in this study is that competition between environmental microbes might select for the evolution of antibiotics able to bypass current antibiotic resistance mechanisms. As a first step, we developed a detailed genomic library construction approach using *Escherichia coli* DH10B as a surrogate host, and demonstrated how vector or genomic DNA dephosphorylation, ligase inactivation, dialysis of the ligation product and vector/genomic DNA ratio greatly influence the transformation efficiency. Furthermore, we described the use of an airbrush device to screen *E. coli* metagenomic libraries for their antibacterial activity against *Staphylococcus aureus*, a method we called *Bacteriospray*. This method facilitates the identification of new antimicrobials among large metagenomic libraries. This bacterial spraying tool greatly improves the functional screening of large genomic libraries, as it conveniently allows the production of a thinner and a more uniform layer of target bacteria compared to the commonly used overlay method, resulting in the screening of 5–10 times more clones per agar plate. Using *Burkholderia thailandensis* E264 genomic DNA as a proof of concept, four clones out of 70,000 inhibited the growth of *S. aureus* and were found to each contain a DNA insert. Analysis of these inserted DNA fragments revealed genomic regions never previously reported to be responsible for the production of antimicrobials, or predicted by bioinformatics tools.

Thereafter, DNA from different hospital biofilms was used to construct and express *Escherichia coli* DH10 B libraries with BAC vectors (Bacterial Artificial Chromosome) using the previously optimized method mentioned above. The functional screening of 150,000 clones against methicillin-resistant *Staphylococcus aureus* (MRSA) using the bacteriospray method allowed the identification of ten active clones. Some clones displayed particular star-shaped morphologies or formed small colonies, but a unique *E. coli* transformant clone was characterized by a growth defect in liquid and solid media. The inserted DNA was 100% identical to a part of gene sequence predicted to contain the cobaltochelate subunit (CobN) of *Mycobacterium* sp. A blast ORF finder analysis using ATG as a start codon revealed two different potential peptides that were synthesized on solid support to validate their activity. Structural modeling, circular dichroism analysis and nuclear magnetic resonance analysis of the two synthetic peptides showed that peptide-1 (derived from ORF1) adopts an α -helix conformation, a structure commonly observed among antibacterial peptides, whereas peptide-2 (derived from ORF2) is unstructured. Results showed that peptide-2 was effective against MRSA whereas peptide-1 was found inactive. Structure-activity relationship studies are proposed to improve the biological activity of peptide-2.

Results obtained in the course of this thesis allowed us to improve our understanding of several factors influencing the heterologous expression of foreign DNA, particularly in the context of functional metagenomics. In addition, the development of the new screening tool presented here will facilitate the discovery of new antibiotics by allowing the screening of large eDNA metagenomic libraries. Finally, the identification of an antimicrobial peptide obtained by exploiting eDNA from multidrug-resistant microorganism-rich environments such as hospital biofilms will likely prompt the scientific community to explore such eDNA (environmental DNA) sources for the discovery of antibiotics able to eradicate multidrug-resistant bacteria.

Table des matières

CHAPITRE 1 : REVUE DE LA LITTÉRATURE	1
1.1. ANTIBIOTIQUES FACE À LA RESISTANCE BACTERIENNE AUX ANTIBIOTIQUES	2
1.1.1. La résistance bactérienne aux antibiotiques	2
1.1.2. Les antibiotiques : le passé, le présent et le futur	6
1.1.3. Les microorganismes comme source de nouveaux antibiotiques	7
1.1.4. L'innovation pour découvrir de nouveaux antibiotiques.....	9
1.2. METAGENOMIQUE : ALTERNATIVE POUR ACCEDER À DE NOUVELLES MOLECULES ANTIBIOTIQUES	10
1.2.1. Construction de banques d'ADN en métagénomique fonctionnelle	14
1.2.1.1. Méthode d'extraction d'ADN	15
1.2.1.2. Source d'échantillonnage de l'ADN.....	17
1.2.1.3. Vecteurs de clonage	18
1.2.1.4. Les souches d'expression hétérologue.....	21
1.2.1.4.1. Souches d'expression <i>Escherichia coli</i>	21
1.2.1.4.2. Souche d'expression <i>Bacillus subtilis</i>	23
1.2.1.4.3. Souche d'expression du genre <i>Streptomyces</i>	24
1.2.1.5. Systèmes d'expression	25
1.2.2. Criblage fonctionnel des banques d'ADN métagénomiques	26
1.2.3. Limites de la métagénomiques fonctionnelle.....	28
1.3 PROBLÉMATIQUE, HYPOTHÈSE ET OBJECTIFS	30
1.3.1. Problématique	30
1.3.2. Hypothèses.....	32
1.3.3. Objectifs.....	33
1.3.3.1 Objectif général.....	33
1.3.3.2. Objectifs spécifiques	34
CHAPITRE 2 : ARTICLE 1	35
2.1. NOVEL 'BACTERIOSPRAY METHOD FACILITATES THE FUNCTIONAL SCREENING OF METAGENOMIC LIBRARIES FOR ANTIMICROBIAL ACTIVITY	36
2.1.1. Résumé	37
2.1.2. Abstract	38

2.1.3. Introduction	39
2.1.4. Materials and Methods.....	41
2.1.4.1. Bacterial Strains and Plasmid Growth Conditions.....	41
2.1.4.2. Library Construction Strategies.....	41
2.1.4.3. Derivation of a Chloramphenicol Resistant Strain.....	43
2.1.4.4. Top Agar Overlay Method	43
2.1.4.5. Bacteriospray Screening Method	44
2.1.4.6. Growth Inhibition Zone Confirmation, DNA Sequencing and Insert Analysis.....	45
2.1.4.7. Statistical Analysis	46
2.1.5. Results.....	47
2.1.5.1. Transformation Efficiency Optimization.....	47
2.1.5.2. Burkholderia thailandensis Library Construction.....	50
2.1.5.3. Top Agar Overlay vs Bacteriospray Assay for the Functional Screening of Metagenomic Libraries	51
2.1.5.4. Sequence Analysis of Inserted Genes	52
2.1.6. Discussion	54
2.1.7. Conclusions	56
2.2. DONNÉES COMPLÉMENTAIRES À L'ARTICLE.....	57
2.2.1. Mise en contexte	57
2.2.2. Méthodologie	57
2.2.3. Résultats.....	57
2.2.3.1. Analyse NCBI.....	57
2.2.3.2. Analyse par antiSMASH.....	61
2.2.3.3. Analyse par BAGEL	61
2.3. RÉSULTATS SUPPLÉMENTAIRES DU CRIBLAGE DE LA BANQUE D'ADN ISSUS DE B. THAILANDENSIS.....	64
2.3.1. Mise en contexte	64
2.3.2. Méthodologie	64
2.3.3. Résultats.....	65
2.3.3.1. Autres clones sélectionnés	65
CHAPITRE 3 : ÉTUDE MÉTAGÉNOMIQUE FONCTIONNELLE	68
3.1.1. Mise en contexte	69
3.1.2. Matériel et méthodes	69
3.1.2.1. Échantillonnage de l'ADNe	69

3.1.2.2. Digestion enzymatique du vecteur pBeloBAC11	70
3.1.2.3. Extraction de l'ADNe des biofilms pour la construction de banques.....	70
3.1.2.4. Digestion de l'ADNe des biofilms	70
3.1.2.5. Extraction sur gel d'agarose de l'ADNe digéré	70
3.1.2.6. Ligation de l'ADNe au vecteur pBeloBAC11	71
3.1.2.7. Transformation	71
3.1.2.8. Criblage fonctionnelle.....	72
3.1.2.9. Analyse des clones sélectionnés	72
3.1.2.9.1. Analyse de la morphologie.....	72
3.1.2.9.2. Confirmation de l'activité.....	72
3.1.2.9.3. Séquençage de l'insert et analyse de la séquence par BLASTn.....	73
3.1.2.10. Synthèse, purification, caractérisation des peptides et test de leur effet sur la croissance de <i>Staphylococcus aureus</i> résistant à la méthicilline.....	73
3.1.3. Résultats	73
3.1.3.1. Prélèvement des biofilms	73
3.1.3.2. Digestion du vecteur pBeloBAC11	74
3.1.3.3. Optimisation de l'extraction d'ADN.....	75
3.1.3.4. Digestion de l'ADN à partir des biofilms.....	76
3.1.3.5. Extraction d'ADN des gels d'agarose.....	77
3.1.3.6. Ligation à bords francs vs ligation à bords cohésifs.....	79
3.1.3.7. Cellules compétentes et efficacité de transformation.....	79
3.1.3.8. Criblage fonctionnel	80
3.1.3.8.1. Procédure détaillée des étapes de construction des banques d'ADN métagénomiques et de criblage fonctionnel.....	82
3.1.3.8.2. Présentation des 10 clones actifs	84
3.1.3.9. Le clone MHW ₅₃₈	87
3.1.3.9.1. Analyse de la séquence du clone MHW ₅₃₈	89
3.1.3.9.2. Analyse des structures prédictives des séquences peptidiques possibles du clone MHW ₅₃₈	91
3.1.3.10. Synthèse, clivage et purification des peptides	92
3.1.3.10.1. Justification de la synthèse de certains peptides.....	95
3.1.3.11. Effet des peptides sur la croissance de SARM	97
3.1.3.11.1. Effet des dérivés du peptide issu de l'ORF1	97
3.1.3.11.2. Effet des peptides issus des ORF2, 5 et 6.....	100
3.1.3.12. Nouvelle synthèse de peptides	100
3.1.3.12.1. Analyse de la composition en carbone, hydrogène et azote (CHN) des peptides...	101

3.1.3.12.2. Test d'activité des peptides ORF1 _h , ORF1 _i , ORF1 _j , ORF 1 _k , ORF2, l'ORF5 et de l'ORF6 sur SARM et d'autres bactéries cibles	102
---	-----

CHAPITRE 4 : ARTICLE 2..... 104

4.1. DISCOVERY OF AN ANTIMICROBIAL PEPTIDE FROM THE FUNCTIONAL METAGENOMIC SCREENING OF A HOSPITAL BIOFILM 105

4.1.1. Résumé	106
4.1.2. Abstract	107
4.1.3. Introduction	108
4.1.4. Materials and methods	110
4.1.4.1. Biofilm sampling	110
4.1.4.2. DNA extraction	110
4.1.4.3. Bacterial strains and plasmids	110
4.1.4.4. Library construction	111
4.1.4.5. Library screening for antimicrobial activity and phenotype confirmation	111
4.1.4.6. DNA insert amplification and sequencing	112
4.1.4.7. Sequence analysis and structure prediction	112
4.1.4.8. Antimicrobial activity assay with cell culture supernatants	113
4.1.4.9. Peptide synthesis, purification and characterization	113
4.1.4.10. Circular dichroism (CD) analysis	115
4.1.4.11. Nuclear magnetic resonance (NMR) analysis	115
4.1.4.12. Antimicrobial activity assay	115
4.1.5. Results and discussion	116
4.1.5.1. DNA extraction from hospital biofilms	116
4.1.5.2. Metagenomic library construction and screening for antimicrobials molecules	118
4.1.5.2. MHW ₅₃₈ selected clone analysis	119
4.1.5.3. Synthesis, activity test and characterisation of ORF1- and ORF2-derived peptides	122
4.1.6. Conclusion	128

CHAPITRE 5 : DISCUSSION GÉNÉRALE 129

5.1.1. Préambule	130
5.1.2. Choix de <i>Burkholderia thailandensis</i> pour les optimisations de construction de banques d'ADN métagénomiques et de criblage fonctionnel	131
5.1.3. Choix des biofilms d'effluents hospitaliers pour l'étude métagénomique fonctionnelle	133
5.1.4. Rôle des facteurs influençant l'efficacité de transformation	135

5.1.5. Intervention des gènes de la bactérie hôte dans l'expression hétérologue	137
5.1.6. Importance du nombre de clones à cribler	139
5.1.7. Analyse des inserts des clones sélectionnés	140
5.1.8. Effet des peptides antimicrobiens découverts à partir de la séquence de l'insert du clone MHW ₅ 38 sur la croissance de <i>S. aureus</i> résistant à la méthicilline.	143
5.1.9. Autres hypothèses expliquant l'effet du peptide-2 à forte concentration	145
CHAPITRE 6 : CONCLUSION ET PERSPECTIVES	149
6.1. CONCLUSION	150
6.2. DIFFICULTÉS INHÉRENTES À CE PROJET ET PERSPECTIVES	152
6.3. PERSPECTIVES D'AVENIR EN MÉTAGÉNOMIQUE FONCTIONNELLE.....	157
CHAPITRE 7 : BIBLIOGRAPHIE.....	159
CHAPITRE 8 : ANNEXES	179
8.1. ANNEXE A: article revue à soumettre prochainement	180
8.1.1. Abstract	181
8.1.2. Introduction	182
8.1.3. Bacteria as a source of antibiotic.....	183
8.1.4. Functional screening of metagenomic libraries	184
8.1.5. Functional metagenomics : challenges and alternatives	187
8.1.7. Future prospects in functional metagenomics	193
8.1.8. Conclusion.....	195
8.1.9 Bibliography	197
8.2. ANNEXE B: séquences nucléotidique des inserts B1, B2, B3, B4, B5, B6 issus du criblage de la banque d'ADN de <i>Burkholderia thailandensis</i> actifs contre <i>Staphylococcus aureus</i> souche Newman	220
8.2.1. Séquence du clone B1	220
8.2.2. Séquence du clone B2	220
8.2.3. Séquence du clone B3 et B4	222
8.2.4. Séquence du clone B5	223
8.2.5. Séquence du clone B6	225
8.3. ANNEXE C : données supplémentaire relatives au chapitre 3.....	226
8.3.1. Séquence nucléotidique des clones 300 et 128.....	226
8.3.1.1. Clone 300.....	226

8.3.1.2. Clone 128.....	226
8.3.2. Spectres HPLC et MALDI-TOF de certains peptides	226
8.3.2.1. Analyse des brutes de clivage par HPLC analytique	226
8.3.2.2. HPLC des peptides après purification sur colonne	229
8.3.2.3. BLASTn et BLASTx de la séquence de l'insert MHW ₅₃₈	230

Liste des tableaux

Tableau 1.1. Liste des pathogènes prioritaires pour le développement de	5
Tableau 1.2. Récapitulatif de la majorité des études de métagénomiques	12
Tableau 1.3. Lieu d'isolement pour la recherche de fonctions par	17
Tableau 1.4. Exemples de certains vecteurs navettes de type BAC, fosmide.	20
Tableau 1.5. Exemples de quelques souches de <i>Bacillus subtilis</i> ayant une	23
Tableau 2.1. Locus tag identified in the four clones derived from <i>B. thailandensis</i> . 53	
Tableau 2.2. Analyse des ORFs possibles des clones actifs sélectionnés	58
Tableau 2.3. Analyse des ORFs possibles des clones sélectionnées.....	66
Tableau 3.1. Comparaison entre les rendements d'extraction d'ADNe de.....	76
Tableau 3.2. Séquence des neuf peptides dérivés de la séquence du clone.....	91
Tableau 3.3. La séquence et masse exacte (incluant la valeur m/z notée par.....	93
Tableau 3.4. Contenu en carbone, hydrogène et azote des peptides ORF1 _h ,.....	101
Tableau 8.1. eDNA library construction for the identification of antimicrobials.....	185
Tableau 8.2. Unusual eDNA sampling in functional-based metagenomics	188

Liste des figures

Figure 1.1. Cycle de propagation des bactéries résistantes aux antibiotiques.....	4
Figure 1.2. Schématisation de certaines raisons de non culture des bactéries.....	9
Figure 1.3. Vue d'ensemble des études de métagénomiques fonctionnelles).	15
Figure 1.4. Séquence du vecteur pBeloBAC11 réalisé par CLC sequence	19
Figure 1.5. Exemple de phénotypes recherchés lors des criblages de	28
Figure 2.1. Sketch of the device used to spray a thin layer of <i>S. aureus</i>	45
Figure 2.2. <i>Burkholderia</i> DNA size after <i>SphI</i> digestion. 0.8% agarose gel	47
Figure 2.3. Influence of various parameters on the transformation efficiency of	49
Figure 2.4. Distribution of inserts sizes of bacterial artificial chromosome (BAC).	50
Figure 2.5. Examples of growth inhibition zones observed from antibiotic	52
Figure 2.6. Confirmation of the antibacterial activity of four positive <i>E. coli</i> clones ...	52
Figure 2.7. Chromosomal context of the DNA insert identified in the four.....	53
Figure 2.8. Analyse des clones : pBeloBAC11-B1, pBeloBAC11-B2	63
Figure 2.9. Contexte chromosomique des inserts identifiés chez les clones	66
Figure 3.1. Prélèvement des biofilms : (A) prélevé d'un conduit d'égout d'un.....	74
Figure 3.2. Temps de digestion du vecteur pBeloBAC11. (A) digestion du.....	75
Figure 3.3. Comparaison entre les trois trousse d'extraction d'ADN testées,.....	76
Figure 3.4. Construction de banques d'ADN avec et sans extraction sur gel.....	78
Figure 3.5. Exemple de la méthode de confirmation des clones. À partir de 5 µL	81
Figure 3.6. Dispositif complet du <i>bacteriospray</i> . Les boîtes de Pétri de 20x20 cm ² ..	81
Figure 3.7. Comparaison du criblage fonctionnel entre la méthode de double	82

Figure 3.8. Procédure détaillée des étapes de construction et de criblage	83
Figure 3.9. Produit PCR de l'amplification des vecteurs : puits 1 1kb échelle.....	85
Figure 3.10. Image de la morphologie en étoile du clone MHW ₂ 178. Image	85
Figure 3.11. Produit PCR du vecteur pBeloBAC11-MHW ₂ 300 (puits 3),	86
Figure 3.12. Morphologie en petites colonies du clone MHW ₃ 128.	87
Figure 3.13. Observation de la morphologie du clone MHW ₅ 38 sur gélose	89
Figure 3.14. Résultat de la recherche BLASTx de la séquence de l'insert du.....	90
Figure 3.15. Analyse des structures prédictives des neuf séquences peptidiques...92	
Figure 3.16. Profil HPLC analytique des peptides bruts ORF 1 (A) et ORF 1 _a (B)....95	
Figure 3.17. Chromatogrammes obtenus par analyse HPLC démontrant.....97	
Figure 3.18. Effet du peptide ORF1 et de ses analogues à une concentration.....98	
Figure 3.19. Effet des peptides dérivés de l'ORF1 à une concentration.....99	
Figure 3.20. Effet des peptides dérivés des ORF2, 5 et 6 à une concentration.....100	
Figure 3.21. Effet des peptides ORF2 (A) et ORF6 (B) sur la croissance de.....102	
Figure 3.22. Effet des peptides ORF2 (A) et ORF6 (B) sur la croissance de.....103	
Figure 4.1. Extracted DNA from MHW hospital wastewater drainage pipe biofilm..	117
Figure 4.2. Picture of control and inhibitory <i>E. coli</i> DH10B clones against MRSA .	119
Figure 4.3. Visualisation of PCR amplification products on the eDNA fragment.	121
Figure 4.4. Representation of the clone WHW38 inserted DNA, (A) in whole.....	122
Figure 4.5. Growth inhibitory activity of synthetic peptides 1 and 2 against.	125
Figure 4.6. Peptide structure models of peptide-1 and peptide-2 as predicted.	125
Figure 4.7. Circular dichroism analysis of peptide-1 and peptide-2. Peptides.....	126
Figure 4.8. Overlay of NMR NOESY spectra of peptide-1 (blue) and peptide-2	126

Figure 5.1. Décalage du cadre de lecture de deux nucléotides entre le peptide.....	144
Figure 5.2. Alignement de la séquence nucléotidique ORF6 avec la partie	146
Figure 5.3. Comparaison entre la séquence peptidique de ORF2 et celle de.....	147
Figure 5.4. Analyse structurale prédictive de la séquence ORFM par le serveur....	147
Figure 5.5. Séquence en acides aminés de la partie en amont et en aval de la	148
Figure 8.1. HPLC analytique du brut de clivage des peptides sur un gradient.....	228
Figure 8.2. HPLC analytique des peptides sur un gradient linéaire de 0 à 100% ...	229

LISTE DES ABRÉVIATIONS

ADNe	ADN environnementale
ADNg	ADN génomique
ADP	Antimicrobial database
ARNm	ARN messenger
BAC	Bacterial artificial chromosome
BGCs	Biosynthetic gene clusters
BMR	Bactéries multirésistantes
CHN	Carbone hydrogène azote
CD	Circular dichroism
CFU	Colony-forming unit
Da	Dalton
DIC	Diisopropylcarbodiimide
DMF	Dimethylformamide
EDT	Ethanedithiol
EHE	Égouts d'hôpital d'Égypte
EPS	Extracellular polymeric substances
HPLC	High-performance liquid chromatography
Kb	Kilo base
LB	Lysogeny Broth
LSW	Laval sanitary wastewater
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionisation – Time-of- Flight
MCS	Multiple cloning site
MHW	Montreal hospital wastewater
MIC	Minimal inhibitory concentration

NRPS	Non ribosomal peptide synthase
OD	Optical density
O/N	Overnight
ORF	Open reading frame
Ppm	Partie par million
Pb	Paire de base
pH	Potentiel d'hydrogène
IPTG	Isopropyl- β -D-1-thiogalactopyranoside
SARM	<i>Staphylococcus aureus</i> résistant à la méthicilline
PCR	Polymerase chain reaction
pmol	Picomole
T	Transformation efficiency
TFA	Acide trifluoroacétique
RPBS	Ressource Parisienne Bioinformatique Structurale
RMN	Résonance magnétique nucléaire
X-gal	5-bromo-4-chloro-indolyl- β -D-thiogalactopyranoside

CHAPITRE 1 : REVUE DE LA LITTÉRATURE

1.1. ANTIBIOTIQUES FACE À LA RESISTANCE BACTERIENNE AUX ANTIBIOTIQUES

1.1.1. La résistance bactérienne aux antibiotiques

La résistance aux antibiotiques n'est pas un problème nouveau. Depuis la découverte du premier antibiotique, un combat s'est annoncé entre les scientifiques développant de nouveaux antibiotiques et les bactéries développant de nouveaux mécanismes de résistance (Demain & Sanchez, 2009). En effet, suite à la découverte de la pénicilline en 1928 par Fleming, des bactéries résistantes à la pénicilline ont rapidement été observées (World Health Organization., 2014). Les facteurs influençant le développement de la résistance aux antibiotiques ont été classés en deux points :

i. Le premier facteur responsable du développement de la résistance aux antibiotiques est : la surutilisation de ces derniers. En effet, près de 50 % de tous les antibiotiques prescrits en médecine humaine ne sont pas nécessaires ou pas efficaces de la manière prescrite (Blaskovich *et al.*, 2017; Control & Prevention, 2013; Wright, 2017). Par exemple la prescription d'antibiotiques lors d'infections virales, expose les bactéries à des concentrations d'antibiotiques sublétales favorisant la croissance d'une sous-population bactérienne résistante (Blaskovich *et al.*, 2017). De plus, les antibiotiques sont aussi utilisés dans l'élevage animal surtout pour favoriser la croissance des animaux producteurs de ferme. Quoique cette pratique ne soit pas nécessaire, plus d'antibiotiques sont utilisés dans la production alimentaire que dans la médecine humaine (Food & Administration, 2014).

ii. Le deuxième facteur causant le développement de la résistance aux antibiotiques est : la transmission des bactéries résistantes entre l'espèce humaine et animale à travers l'environnement ou l'alimentation (Figure 1.1). Ce réseau de transmission est d'autant plus alarmant en raison d'un phénomène non prédit par Fleming qui est le transfert horizontal de gènes de résistance entre les bactéries, qui leur confère un haut niveau de résistance à plusieurs antibiotiques donnant lieu aux bactéries multirésistantes (BMR) (Davies & Davies, 2010).

Ces BMR se propagent à une vitesse remarquable. Par exemple l'IDSA (Infectious Diseases Society of America) indique que chaque année aux États-Unis, au moins 2 millions de personnes contractent des infections aux BMR graves. Une étude réalisée en 2008 par Roberts et collaborateurs estime que cela coûte environ 20 milliards de dollars au système de santé (aux USA) avec une perte de productivité économique estimée à 35 milliards de dollars par année (Roberts *et al.*, 2009).

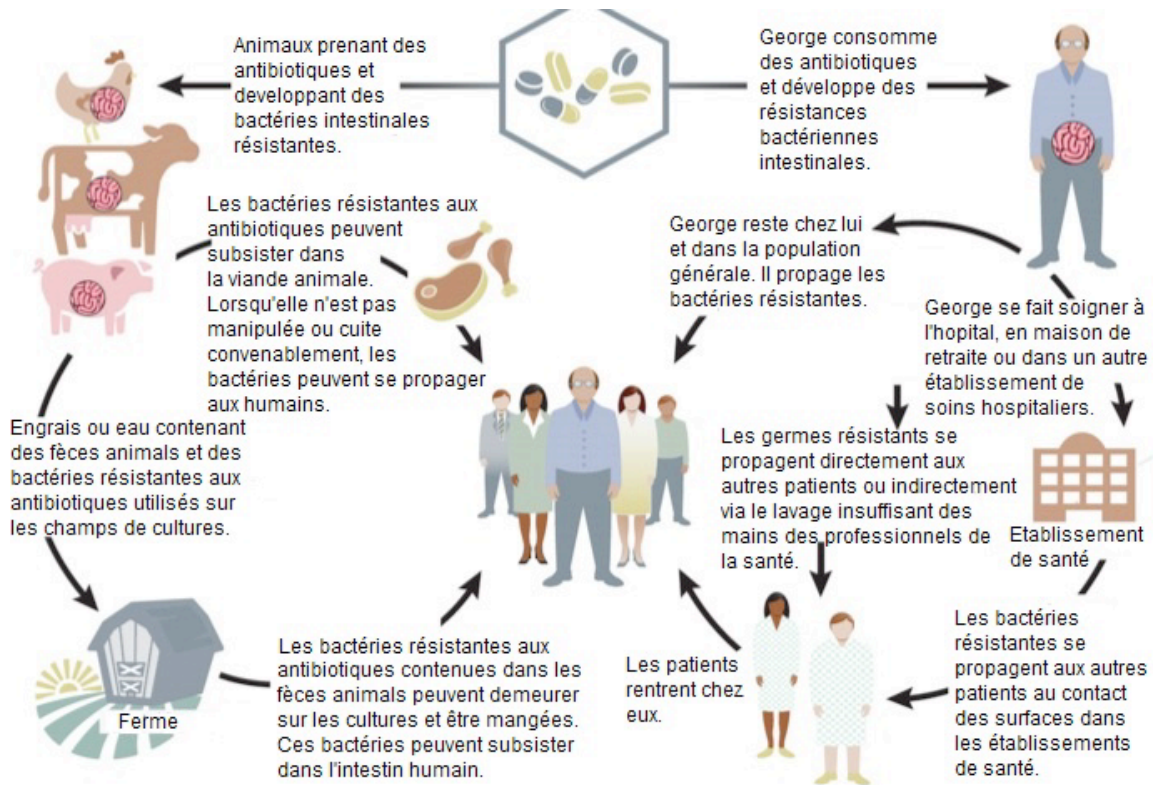


Figure 1.1. Cycle de propagation des bactéries résistantes aux antibiotiques adapté de CDC (Center for Disease Control and Prevention).

1.1.1.1. Liste de bactéries pathogènes

L'organisation mondiale de la santé (OMS) a émis une liste de bactéries pathogènes multirésistantes en fonction des niveaux d'urgence allant de priorité moyenne, élevée à critique, résumée dans le tableau 1.1. Le nombre de ces bactéries multirésistantes est en continuelle augmentation surtout dans les hôpitaux où des patients succombent à des infections après une simple intervention chirurgicale, ou encore où des patients succombent à des infections bactériennes autrefois traitables (Roberts *et al.*, 2009; Wright, 2017; Zhu *et al.*, 2014). Ce fléau de multirésistance aux antibiotiques est particulièrement problématique en raison du manque d'antibiotiques efficaces, surtout pour le traitement des infections causées par les bactéries multirésistantes à Gram négatif, qui ont une membrane externe supplémentaire difficile à pénétrer avec des pompes à efflux empêchant l'antibiotique d'atteindre la concentration requise pour inhiber la croissance des

bactéries (Blaskovich *et al.*, 2017). Plus inquiétant encore, certaines études indiquent l'apparition de souches résistantes à la colistine qui est normalement l'antibiotique d'utilisation de derniers recours (Liu *et al.*, 2016; Poirel *et al.*, 2016). Cela nous conduit petit à petit vers une ère post-antibiotique (World Health Organization., 2014).

Tableau 1.1. Liste des pathogènes prioritaires pour le développement de nouveaux antibiotiques selon l'OMS (World Health Organization, 2017).

Priorité	Espèces bactériennes
1- Priorité critique	<i>Acinetobacter baumannii</i> , résistante aux carbapénèmes
	<i>Pseudomonas aeruginosa</i> , résistante aux carbapénèmes
	<i>Enterobacteriaceae*</i> , résistante aux carbapénèmes et aux céphalosporines de 3eme génération
2- Priorité élevée	<i>Enterococcus faecium</i> , résistante à la vancomycine
	<i>Staphylococcus aureus</i> , résistante à la méthicilline, vancomycine intermédiaire et résistance
	<i>Helicobacter pylori</i> , résistante à la clarithromycine
	<i>Campylobacter</i> , résistante à la fluoroquinolone
	<i>Salmonella spp.</i> , résistante à la fluoroquinolone
	<i>Neisseria gonorrhoeae</i> , résistante aux céphalosporines de 3eme génération et à la fluoroquinolone
3- Priorité moyenne	<i>Streptococcus pneumoniae</i> , pénicilline-non-susceptible
	<i>Haemophilus influenzae</i> , résistance à l'ampicilline
	<i>Shigella spp.</i> , résistante à la fluoroquinolone

Le groupe des Mycobactéries (y compris *Mycobacterium tuberculosis*) n'a pas été intégré dans cette hiérarchisation car il s'agit déjà d'une priorité pour laquelle de nouveaux traitements innovants sont nécessaires de toute urgence.

Le groupe des *Enterobacteriaceae** comprend aussi : *Klebsiella pneumoniae*, *Escherichia coli*, *Enterobacter spp.*, *Serratia spp.*, *Proteus spp.*, *Providencia spp.*, and *Morganella spp.*

1.1.2. Les antibiotiques : le passé, le présent et le futur

Les premiers antibiotiques utilisés en médecine humaine étaient des composés synthétiques ayant une haute toxicité comme par exemple le Salvarsan à base d'arsenic. Mais depuis la découverte de la pénicilline, les antibiotiques dérivés de molécules bioactives naturelles produites par les microorganismes composent la majorité de l'arsenal d'antibiotiques utilisés aujourd'hui. La plupart des antibiotiques naturels utilisés en médecine humaine ont été découverts dans les années 1940 à 1960 par la Plateforme Waksman. Cette plateforme consistait à l'isolement des streptomycètes du sol et à tester leur activité en milieu liquide ou en milieu solide (Lewis, 2013). Depuis la découverte de la streptomycine en 1943, le criblage des Actinomycètes a permis la découverte d'environ 3,000 antibiotiques (90% produits par le genre *Streptomyces*), ce qui reflète à la fois leur relative facilité de culture et leur importance dans la découverte de molécules bioactives (Clardy *et al.*, 2006). À l'époque cette approche a été largement adoptée par les compagnies pharmaceutiques, ce qui a permis d'identifier plus de 28,000 composés (Wright, 2017). Parmi ceux-ci, environ 200 molécules ont directement été utilisées pour le développement pharmaceutique où les antibiotiques sont directement produits par fermentation en milieu liquide (Demain & Sanchez, 2009), par exemple : la pénicilline (*genre Penicillium*), la tétracycline (*genre Streptomyces*), la fosfomycine (*genre Streptomyces*), la streptomycine (*Streptomyces griseus*), l'érythromycine (*Saccharopolyspora erythraea*), etc. Entre 200 à 300 autres molécules ont subi certaines modifications pour augmenter leur efficacité et élargir leur spectre d'action pour le développement de molécules antibiotiques semi-synthétiques comme par exemple : l'amoxicilline, l'azithromycine, la minocycline, etc. Et le reste des antibiotiques ne sera jamais commercialisé principalement à cause de leur efficacité réduite, leur toxicité, leur stabilité, etc. (Wright, 2017). Par la suite, le déclin des

découvertes de nouveaux antibiotiques a duré plus de 30 ans (1980-2010) (The Pew Charitable Trusts, 2016). Ce déclin est principalement dû d'une part à la redécouverte des mêmes antibiotiques produits par plusieurs espèces différentes étudiées par culture bactérienne au laboratoire (The Pew Charitable Trusts, 2016) et d'autre part au désengagement des compagnies pharmaceutiques (Li & Vederas, 2009) qui concentrent maintenant leurs efforts vers les maladies chroniques beaucoup plus rentables en retours d'investissement (Wright, 2017). En effet, lorsqu'un antibiotique est prêt à être développé, d'autres défis réglementaires sont à prévoir en plus des faibles rendements de retours d'investissement (The Pew Charitable Trusts, 2016).

1.1.3. Les microorganismes comme source de nouveaux antibiotiques

La diversité microbienne sur la biosphère terrestre est estimée à plus de 10^{30} cellules et à environ 10^7 - 10^9 bactéries par gramme de sol. La diversité procaryote cultivable fait partie de 30 des 80 phyla bactériens actuellement reconnus et 3 des 26 phyla Archaeal (Overmann *et al.*, 2017). C'est-à-dire que le reste de cette diversité est encore très peu connu car seulement 1% de cette dernière demeure cultivable par les méthodes classiques *in vitro* (Culligan *et al.*, 2014). Cette théorie de non-cultivabilité est d'autant plus appuyée par des analyses de l'ARN ribosomique qui nous indiquent qu'environ un tiers des divisions bactériennes n'ont pas de représentants de culture et sont connues seulement par séquençage de l'ADNr 16S, suggérant que le nombre d'antibiotiques à découvrir à partir de sources non cultivables est de deux à trois ordres de grandeur supérieurs à celui des sources cultivables (Chang & Brady, 2013a; Clardy *et al.*, 2006). On appelle «plateforme Waksman» l'approche classique d'identification de nouveaux antibiotiques par la culture des microorganismes (Lewis, 2012). Toutefois, cette approche n'est pas phylogénétiquement dirigée car elle sélectionne la diversité microbienne à multiplication rapide (Figure 1.2).

Pour isoler de nouvelles lignées environnementales, les conditions de culture en laboratoire doivent imiter les conditions dans lesquelles les bactéries se retrouvent dans leurs conditions naturelles, en matière de nutriments, de concentration en oxygène, de pH et surtout de leur association avec les autres organismes et/ou du besoin de l'utilisation de facteurs de croissances particuliers qui sont produits par les microorganismes voisins (D'Onofrio *et al.*, 2010; Hugenholtz, 2002). Par exemple, remplacer l'agar par la gomme gellane permet d'augmenter la diversité des bactéries cultivées (Cardenas & Tiedje, 2008; Tamaki *et al.*, 2005). Aussi, l'utilisation de milieux de cultures pauvres en oligo-éléments permet de retarder la croissance des bactéries à croissance rapide, laissant la place aux bactéries à croissance lente (Daniel, 2004).

Dans cette optique, une approche ingénieuse a été rapportée par Kaeberlein et collaborateurs (Kaeberlein *et al.*, 2002). En se basant sur le fait que les bactéries vivent rarement en culture pure et que certaines bactéries ne peuvent être mises en culture par des techniques classiques de laboratoire, les auteurs ont conçu un dispositif de culture sur des membranes semi-perméables déposées directement dans le sol. Les nutriments et les facteurs de croissance de l'environnement naturel peuvent ainsi diffuser, et ce en respectant les interactions entre les microorganismes du sol. Ces dispositifs sont conçus afin qu'une seule souche bactérienne soit cultivée dans une petite chambre de diffusion. De cette façon, les bactéries incultivables peuvent être plus facilement cultivées en culture pure. Cette approche permet d'améliorer le taux de croissance de ce type de bactéries d'environ 50% et a récemment conduit à la découverte d'une nouvelle molécule antibiotique : la Teixobactine (Ling *et al.*, 2015). Néanmoins, l'accès à ce type de bactéries demeure difficile car le développement de méthodes pour cultiver l'énorme diversité de la microflore risque d'être lent et requiert une meilleure compréhension de la physiologie de ces bactéries.

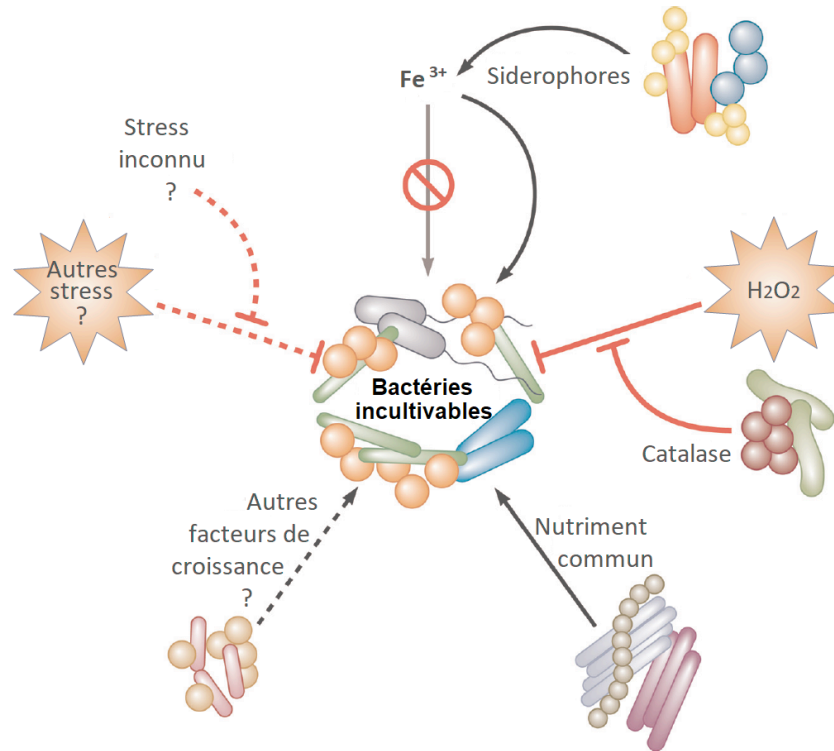


Figure 1.2. Schématisation de certaines raisons de non culture des bactéries adapté de Stewart *et al.*, (Stewart, 2012).

1.1.4. L'innovation pour découvrir de nouveaux antibiotiques

L'étude des génomes des microorganismes producteurs d'antibiotiques tels que le groupe des Actinomycètes a montré que leur capacité de production d'antibiotiques surpasse leur potentiel connu par la présence de nombreux gènes silencieux qui ne sont pas exprimés en conditions usuelles de laboratoire (Alduina & Gallo, 2012a).

Des approches innovantes et encourageantes ont émergé dans le but d'explorer d'autres horizons de découvertes comme par exemple activer les gènes cryptiques chez ce groupe des Actinobactéries (Jiang *et al.*, 2015; McKenzie *et al.*, 2010; Yamanaka *et al.*, 2014).

De plus considérant, que seulement 1% de la diversité microbienne soit facilement cultivable en laboratoire, la solution pour découvrir de nouveaux

antibiotiques viendrait aussi de l'étude de ces bactéries non-cultivables. Une approche très prometteuse impliquant l'extraction directe de l'ADN d'un environnement en vue de son expression hétérologue a été développée. Cette technique appelée métagénomique fonctionnelle permet d'accéder aux métabolites secondaires produits par tous types de bactéries. Cette approche existant depuis une vingtaine d'années est très prometteuse pour la découverte de nouvelles molécules bioactives (Handelsman *et al.*, 1998a; Mullany, 2014; Wooley *et al.*, 2010).

D'autres stratégies intéressantes peuvent être entreprises comme par exemple l'isolement de certains composés qui agissent comme éliciteurs améliorant l'efficacité des antibiotiques déjà existants (Wright, 2017).

1.2. METAGENOMIQUE : ALTERNATIVE POUR ACCEDER À DE NOUVELLES MOLECULES ANTIBIOTIQUES

Les études basées sur la métagénomique peuvent être faite de deux manières. Dans la première, l'ADN génomique environnemental est directement séquencé, puis les séquences résultantes sont assemblées, analysées et comparées aux séquences déposées dans les banques de données telles que GenBank. La fonction du gène et le type de protéines qui pourraient être produites sont prédits. Avec le développement des nouvelles techniques de séquençages à haut débit (next generation sequencing (NGS)) à prix abordable, cette approche est favorisée d'autant plus qu'aucun travail au laboratoire n'est nécessaire à part l'étape d'extraction d'ADN utile au séquençage. Néanmoins, relier les séquences de gènes à leurs fonctions est la clé des découvertes car avec la métagénomique par séquençage, l'annotation des gènes est limitée aux séquences déposées dans les

bases de données (Mullany, 2014). Des outils tels antiSMASH restent en effet contraint par les informations déjà disponibles (Blin *et al.*, 2017). La deuxième méthodologie est nommée métagénomique fonctionnelle. Elle est basée directement sur le produit des gènes. L'ADN environnemental (ADNe) est extrait, cloné dans un vecteur, puis transformé dans des hôtes d'expressions hétérologues pour générer des banques d'ADN métagénomique. Les clones sont alors testés pour leurs fonctions, activités antibiotiques, enzymatiques ou autre, puis la séquence d'ADN cloné responsable du phénotype est ensuite déterminée (Courtois *et al.*, 2003; Handelsman, 2007; Lussier *et al.*, 2011; Mullany, 2014; Rondon *et al.*, 2000b). Le clonage de l'ADNe dans un hôte a été proposé pour la première fois en 1985 par Pace *et al.* dans le but de cloner l'ADN provenant des picoplanktons, à des fins de séquençage de l'ARNr 16S (Simon & Daniel, 2011). Puis dans les années 1990, Jo Handelsman a proposé l'approche fonctionnelle pour étudier le rôle des bactéries de l'environnement qui ne peuvent être cultivées en laboratoire (Handelsman *et al.*, 1998b). Cette démarche a révolutionné la façon de rechercher de nouveaux gènes et/ou des produits d'expression de gènes conduisant à la découverte de nouvelles molécules bioactives sans avoir à cultiver les microorganismes. L'avantage de cette méthode est que des gènes inconnus peuvent être étudiés et leurs fonctions reliées directement aux gènes de la séquence insérée. La principale limite se situe au niveau du type d'hôte à utiliser (Handelsman, 2007). Quelques antibiotiques ont été découverts par métagénomique fonctionnelle (Tableau 1.2).

Tableau 1.2. Récapitulatif de la majorité des études de métagénomiques fonctionnelles pour la découverte de nouveaux antibiotiques en fonction de l'origine de l'ADNe, la taille de l'insert, le type de vecteur choisi et l'hôte d'expression hétérologue.

Antimicrobiens	Origine de l'ADN environnementale	Taille de l'insert (kb)	Type de vecteur	L'hôte d'expression hétérologue	Clones actifs/clones criblés	L'activité biologique	Références
Protéine de 55 KDA	Sol de l'agriculture	27	BAC pBeloBAC11	<i>E. coli DH10B</i>	1/28224	<i>B. subtilis</i> , <i>S. aureus</i>	(Rondon <i>et al.</i> , 2000b)
Terragine A, B, C, D, E	Sol (ND)	-	Vecteur (ND)	<i>S. lividans</i>	2/1020	<i>E. coli</i> , <i>B. subtilis</i>	(Wang <i>et al.</i> , 2000)
Indigo & Indirubine	Sol issu de Nouvelle-Angleterre	37	BAC pBTP2	<i>E. coli DH10B</i>	4/12 000	<i>B. subtilis</i> , <i>S. aureus</i>	(MacNeil <i>et al.</i> , 2001b)
Violaceine	Sol issu de Ithaca, New York	-	Cosmide SuperCos1	<i>E. coli</i>	-	<i>B. subtilis</i> , <i>S. aureus</i> , <i>Streptococcus</i> sp.	(Brady <i>et al.</i> , 2001a)*
Turbomycine	Sol (ND)	44.5	BAC pBeloBAC11	<i>E. coli DH10B</i>	-/24 546	<i>E. coli</i> , <i>B. subtilis</i> , <i>S. aureus</i> , ...	(Gillespie <i>et al.</i> , 2002)
N-acyl tyrosine	Sols issu de Ithaca, New York	-	Cosmide	<i>E. coli</i>	-	-	(Brady <i>et al.</i> , 2002)*
Molécule antimicrobienne	Sol arable de France	50	Cosmide navette pOS7001	<i>E. coli DH10B</i> , <i>S. lividans</i> <i>TK24</i>	1/5 000	<i>B. subtilis</i>	(Courtois <i>et al.</i> , 2003)
Palmitoylputrescine	Réservoir de broméliacées Costa Rica	-	Cosmide pCSLF16	<i>E. coli</i>	-	<i>B. subtilis</i>	(Brady & Clardy, 2004)*
Long-chain N-acyl tyrosine	Boston, Ithaca & Costa Rica	-	Cosmide	<i>E. coli</i>	-	<i>B. subtilis</i>	(Brady <i>et al.</i> , 2004)*

Indigo & indirubine	Sols issus de Jindong Valley, Corée	35	Fosmide pEpiFOS-5	<i>E. coli</i>	2/113700	<i>B. subtilis</i>	(Lim <i>et al.</i> , 2005)
Isocyanide	Sol de Boston	-	Cosmide pCSLG18	<i>E. coli</i> ,	-	<i>B. subtilis</i>	(Brady & Clardy, 2005a)*
Long-chain N-acyl tryptophane et arginine	Sol (ND)	-	Cosmide pSuperCosI	<i>E. coli</i>	-	<i>B. subtilis</i>	(Brady & Clardy, 2005b)
Molécule antibactérienne	Sol environnemental	-	Cosmide pWEB, pWEB::TNC	<i>E. coli</i> XL-1Blue MR	1/20 000	<i>B. subtilis</i>	(Brady, 2007)
Type III Polykétide synthase	Sol de Oregon et Pennsylvanie	-	Cosmide navette pTR101	<i>E. coli</i> , <i>R. metallidurans</i>	ND	<i>B. subtilis</i>	(Craig <i>et al.</i> , 2009)
Molécule antibactérienne	Sol de Pennsylvanie et Oregon	-	Cosmide navette pJWC1	<i>E. coli</i> , <i>A. tumefaciens</i> , <i>B. graminis</i> , <i>C. vibrioides</i> , <i>E. coli</i> , <i>P. putida</i> , <i>R. metallidurans</i>	8/-	<i>B. subtilis</i>	(Craig <i>et al.</i> , 2010)
Molécule antibactérienne	Forêt, Belgique	12	Plasmide navette PHT01	<i>E. coli</i> , <i>B. subtilis</i>	1/25 000	<i>B. cereus</i>	(Biver <i>et al.</i> , 2013)
Molécule antibactérienne	Sol de l'Arizona	-	Cosmide pJWC1, pJSS	<i>E. coli</i> , <i>Ralstonia metallidurans</i>	19/700 000	<i>B. subtilis</i> tétracycline-résistante	(Iqbal <i>et al.</i> , 2014)
N-acyltyrosine	Biofilm de drainage d'abattoir	2-10	Vecteur navette pEBP18	<i>E. coli</i> DH10B	-	<i>B. subtilis</i>	(Thies <i>et al.</i> , 2016)
Colicine V	Microbiome humain	35	Cosmide navette pJWC1	<i>E. coli</i> EC100	12/350 000	<i>E. coli</i>	(Cohen <i>et al.</i> , 2018)

1.2.1. Construction de banques d'ADN en métagénomique fonctionnelle

La réussite d'une étude métagénomique fonctionnelle dépend de plusieurs facteurs (Figure 1.4). La construction des banques d'ADN métagénomiques est l'étape cruciale de la métagénomique fonctionnelle. Elle se subdivise en plusieurs petites étapes toutes importantes pour le bon fonctionnement de la technique. Cela débute du lieu d'échantillonnage de l'ADN au vecteur de clonage choisi, à la souche d'expression hétérologue adaptée et au type d'induction choisi. Toutes ces étapes doivent être pensées avant de débiter une étude de métagénomique fonctionnelle en fonction de la molécule bioactive recherchée (Lam *et al.*, 2015).

Le type de banque d'ADN métagénomique dépend en premier lieu de la qualité et de la taille de l'ADN obtenu après extraction, du vecteur employé, de l'hôte utilisé, de la molécule bioactive ciblée et de la méthode de criblage choisie (Simon & Daniel, 2010).

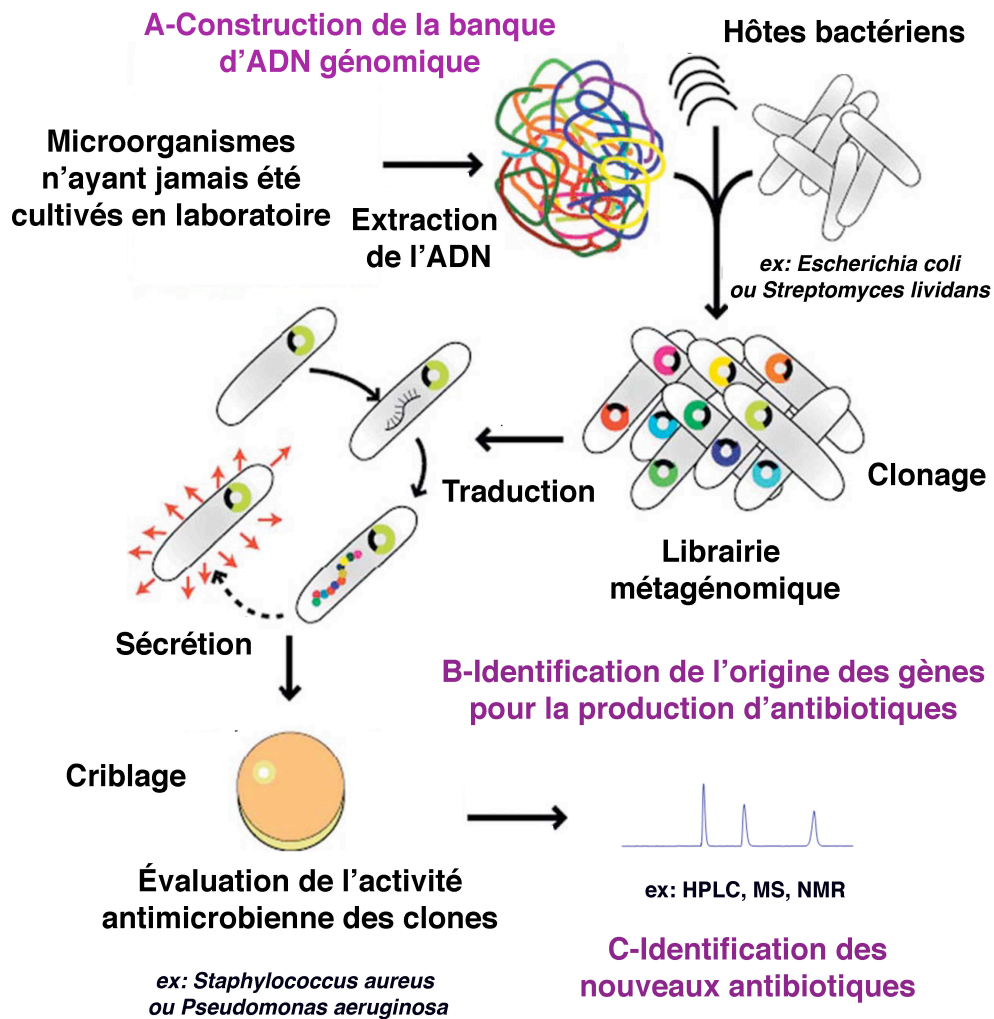


Figure 1.3. Vue d'ensemble des études de métagénomiques fonctionnelles adapté de Sabree *et al.* (Sabree *et al.*, 2009).

1.2.1.1. Méthode d'extraction d'ADN

L'extraction de l'ADN est la première étape dans toutes études métagénomiques. Deux méthodes d'extractions d'ADN sont citées dans la littérature : directe et indirecte. La méthode d'extraction indirecte consiste en la dispersion entre les particules du sol, la séparation des cellules du sol par centrifugation à faible vitesse

ou par gradient de densité et enfin la lyse bactérienne pour l'extraction et la purification de l'ADN. Cette méthode est nécessaire lors de l'isolement à partir de sol riche en composés qui interfèrent avec l'extraction d'ADN car durant cette dernière, les agrégats sont éliminés. Or, certaines bactéries forment des agrégats avec le sol, cette diversité microbienne sera donc perdue lors de l'extraction indirecte (Robe *et al.*, 2003).

La méthode d'extraction directe est la plus utilisée (Lussier *et al.*, 2011; Owen *et al.*, 2013; Rondon *et al.*, 2000b). Elle consiste à une lyse *in situ* des cellules bactériennes, suivie de l'extraction et de la purification de l'ADN. Elle permet d'atteindre des rendements d'ADN au moins 10 fois supérieurs à l'extraction indirecte (Robe *et al.*, 2003).

Le facteur principal qui influence la taille de l'ADN lors de son extraction est la méthode de lyse de la paroi bactérienne qui peut être physique ou chimique et enzymatique. La méthode physique comprend par exemple la congélation/décongélation et le «Bead-Beating». Ces techniques sont agressives et peuvent causer des dommages physiques à l'ADN, en générant de petits fragments. Cependant, elles permettent d'accéder à l'ADN de tous types de microorganismes y compris les plus récalcitrants. Une étude publiée par Burgmann et collaborateurs démontre que le rendement d'extraction d'ADN augmente avec l'augmentation du temps et de la vitesse du «Bead-Beating», mais que des fragments d'ADN de petites longueurs sont générés (Bürgmann *et al.*, 2001). La combinaison entre les méthodes chimiques et enzymatiques (SDS-lysozyme) peut cependant être mieux adaptée pour générer des fragments d'ADN d'une longueur entre 40-90 kb avec un rendement acceptable (Krsek & Wellington, 1999; Robe *et al.*, 2003).

1.2.1.2. Source d'échantillonnage de l'ADN

Le choix du lieu d'échantillonnage de l'ADN en vue d'une étude métagénomique est l'étape clé d'une étude métagénomique fonctionnelle (Culligan *et al.*, 2014). Un lieu d'échantillonnage réfléchi, en fonction du phénotype recherché augmente la probabilité de découvrir de nouveaux gènes en raison des capacités d'adaptation physiologique des microorganismes aux conditions environnementales dominantes (Vercammen *et al.*, 2013; Zhang *et al.*, 2017). Le tableau 1.3 résume les lieux d'isolement de quelques études qui ont mené à la découverte d'enzymes et de gènes de résistance à partir de lieux d'isolement d'ADN hostiles.

Tableau 1.3. Lieu d'isolement pour la recherche de fonctions par métagénomique fonctionnelle.

Type de molécule	Lieu d'isolement	Références	
Enzymes	β -glucosidase	Microflore du nématode <i>Bursaphelenchus</i> <i>xylophilus</i>	(Zhang <i>et al.</i> , 2017)
	β -galactosidase	Champs de maïs	(Cheng <i>et al.</i> , 2017)
	β -glucanase	Sol volcanique	(Angelov <i>et al.</i> , 2017)
	Exonucléase (ExoMeg1)	Biome semi-aride au Brésil	(Silva-Portela <i>et al.</i> , 2016)
	Estérase	Sources chaudes	(Leis <i>et al.</i> , 2015)
	Amylases, cellulases	Rumen des vaches	(Wang <i>et al.</i> , 2013)
	Flavonoïdes	Excrément d'éléphants	(Rabausch <i>et al.</i> , 2013)
	Activité hydrolytique	lléon humain	(Cecchini <i>et al.</i> , 2013)
	ADN polymérase	Glaciers	(Simon <i>et al.</i> , 2009)

Gènes de résistance aux antibiotiques	Tétracycline	Production de fromages	(Florez <i>et al.</i> , 2017)
	Nouvelle classe de β -lactamases	Rivière polluée	(Vercammen <i>et al.</i> , 2013)
	β -lactamases	Eaux usées	(Uyaguari <i>et al.</i> , 2011)
	Ampicilline, amoxicilline, pénicilline	Excréments de goélands	(Martiny <i>et al.</i> , 2011)
	Tétracycline	Métagénome oral	(Diaz-Torres <i>et al.</i> , 2003)
	β -lactamases	Sol d'Alaska	(Allen <i>et al.</i> , 2009)

1.2.1.3. Vecteurs de clonage

En fonction du vecteur de clonage utilisé pour la construction de banques d'ADN métagénomiques, deux types de banques d'ADN peuvent être construites : les banques d'ADN à petits inserts et les banques d'ADN à larges inserts.

Les banques d'ADN à petits inserts sont de type plasmidique et peuvent généralement intégrer des fragments d'ADN d'une taille inférieure à 10 kb. Elles sont utilisées pour identifier de nouvelles molécules bioactives codées par un gène unique ou un petit opéron (Biver *et al.*, 2013). L'avantage de ce type de banques est leur nombre élevé de copies par cellules, ce qui permet la détection de gènes faiblement exprimés et la possibilité d'une transcription par le promoteur du vecteur. Son inconvénient est le nombre important de clones à tester pour déterminer celui ou ceux qui sont actifs (Culligan *et al.*, 2014; Mullany, 2014).

Les banques d'ADN à larges inserts sont de trois types: cosmide, fosmide ou chromosome bactérien artificiel (BAC). Le cosmide est un hybride entre un plasmide et un phage contenant une séquence *cos*. Il peut intégrer des fragments d'ADN de 15 à 40 kb. Le fosmide quant à lui est similaire au cosmide, mais basé

sur le facteur F de *E. coli* (Kakirde *et al.*, 2010) et peut intégrer des fragments d'ADN d'une taille entre 25 à 45 kb. Les BAC ont initialement été développés pour le clonage de l'ADN humain (Kim *et al.*, 1996). Ils peuvent intégrer des fragments d'ADN jusqu'à 100 à 200 kb (Sosio *et al.*, 2000). Les BAC sont maintenus en faibles nombres de copies dans l'hôte, ce qui peut être un avantage sachant que les vecteurs à faibles copies sont plus stables (Culligan *et al.*, 2014). Un exemple de vecteur de type BAC le plus utilisé dans les études métagénomiques fonctionnelles est le vecteur pBeloBAC11 (Figure 1.5) (Gillespie *et al.*, 2002; Liles *et al.*, 2003; MacNeil *et al.*, 2001a; Riesenfeld *et al.*, 2004; Rondon *et al.*, 2000a; Rondon *et al.*, 2000b; Tao & Zhang, 1998). Il a été développé à partir du vecteur pBAC108L pour la construction de bibliothèques d'ADN humains à larges inserts (Kim *et al.*, 1996). Ce vecteur peut seulement se maintenir chez l'hôte d'expression hétérologue *E. coli*, il comprend une sélection bleu/blanc lors de la construction de banques d'ADN métagénomiques, un promoteur T7, un marqueur de résistance au chloramphénicol (CM^r) et un site de clonage multiple avec trois sites de restriction uniques : *Bam*HI, *Hind*III et *Sph*I.

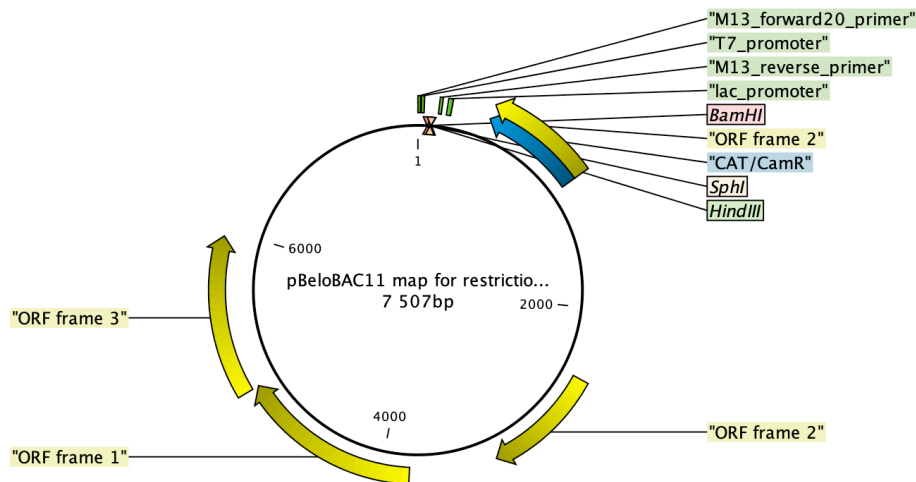


Figure 1.4. Séquence du vecteur pBeloBAC11 réalisé par CLC sequence viewer 7.

Un autre exemple de vecteurs utile à intégrer dans les études métagénomiques est le vecteur de type navette qui permet la construction et l'expression des banques d'ADN chez plus d'un hôte d'expression hétérologue (Daniel, 2004; Martinez *et al.*, 2004; Rondon *et al.*, 2000b). De nombreux vecteurs navette de type plasmide, cosmide, ou BAC ont été développés et utilisés en métagénomique fonctionnelle (tableau 1.4). Cependant, le transfert du vecteur navette d'un hôte à un autre par conjugaison pose parfois certaines difficultés, surtout lorsque le vecteur en question est un BAC (Biver *et al.*, 2013; Lussier, 2011).

Tableau 1.4. Exemples de certains vecteurs navettes de type BAC, fosmide, cosmide et plasmide utilisés en métagénomique fonctionnelle.

Vecteur	Nom	Hôtes	Références
BAC	SuperBac	<i>E.coli</i> , <i>Bacillus</i> , <i>Enterococcus</i> , <i>Staphylococcus</i> , <i>Streptococcus</i>	(Handelsman <i>et al.</i> , 2002)
	pMBD14	<i>S. lividans</i> , <i>P. putida</i>	(Banik & Brady, 2008; Martinez <i>et al.</i> , 2004)
	pGNS-BAC	Diverses Gram négatif	(Kakirde <i>et al.</i> , 2011)
Fosmide	pSTREPTOBAC V	<i>E. coli</i> , <i>S. lividans</i>	(Alduina & Gallo, 2012a)
	pRS44	<i>P. fluorescens</i> , <i>X.</i> <i>campestris</i>	(Aakvik <i>et al.</i> , 2009)
	pOS7001	<i>E. coli</i> , <i>S. lividans</i>	(Courtois <i>et al.</i> , 2003)

Cosmide	pJWC1	<i>R. metallidurans</i> , <i>E. coli</i>	(Craig <i>et al.</i> , 2009)
	pFX583	<i>E. coli</i> , <i>S. lividans</i>	(Lussier <i>et al.</i> , 2011)
	pFX64	<i>E. coli</i> , <i>B. subtilis</i>	(Lussier, 2011)
	pZRJ362	<i>S. lividans</i> , <i>E. coli</i>	(Li <i>et al.</i> , 2013)
	pOJ436	<i>E. coli</i> , <i>S. lividans</i>	(Chang & Brady, 2013a)
Plasmide	pEBP	<i>E. coli</i> , <i>B. subtilis</i> , <i>S. lividans</i>	(Troeschel <i>et al.</i> , 2012)
	pHT01	<i>E. coli</i> , <i>B. subtilis</i>	(Biver <i>et al.</i> , 2013)

1.2.1.4. Les souches d'expression hétérologue

1.2.1.4.1. Souches d'expression *Escherichia coli*

Malgré le fait que *E. coli* soit connue pour être un hôte d'expression pauvre (Alduina & Gallo, 2012b), elle demeure la plus utilisée dans les études de métagénomiques fonctionnelles (Brady *et al.*, 2001b; Brady *et al.*, 1999; Gillespie *et al.*, 2002; Iqbal *et al.*, 2016; Lim *et al.*, 2005; Thies *et al.*, 2016) puisqu'elle croît rapidement sur des milieux peu coûteux, est facile à manipuler génétiquement et est très utile en fermentation. Les banques d'ADN métagénomiques sont généralement d'abord construites chez *E. coli* avant d'être transférées chez un autre hôte d'expression (Iqbal *et al.*, 2016). Quelques projets en métagénomique fonctionnelle avec *E. coli* comme seule souche d'expression ont permis de découvrir de nouvelles molécules bioactives (Brady, 2007; Brady *et al.*, 2002; Rondon *et al.*, 2000a; Thies *et al.*, 2016).

La majorité des souches de *E. coli* utilisées en laboratoire dérivent des isolats classiques K-12 et B (Delbruck, 1945; Tatum & Lederberg, 1947). La comparaison entre ces deux types de souches est réalisée avec les descendances de K-12 (MG1655 et W3110) et de B (REL606 et BL21) (Daegelen *et al.*, 2009). La comparaison des génomes des souches *E. coli* K-12 et *E. coli* B indique une similarité de 99% accompagnée de certaines divergences (Jeong *et al.*, 2009). La différence entre ces deux souches est liée majoritairement au nombre de séquences d'insertion (IS). La séquence d'insertion *IS1* est majoritairement présente chez les souches B avec 20 copies contre seulement 6 pour les souches K-12. La séquence d'insertion *IS5* est seulement présente chez les K-12 (Schneider *et al.*, 2002). L'intégration de ces séquences d'insertions dans un gène l'inactive. En effet, la souche BL21 ne produit pas de protéase Lon suite à l'insertion de *IS186* dans le promoteur du gène, et ne produit pas non plus la protéase *ompT* suite à une délétion spontanée, ces deux protéases ayant pour fonction de dégrader les protéines sécrétées. Cela permet à la souche BL21 d'être mieux adaptée pour l'expression de gènes hétérologues, prévenant la dégradation des molécules produites (Jeong *et al.*, 2009; Studier *et al.*, 2009). Une caractéristique intéressante de certains dérivés de la souche BL21 est l'intégration du phage DE3 dans le site d'attachement λ , possédant dans le génome du phage, le gène de l'ARN polymérase T7 sous le contrôle du promoteur lacUV5, donnant lieu aux souches BL21(DE3) (Schneider *et al.*, 2002). Cependant BL21 n'est pas un bon choix pour le clonage car elle est *recA+*. Une autre différence importante est la délétion d'un groupe de gènes (*gspD-K*) chez *E. coli* K-12 impliqué dans le système de sécrétion de type II, ce qui laisse croire que les souches de type B ont des capacités de sécrétion absentes chez le type K-12 (Jeong *et al.*, 2009).

1.2.1.4.2. Souche d'expression *Bacillus subtilis*

Bacillus subtilis est parfois utilisé pour l'expression hétérologue d'un ADN en vue d'une étude métagénomique fonctionnelle. Cette bactérie à Gram-positif possède différentes propriétés d'expression, notamment la capacité à produire de nombreux enzymes excrétés directement dans le milieu, ne présente pas de biais dans l'utilisation des codons, et est très facile à manipuler génétiquement notamment pour son utilisation en fermentation (Wong, 1995). Néanmoins, l'expression hétérologue chez *Bacillus* présente certaines limites, surtout pour l'expression hétérologue d'enzymes. En effet, le principal problème est sa production d'un nombre élevé de protéases extracellulaires causant la dégradation des protéines hétérologues produites, réduisant considérablement le taux de production. Jusqu'à présent, 8 protéases ont été rapportées dans la littérature AprE, NprE, NprB, Mpr, Epr, Bpr, Vpr, HtrA, ce qui a permis la construction de souches de *B. subtilis* ayant une très basse activité protéolytique extracellulaire, dans le but d'améliorer l'expression hétérologue (Westers *et al.*, 2004).

L'expression d'un ADNe sous le contrôle du promoteur de l'ARN polymérase T7 chez *B. subtilis* est devenue possible grâce à l'intégration du gène codant pour la T7 polymérase au niveau du locus *amyE* de la souche *B. subtilis* BD170, sous l'induction du promoteur inductible xylose (Conrad *et al.*, 1996). Le tableau 1.6 présente quelques exemples de souches *B. subtilis* ayant une activité protéolytique réduite.

Tableau 1.5. Exemples de quelques souches de *Bacillus subtilis* ayant une activité protéolytique réduite.

<i>Bacillus subtilis</i>	Activité protéolytique inactive	Références
WB800, WB700N, WB700NHM, WB800, et WB800HM	<i>nprE nprB aprE epr mpr bpr vpr wprA</i>	(Wu <i>et al.</i> , 2002)
Dpr8	<i>nprE nprB aprE epr mpr bpr vpr wprA</i>	(Kodama <i>et al.</i> , 2007)

1.2.1.4.3. Souche d'expression du genre *Streptomyces*

Appartenant aux actinobactéries, le genre *Streptomyces* produit une gamme variée de métabolites secondaires tels que des antibiotiques, antifongiques, antiviraux, antitumoraux, antihypertensifs, immunosuppresseurs, enzymes, etc. Parmi ce genre, *S. coelicolor*, *S. albus* et de *S. lividans* (tableau 1.1) sont couramment utilisés pour l'expression hétérologue d'un ADN exogène (Courtois *et al.*, 2003; Martinez *et al.*, 2004) surtout si celui-ci provient d'actinobactéries (Iqbal *et al.*, 2016). Par exemple l'utilisation de *S. lividans* comme souche d'expression hétérologue a permis de découvrir la terragine (Wang *et al.*, 2000). L'un des principaux avantages des actinobactéries comme hôte est qu'ils comprennent généralement les précurseurs primaires nécessaires pour la synthèse de produits naturels à partir d'ADN exogène, vu qu'ils sont responsables de la production de nombreux métabolites secondaires. De plus, l'excrétion du métabolite produit se fait directement dans le milieu de culture (Iqbal *et al.*, 2016). Dans certains cas, le produit de l'expression hétérologue étant absent ou produit en très petite quantité, des modifications au niveau de l'hôte sont alors nécessaires pour améliorer cette production (Alduina & Gallo, 2012b). Néanmoins, l'analyse génomique des

Streptomycètes révèle qu'ils comprennent plus de 20 *clusters* de gènes responsables de la production de précurseurs et/ou responsables de la production de métabolites secondaires qui pourrait interférer avec la production hétérologue du métabolite voulue. La délétion de certains de ces groupes de gènes comme celui qui code pour la production de l'actinorhodine chez *S. lividans* améliore la production hétérologue (Mary & Betlach, 1999).

1.2.1.5. Systèmes d'expression

L'expression dans un hôte tel que *E. coli* est souvent compromise par une transcription inefficace ou incomplète des gènes regroupés en *clusters*, attribuable soit à une faible reconnaissance des promoteurs par l'ARN polymérase native ou bien par la présence de sites de terminaison de transcription. Un moyen efficace employé pour augmenter l'expression hétérologue consiste à stimuler l'expression génique au niveau transcriptionnel via des systèmes de promoteurs puissants (Arvani *et al.*, 2012). Le promoteur fort le plus utilisé pour l'expression hétérologue est le promoteur T7 (Arvani *et al.*, 2012; Brady & Clardy, 2000; Lussier *et al.*, 2011; MacNeil *et al.*, 2001b) dans lequel l'ARN polymérase du bactériophage T7 est exprimé hors du chromosome de l'hôte sous la régulation du promoteur *lacUV5* régulé par le répresseur *lacI*. Ce système est inductible à l'ajout d'un analogue du lactose : l' isopropyl- β -D-1-thiogalactopyranoside (IPTG) dans le milieu (Studier & Moffatt, 1986). De nombreux vecteurs ont été construits comprenant le promoteur T7, un site de liaison des ribosomes (RBS), suivi d'un site de clonage multiple (MCS) et d'une séquence de terminaison (Gabor *et al.*, 2004). Un autre système d'expression intéressant est celui qui se base sur le gène *sacB* qui code pour la levansucrase qui convertit le substrat sucrose en levane, un composé toxique pour *E. coli*. L'avantage de ce système est que seuls les clones recombinants peuvent croître. D'autres systèmes de promoteurs avec divers facteurs d'induction tels que l'arabinose, la tétracycline, peuvent aussi être employés (Lutz & Bujard, 1997).

1.2.2. Criblage fonctionnel des banques d'ADN métagénomiques

Les méthodes de criblage fonctionnel sont basées sur l'expression hétérologue des gènes des banques d'ADN métagénomiques pour sélectionner le phénotype recherché (Figure 1.6). L'avantage de cette approche est d'identifier de nouvelles fonctions pas facilement prévisibles par les séquences, ce qui permet de découvrir de nouveaux gènes qui n'ont pas d'homologie de séquences avec des gènes déjà connus (Cheng *et al.*, 2017). Ce criblage fonctionnel peut cibler différentes fonctions classées en trois catégories :

i. Le phénotype recherché permet d'identifier directement les clones recombinants actifs. Par exemple les clones recombinants de couleur ont souvent été associés à la production de métabolites secondaires comme c'est le cas des antibiotiques pigmentés : violaceine (Brady *et al.*, 2001a), turbomycine (Gillespie *et al.*, 2002) et indirubine (MacNeil *et al.*, 2001b). Le criblage fonctionnel pour la recherche de molécules antibactériennes non pigmentés se fait par double couche d'agar où les clones recombinants produisent une zone d'inhibition de croissance autour des colonies ce qui indique la production d'une molécule antibactérienne (Biver *et al.*, 2013; Brady, 2007; Chang & Brady, 2013b; Courtois *et al.*, 2003; Craig *et al.*, 2010; Martinez *et al.*, 2004; Rondon *et al.*, 2000b). Ce type de criblage implique un travail considérable car le nombre de clones actifs varie approximativement de 1/4000 (Rondon *et al.*, 2000b) jusqu'à 1/100 000 (Brady & Clardy, 2000).

ii. Le phénotype recherché est responsable de la détection du clone par complémentation hétérologue de l'hôte. Cette fonction est plus simple à détecter car seuls les clones ayant l'insert du phénotype recherché peuvent croître dans les conditions du criblage. Par exemple, dans l'étude de Simon et collaborateurs, le criblage d'une banque d'ADN métagénomique à partir des glaciers, chez une souche de *E. coli* qui comprend une mutation au niveau du gène de l'ADN

polymérase qui devient inactive aux températures inférieures à 20°C a été réalisée. Lors du criblage de cette banque à 18°C, seuls les clones capables de compléter cette mutation ont eu la capacité de croître. Cette étude a permis de découvrir une nouvelle ADN polymérase à partir de bactéries non cultivables (Simon *et al.*, 2009).

iii. Le phénotype recherché est induit par un substrat ajouté au milieu de culture. Dans ce cas, seuls les clones recombinants ayant le phénotype recherché pourront croître sur le milieu de sélection ou exprimer le phénotype. Ce type de criblage conduit à la recherche de gènes de résistance à un antibiotique particulier, comme la résistance à la tétracycline (Riesenfeld *et al.*, 2004) ou aux β -lactames (Allen *et al.*, 2009), où l'antibiotique est directement ajouté dans le milieu de croissance. Seuls les clones ayant le gène de résistance à cet antibiotique pourront croître. Ce criblage permet aussi de rechercher par exemple la fonction enzymatique par l'emploi de colorants, de dérivés insolubles ou de chromophores comme substrats d'enzymes, incorporés dans le milieu de croissance. Les clones recombinants qui produisent l'enzyme recherchée seront visibles à l'œil grâce à la dégradation du substrat présent dans le milieu comme c'est le cas pour la lipase (Ranjan *et al.*, 2005), l'estérase (Chu *et al.*, 2008), ou encore la β -galactosidases (Zhang *et al.*, 2017).

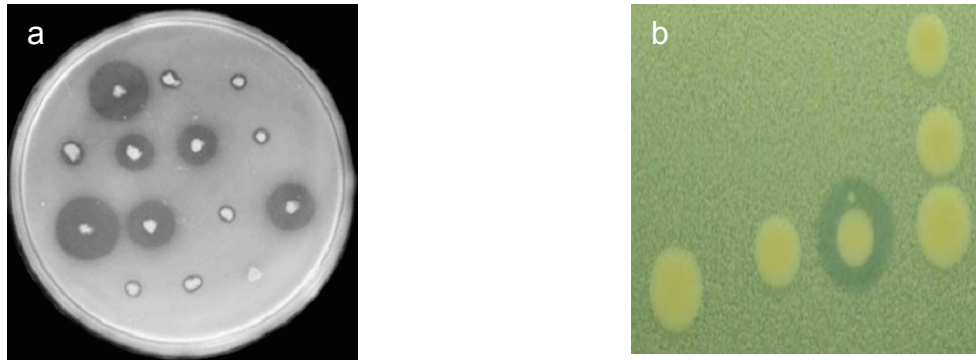


Figure 1.5. Exemple de phénotypes recherchés lors des criblages de banques d'ADN métagénomiques; (a) activité enzymatique par dégradation d'un substrat présent dans le milieu de culture (lipotylique) (Ranjan *et al.*, 2005), (b) activité antibactérienne par production d'une molécule antibiotique (Brady, 2007).

1.2.3. Limites de la métagénomiques fonctionnelle

La métagénomique fonctionnelle fait face à de nombreux facteurs limitants qui peuvent influencer le résultat final de l'étude.

i. L'une des plus grandes limites de cette technique est la dépendance de l'expression de l'ADN métagénomique, de l'hôte d'expression hétérologue utilisé (Daniel, 2004; Uchiyama & Miyazaki, 2009). Ainsi, *E. coli* est l'hôte d'expression hétérologue le plus utilisé (Brady *et al.*, 2001b; Brady *et al.*, 1999; Gillespie *et al.*, 2002; Iqbal *et al.*, 2016; Lim *et al.*, 2005; Thies *et al.*, 2016). Cependant, l'utilisation de *E. coli* comme hôte unique limite l'expression de l'ADN métagénomique (Mullany, 2014). Une estimation indique qu'elle peut exprimer au mieux 40% de gènes provenant d'organismes de diverses origines (Gabor *et al.*, 2004). Cela ne veut cependant pas dire que l'expression des gènes hétérologues chez *E. coli* soit limitée à l'ADN provenant de protéobactéries. Par exemple, l'expression de gènes du chromosome *Bacillus cereus* intégré dans un vecteur BAC maintenu chez *E. coli* est possible (Rondon *et al.*, 1999).

ii. De plus, les clusters de gènes responsables de la production de molécules antibiotiques comprennent souvent le gène de résistance à l'antibiotique produit. Mais, lors des études métagénomiques, l'ADN est digéré aléatoirement, l'insert pourrait alors ne pas contenir le gène de résistance (Steffensky *et al.*, 2000). Dans ce cas, la molécule antibiotique produite peut être toxique pour l'hôte (Rondon *et al.*, 2000b).

iii. Aussi, l'expression hétérologue des voies organisées en groupes de gènes de type PKS (polyketides synthase) et NRPS (non ribosomal peptide synthase) dont la taille est d'environ 100 kb nécessite des conditions d'inductions optimales. Étant donné que ces voies sont codées par plusieurs unités transcriptionnelles polycistroniques, l'expression hétérologue du produit recherché implique l'insertion de l'ensemble du *cluster* chez un seul clone. En théorie, les vecteurs de type BAC sont capables de maintenir des inserts de cette taille en faible copie (Sosio *et al.*, 2000), mais en pratique les vecteurs BAC peuvent maintenir des fragments d'ADN d'une taille avoisinant les 40 kb (Gillespie *et al.*, 2002; Rondon *et al.*, 2000a). En même temps, maintenir des fragments d'ADN de cette taille représente une grande limite technique (Delmont *et al.*, 2011a; Williamson *et al.*, 2011).

iv. La majorité des microorganismes ont une préférence particulière pour certains codons (Saier, 1995). Dans le cas de *E. coli* l'expression d'ADN contenant des codons rares, peut aboutir à des problèmes de traduction comme l'incorporation d'un mauvais acide aminé, le changement du cadre de lecture ou encore une terminaison prématurée (Bisant & Maizel, 1995; Villegas & Kropinski, 2008b). Par conséquent, la production de la molécule se retrouve considérablement réduite, voir biaisée. Ce même problème est moins retrouvé avec l'expression hétérologue chez l'hôte *Bacillus* (Wong, 1995).

v. Dans certains cas, il est possible que le produit final ne soit pas celui attendu. Par exemple, la production d'une protéine active s'appuie beaucoup sur les

facteurs *trans* qui sont fournis par l'organisme-hôte tel que les facteurs de transcription, les inducteurs, les chaperons, les cofacteurs, les enzymes de modification des protéines, etc. Dans le cas où la protéine ne prend pas sa conformation active parce qu'un facteur n'est pas fourni par l'organisme-hôte et qu'en plus celui-ci n'est pas porté en facteur *cis*, aucune activité ne sera observée (Gabor *et al.*, 2004).

vi. Pour terminer, l'expression chez l'hôte dépend beaucoup de la stabilité de l'ARNm. Plus l'ARNm est présent dans la cellule, plus il y aura expression de protéines hétérologues. La principale enzyme responsable de sa dégradation est l'exonucléase 3'-5' (RNase). Certaines structures telles que la polyadénylation de l'ARNm en 3' permettent de protéger l'ARNm des enzymes de dégradation. Néanmoins, il est préférable d'utiliser des souches d'expression mutantes en ce type d'enzymes pour augmenter la stabilité de l'ARNm (Lopez *et al.*, 1999).

1.3 PROBLÉMATIQUE, HYPOTHÈSE ET OBJECTIFS

1.3.1. Problématique

Un tsunami de résistance aux antibiotiques se propage dans le monde entier. L'utilisation massive et abusive des antibiotiques pour le contrôle des maladies infectieuses sélectionne inévitablement les bactéries résistantes et provoque l'émergence de BMR, qui menacent maintenant de redevenir une cause majeure de mortalité (Harris, 2014). Étant donné l'impressionnante diminution du nombre d'antibiotiques ayant été commercialisés au cours des dernières décennies, il y a urgence de découvrir de nouvelles classes d'antibiotiques afin de vaincre ces BMR. Ce problème a d'ailleurs récemment été reconnu comme une priorité par l'Organisation Mondiale de la Santé (World Health Organization., 2014).

Un exemple de ces BMR est : *Staphylococcus aureus* résistant à la méthicilline, une bactérie appartenant à la liste des 'ESKAPE', un groupe de bactéries à Gram-positif et à Gram-négatif composé de *Enterococcus faecium*, *S. aureus* (SARM), *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *P. aeruginosa*, et *Enterobacter*, responsable de la majorité des infections nosocomiales répertorié par l'IDSA (Infectious Diseases Society of America) (Santajit & Indrawattana, 2016). Par exemple, beaucoup plus de patients dans les hôpitaux américains meurent d'infections dues à SARM qu'au SIDA et à la tuberculose combinés (Boucher *et al.*, 2009; Klevens *et al.*, 2006).

De plus, au Canada comme ailleurs, les êtres humains polluent de nombreux environnements en rejetant dans la nature plusieurs produits chimiques, incluant de grandes quantités d'antibiotiques. Cette forte pression de sélection des antibiotiques conduit à l'évolution/adaptation des bactéries par mutation en plus du transfert de gènes de résistance portés sur des plasmides par conjugaison (Katouli *et al.*, 2012). Cela conduit à la propagation de bactéries multirésistantes (MDR). Les chercheurs estiment que dans le monde, la quantité d'antibiotiques utilisée pour l'usage médical et agricole équivaut à environ 250000 tonnes par an (Gillings, 2015). La majorité de ces médicaments aboutit dans les eaux usées et se retrouve dans les rivières et les lacs, incluant les sources d'eau potable pour les humains et animaux.

La grande majorité des antibiotiques utilisés aujourd'hui sont produits ou dérivés de souches microbiennes telles que *Penicillium* ou *Streptomyces*, cultivables par des méthodes classiques (Asnaashari *et al.* ; Bentley *et al.*, 2002). Ces méthodes impliquent d'isoler des microorganismes de l'environnement, d'effectuer leur culture en laboratoire, et d'analyser leur surnageant afin d'isoler de nouvelles molécules naturelles. Cependant, il est largement reconnu qu'au mieux 1% de la diversité microbienne existante dans l'environnement terrestre et

aquatique soit cultivable en laboratoire par des méthodes traditionnelles (Banik & Brady, 2010).

L'une des méthodes les plus prometteuses pour exploiter la diversité microbienne encore inconnue, et pouvant permettre de découvrir de nouveaux antibiotiques est la métagénomique fonctionnelle (Mullany, 2014). L'ADN de ces microorganismes est directement isolé de l'environnement, cloné dans des vecteurs adaptés, puis exprimé dans des hôtes d'expressions hétérologues pouvant être manipulés génétiquement. Le lieu d'isolement de l'ADN environnementale en vue d'une étude métagénomique fonctionnelle étant fort important, l'étude proposée cible un lieu d'isolement à fort antagonisme bactérien, afin d'augmenter la probabilité de découvrir un nouvel antibiotique (ou une famille d'antibiotiques) à des fins thérapeutiques.

1.3.2. Hypothèses

Nous savons que plusieurs environnements sont pollués par de grandes quantités d'antibiotiques. Nous savons aussi que cette forte pression de sélection créée par ces antibiotiques conduit à l'évolution des bactéries présentes dans cet environnement vers des bactéries résistantes, voire multirésistantes.

De plus il est maintenant bien accepté que les bactéries produisent des antibiotiques notamment pour tuer d'autres bactéries afin que les mêmes espèces dominent un environnement donné (Bleich *et al.*, 2015). Imaginons alors ce qui pourrait se produire entre des bactéries multirésistantes présentes dans un environnement riche en antibiotiques tels que les effluents hospitaliers. L'évolution/adaptation de ces populations microbiennes par mutation sélectionnera des bactéries résistantes et compétitives engendrée par cette pression d'antibiotique. Ces bactéries pourraient alors produire des molécules inhibant la croissance des compétiteurs par antibiose.

On peut émettre l'hypothèse qu'un environnement enrichi en BMR sélectionnera positivement par antibiose la croissance de bactéries capables de compétitionner grâce à la production d'antibiotiques efficaces contre ces BMR.

Aussi, vu que la majorité des criblages fonctionnels dans les études métagénomiques sont réalisés contre des bactéries cibles non BMR (sauf pour l'étude de Iqbal et al., 2014), l'emploi de BMR pendant le criblage favorisera la sélection de candidats intéressants portant des inserts d'ADN métagénomiques comprenant des gènes responsables de la production d'antibiotiques efficaces.

Mon projet de doctorat vise à trouver un nouvel antibiotique, ou une nouvelle classe d'antibiotiques. Nous proposons d'isoler, de caractériser et de produire avec un nouvel antibiotique à partir d'échantillons environnementaux par une approche multidisciplinaire combinant métagénomique fonctionnelle, criblage à haut débit contre des BMR, identification des gènes de synthèse par séquençage à haut débit et synthèse/caractérisation exhaustive du métabolite d'intérêt.

1.3.3. Objectifs

1.3.3.1 Objectif général

Isoler et caractériser un nouvel antibiotique, ou une nouvelle classe d'antibiotique à partir d'un ADN environnemental (ADNe) isolé à partir de biofilms hospitaliers, par métagénomique fonctionnelle.

1.3.3.2. Objectifs spécifiques

Objectif 1 : Construction des banques d'ADN métagénomiques

- Développement de méthodes de construction de banques d'ADN métagénomique et de criblage fonctionnel
- Collecte des échantillons de biofilms
- Construction des banques d'ADN métagénomiques

Objectif 2 : Criblage fonctionnel contre SARM

- Sélection des clones positifs inhibant la croissance de la bactérie cible
- Confirmation du phénotype observé et étude des clones sélectionnés

Objectif 3 : Identification du gène responsable du phénotype observé

- Séquençage de l'insert
- Analyse prédictive de la séquence insérée

Objectif 4 : Synthèse/caractérisation du métabolite actif

- Synthèse du métabolite hypothétique
- Évaluation d'activité *in vitro* contre la bactérie cible du criblage

CHAPITRE 2 : ARTICLE 1

2.1. NOVEL 'BACTERIOSPRAY METHOD FACILITATES THE FUNCTIONAL SCREENING OF METAGENOMIC LIBRARIES FOR ANTIMICROBIAL ACTIVITY

Anissa Brahami, Annie Castonguay * and Éric Déziel *

INRS-Institut Armand-Frappier, Laval, Quebec, H7V 1B7, Canada;

Journal: MDPI methods and protocols

Submitted: 7 December 2018

Acceptation: 1 January 2019

Publication: 7 January 2019

Citation: Brahami, A., A. Castonguay and É. Déziel (2019). "Novel 'Bacteriospray' Method Facilitates the Functional Screening of Metagenomic Libraries for Antimicrobial Activity." *Methods and Protocols* **2**(1): 4.

Contribution: Methodology, conceptualization, data curation, writing the manuscript: all authors; performed the experiments: A.B; supervision and funding acquisition: A.C. and E.D.

Acknowledgments: The authors would like to thank Philippe Constant (INRS) and Marie-Christine Groleau for helpful discussions, and Sylvain Milot for technical assistance.

2.1.1. Résumé

La métagénomique fonctionnelle consiste en l'expression d'un ADN environnemental (ADNe) à l'aide de souches d'expression hétérologues. Un criblage fonctionnel est ensuite entrepris pour sélectionner les transformants responsables de la production de métabolites secondaires. Dans cette étude, nous décrivons une méthode de criblage fonctionnel permettant de détecter la plus petite zone d'inhibition produite par les clones transformants. Notamment, nous rapportons un protocole détaillé sur la construction des banques d'ADN génomiques en utilisant l'hôte d'expression *Escherichia coli* DH10B. Nous démontrons comment la déphosphorylation de l'ADN du vecteur / déphosphorylation de l'ADNe, l'inactivation de la ligase après ligation, la dialyse du produit de ligation et le rapport ADN vecteur / génomique influencent l'efficacité de transformation. De plus nous décrivons l'utilisation d'un aérographe pour le criblage fonctionnel de cette banque d'ADN génomique construite en vaporisant la bactérie cible *Staphylococcus aureus*. Cette méthode nommée bacteriospray permet un criblage plus efficace puisqu'une fine et uniforme couche de bactéries cible est déposée, ce qui permet de criblage de 5 à 10 fois plus de clones que la méthode de double couche d'agar conventionnelle. Afin de valider cette méthode, l'ADN génomique de *Burkholderia thailandensis* E264 a servi à la construction d'une banque d'ADN, où quatre clones inhibaient la croissance de *S. aureus* sur 70 000 criblés. L'analyse des inserts de ces clones actifs a révélé que ces régions génomiques n'avaient jamais été déclarées responsables de la production de molécules antimicrobiennes.

2.1.2. Abstract

Metagenomic techniques, notably the cloning of environmental DNA (eDNA) into surrogate hosts, have given access to the genome of uncultured bacteria. However, the determination of gene functions based on DNA sequences alone remains a significant challenge. The functional screening of metagenomic libraries represents an interesting approach in the discovery of microbial metabolites. We describe here an optimized screening approach that facilitates the identification of new antimicrobials among large metagenomic libraries. Notably, we report a detailed genomic library construction protocol using *Escherichia coli* DH10B as a surrogate host, and demonstrate how vector or genomic DNA dephosphorylation, ligase inactivation, dialysis of the ligation product and vector/genomic DNA ratio greatly influence clone recovery. Furthermore, we describe the use of an airbrush device to screen *E. coli* metagenomic libraries for their antibacterial activity against *Staphylococcus aureus*, a method we called bacteriospray. This bacterial spraying tool greatly facilitates and improves the functional screening of large genomic libraries, as it conveniently allows the production of a thinner and more uniform layer of target bacteria compared to the commonly used overlay method, resulting in the screening of 5–10 times more clones per agar plate. Using the *Burkholderia thailandensis* E264 genomic DNA as a proof of concept, four clones out of 70,000 inhibited the growth of *S. aureus* and were found to each contain a DNA insert. Analysis of these chromosomal fragments revealed genomic regions never previously reported to be responsible for the production of antimicrobials, nor predicted by bioinformatics tools.

2.1.3. Introduction

Bacterial resistance to the arsenal of clinically approved antibiotics is currently threatening the efficient treatment and prevention of infectious diseases [1–3]. Bacteria are an important source of natural products [4] and their genomes contain biosynthetic gene clusters (BGCs) that are predicted to encode the biosynthesis of interesting compounds that are not yet identified and/or characterized. These clusters might represent pathways for the production of potentially highly active antibiotic compounds and, thus, their identification could have an important positive impact on drug discovery [5]. As only 1% of all microbial species can be cultured using standard laboratory conditions [6], functional metagenomics is considered a key method for the identification of new metabolites, notably via the heterologous expression of DNA in engineered host strains [7]. Using this strategy, bacterial DNA is directly extracted from the environment and cloned in cultured bacterial hosts, providing access to the BGCs present in uncultured bacteria. Several examples of the heterologous expression of BGCs from environmental DNA (eDNA) led to the identification of new metabolites, including terragine from eDNA expressed in *Streptomyces lividans* [8], indolotriptoline from eDNA expressed in *Streptomyces albus* [9] and indirubin from eDNA expressed in *Escherichia coli* [10]. Despite their numerous advantages, functional approaches are very laborious and their reproducibility from one laboratory to another relies on the availability of detailed procedures.

Numerous detailed protocols were previously reported to greatly improve the extraction of eDNA [11–15] and the construction of metagenomic libraries [16–18]. Indeed, several factors are believed to influence transformation efficiencies [16] but surprisingly little is known about the actual effect of some of those factors. Consequently, identifying the cause(s) of low transformation efficiencies remains difficult. Factors include (i) metagenomic DNA sampling (which should be representative of the microbial community), (ii) DNA extraction and digestion (which

should limit DNA shearing while allowing the efficient recovery of large DNA fragments [19]), (iii) vector-to-genomic DNA ratio, (iv) vector or genomic DNA dephosphorylation, and (v) vector ligation conditions (ligase can be heat-inactivated after ligation and/or ligation product can be dialyzed to eliminate buffer).

When metagenomics libraries are screened for antibiotic activity, the choice of the method used for the functional screening for clones that produce the smallest zone of growth inhibition is also crucial to the success of the study. The overlay method is the commonly used method for metagenomic library screening [16,20,21]. But, due to the large size of metagenomics libraries to be screened, this method is time-consuming and hinders the discovery of novel antimicrobials by limiting the number of clones to be tested per experiment (in addition to some technical issues, vide infra). Therefore, the development of an efficient and rapid screening method is highly needed in the field in order to promote the use of functional metagenomics approaches for the identification of novel microbial secondary metabolites.

Herein, we report an optimized and detailed protocol that can be useful for the identification of BGCs from metagenomic libraries using *Burkholderia thailandensis* genomic DNA as a model, by exploiting the heterologous expression of its genetic material into surrogate host *E. coli* DH10B. Moreover, we describe a new method for the formation of target bacteria layers that we named bacteriospray. This technique is more sensitive and allows the screening of larger clone libraries compared to the traditional top agar overlay method. The implementation of the optimized method has led to the construction of a library of size 2.7×10^5 colony forming units (CFU)/ μg with a BAC-inserted DNA transformation efficiency of 99%, allowing the screening of 70,000 clones from which four were found to carry an insert providing inhibitory activity against *Staphylococcus aureus*.

2.1.4. Materials and Methods

2.1.4.1. Bacterial Strains and Plasmid Growth Conditions

B. thailandensis E264 (ATCC 700388, American Type Culture Collection, Manassas, VI, USA) was used for this study. This strain was grown at 30 °C in Lysogeny broth (LB) medium. Electrocompetent *E. coli* strain DH10B used for the cloning and the heterologous expression was purchased from New England Biolabs (NEB), Boston, MA, USA. The library was constructed in the bacterial artificial chromosome pBeloBAC11 (7.5 kb) [22] maintained in *E. coli* K12 ER2420 and can replicate in *E. coli* DH10B with 20 µg/mL chloramphenicol (obtained from NEB). The vector was purified from an overnight LB culture incubated at 37 °C with the Qiagen maxiprep kit. The test strain used for antimicrobial production screening was *Staphylococcus aureus* Newman (laboratory collection), grown at 37 °C in LB medium.

2.1.4.2. Library Construction Strategies

B. thailandensis E264 was grown overnight at 30 °C in LB broth. DNA extraction was done with the FastDNA kit (MP Biomedicals) as recommended by the manufacturer. Briefly, 1.5 mL of bacterial culture was centrifuged, placed in a Lysing Matrix A tube and the appropriate cell lysis solution was added. The mixture was homogenized by vortexing for 30 s (2 times) and DNA extraction was performed as described in the MP protocol. DNA was then partially digested with the restriction enzyme *SphI* (NEB) for 15 min. The enzyme was then heat inactivated for 15 min at 65 °C. Less than 1 µg of the pBeloBAC11 vector was linearized with one unit of the *SphI* enzyme (10,000 units/mL) during a 4 h digestion in a 37 °C water bath. The vector was dephosphorylated when specified (vide infra). One µL of antarctic

phosphatase (NEB) was added to every 1 μg of *digested DNA* reaction mixture with the Antarctic phosphatase buffer (10 \times), before incubation for 30 min at 37 °C. The enzyme was then heat inactivated at 65 °C for 15 min.

Bacterial cohesive-ends DNA was purified by 0.8% agarose gel electrophoresis prepared with 1 \times TAE (Tris-Acetate-EDTA) buffer and a DNA size marker. The gel included GelRed stain (Biotium) and DNA was visualized with UV light (260 nm). Outer lanes containing the DNA ladder were cut off from the gel, the position of the desired fragment sizes was marked, and DNA larger than 3 kb was excised [16]. Gel slices were weighted and DNA was extracted with a QIAEX II gel extraction kit (Qiagen). Extracted DNA fragments were ligated overnight to the *SphI*-linearized pBeloBAC11 vector at 16 °C using NEB T4 DNA ligase. The ligation reaction was optimized by varying the vector and genomic DNA ratio (BAC/gDNA). For that, four BAC/gDNA ratios were tested; 1:3, 1:6, 1:9, 1:12 with 50 ng/ μL as a fixed vector concentration (and also at 100 ng/ μL in the case of the 1:9 ratio). The DNA ratio was calculated using the following formula:

$$\text{DNA ratio} = \frac{(\text{ng vector pBeloBAC11}) \times (\text{DNA size insert})}{(\text{Kb vector size})} \times \left(\frac{\text{BAC}}{\text{gDNA}} \right) \quad (1)$$

The ligase was heat inactivated at 65 °C for 15 min before dialysis for 1 h by adding 20 μL on a millipore mixed cellulose esters (MCE) membrane filter (0.025 μM porosity, VSWP01300) floating on sterile distilled water into a Petri plate. Next, for transformation by electroporation, 2 μL of this dialyzed ligation solution was added to 25 μL *E. coli* DH10B electrocompetent cells, directly transferred to an ice-cold electroporation cuvette (1 mm electrode gap) and electroporated (25 μF , 200 Ω , 2.2 kV). After 1 h incubation in 1 mL NEB 10-beta/stable outgrowth recovery medium, at 37 °C under agitation, 100 μL , 10 $^{-1}$, 10 $^{-2}$ and 10 $^{-3}$ dilutions were spread on LB plates containing 40 $\mu\text{g}/\text{mL}$ 5-bromo-4-chloro-indolyl- β -D-thiogalactopyranoside (X-gal), and 20 $\mu\text{g}/\text{mL}$ chloramphenicol and then incubated overnight at 37 °C. Briefly, colonies

formed from non-recombinant cells appeared blue in color whereas colonies formed from the recombinant ones appeared white. The transformation efficiency (T) was calculated using formula (2) and the average insert size of the library was determined by analyzing different white and blue clones by vector purification followed by *SphI* digestion and agarose gel electrophoresis.

$$T = \frac{\text{CFU control plate}}{\text{pg BAC DNA}} \times \frac{1 \times 10^6}{\mu\text{g}} \times \frac{\text{volume of transformants}}{\text{volume plated}} \times \text{dilution factor} \quad (2)$$

2.1.4.3. Derivation of a Chloramphenicol Resistant Strain

Since *S. aureus* Newman is unable to grow on 20 µg/mL chloramphenicol, a concentration required for overlay and bacteriospray experiments, a resistant strain was derived by selection on antibiotic gradient LB agar plate. First the Petri dish was tilted to create a slope and 10 mL of LB agar containing 20 µg/mL chloramphenicol was poured. After agar drying, a second 10 mL of LB agar with no antibiotic was added, creating a zero to 20 µg/mL gradient. The inoculation was started from the position of no antibiotic side [23] and a resistant mutant was picked for further experiments.

2.1.4.4. Top Agar Overlay Method

The top agar overlay method was used according to Brady (2007). Screening for antimicrobial activity was carried out using the two-layer overlay method with the chloramphenicol-resistant *S. aureus* as the test strain. *E. coli* DH10B clones carrying genomic DNA fragments were grown for 24 h at 37 °C on 20 × 20 mm² square plates containing LB agar with 40 µg/mL (X-gal), 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 20 µg/mL chloramphenicol. The number of *E. coli*

colonies per plate was optimized by using different amounts of a 10^{-2} dilution of the initial library to allow detection of very small zones of growth inhibition. Plates were then incubated at room temperature for an additional five days to promote the accumulation of metabolites around colonies while preventing their excessive development. Before overlaying the plates with 10 mL LB solidified with 0.75% agar, the *S. aureus* assay strain grown overnight in LB medium was suspended to a final $OD_{600} = 0.05$ in this top agar cooled at 50 °C. Carefully, the top agar was poured on the top of the plates (so that the *E. coli* colonies would not be disturbed). Plates were incubated overnight at 30 °C for 24 h–48 h and were then visually screened for *E. coli* colonies that produce clear zones of growth inhibition in the *S. aureus* hazy layer.

2.1.4.5. Bacteriospray Screening Method

As an alternative to the top agar overlay method, an airbrush tool (Anest Iwata-Medea model Revolution BCR, Anest Iwata-Medea, Inc, Portland, OR, USA) was used to spray a mist of bacteria on agar plates containing colonies of the *E. coli* DH10B, transformed with the BAC library prepared as described above (Figure 1). The spray inoculation was performed under a class II laminar flow hood. Plates were placed inside a lidless plastic box to limit aerosols exposure. The airbrush was connected to a compressor delivering air at 20 psi. A Whatman Nylon 0.2 µm filter (L # 3066 Life Sciences, Sigma-Aldrich, ON, Canada) was placed at the outlet of the compressor to avoid any contamination coming from the air source. The airbrush was equipped with a sterile 20 mL plastic tank containing the bacterial suspension. The airbrush needle tip was adjusted so that it would not quite touch the cap, leaving just enough room for air to escape. The airbrush and the tank were decontaminated with ethanol 70% and washed with sterile water before use.

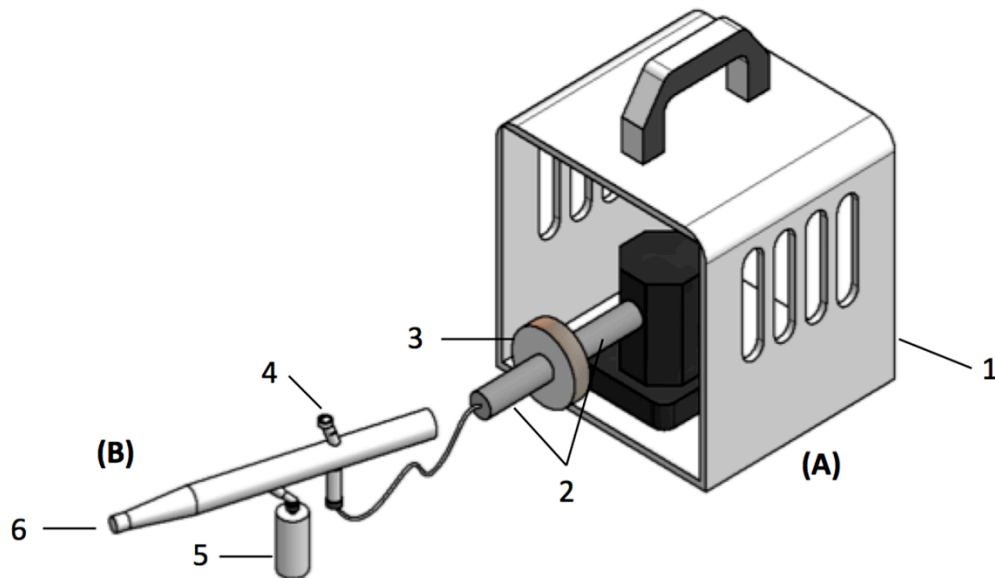


Figure 2.1. Sketch of the device used to spray a thin layer of *S. aureus* suspension using an airbrush, consisting of the following components: (A) an air delivery system, including a compressor (1), a connector (2), a 0.2 μ M filter (3); and (B) an airbrush, including a needle cap (4), a tank (5), and a nozzle for bacterial suspension mist (6).

The screening plates were prepared as above, with an optimized number of clones per plate grown for 24 h. About 5 mL of a 10^5 *S. aureus* suspension in LB was sprayed as a mist on the plates. Plates were placed vertically at a 60° angle from the spray direction and the airbrush was placed at about 20 cm from the plate. The flow rate was adjusted to obtain a very thin layer of the bacterial suspension, at the eye detection limit. Plates were then incubated at 30 °C for 24-48 h before being screened visually, as described above.

2.1.4.6. Growth Inhibition Zone Confirmation, DNA Sequencing and Insert Analysis

Clones producing a zone of growth inhibition, presumptive of diffusible antibiotics [24–26], were picked, purified onto LB containing 40 μ g/mL chloramphenicol and

tested again for confirmation of *S. aureus* growth inhibition. Briefly, 5 µL of overnight preculture of purified clones were spotted on LB plates containing 20 µg/mL chloramphenicol and 1 mM IPTG, incubated overnight at 37 °C followed by five days at room temperature. The antibacterial activity was then tested using the bacteriospray method, as described above.

For DNA sequencing, polymerase chain reaction (PCR) amplification of confirmed clones was performed using the extracted BAC DNA as template with the following primer pairs: M13 F (-40) (5'-GTTTTCCAGTCACG-3') and M13 R (-26) r (5'-CAGGAAACAGCTATG-3'). The PCR reaction was performed in a 60 µL reaction volume containing 10 µL of 5X Q5 buffer (NEB), 10 µL of Q5 high GC enhancer 5X, 2.5 µL of 10 pmol/µL of each forward and reverse primers, 1 µL of 0.5 ng DNA template, 1 µL of 10 µM dNTPs and 0.5 µL of Q5 Taq. Reaction conditions consisted of an initial denaturation by heating at 98 °C for 30 s, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 57 °C for 30 s, extension at 72 °C for 1 min, and final extension at 72 °C for 5 min. Amplified products were visualized by 1% agarose gel electrophoresis. Amplicons were dialyzed as described above before Sanger DNA sequencing (Institut de recherches cliniques de Montréal; IRCM). Sequence analysis was done using the CLC sequence viewer (version 7). Since the genome of *B. thailandensis* E264 was previously sequenced (with data available from the www.burkholderia.com website), the exact nature of cloned genes could be identified using bioinformatics tools. Sequence-based homology searches in the GenBank database were carried out using the protein BLAST (BLASTp) program (BLOSUM62, existence: 11, extension: 1, conditional compositional score matrix adjustment).

2.1.4.7. Statistical Analysis

Each experiment was performed at least in triplicates (n = 3) and experimental values are presented as means +/- standard deviation. Statistical analysis was done

using Prism 6 (GraphPad Prism version 6.00, GraphPad Software, Inc, La Jolla CA, USA). A one-way analysis of variance followed by a Tukey's post hoc test for multiple comparison groups were computed to evaluate the effect of dialysis, x, y, on transformation efficiency ($\alpha < 0.05$). As an alternative, a Kruskal–Wallis test followed by a Dunn's post hoc test for multiple comparison groups were used to compare transformation efficiency as a function of ligation product dialysis due to unequal variance between the three different treatment groups.

2.1.5. Results

2.1.5.1. Transformation Efficiency Optimization

DNA fragments between 0.1–10 kb were obtained after *SphI* digestion of the extracted *B. thailandensis* E264 DNA (Figure 2). Fragments larger than 3 kb were purified and used for all cloning experiments involving pBeloBAC11 as the expression vector and *E. coli* DH10B as the surrogate host.

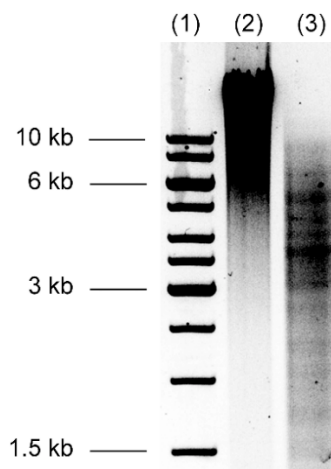


Figure 2.2. *Burkholderia* DNA size after *SphI* digestion. 0.8% agarose gel loaded with Thermo Scientific GeneRuler 1 kb DNA Ladder molecular weight standard (lane 1), *B. thailandensis* DNA extraction sample (lane 2) and genomic DNA following *SphI* digestion sample (lane 3).

Optimization of the cloning efficiency was performed by varying several parameters, starting with the vector to genomic DNA ratio (BAC/gDNA). Surprisingly, varying digested DNA ratio from 1:3 to 1:12 did not have a considerable impact on the resulting transformation efficiency (Figure 3A columns (1), (2), (3), (4)). Nevertheless, best results were obtained when a BAC/gDNA ratio of 1:9 was used (3.2×10^4 CFU/ μ g). It is noteworthy that increasing the BAC vector DNA concentration from 50 ng/ μ L to 100 ng/ μ L led to a 5-fold decrease in transformation efficiency (Figure 3A column (5)). Thus, for all further optimization experiments, a 1:9 BAC/gDNA ratio and a vector DNA concentration of 50 ng/ μ L were selected.

The importance of vector dephosphorylation was then assessed, a step that was already integrated in the protocol, prior to the ligation. As previously reported for other systems [27] this parameter was found to have a significant influence on the cloning efficiency. When 5'-phosphate groups of the vector DNA were not removed, a 1000-fold decrease (from 10^5 to 10^2 CFU/ μ g) in cloning efficiency was observed (Figure 3B). The dephosphorylation of the genomic DNA alone was found to lead only to a small increase in transformation efficiency (from 8.6×10^4 to 2.2×10^5 CFU/ μ g). Interestingly, under those conditions, no blue colonies were observed even after an overnight incubation at 4°C compared to 10% blue colonies when genomic DNA was not dephosphorylated. To further confirm that the maximum transformation efficiency was achieved, the effect of the dialysis of the ligation product and of ligase inactivation on the transformation efficiency were investigated. These two steps were found to be important for the efficiency of the protocol reported here, the dialysis of the ligation product being even more influential. Omission of dialysis of the ligation product or inactivation of the ligase before dialysis led to a 1000-fold and a 10-fold decrease in transformation efficiency, respectively (Figure 3C).

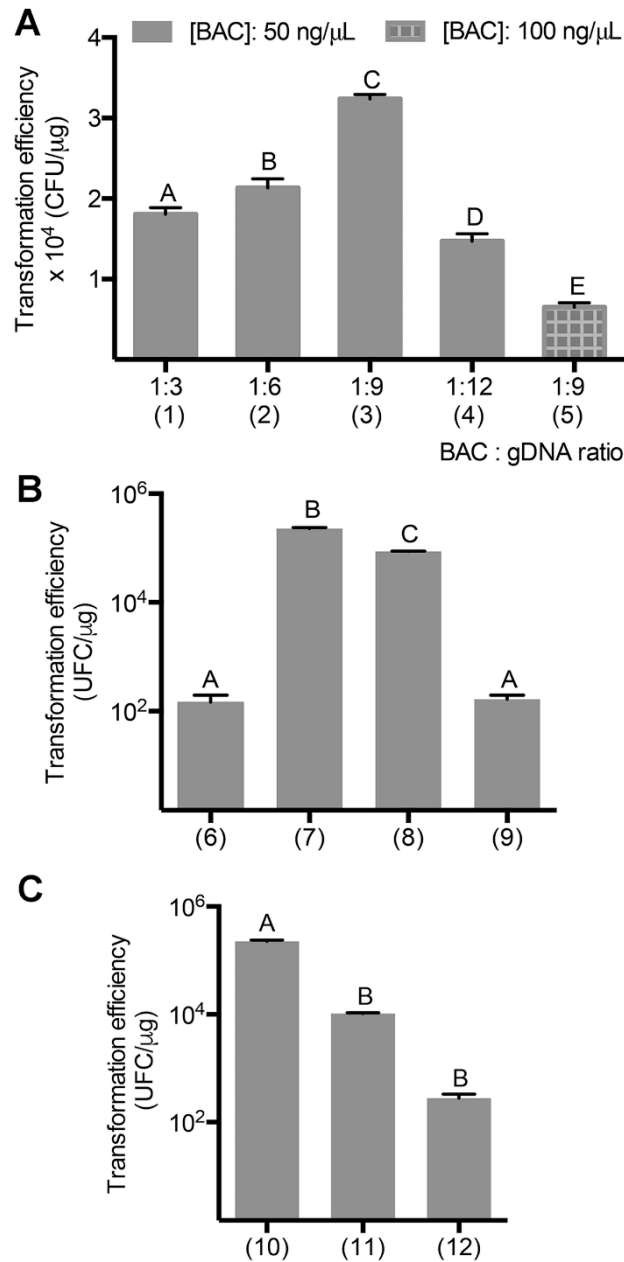


Figure 2.3. Influence of various parameters on the transformation efficiency of pBeloBAC11 in *E. coli* DH10B. (A) Effect of vector to genomic DNA ratio (BAC:gDNA) and vector concentration (ng/μL): (1), (2), (3), (4) DNA and vector ratio from 1:3 to 1:12 respectively under 50 ng/μL vector concentration; (5) DNA and vector ratio 1:9 under 100 ng/μL vector concentration (conditions: vector DNA was dephosphorylated before ligation and ligation product underwent dialysis). (B) Effect of vector (BAC) and genomic DNA (gDNA) dephosphorylation on the transformation efficiency: (6) BAC and gDNA not dephosphorylated; (7) BAC or gDNA dephosphorylated; (8) BAC dephosphorylated and gDNA not; (9) gDNA

dephosphorylated and BAC not (conditions: 1:9 BAC:gDNA ratio was used, ligase was inactivated and ligation product underwent dialysis). (C) Effect of ligase inactivation and dialysis of the ligation product: (10) ligase inactivation and dialysis; (11) dialysis only; (12) ligase inactivation only (conditions: 1:9 BAC:gDNA ratio was used, BAC and gDNA were dephosphorylated). Error bars indicate the standard deviation of three independent experiments, and statistical significance was determined using Tukey post hoc test. A, B, C, D and E represent Tukey's multiple comparison groups ($\alpha = 0.05$)

2.1.5.2. *Burkholderia thailandensis* Library Construction

Several independent libraries were constructed in the pBeloBAC11 vector using genomic DNA from *B. thailandensis* E264 for zone of growth inhibition screening. The best transformation efficiency was obtained under the optimized conditions described previously: BAC/gDNA ratio of 1:9, dephosphorylation of vector DNA (50 ng/ μ L), or dephosphorylation of genomic DNA, T4 ligase inactivation and dialysis of the ligation product. The best transformation efficiency obtained was $2.4 \times 10^5 \pm 0.2$ CFU/ μ g. *SphI* enzyme-assisted restriction digestion of 12 random clones demonstrated that insert DNA size varied between 2–5 kb (Figure 4).

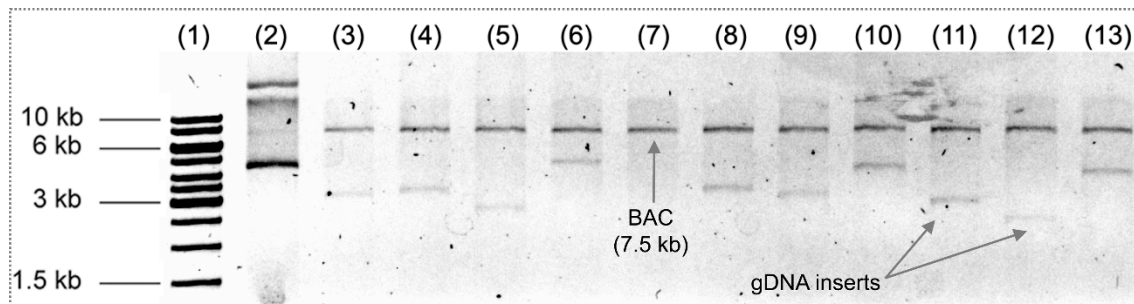


Figure 2.4. Distribution of inserts sizes of bacterial artificial chromosome (BAC) clones. 1% agarose gel loaded with Thermo Scientific GeneRuler 1 kb DNA Ladder molecular weight standard (lane 1), pBeloBAC11 DNA not digested sample (lane 2), pBeloBAC11 DNA control sample digested with *SphI* (7.5 kb) (lane 7), confirmation of the presence of a DNA insert from the *B. thailandensis* genome (lanes 3–6 and 8–13).

2.1.5.3. Top Agar Overlay vs Bacteriospray Assay for the Functional Screening of Metagenomic Libraries

To investigate the functional diversity of the genomic DNA libraries constructed, a distribution of the clones was carried out in $20 \times 20 \text{ mm}^2$ LB square plates. The first method used for antimicrobial screening was the traditional top agar overlay method. Using the method, we could accommodate a maximum of 300–500 clones per plate. In our experience, *E. coli* colonies were disturbed when the top agar layer containing *S. aureus* bacterial suspension was poured (Figure 5A), causing a non-homogenous top agar suspension layer and colonies streaking. Moreover, due to this lack of homogeneity of the resulting bacterial layer, the visual screening for small zones of growth inhibition was very difficult, leading to ambiguous results. To overcome the aforementioned well-known problems, we therefore developed a novel bacteriospray method for the functional screening of genomic libraries (Figure 5B). A pilot study was performed to optimize the airbrush suspension volume to be used, as well as to ensure no contaminant would come from the air (data not shown). The most adequate airbrush spraying volume was estimated to be 5 mL of *S. aureus* bacterial suspension for eight $20 \times 20 \text{ mm}^2$ square plates. The *S. aureus* suspension was sprayed as a mist to obtain a very thin bacterial layer. A pressure of 20 psi was found to be optimal. Whereas 10 psi were not sufficient to properly disperse the bacteria, 30 psi were too strong, causing an unstable air flow and consequently an heterogenous bacterial film. As a pressure of 20 psi allowed proper bacterial growth after using the spraying tool, possible cell damage was assumed to be minimal. Notably, using this method, each plate could accommodate up to 7000 clones. The bacteriospray method was found to be more efficient and sensitive than the top agar overlay method. Using this strategy, 70,000 clones were screened for antibacterial activity, and four clones (B1–B4) led to the formation of a small zone of growth inhibition. Finally, their activity against *S. aureus* was confirmed, as shown in Figure 6.

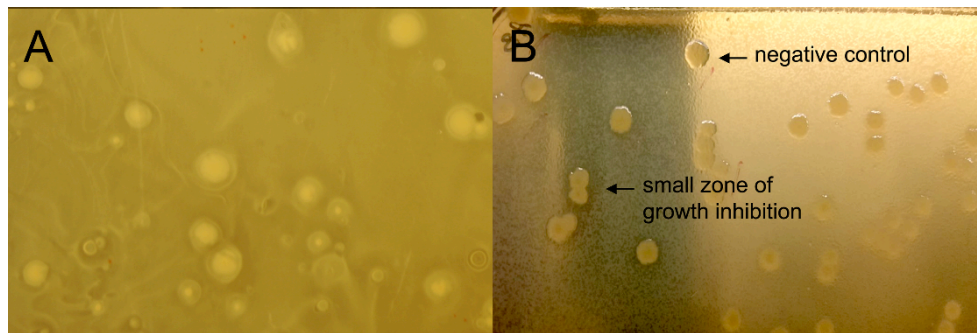


Figure 2.5. Examples of growth inhibition zones observed from antibiotic-producing clones from *E. coli* library. (A) the top agar overlay method vs (B) the airbrush assay.

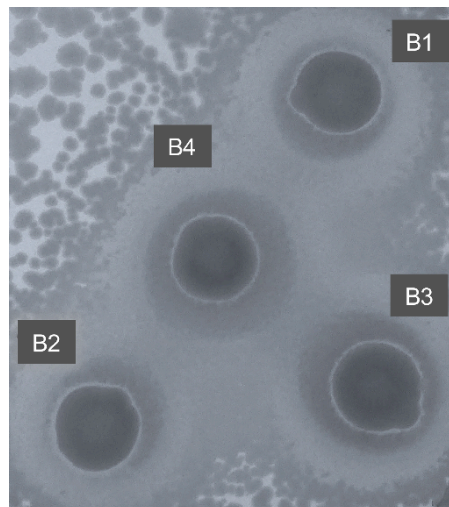


Figure 2.6. Confirmation of the antibacterial activity of four positive *E. coli* clones B1–B4 against *S. aureus*, using the airbrush assay.

2.1.5.4. Sequence Analysis of Inserted Genes

Following PCR amplification, sequencing of the DNA contained in the four active clones B1–B4 was achieved. By comparison with the whole *B. thailandensis* E264 genome, the insert sequence of each clone was identified (Figure 7). Two of them were found to be identical (B3 and B4) but, importantly, none of them were

previously reported to be responsible for the production of an antibiotic, nor predicted by bioinformatics tools (Table 1).

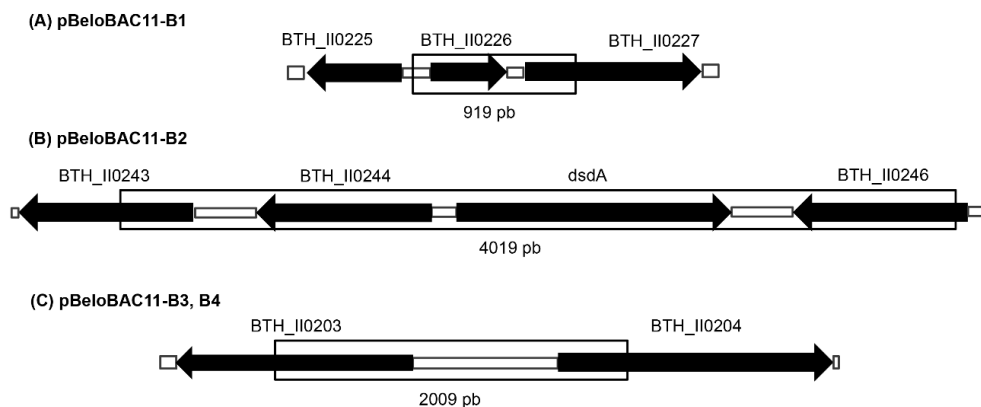


Figure 2.7. Chromosomal context of the DNA insert identified in the four selected clones derived from *B. thailandensis* E264 chromosome II. Inserted region of the clone (A) pBeloBAC11-B1, (B) pBeloBAC11-B2 and (C) pBeloBAC11-B3, B4.

Tableau 2.1. Locus tag identified in the four clones derived from *B. thailandensis* E264 chromosome II metagenomic library in pBeloBAC11 vector

Clone Locus Tags	Similar Protein (Accession Number)
pBeloBAC11 B1	
<i>BTH_II0226</i>	Snoal-like protein (3H3H A)
<i>BTH_II0227</i>	Conserved hypothetical protein (ABC35626.1)

pBeloBAC11 B2	
<i>BTH_II0243</i>	LysR family transcriptional regulator (WP_009894988.1)
<i>BTH_II0244</i>	LysR family transcriptional regulator (WP_009894990.1)
<i>BTH_II0245</i>	<i>dsdA</i> D-serindeshydratase (Q2T8Q2.1)
<i>BTH_II0246</i>	LysR family transcriptional regulator (WP_011400849.1)

pBeloBAC11 B3 & B4	
<i>BTH_II0203</i>	Carbohydrate diacid regulator (ABC35448.1)
<i>BTH_II0204</i>	Peptide synthase, putative (ABC34379.1)

2.1.6. Discussion

In this study, we developed and optimized a method to extract and clone DNA fragments to construct and efficiently screen a genomic library for antimicrobial activity in a surrogate host, here *E. coli* DH10B. This functional genomic approach allowed the identification of gene functions that could not have been predicted from DNA sequences alone [28], or that might not have been expressed under standard culture conditions [29]. The pBeloBAC11 vector was selected for this investigation, as it includes a T7 promoter that allows the use of the highly effective T7 RNAPol, facilitating the strong transcription of foreign DNA inserts [30] and consequently increasing the probability of detecting clones expressing antibacterial activity. Moreover, this vector can accommodate large DNA inserts, making the developed method suitable for the screening of very large size metagenomic libraries, allowing a reduced number of clones to be tested while increasing the probability of identifying new bioactive molecules for which large BGC are used for their biosynthesis [31,32].

Dephosphorylating the vector or the DNA insert resulted in virtually no blue colonies after vector transformation, indicating the absence of empty vectors (and that the highest transformation efficiency was achieved). Besides BACs, plasmid and cosmid vectors can also be used with this protocol. The choice of vector is based on several criteria, including the host bacteria, the expected size of DNA inserts, the type of function targeted by the screening and the screening strategy. Large-insert libraries are required to recover large gene clusters coding a complex pathway [8]. However, small-insert libraries can be employed for the identification of novel molecules encoded by a single gene or small operon. Here, the size of DNA fragments inserted in the clones that expressed antimicrobial activity varied from 0.9 to 4 kb. Since the size of digested fragments did not exceed 10 kb (Figure 2), it was not possible to construct a larger insert library. Whereas this insert size is large enough to isolate antimicrobial activities encoded by small operons, it is not sufficiently large for the recovery of large gene clusters [33].

Sequencing the selected clones revealed that two were identical (B3 and B4) and that gene BTH_II0204 was included in the insert. Curiously, this gene is part of a four gene operon predicted to contain a single nonribosomal peptide synthetase (NRPS) gene BTH_II0204-0207 in the *B. thailandensis* E264 and *B. pseudomallei* K96243 genomes [34,35]. These genes are annotated as NRPS, a dehydratase, an isomerase, and a methyltransferase [34]. Since the nature of the complete operon is unknown, it was not possible to conclude on the origin of the inhibitory activity. Clone B1 contains the gene BTH_II0226 that is predicted to encode a Snoal-like protein in *B. thailandensis* E264. These proteins encode small polyketides cyclases, which catalyze the ring closure step in the biosynthesis of polyketide antibiotics produced in *Streptomyces* [36], but the function is not determined in *B. thailandensis*. It is noteworthy that a search was also performed with the BAGEL database [37], which suggested a possibility for the three distinct DNA sequences to be responsible for the production of an antimicrobial peptide. Further analysis and plasmid gene subcloning of these

clones will be pursued in order to identify the origin of the observed antimicrobial activity.

This study also highlights that the success/failure of a functional genomic study depends significantly on the phenotype screening method used, a typically overlooked aspect of this promising strategy. The use of the bacteriospray method has led to significantly enhanced screening efficiencies compared to the traditional top agar overlay method, which causes the lack of homogenous top agar layer and colonies streaking, making identification of the active clones difficult. For instance, with this newly developed method, an average of 5,000–7,000 clones per plate could be tested at a time, compared to a maximum of 300–500 clones when the traditional top agar overlay method was used. Our results are in agreement with other genomic functional screening efforts for which a maximum of 500 to 1500 colonies per plate could be screened using the overlay method [10,16,21], making the bacteriospray method approximately 5–10 times more efficient for the screening of large-size genomic libraries. A similar method was previously used for spraying monolayers of *Bacillus subtilis* spores [38–40], but to our knowledge, this technique was never used to spray live bacteria. The spraying of a bacterial suspension provides a homogenous monolayer lawn of bacteria that allows the sensitive detection of the thinnest bacterial inhibition zones, which will greatly benefit functional metagenomic studies.

2.1.7. Conclusions

In this article, we described an improved protocol for the construction and screening of functional metagenomic libraries. We also developed a new bacterial spraying method, which will be useful for the screening of large libraries. We demonstrated that this new tool allows the identification of small zones of growth inhibition and is effective and sensitive enough to be used for the screening of eDNA metagenomic libraries. We believe this protocol could play an important role in the discovery of novel antibiotics.

2.2. DONNÉES COMPLÉMENTAIRES À L'ARTICLE

2.2.1. Mise en contexte

Afin d'explorer l'origine de l'activité observée chez les clones B1, B2, B3 et B4, sélectionnés lors du criblage fonctionnel de la banque d'ADN génomique de *B. thailandensis*, les séquences d'ADN ont été analysées par recherche d'ORFs, suivie d'un BLASTp dans NCBI, d'une recherche dans la base de données BAGEL et antiSMASH.

2.2.2. Méthodologie

La recherche des ORF des séquences a été réalisée en utilisant les conditions standard d'ORF Finder NCBI avec un code génétique standard et avec ATG comme codon d'initiation de la traduction. Pour le clone B1 une séquence minimale de 75 pb a été choisie, alors que pour les clones B2 et B3 (B4), une séquence minimale de 150 pb a été choisie étant donnée la taille de l'insert de ces clones. Évidemment, le site de Burkholderia.com a été utilisé comme point de départ de ces analyses. Les ORFs résultants de cette recherche ont subi un BLAST sur NCBI UniProtKB/Swiss-Prot (swissprot) et RefSeq-proteins. Les séquences ont aussi été analysées sur la base de donnée antiSMASH 3.0 pour une annotation des métabolites secondaires (Weber *et al.*, 2015) et sur la base de donnée BAGEL 3 pour la recherche d'éventuels peptides et bactériocines (de Jong *et al.*, 2006).

2.2.3. Résultats

2.2.3.1. Analyse NCBI

L'analyse de l'insert du clone B1 sur le site de Burkholderia.com indique que la région insérée s'étend de : 270994 à 271913bp et comprend le locus *BTH_II0226*,

une partie d'une région intergénique en amont et en aval de ce gène ainsi qu'une partie d'une protéine hypothétique inconnue nommée *BTH_II0227* (séquences en annexe B). Les ORF obtenus par ORF Finder nous indiquent la présence de 9 ORF possibles, ces derniers ont tous été rapportés dans le tableau 2.2. L'ORF 4 et l'ORF 5 sont respectivement les locus *BTH_II0226* et *BTH_II0227*. Les ORF 2, 3, 6 et 9 n'ont aucune homologie sur NCBI, et les ORF 1, 7 et 8 sont comparées aux protéines hypothétiques chez d'autres espèces. L'insert du clone B2 s'étend de : 295870 à 299889 pb. Il inclut le locus *BTH_II0244*, *BTH_II0245*, *BTH_II0246* et une partie du locus *BTH_II0243*. Avec une taille minimum de 150 pb, 19 ORFs sont proposés (tableau 2.2). Parmi ces ORFs, la moitié n'a aucune homologie sur NCBI. L'insert des deux clones identiques B3 et B4 s'étend de 237640 à 239649 pb avec 10 ORFs possibles. Tel qu'observé chez le clone B1, les ORFs reliés à des gènes connus de *Burkholderia thailandensis* ont été identifiés. Aucune information ne peut être déduite quant à l'origine du(es) gène(s) responsable(s) de l'activité antimicrobienne observée. Mais il est à noter que tous les inserts de ces clones (B1, B2, B3 & B4) proviennent du chromosome II de la souche *B. thailandensis*.

Tableau 2.2. Analyse des ORFs possibles des clones actifs sélectionnés pBeloBAC11 B1-B4 issus du criblage fonctionnel de l'ADN génomique de *Burkholderia thailandensis* décrit dans l'article #1.

ORF des clones	Taille en acides aminés	Protéines homologues (numéro d'accèsion)	Couverture %	Identité %
pBeloBAC11 B1				
ORF1	141	Protéine hypothétique de <i>Vibrio harveyi</i> (WP 033006950.1)	60	29
ORF2	115	Aucune homologie		
ORF3	33	Aucune homologie		
ORF4	121	Protéine de la famille du facteur de transport nucléaire (WP	100	100

		009894939.1)		
ORF5	100	Protéine hypothétique (WP 020415513.1)	19	100
		<i>BTH_II0226</i>		
ORF6	50	Aucune homologie		
ORF7	132	PadR régulateur transcriptionnel <i>Rhodococcus</i> sp. OK269 (WP 115968088.1)	55	41
ORF8	25	Protéine hypothétique <i>Pseudomonas</i> sp. 286 (WP 122587536.1)	44	91
ORF9	31	Aucune homologie		
pBeloBAC11 B2				
ORF1	180	Protéine de la famille des recombinases	42	38
ORF2	444	D-serine déshydratase, <i>Burkholderia thailandensis</i> (WP 025370071.1)	99	100
ORF3	160	Aucune homologie		
ORF4	54	Régulateur transcriptionnel LysR, <i>Burkholderia thailandensis</i> (WP 009894988.1)	100	100
		<i>BTH_II0243</i>		
ORF5	50	Glycoside hydrolase Nocardioïdes (WP 056710377.1)	76	45
ORF6	98	Aucune homologie		
ORF7	120	Aucune homologie		
ORF8	159	Aucune homologie		
ORF9	184	Aucune homologie		
ORF10	161	Transporteur de la famille DMT <i>Labrys</i> sp (WP 082980045.1)	58	39
ORF11	120	Aucune homologie		
ORF12	410	Protéine hypothétique <i>Micromonospora</i> sp (WP	16	38

123602669.1				
ORF13	124	Régulateur transcriptionnel LysR, <i>Burkholderia thailandensis</i> (WP 009894990.1) <i>BTH_I10244</i>	100	100
ORF14	83	Aucune homologie		
ORF15	131	Facteur d'élongation G Candidatus <i>Nitrospira inopinata</i> (WP 062486297.1)	38	36
ORF16	57	Aucune homologie		
ORF17	137	Aucune homologie		
ORF18	191	Aucune homologie		
ORF19	60	Aucune homologie		
pBeloBAC11 B3 & B4				
ORF1	219	Régulateur transcriptionnel <i>Burkholderia thailandensis</i> (WP 009894882.1)	100	100
ORF2	73	MROH8 protéine <i>Vulpes vulpes</i> (XP 025865510.1)	54	44
ORF3	51	Protéine hypothétique <i>Burkholderia pseudomallei</i>	72	65
ORF4	90	Peptide synthase <i>Burkholderia stagnalis</i> (WP 059960434.1)	25	83
ORF5	114	Aucune homologie		
ORF6	101	Aucune homologie		
ORF7	89	Protéine hypothétique <i>Burkholderia pseudomallei</i> (WP 103237121.1)	93	35
ORF8	68	Hydrolase alpha/bêta <i>Burkholderia thailandensis</i> (WP 009894888.1)	77	98
ORF9		Phosphoerythronate dehydrogenase PdxB <i>Pseudomonas parafluva</i>		
ORF10	105	Protéine hypothétique <i>Leptonema illini</i> (WP 002773471.1)	59	31
ORF11	60	Aucune homologie		

ORF12	123	Histidine kinase Allokutzneria albata (WP 030426657.1)	45	32
ORF13	60	Pectate lyase P59 <i>Capsicum annuum</i> (XP 016561434.1)	66	38
ORF14				
ORF15	83	Protéine hypothétique <i>Burkholderia</i> (WP 124653589.1)	57	52
ORF16	111	Aucune homologie		
ORF17	114	Alpha-L_rhamnosidase (WP 102978152.1)	37	49
ORF18	97	Beta-ketoacyl-ACP synthase <i>Fischerella</i> sp (WP 026732139.1)	48	40
ORF19	237	Aucune homologie		

2.2.3.2. Analyse par antiSMASH

L'analyse antiSMASH du clone B1 n'a pas été possible étant donné que la taille minimum d'une séquence d'ADN acceptée pour l'analyse est de 1000 pb. Pour les clones B2, B3 (B4), aucune prédiction de métabolites secondaires n'a pu être trouvée.

2.2.3.3. Analyse par BAGEL

L'analyse BAGEL 3 (de Jong *et al.*, 2006) des clones a généré un résultat identique pour chacun des clones, indiquant la présence d'un peptide similaire à celui de la bottromycine (séquence en figure 2.8). La bottromycine est un peptide ayant une activité antibiotique produite par *Streptomyces bottropensis* (Waisvisz *et al.*, 1957). Ce peptide s'est avéré actif contre SARM et *Enterococci* résistante à la vancomycine (VRE) (Shimamura *et al.*, 2009). Même si ces clones n'avaient aucun pourcentage de similarité entre eux, il est à noter que ce peptide n'a jamais été rapporté dans la littérature comme étant produit chez le genre *Burkholderia*. Étant

donné que cette application ne permet pas de déterminer le pourcentage de similarité, un BLASTp du peptide proposé a été réalisé, mais aucune similarité n'a été déterminée avec le peptide bottromycine. Le résultat de l'application BAGEL n'apporte finalement aucune information sur l'origine de cette séquence.

Un alignement de chaque insert des clones B1, B2, B3-B4 (figure 2.8 B, C, D) avec le groupe de gène de production de la bottromycine indique que l'insert du clone B1 est similaire à 58% à celui de la méthyle-transférase (*bmbF*) avec une couverture de 45%. Les résultats sont similaires pour les inserts des clones B2, B3-B4. Le pourcentage de similarité obtenu n'est pas suffisant pour assigner une fonctionnalité aux inserts de ces clones.

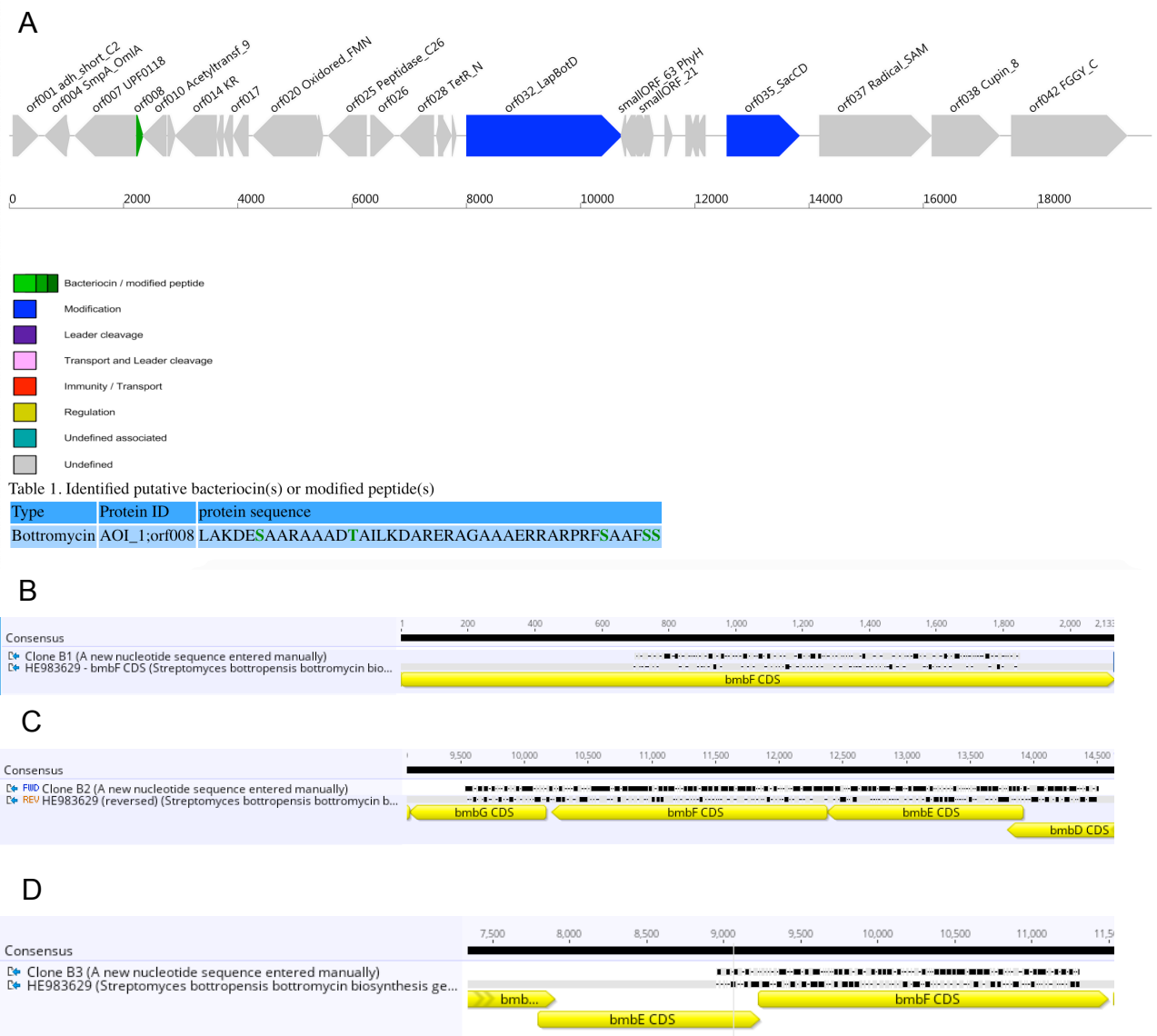


Figure 2.8. Analyse des clones pBeloBAC11-B1, pBeloBAC11-B2, pBeloBAC11-B3 (B4) dérivés de la banque d'ADN génomique de *B. thailandensis* E 264 chromosome II. (A) analyse des clones à l'aide du logiciel BAGEL3, (B, C et D) recherche de correspondance du clone B1, B2, B3-B4 avec le groupe de gènes de production de la bottromycine réalisé à l'aide de Geneious. *bmbF* : méthyle-transférase, *bmbE* : thiazoline biosynthesis, *bmbG* : hydrolase for cyclic tetrapep formation.

2.3. RÉSULTATS SUPPLÉMENTAIRES DU CRIBLAGE DE LA BANQUE D'ADN ISSUS DE *B. THAILANDENSIS*

2.3.1. Mise en contexte

D'autres clones issus du criblage fonctionnel de la banque d'ADN génomique de *Burkholderia thailandensis* se sont avérés actifs. Mais, en raison de certains problèmes lors de la confirmation de l'activité observée, ces clones n'ont pas été présentés dans l'article précédent, mais sont rapportés dans cette section

2.3.2. Méthodologie

Les clones suivants ont été sélectionnés par le même criblage fonctionnel qui a permis d'identifier les clones B1-B3 (B4) (article #1). Ces clones ont par la suite subi un test de confirmation de l'activité tel qu'indiqué dans l'article #1 (section 2.1.4.6). Les ORFs résultants de la recherche par ORF Finder ont subi un blast sur NCBI UniProtKB/Swiss-Prot (swissprot) et RefSeq-proteins. Les séquences ont aussi été analysées sur la base de donnée antiSMASH 3.0 pour une annotation des métabolites secondaires (Weber *et al.*, 2015) et sur la base de données BAGEL 3 pour la recherche d'éventuels peptides et bactériocines (de Jong *et al.*, 2006).

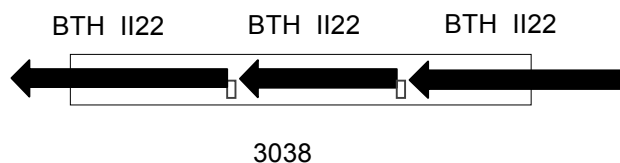
2.3.3. Résultats

2.3.3.1. Autres clones sélectionnés

Suite au criblage fonctionnel réalisé sur la banque d'ADN génomique de *Burkholderia thailandensis*, deux autres clones nommés B5 et B6 ont été sélectionnés. Ces clones étaient actifs contre *Staphylococcus aureus* Newman.

L'insert du clone B5 s'étend de : 2739816 à 2742854 pb avec 3038 pb (figure 2.9 A). Il comprend le gène *BTH_II2232* qui est une protéine hypothétique inconnue retrouvée chez l'espèce *B. thailandensis* ainsi qu'une partie du locus *BTH_II2231* qui est une protéine de la famille des transporteurs et *BTH_II2233* qui est une protéine de la famille CAIB et BAIF. L'insert de ce clone B6 s'étend de 446857 à 447934 pb avec 1077 pb (figure 2.9 B). Il comprend le locus *BTH_II0377* qui est une protéine hypothétique et le gène *BTH_II0376* qui est un régulateur de la réponse de liaison à l'ADN. La confirmation de l'activité de ces deux clones était problématique. En effet, suivant le test de confirmation de l'activité, un seul réplicat a pu être reproduit pour chacun de ces deux clones. L'activité notée était variable d'un réplicat à l'autre, ce qui explique la raison pour laquelle ces clones n'ont pas été intégrés dans la publication ci-haut. Néanmoins, l'analyse des ORFs par ORF Finder a été réalisée (tableau 2.3). Les protéines hypothétiques retrouvées chez les deux clones pBeloBAC11-B5, pBeloBAC11 B6, pourraient être à l'origine de l'activité observée. Toutefois, des tests de confirmation de l'activité par retransformation du vecteur pBeloBAC11-B5 et pBeloBAC11-B6 chez *E. coli* devraient être réalisés.

(A) pBeloBAC11-B5



(B) pBeloBAC11-B6

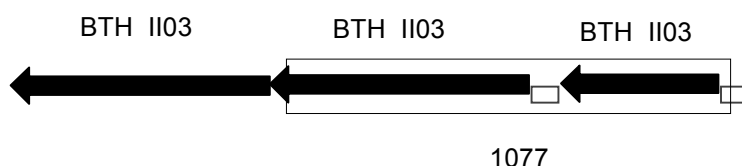


Figure 2.9. Contexte chromosomique des inserts identifiés chez les clones : pBeloBAC11-B5 et pBeloBAC11-B6 dérivés de la banque d'ADN génomique de *B. thailandensis* E 264 chromosome II.

Tableau 2.3. Analyse des ORFs possibles des clones sélectionnés pBeloBAC11-B5-B6 issus du criblage fonctionnel de l'ADN génomique de *Burkholderia thailandensis*.

ORF des clones	Taille en acides aminés	Protéines homologues (numéro d'accension)	Couverture %	Identité %
pBeloAC11-B5				
ORF1	118	Protéine de transcription (WP044298559.1)	50	36
ORF2	278	Aucune homologie		
ORF3	259	Aucune homologie		
ORF4	88	Antigène O <i>Rhizobiales bacterium</i> (WP112397002.1)	32	52
ORF5	82	Protéine hypothétique (WP043467630.1)	31	56
ORF6	231	Aucune homologie		
ORF7	261	Aucune homologie		
ORF8	139	CoA transférase (WP009894225.1)	66	100
ORF9	168	Cyclase <i>Burkholderia</i>	80	93

<i>pseudomallei</i>				
(WP122985119.1)				
ORF10	118	MFS transporter (WP009901476.1)	44	100
ORF11	105	Protéine RlpA transglycosylase (WP025001967.1)	84	30
ORF12	168	Aucune homologie		
ORF13	172	MFS transporter (WP009894224.1)	95	99
ORF14	51	Protéine hypothétique (WP122835898.1)	100	98
ORF15	193	Régulateur transcriptionnel (WP011401565.1)	100	100
ORF16	230	MFS transporteur (WP009901476.1)	84	100
pBeloBAC11-B6				
ORF1	41	Aucune homologie		
ORF2	52	Aucune homologie		
ORF3	34	Aucune homologie		
ORF4	30	Termineur transcriptionnel	96	62
<i>Actinomyces canis</i>				
(WP124871490.1)				
ORF5	66	Aucune homologie		
ORF6	81	Protéine metallophosphoesterase	39	41
<i>Aerococcus urinaehominis</i>				
(WP067979255.1)				
ORF7	64	Aucune homologie		
ORF8	165	Régulateur de réponse	32	100
<i>Burkholderia</i> sp				
(WP066486038.1)				
ORF9	171	Régulateur de réponse (WP009895231.1)	99	100
ORF10	94	Protéine hypothétique (WP009907199.1)	100	100

CHAPITRE 3 : ÉTUDE MÉTAGÉNOMIQUE FONCTIONNELLE

3.1.1. Mise en contexte

De nombreux indices semblent indiquer que *i*) le lieu d'isolement de l'ADN en vue d'une étude métagénomique détermine le type de molécules retrouvées lors du criblage fonctionnel (tableau 1.2), et que *ii*) la méthode de criblage utilisée doit être très sensible afin de permettre la détection de la plus petite zone d'inhibition produite par les clones transformants. Récemment, nous avons développé une méthode de criblage très sensible nommée *bacteriospray* qui permet le criblage de 5 à 10 fois plus de clones par rapport à la méthode traditionnelle de double couche d'agar utilisée dans les études métagénomiques. Pour faire suite à ces travaux, nous avons prélevé trois différents biofilms pour la construction de plusieurs banques d'ADN métagénomiques dans le but d'un criblage fonctionnel par *bacteriospray* de ces banques d'ADN contre SARM.

3.1.2. Matériel et méthodes

3.1.2.1. Échantillonnage de l'ADNe

Trois différents biofilms issus de conduits d'égouts sanitaires d'hôpitaux ont été prélevés. Le premier biofilm nommé EHW (Egypt Hospital Wastewater) provient d'un hôpital d'Égypte, le deuxième biofilm nommé LSW (Laval Sanitary Wastewater) a été prélevé d'un égout sanitaire où se déversent les eaux sanitaires urbaines à celles d'hôpitaux, et le troisième biofilm nommé MHW (Montreal Hospital Wastewater) a été prélevé directement d'un conduit d'égout sanitaire d'un hôpital de la région de Montréal. Les échantillons de biofilms ont été conservés dans des sacs en plastique stériles et transportés au laboratoire sur glace pour une extraction directe d'ADN (le même jour).

3.1.2.2. Digestion enzymatique du vecteur pBeloBAC11

La digestion du vecteur pBeloBAC11 a été réalisée selon les conditions suivantes: 2 μ L du tampon de l'enzyme, 1 μ L d'une solution aqueuse d'ADN du vecteur (0,5 à 1 μ g/ μ L), et 1 μ L d'enzymes (NEB, New England Biolabs) ont été dilués avec 16 μ L d'eau milliQ.

3.1.2.3. Extraction de l'ADNe des biofilms

Le biofilm nommé MHW a été utilisé pour une optimisation de l'extraction d'ADN en comparant trois différentes trousse d'extraction d'ADN: le FastDNA spin de MP Biomedicals, la trousse d'extraction d'ADN PowerSoil de MoBio de Qiagen et la trousse GeneJET, de purification d'ADN génomique de Thermo Scientific, en suivant le protocole proposé par chaque trousse d'extraction. La pureté de l'ADN extrait a été évaluée en déterminant le rapport A260/A280 (rapport optimal compris entre 1,8 et 2) et le rapport A260/A230 (rapport optimal à 1,8). L'ADN a par la suite été visualisé par électrophorèse sur gel d'agarose 0,8%.

3.1.2.4. Digestion de l'ADNe des biofilms

L'ADN extrait des biofilms a été digéré à l'aide de trois enzymes de restriction du site de clonage multiple du vecteur pBeloBAC11 (*Bam*HI, *Hind*III et *Sph*I) afin de déterminer le profil de digestion de chaque enzyme, pour ensuite sélectionner l'enzyme permettant d'obtenir les fragments d'ADN les plus longs après l'étape de digestion. Le temps de digestion a ensuite été optimisé (*Sph*I seulement, 30 min, 1h à 2h), et l'ADN digéré a été vérifié par électrophorèse sur gel après 30 min.

3.1.2.5. Extraction sur gel d'agarose de l'ADNe digéré

Après digestion de l'ADN, un gel d'agarose (4 h) a permis de séparer les fragments d'ADN digérés en fonction de leur taille (Brady, 2007). Dans le but de

construire des banques d'ADN à larges inserts, l'ADN dont la taille était supérieure à 10 kb a été extraite sur gel d'agarose à l'aide de différentes trousse d'extractions d'ADN (Qiagen et Zymoclean) en suivant le protocole proposé par les deux fournisseurs.

3.1.2.6. Ligation de l'ADNe au vecteur pBeloBAC11

Deux types de ligations d'ADN à l'aide de la T4 ligase ont été testés : une ligation avec des extrémités d'ADN à bords cohésifs où les enzymes de restrictions *BamHI*, *HindIII* et *SphI* ont été utilisés, et une ligation avec des extrémités d'ADN à bords francs, où l'enzyme *SmaI* (site de restriction juste avant l'enzyme *BamHI* sur le vecteur pBeloBAC11) a été utilisé pour la digestion du vecteur pBeloBAC11 et de l'ADNe. Pour la ligation à bords francs, la T4 polymérase a été utilisée pour le polissage de l'ADN après digestion avec l'enzyme *SmaI* selon les recommandations proposées par le fournisseur des enzymes (NEB, New England Biolabs).

3.1.2.7. Transformation

Trois types de cellules compétentes *E. coli* ont été testés pour sélectionner celles qui permettaient d'obtenir la meilleure efficacité de transformation. Les premières cellules compétentes utilisées ont été préparées au laboratoire en suivant le protocole de préparation de cellules compétentes au CaCl_2 , suivi d'une transformation par chauffage (Chang *et al.*, 2017). Les deux autres types de cellules compétentes utilisées par la suite provenaient de sources commerciales (*E. coli* DH10B de NEB et MegaX DH10B de Thermo Scientific). Les cellules compétentes commerciales étaient transformées par électroporation suivant les conditions proposées par les fournisseurs.

3.1.2.8. Criblage fonctionnelle

Deux méthodes de criblage ont été utilisées: en premier lieu, la méthode de double couche d'agar et par la suite la méthode *bacteriospray*, tel qu'indiqué dans les chapitres 2, 3 et 4.

3.1.2.9. Analyse des clones sélectionnés

3.1.2.9.1. Analyse de la morphologie

Une première analyse de la morphologie des clones sélectionnés a été réalisée en déposant 5 µL de pré-culture dans le milieu LB de chaque clone sur gélose LB (40 µg/mL X-gal) contenant 1 mM d'IPTG et 20 µg/mL de chloramphénicol. Après une nuit (O/N) d'incubation à 37°C, la couleur des clones a été vérifiée, puis toute morphologie particulière notée.

3.1.2.9.2. Confirmation de l'activité

Les clones ayant la plus petite zone d'inhibition ont été repiqués pour confirmer l'activité. Sur une plaque de 20 x 20 cm² contenant le milieu LB + 20 µg/mL chloramphénicol, 5 µL d'une pré-culture du clone sélectionné a été déposé. Après un jour d'incubation O/N à 37°C suivi de 5 jours d'incubation à la température de la pièce, la confirmation de l'activité a été effectuée par *bacteriospray*. Ce test a été effectué en triplicat. Seuls les clones ayant reproduit le phénotype observé au moins trois fois ont été sélectionnés pour confirmer l'activité par retransformation du vecteur chez une autre souche *E. coli* DH10B (tel que détaillé dans le chapitre 4 section 4.1.4.5.).

3.1.2.9.3. Séquençage de l'insert et analyse de la séquence par BLASTn

L'amplification par PCR des vecteurs de chaque clone ayant été extraits par miniprep avec une trousse d'extraction de plasmide (BioBasic) a été effectuée en utilisant les amorces universelles M13F (-40) et M13R (-26) en suivant les conditions indiquées précédemment (section 3.1.5.6). Ce produit PCR a été envoyé au séquençage Sanger. Les séquences d'ADN obtenues ont été analysées avec le logiciel CLC (version 7), et avec BLASTn pour une recherche d'homologie de séquences.

La séquence de l'insert d'ADNe du clone MHW₅₃₈ (qui fait l'objet de l'article #2 en chapitre 4) a aussi été analysée par ORF Finder (BLAST, NCBI), afin d'identifier les cadres de lecture de cette séquence. BLASTp a été utilisé pour effectuer une recherche d'homologie de séquences protéiques à partir des séquences résultantes, tandis que BAGEL et Pfam ont été utilisés pour effectuer une recherche d'homologie de séquences peptidiques. Le serveur PEP-FOLD a été utilisé pour les analyses de prédiction de structures des peptides.

3.1.2.10. Synthèse, purification, caractérisation des peptides et test de leur effet sur la croissance de *Staphylococcus aureus* résistant à la méthicilline

Cette partie est détaillée en chapitre 4 sections 4.1.4.9 et 4.1.4.12.

3.1.3. Résultats

3.1.3.1. Prélèvement des biofilms

Le prélèvement du biofilm nommé MHW a été réalisé par moi-même à partir d'un conduit d'égout sanitaire d'un hôpital situé à l'ouest de la région de Montréal (Figure 3.1A). Le biofilm provenant de l'hôpital d'Égypte (EHW) a été prélevé par un ancien étudiant du laboratoire du professeur Éric Déziel, soit le professeur Ahmad M. Abdel-Mawgoud M. Saleh (Figure 3.1B). Enfin, le biofilm nommé LSW a été prélevé

par des collègues de la Chaire industrielle CRSNG en eau potable de Polytechnique d'un conduit d'égout au nord de Montréal où se déversent aussi bien des égouts sanitaires que des égouts hospitaliers (Figure 3.1C). Par souci de confidentialité, nous ne nommerons pas les hôpitaux qui ont bien accepté de nous donner l'accès pour effectuer les prélèvements.



Figure 3.1. Prélèvement des biofilms : (A) prélevé d'un conduit d'égout d'un hôpital de la région de Montréal, (B) prélevé d'un conduit d'égout d'un hôpital d'Égypte (C) prélevé d'un conduit d'égout près d'un hôpital du nord de Montréal.

3.1.3.2. Digestion du vecteur pBeloBAC11

Étant donné la taille du vecteur pBeloBAC11 qui est de 7,5 kb, la digestion avec les trois enzymes de restriction *BamHI*, *HindIII* et *SphI* retrouvés dans le site de clonage multiple a été problématique. Aucune digestion complète n'était possible en-dessous d'un temps de digestion de 4 h (Figure 3.2).

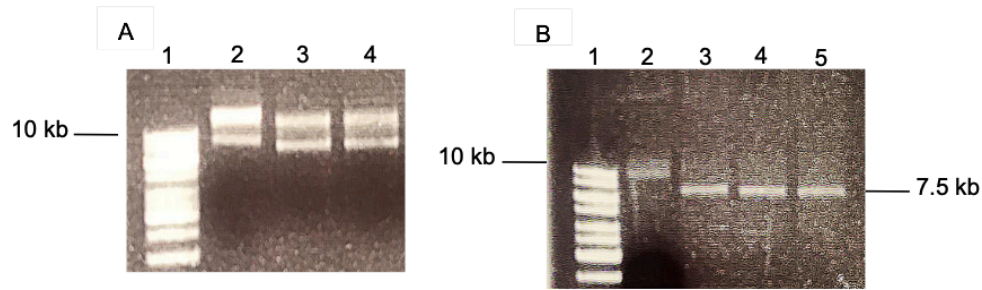


Figure 3.2. Temps de digestion du vecteur pBeloBAC11. (A) digestion enzymatique non complétée du vecteur pendant 2 h, puits 1 : échelle de poids moléculaire 1 kb (NEB), puits 2, 3, 4 : digestion du vecteur pBeloBAC11 avec les enzymes *BamHI*, *HindIII*, *SphI*, respectivement, (B) digestion enzymatique du vecteur pendant 4 h, puits 1 : échelle de poids moléculaire 1 kb (NEB), puits 1 : témoin vecteur non digéré, puits 3, 4, 5 : digestion du vecteur pBeloBAC11 avec les enzymes *BamHI*, *HindIII*, *SphI*, respectivement.

3.1.3.3. Optimisation de l'extraction d'ADN

En comparant les trois trousse d'extraction d'ADN testées, le rendement le plus élevé a été obtenu avec la trousse MP Biomedicals (Tableau 3.1), comparativement à l'ADN extrait à l'aide de la trousse MoBio. L'ADN obtenu avec la trousse d'extraction d'ADN génomique Gene JET était peu abondant et très dégradé (Figure 3.3). La pureté de l'ADN en analysant le ratio A260 / A280 et A260 / A230 était satisfaisante à la fois avec les kits MP Biomedicals et MoBio. Étant donné que la taille de l'ADN obtenue à l'aide de la trousse d'extraction d'ADN MP était supérieure à 48,5 kb (Figure 4.3 (A) puits : 3) comparé à l'ADN obtenu avec la trousse d'extraction d'ADN MoBio (Figure 3.3 (A) puits : 2), ce kit fut choisi pour toutes les extractions d'ADN subséquentes réalisées à partir des biofilms.

Tableau 3.1. Comparaison entre les rendements d'extraction d'ADNe de biofilms par différentes trouses.

Kits d'extraction d'ADN	Rendement d'ADN ($\mu\text{g/g}$)	A_{260}/A_{280}	A_{260}/A_{230}
FastDNA spin kit de MP	47.6 ± 10.7	1.99 ± 0.2	2.00 ± 0.13
MoBio PowerSoil, kit d'extraction d'ADN de Qiagen	35.47 ± 7.6	1.95 ± 0.42	1.75 ± 0.28
Gene Jet génomique, kit d'extraction DNA de Thermo Scientific	23.7 ± 5.13	1.62 ± 0.5	1.6 ± 0.35

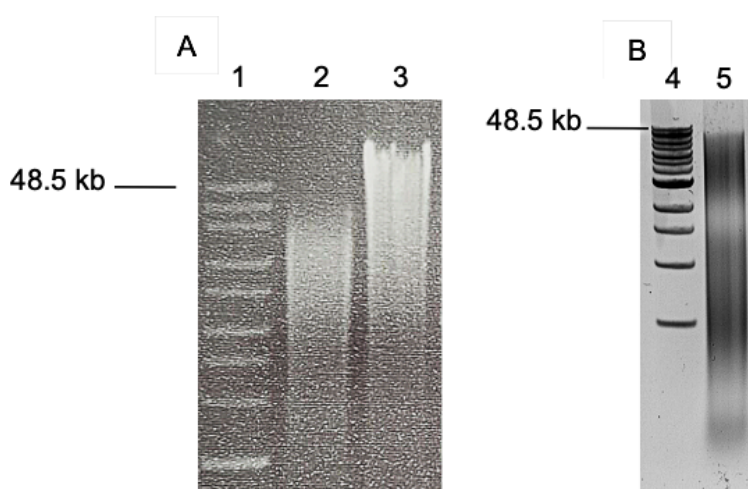


Figure 3.3. Comparaison entre les trois trouses d'extraction d'ADN testées, (A) comparaison entre les kits d'extraction d'ADN, puits 2 : MP Biomedicals, puits 3 : MoBio et (B), puits 5 : le kit d'extraction d'ADN Gene Jet, puits 1 & 4 : extension de l'échelle de poids moléculaire d'ADN 1kb (NEB).

3.1.3.4. Digestion de l'ADN à partir des biofilms

Les optimisations du temps de digestion de l'ADN de biofilms ont démontré que, quoique partiel, le niveau de digestion de l'ADN était suffisant après 30 min (Figure 3.4). Donc, une digestion partielle fut choisie pour toutes les digestions de biofilms. De plus, il n'y avait pas de différence de taille d'ADN notable entre la digestion à partir de chacun des trois enzymes de restriction, *BamHI*, *HindIII*, *SphI* de façon indépendante. C'est pour cette raison et pour favoriser la diversité des banques

d'ADN métagénomiques que plusieurs de ces banques d'ADN ont été générées avec les trois enzymes.

3.1.3.5. Extraction d'ADN des gels d'agarose pour la construction de banques d'ADN métagénomiques

Dans le but de créer des banques d'ADN à larges inserts, nous avons tenté d'extraire sur gel l'ADN digéré dont la taille était supérieure à 10 kb suivant le protocole établi par les différentes trousse d'extraction d'ADN utilisées. Le rendement de l'extraction d'ADN sur gel d'agarose (Figure 3.4 A) obtenu à partir de la trousse d'extraction sur gel Qiagen était faible alors que celui obtenu en utilisant le kit d'extraction sur gel Zymoclean était plus élevé. Néanmoins, ce dernier ne s'est pas avéré suffisant pour atteindre le ratio BAC:ADN proposé dans l'optimisation de l'efficacité de transformation, soit 1:9 (chapitre 2 section 2.1.5.1). Une concentration de plusieurs des produits d'extraction sur gel a été réalisée afin d'obtenir les ratios nécessaires. En raison des problèmes rencontrés l'étape d'extraction sur gel n'a pas été réalisée lors de la construction des banques d'ADN métagénomiques de ce chapitre et du chapitre 4.

En effet, la figure B (puits 3, 4, 5, 6) montre la digestion de quatre clones sélectionnés aléatoirement lors de la construction d'une banque d'ADN sans extraction sur gel. La taille des inserts varie de 2 kb (puit 5) à un insert dont la taille était supérieure à 10 kb (puit 3). Le clone du puit 6 ne comprenait aucun insert. Toutefois, pour certains clones indiqués en figure 3.4 C, des problèmes de digestion ont été rencontrés (puits 3, 5, 6, 7).

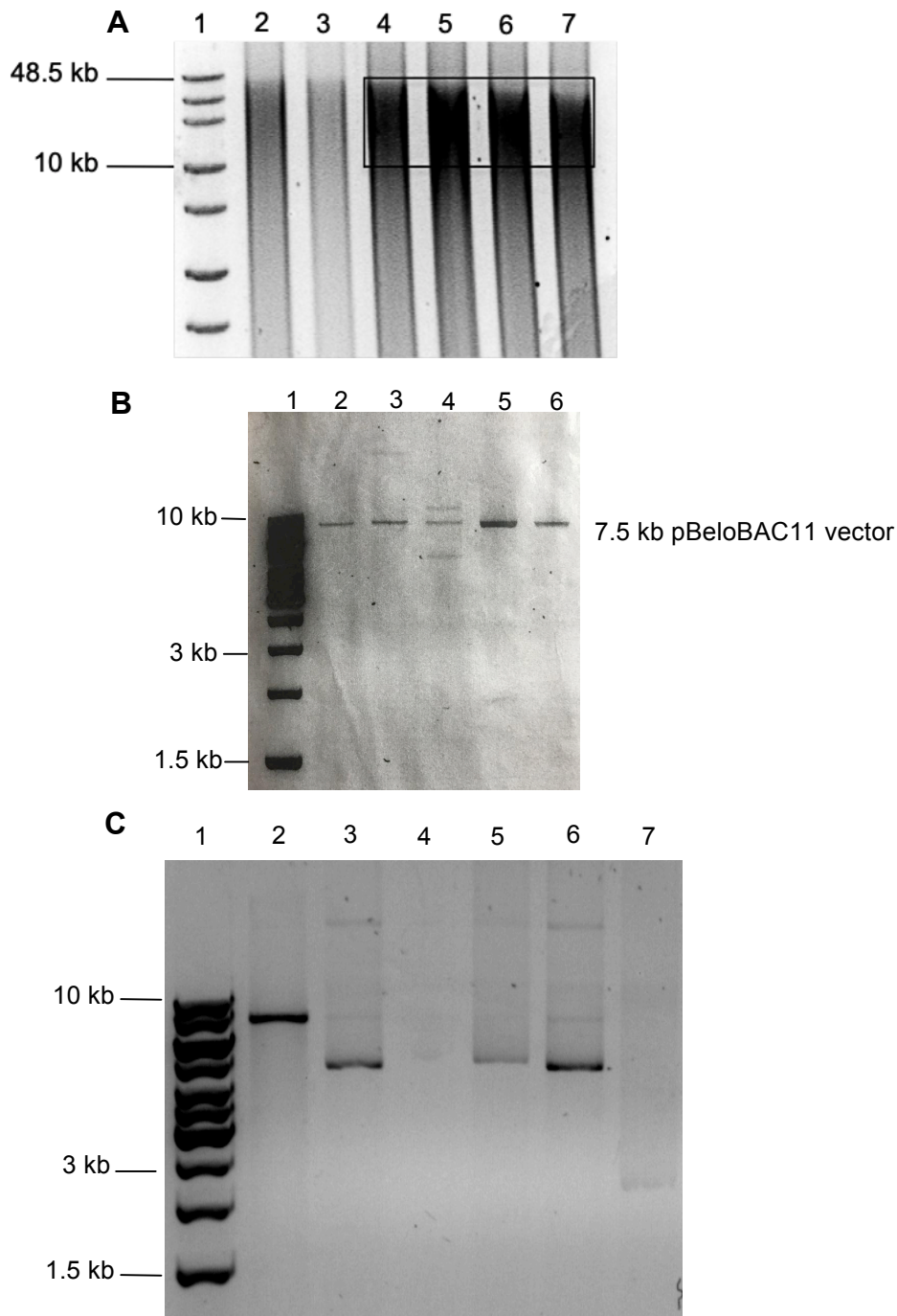


Figure 3.4. Construction de banques d'ADN avec et sans extraction sur gel. (A) migration de l'ADN digéré et extraction de la bande de gel d'agarose de l'ADN d'une taille supérieure à 10 kb. Puit 1 : extension de l'échelle de poids

moléculaire d'ADN 1kb (NEB), puits 2, 3 : ADN insuffisant dans les puits pour réaliser une extraction sur gel, puits 4, 5, 6, 7 : ADN digéré par l'enzyme SphI. NB : un puits témoin d'ADN non digéré aurait pu être ajouté. **(B) digestion de 4 clones aléatoires après construction de banque sans l'étape extraction sur gel de l'ADN après digestion.** Puits 1 : échelle de poids moléculaire d'ADN 1kb (NEB), puit 2 : digestion du vecteur vide, puit 3 : digestion d'un clone dont la taille de l'insert est supérieur à 10 kb, puit 4 : digestion d'un clone ayant deux inserts un supérieur à 10 kb et l'autre de 5 kb environ, puit 5 : digestion d'un clone dont l'insert a une taille de 2 kb, puit 6 : digestion d'un clone n'ayant pas d'insert, **(C) digestion de 4 inserts ayant un problème de digestion.** Puits 1 : échelle de poids moléculaire d'ADN 1kb (NEB), puit 2 : digestion du vecteur vide, puit 3, 5, 6, 7 : clones ayant un problème de digestion, puit 4 : vide. Il est à noter que pour la figure (B) et (C) un témoin du vecteur non digéré aurait pu être ajouté.

3.1.3.6. Ligation à bords francs vs ligation à bords cohésifs

En suivant les étapes de construction de banques d'ADN optimisées (chapitre 2 section 2.1.4.2), l'efficacité de transformation notée après une ligation à bords francs était aux alentours de 6×10^2 UFC/ μ g d'ADN, tandis que l'efficacité de transformation après une ligation à bords cohésifs était de 2.5×10^5 UFC/ μ g (Chapitre 2 section 2.1.5.1). C'est pour cette raison que toutes les ligations effectuées par la suite au cours de ce projet furent des ligations à bords cohésifs.

3.1.3.7. Cellules compétentes et efficacité de transformation

Les cellules compétentes commerciales se sont avérées plus efficaces que celles préparées au CaCl_2 . De plus l'efficacité de transformation obtenue en utilisant les cellules compétentes de la compagnie NEB était comparable à celle obtenue à partir des cellules compétentes de la compagnie Thermo Scientifique. Les cellules compétentes *E. coli* DH10B de NEB ont donc été choisies pour toutes les transformations qui ont par la suite été effectuées au cours de ce projet.

3.1.3.8. Criblage fonctionnel

À partir de la méthode de double couche d'agar, environ 20,000 clones ont été criblés puis 1,134 clones ont été sélectionnés pour une première confirmation de l'activité, tel qu'illustré à la figure 3.5. Il est à noter qu'aucun clone n'a reproduit le phénotype d'origine : tous correspondaient à des faux positifs. Avec la méthode *bacteriospray* (figure 3.6), 150,000 clones ont été testés et 320 clones ont été sélectionnés pour la première étape de confirmation du phénotype. Seuls dix clones ont retrouvé leur phénotype de départ (en triplicat) et se sont avérés actifs contre SARM. On peut remarquer de ces deux méthodes de criblage que plus de clones ont été sélectionnés pour la confirmation de l'activité par la méthode de double couche d'agar comparativement à la méthode *bacteriospray* (1,134 contre 320 respectivement). La principale raison est qu'il était très difficile avec la méthode de double couche d'agar de distinguer entre un clone positif ayant une zone d'inhibition de croissance et un clone avec un halo d'agar (Figure 3.7). Ces clones étaient alors sélectionnés pour confirmer une activité inexistante. Tel que démontré au chapitre 2, l'efficacité et la fiabilité de la méthode *bacteriospray* permet le criblage de 5 à 10 fois plus de clones que la méthode de double couche d'agar conventionnelle.

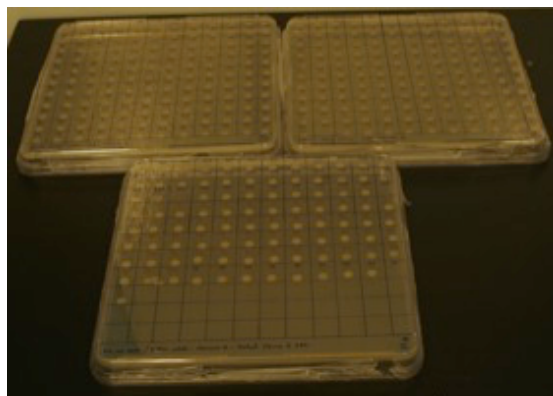


Figure 3.5. Exemple de la méthode de confirmation des clones. À partir de 5 μL d'une pré-culture, chaque clone issu du criblage est déposé sur gélose X-gal-chloramphénicol-IPTG pour une première confirmation d'activité. 1,134 clones sélectionnés sur 20,000 clones furent criblés à partir de la banque d'ADN métagénomique du biofilm MHW par la méthode de double couche d'agar versus 320 clones sélectionnés sur environ 150,000 clones ayant été criblés à partir de la banque d'ADN métagénomique du biofilm MHW par *bacteriospray*.



Figure 3.6. Dispositif complet du *bacteriospray*. Les boîtes de Pétri de 20x20 cm^2 utilisées pour le criblage fonctionnel étaient entourées d'une boîte en plastique pour limiter les aérosols émis par l'aérographe.

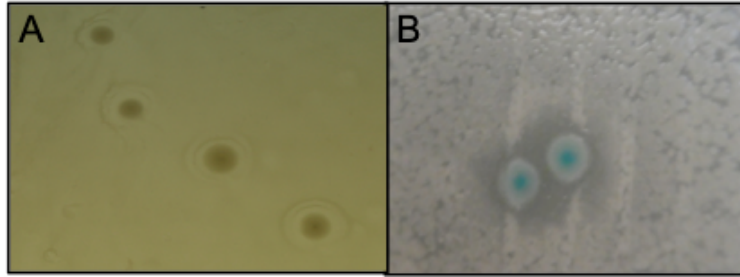


Figure 3.7. Comparaison du criblage fonctionnel entre la méthode de double couche d'agar et de *bacteriospray*. (A) exemple d'un halo d'agar autour de certains clones testés par la méthode traditionnelle, considérés de façon erronée comme étant actifs, (B) exemple de clones actifs observés au cours du criblage fonctionnel par bacteriospray.

3.1.3.8.1. Procédure détaillée des étapes de construction des banques d'ADN métagénomiques et de criblage fonctionnel

La construction d'une banque d'ADN métagénomique nécessite une planification minutieuse. La figure 3.8 résume toutes les procédures ainsi que le temps approximatif requis pour chaque étape.

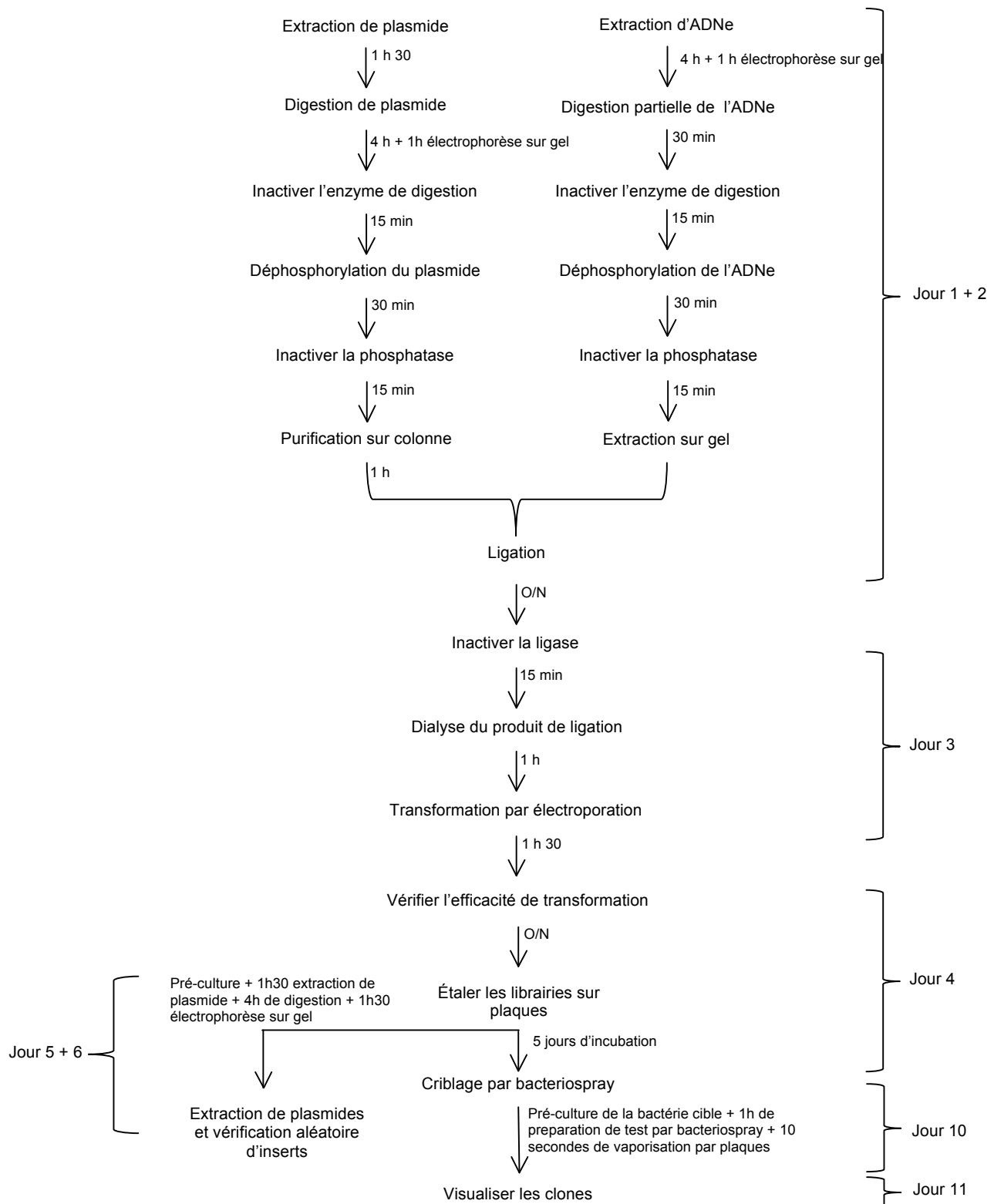


Figure 3.8. Procédure détaillée des étapes de construction et de criblage fonctionnel des banques d'ADN métagénomiques.

3.1.3.8.2. Présentation des 10 clones actifs

Le criblage fonctionnel des 150,000 clones par *bacteriospray* a mené à la sélection des dix clones actifs suivants : MHW₁23, MHW₁64, MHW₂105, MHW₂178, MHW₂300, MHW₃128, MHW₄78, MHW₄203, MHW₄317, MHW₅38 (MHW_x Y, X: représente le numéro de la banque criblée, Y: représente le numéro donné au clone).

Clones MHW₁23, MHW₁64 et MHW₂105

Ces colonies n'avaient aucune morphologie particulière. Au cours de la première étape de confirmation de l'activité antimicrobienne, ils étaient toujours actifs contre SARM. La confirmation de l'activité par retransformation des vecteurs pBeloBAC11-MHW₁23, pBeloBAC11-MHW₁64, pBeloBAC11-MHW₂105 chez *E. coli* DH10B a été problématique en raison de la difficulté d'obtenir au moins trois réplicats biologiques reproduisant tous le phénotype de départ. En dépit de cela, le produit PCR des vecteurs de chacun de ces clones a été visualisé sur gel d'agarose, et le clone MHW₁23 avait un insert d'une taille d'environ 1,650 pb, et les clones MHW₁64 et MHW₂105 avaient un insert d'environ 500 pb chacun (Figure 3.9).

D'après la figure 3.9 on peut remarquer que le produit PCR de ces trois clones contenait aussi bien le vecteur vide (bande d'ADN dont la taille avoisine les 250 pb), que le vecteur avec insert (bande d'ADN dont la taille est supérieure à 250 pb). Il est à noter que la pré-culture pour chaque clone a été réalisée à partir d'une seule colonie blanche (blanche en référence au lacZ négatif) indicatif de la présence d'un insert. Le produit PCR purifié a quand même été envoyé au séquençage Sanger, mais les séquences obtenues n'étaient pas de bonne qualité.

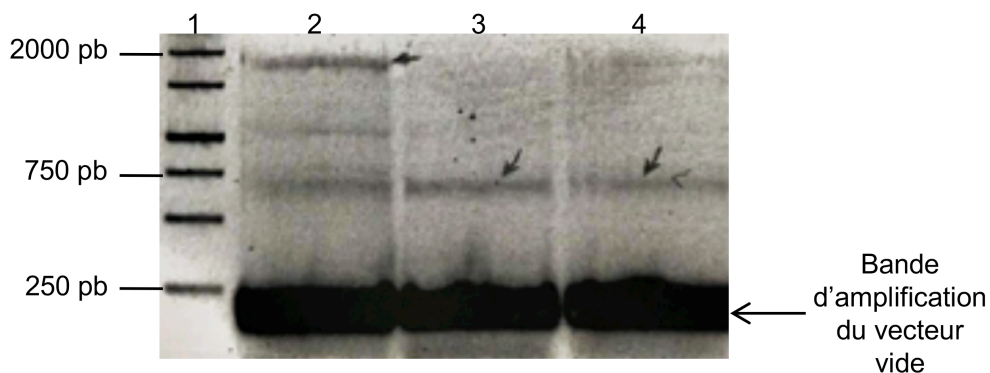


Figure 3.9. Produit PCR de l'amplification des vecteurs : puits 1 1kb échelle de poids moléculaire d'ADN (NEB), puits 2 : pBeloBAC11-WHW₁₂₃, puits 3 : pBeloBAC11-WHW₁₆₄ (puits 3), pBeloBAC11-WHW₂₁₀₅ (puits 4) visualisé sur gel d'agarose.

Clone MHW₂₁₇₈

Cette colonie avait une morphologie particulière en étoile (Figure 3.10). Au cours de la première étape de confirmation de l'activité, ce clone était toujours actif contre SARM mais cette activité ne fut plus observée aussitôt que le vecteur pBeloBAC11-MHW₂₁₇₈ fut retransformé chez *E. coli* DH10B. Le clone d'origine a été conservé car il est possible que l'insert se soit intégré dans le génome de la souche électrocompétente *E. coli* DH10B et que ce soit cette dernière qui confère l'activité antimicrobienne observée ainsi que la morphologie en étoile. L'amplification PCR à l'aide des amorces universelles du vecteur pBeloBAC11 n'a indiqué la présence d'aucun insert.

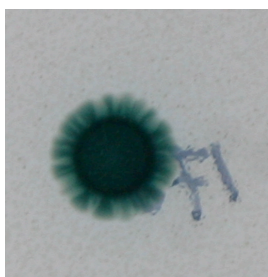


Figure 3.10. Image de la morphologie en étoile du clone MHW₂₁₇₈. Image enregistrée suivant le test d'activité de ce clone contre SARM, pour laquelle une toute petite zone d'inhibition était visible.

Clone MHW₂300

Ce clone n'avait pas de morphologie particulière. Après confirmation de l'activité contre SARM par la présence d'une petite zone d'inhibition, la visualisation du produit PCR du vecteur pBeloBAC11-MHW₂300 de ce clone (Figure 3.11) indiquait la présence d'un insert d'environ 400 pb (en fait 405 pb, d'après la séquence en annexe C). Après séquençage Sanger du produit PCR purifié, le résultat BLASTn de cette séquence n'indique aucune homologie avec les séquences existantes dans la base de données NCBI. La confirmation de l'activité par retransformation du vecteur chez *E. coli* DH10B n'a pas encore été réalisée.

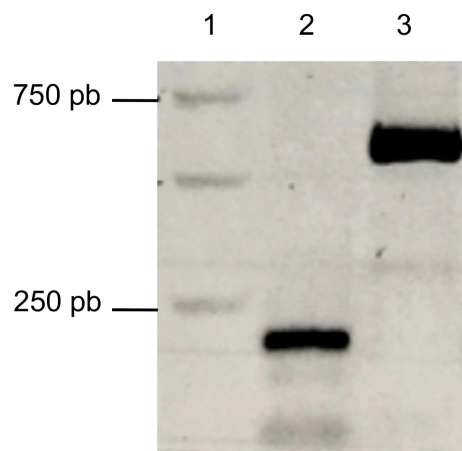


Figure 3.11. Produit PCR du vecteur pBeloBAC11-MHW₂300 (puits 3), et du témoin pBeloBAC11 vide (puits 2), (puits 1) 1kb échelle de poids moléculaire d'ADN (NEB).

Clone MHW₃128

Ce clone avait aussi une morphologie particulière, après le dépôt de 5 µL de ce clone sur gélose LB X-gal IPTG et 20 µg/mL de chloramphénicol, de petites colonies étaient formées. Cela pourrait être le résultat d'une difficulté de croissance du clone *E. coli* suite à la production d'un métabolite (Figure 3.12). Le test d'activité de ce clone contre SARM indique la présence d'une petite zone d'inhibition qui

disparaissait après 16 h d'incubation, probablement une activité bactériostatique provenant du métabolite produit. Le phénotype d'activité contre SARM a été retrouvé en retransformant le vecteur pBeloBAC11-MHW₃128 chez *E. coli* DH10B encore une fois, mais le phénotype de morphologie en petite colonie n'était pas reproduit. Le séquençage du produit PCR de ce clone indique la présence d'un petit insert de 189 pb (voir la séquence en annexe C). Le BLASTn de cette séquence n'a fourni aucune information quant à l'origine et à la fonction de cette séquence insérée.

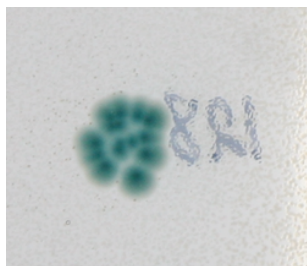


Figure 3.12. Morphologie en petites colonies du clone MHW₃128.

Clones MHW₄78, MHW₄203 et MHW₄317

Ces clones n'avaient aucune morphologie particulière. Leur activité a été confirmée contre SARM. Cependant, la reconfirmation de l'activité par retransformation du vecteur chez *E. coli* DH10B n'a pas encore été réalisée.

3.1.3.9. Le clone MHW₅38

Ce clone avait une morphologie particulière, et tout comme le clone MHW₃128, en déposant 5 µL de ce clone sur gélose LB X-gal IPTG à 20 µg/mL de chloramphénicol, de très petites colonies étaient formées. Le test d'activité de ce clone contre SARM indique la présence d'une petite zone d'inhibition qui était en fait la plus grande des zones d'inhibition qui a pu être observée durant tous les criblages fonctionnels des clones actifs. L'activité a pu être confirmée par retransformation du vecteur pBeloBAC11 MHW₅38 (figure 4.2 du chapitre 4). Sur cette figure, on peut observer que comparativement au témoin *E. coli*-pBeloBAC11

vide qui présente une meilleure croissance sur cette gélose LB, le clone *E. coli*-pBeloBAC11-MHW₅38 montre un défaut de croissance. Ce défaut de croissance est surtout visible en milieu liquide. En effet après une culture O/N de 16 h de ce clone une DO₆₀₀ de 0.33 ± 0.013 a été observée contre une DO₆₀₀ de 4.67 ± 0.01 pour le témoin (pBeloBAC11 vide).

De plus une observation morphologique particulière de ce clone en milieu solide a été identifiée, toujours par dépôt de 5 µL du clone sur gélose LB X-gal IPTG à 20 µg/mL de chloramphénicol. Après une incubation O/N à 37°C, la figure 3.13a montre deux types de colonies : deux clones de couleur bleue qui pourraient être associés au vecteur vide (repiquage 1), et tous les autres clones blancs qui pourraient être associés au vecteur comprenant l'insert (repiquage 2). Après pré-culture de ces deux repiquages, l'extraction du vecteur suivi d'une amplification PCR avec les amorces universelles M13 et d'une visualisation sur gel d'agarose, il a été constaté que l'insert était aussi bien présent chez le clone issu du repiquage 1 que celui issu du repiquage 2, mais que l'activité antimicrobienne était seulement observée pour le repiquage 2.

Après retransformation du vecteur pBeloBAC11-MHW₅38 chez *E. coli* DH10B, le défaut de croissance était toujours visible, mais moins prononcé que pour le clone d'origine. Après une culture O/N du transformant, une DO₆₀₀ de 2.4 ± 0.01 (absorbance du clone témoin : 4.67 ± 0.01) était observée mais la croissance ralentissait aussitôt pour atteindre une DO₆₀₀ de 1.75 ± 0.1 (DO du clone témoin 3.72 ± 0.18) après 24 h de culture.

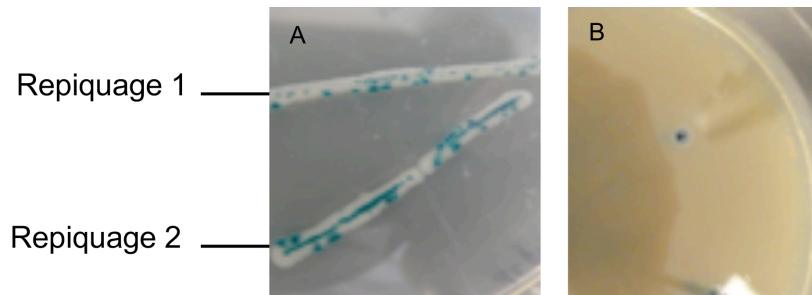


Figure 3.13. Observation de la morphologie du clone MHW₅₃₈ sur gélose x-gal IPTG à 20 µg/mL chloramphénicol, (A) aspect du clone MHW₅₃₈ juste après le criblage par bacteriospray, (B) test d'activité du repiquage 1 par bacteriospray contre SRAM.

3.1.3.9.1. Analyse de la séquence du clone MHW₅₃₈

Le résultat BLASTn de la séquence de l'insert de 100 pb obtenue indique une identité de 95% (avec 100% de couverture) avec *Mycobacterium* sp NRRL B-3805 (numéro d'accèsion CP011022.1). Le résultat BLASTx de la séquence indique 100% de similarité avec la séquence de la sous-unité de la cobaltochélatase CobN de *Mycobacterium* sp URHD0025 (numéro d'accèsion WP_029108886.1) (figure 3.13) (résultat BLASTn et BLASTx en Annexe C).

MADSAVTNDAAAPVLLLLSTSDTLITARASGARYRWANPSRLVSGELEELLDGADIAVVRILGGYRSW
 QDGIDTVVASGLPTVVVSGEQSPDAELMGCSTAPQGAALQTHIYLAQGGVENLRNLHSFLSDTLLMTGF
 GFAAPVSTPTWGILERPAPTAGPTVAVLYYRAQELAGNTGYVEALCSAIEAAGGRALPVFCASLRTAEP
 ELLELLGSADTLITVLAAGGATPATVVGAGNDDSWNVAHLAALDIPILQGLCLTSRSDWAAGDDGMS
 PLDVATQVAVPEFDGRIITVPFSFKEIDSEGLISYVADPERCARVAGLAVRHARLRAIPAAEKRVLVFSAY
 PTKHARIGNAVGLDTPASAVALLRAMRDVGYDIGEAAAAPGDLATIVDSDGDALIHSLIERGGQDPDWLT
 DEALAANPIRVPAKDYRAWFATLPSELADAVVEHWGPPPGELFVDRSRDPDGEIVIAMQAGNVVLIVQ
 PPRGFGENPVAIYHDPDLPPSHHYLAAYRWLDSSFPDSFRADAVVHLGKHGNLEWLPGKTLGMSA**AC**
GTDAALGDLPLIYPFLVNDPGEQTQAKRRAHATLVDHLIPPMARAETYGDIKLEQLLDEHANVSALDP
 GKLPAIRQQIWTLMRAAKMDHDLGLEDRPDEDSFDDMLLHVDGWLCEIKDVQIRDGLHVLGQKPTGAG
 ELDLVLAAILRARQLFGGEQTVPLRQALGLIEDGSDERTAVDAAEAGARELVAALQESGWDASAVEKITD
 NPEVARILRFAATEVVPRLAGTAGEIDQVLRALAGGFIESGSPGSPRLRGLVNVLPTRNFYSVDPKAVPS
 RLAWETGVAMADSLLEYRTDYGRWPQSVGLSVWGTSAMRTAGDDIAEVLALLGVRPVWDDASRRV
 SLESIDLAEGRPRIDVTVRISGFFRDAFPHVVTMLDDAVALVAGLDEPAEDNYVRAHAQADLAEHGQD
 RRATTRIFGSKPGTYGAGLLQLIDSRNWRDDADLAEVYTAWGGFAYGRDLDRPASDDMNRAYRRIAV
 AAKNTDTREHDIADSDDYFQYHGGMVATVRALTGEAPAAYIGDNTRPDAVRTRTLSEETNRVFRARVV
 NPRWITAMRRHGYKGAPEMAATVDYLFYGYDATAHVMADWMYERLSAEYVLDENRKFMSSENPWAL
 HGMAERLLEAAGRGMWAAPEQATLDGLRQVLLETEGELEG

Figure 3.14. Représentation de la séquence de l'insert de 100 pb du clone MHW₅₃₈ à l'intérieur de la séquence d'acides aminés de la sous-unité cobaltochélatase. La partie en gras représentant le peptide identique à la séquence de l'insert.

Afin de considérer toutes les options de cadres de lecture possibles de la séquence nucléotidique du clone MHW₅₃₈, une recherche d'ORFs à l'aide d'ORF Finder a été réalisée en considérant tous sens de codons et tous types de codons de départ. De cette recherche, neuf ORFs ont été proposés. Une fois ces ORFs traduits en acides aminés, neuf peptides dont les tailles variaient de 12 à 33 acides aminés ont été prédits (tableau 3.2 ci-dessous). Ces séquences peptidiques sont de nouvelles séquences de peptides étant donné qu'aucune homologie de séquence n'a été identifiée à partir des bases de données BLASTp de NCBI, Pfam et BAGEL sauf pour la séquence de la figure 3.14.

Tableau 3.2. Séquence des neufs peptides dérivés de la séquence du clone MHW₅₃₈ en considérant les orientations et codons de départ possibles.

ORFs	Nombre d'acides aminés	Séquences
ORF1	33	H- MRPPLGLGALARVVDQERVVDQQRQIAEGSVGPAC-NH ₂
ORF2	32	H- MRDRRCPRRSAADLPVPGQRPGRRGHGQAAGA-NH ₂
ORF3	12	H- MSGRSPRAASVPH-NH ₂
ORF4	13	H- APAAWPGCPRPGR-NH ₂
ORF5	17	H- MGTGRSAADRRGQRRSRM-NH ₂
ORF6	32	H- ACGTDAALGDLPLIYPFLVNDPGEQTQAKRRRAH-NH ₂
ORF7	12	H- HAGPTLPSAICR-NH ₂
ORF8	18	H- MTRSWSTTRARAPRPSGGR-NH ₂
ORF9	18	H- CARRLAWVSPGSLTRNG-NH ₂

Aa : acides aminés à charges positives

Aa : acides aminés à charges négatives

3.1.3.9.2. Analyse des structures prédictives des séquences peptidiques possibles du clone MHW₅₃₈

La prédiction de la structure de ces peptides hypothétiques selon le serveur PEP-FOLD est présentée à la figure 3.15. La prédiction du logiciel est basée sur l'alphabet structural (*structure alphabet*) qui est composé de structures typiques formée par quatre acides aminés en tenant compte du profil énergétique des liens covalents, des ponts hydrogènes et de l'environnement immédiat. Les structures prédictives correspondent aux meilleures structures obtenues après 200 simulations. Les peptides issus des ORFs 1 et 5 sont deux peptides en hélice, avec 4 et 6 charges positives respectivement. Les charges du peptide issu de l'ORF 6 sont relativement équilibrées, avec 4 charges négatives pour 3 charges positives. Ce peptide comprend une partie structurée et une autre partie non structurée. Les

peptides issus des ORF 9, 3, 4, 7 et 8 sont de petits peptides de moins de 18 acides aminés et leur conformation varie de non-structurée à un ou deux tours d'hélice. Le peptide issu de l'ORF 2 est non structuré, mais il s'agit du plus chargé de tous avec 7 charges positives et 2 charges négatives.

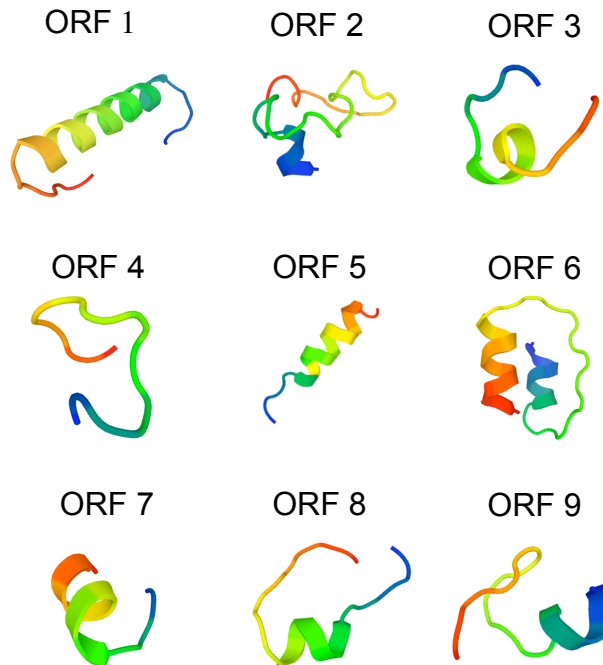


Figure 3.15. Analyse des structures prédictives des neuf séquences peptidiques possibles du clone MHW₅₃₈ par le serveur PEP-FOLD. Dégradé de couleur allant du N-terminal (bleu) au C-terminal (rouge).

3.1.3.10. Synthèse, clivage et purification des peptides

En raison des problèmes de croissance en milieu liquide rencontrés pour le clone MHW₅₃₈, il n'était pas envisageable de purifier le(s) peptide(s) naturellement produit(s) par ce clone. La synthèse chimique de ces séquences via un protocole de synthèse en phase solide a été choisie pour la suite du projet. Étant donné que de multiples peptides antimicrobiens possèdent une structure en hélice amphiphile,

les peptides correspondant aux ORF 1, 5 et 6 ont été considérés prioritairement pour la synthèse. Par ailleurs, il existe des peptides antimicrobiens pour lesquels leur principale caractéristique est leur nombre élevé de charges positives. Ainsi la séquence ORF 2 a été également considérée pour la synthèse. Enfin, des analogues de la séquence ORF 1 ont également été envisagés afin de pouvoir établir des liens entre la structure et l'activité, le peptide prédit par cet ORF étant le plus probable d'après la littérature.

Tableau 3.3. La séquence et masse exacte (incluant la valeur m/z notée par MALDI-TOF MS) des peptides synthétisés, dérivées de la séquence de l'insert du clone MHW₅₃₈.

ORFs	aa	Séquence	m/z	
			Calculé ^a	Observé
ORF1	33	H- MRPPLGLGALARVVDQERVDQRQIAEGSVGPAC -NH ₂	3490,02	3489,05
ORF1 _a	33	H- MRPPLGLGALARVVDQERVDQRQIAEGSVGPAS -NH ₂	3473,96	3472,99
ORF1 _b	33	H- NIeRPPLGLGALARVVDQERVDQRQIAEGSVGPAC -NH ₂	3472,93	3471,01
ORF1 _c	33	H- NIeRPPLGLGALARVVDQERVDQRQIAEGSVGPAS -NH ₂	3456,87	3454,95
ORF1 _d	32	H- MRPPLGLGALARVVDQERVDQRQIAEGSVGPA -NH ₂	3386,88	3385,90
ORF1 _e	33	H- RPPLGLGALARVVDQERVDQRQIAEGSVGPACK -NH ₂	3487,00	3486,02
ORF1 _f	34	H- MRPPLGLGALARVVDQERVDQRQIAEGSVGPACK -NH ₂	3618,19	3617,23
ORF1 _g	34	H- NIeRPPLGLGALARVVDQERVDQRQIAEGSVGPACK -NH ₂	3601,10	3599,19
ORF1 _h	31	H- RPPLGLGALARVVDQERVDQRQIAEGSVGPA -NH ₂	3255,68	3254,7
ORF1 _i	18	H-LGLGALARVVDQERVDQR -NH ₂	1995,27	1994,00
ORF1 _j	16	H- MRPPLGLGALARVVDQ -NH ₂	1693,04	1692,05
ORF1 _k	17	H-ERVDQRQIAEGSVGPAC -NH ₂	1814,99	1812,01
ORF2	31	H- RDRCPRRSAADLPVPGQRPGRGHPGQAAGA -NH ₂	3263,65	3263,00
ORF2 _a	31	H- RDRRSPRSAADLPVPGQRPGRGHPGQAAGA -NH ₂	3247,59	3245,92
ORF5	18	H- GTGRSAADRRGQRRSR -NH ₂	1786,93	1785,93
ORF6	33	H- ACGTDAALGDLPLIYPFLVNDPGEETQAKRRAH -NH ₂	3467,90	3466,92

^a Masse exacte de l'ion pseudomoléculaire monochargé ($[M + H]^+$)

Étant donné la taille considérable des peptides à synthétiser (jusqu'à 33 aa), les conditions réactionnelles ont été optimisées afin de s'assurer de l'obtention d'une

quantité de peptide adéquate pour la suite de l'étude. Pour le peptide ORF1, la synthèse a d'abord été effectuée en ne réalisant qu'une seule réaction par étape de couplage. Or, d'après l'analyse du chromatogramme obtenu par HPLC du mélange réactionnel final (figure 3.16a), il s'est avéré qu'aucun produit majoritaire n'avait été formé. De plus, l'analyse du mélange réactionnel par spectrométrie de masse MALDI-TOF n'a pas permis d'observer la masse moléculaire attendue pour le peptide en question, soit de 3490,02 Da.

Une autre tentative de synthèse a donc été réalisée, celle du peptide ORF1_a, un analogue du peptide ORF1 pour lequel la sérine est située en position C-terminale, à la place d'une cystéine. Cette fois, deux réactions subséquentes ont été réalisées pour chaque couplage. Une fois de plus, le mélange réactionnel brut final a été vérifié par HPLC et MALDI-TOF MS. Malgré le fait que l'analyse par spectrométrie de masse ait permis d'observer la présence de la masse recherchée (3472.99 Da), la présence d'autres pics fut notée par HPLC. Nous en avons donc conclu qu'un triple couplage serait nécessaire afin d'augmenter le rendement de synthèse du peptide désiré. La figure 3.16b démontre la formation d'un produit majoritaire correspondant au peptide désiré en effectuant chaque réaction de couplage à trois reprises.

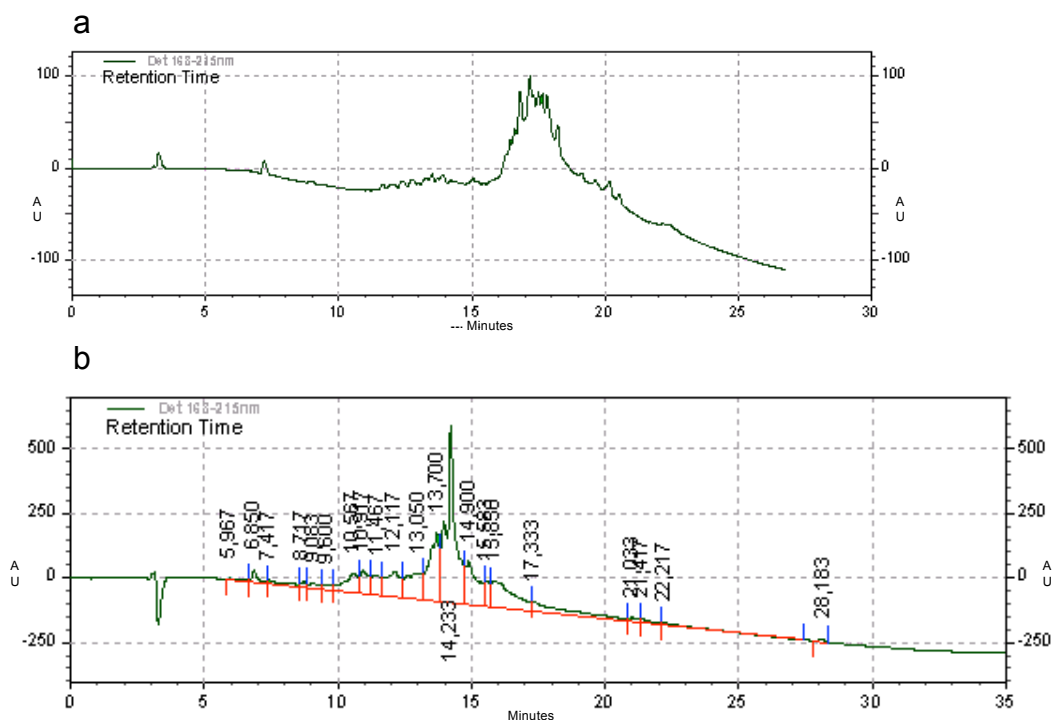


Figure 3.16. Profil HPLC analytique des peptides bruts ORF 1 (A) et ORF 1_a (B). Les analyses ont été effectuées en utilisant une colonne C₁₈ CSC-Kromasil 100 Å, 5 µm, 250 x 4,6 mm, en suivant un gradient linéaire de 0 à 60% acétonitrile/H₂O 0.1% TFA en 30 minutes.

3.1.3.10.1. Justification de la synthèse de certains peptides

Les peptides synthétisés dans le cadre de ce travail ont été préparés dans le but d'évaluer leur effet sur la croissance de la bactérie cible SARM en milieu liquide au moyen de l'appareil BioScreen. Les analogues du peptide ORF1 ont été sélectionnés avec un but précis. Comme mentionné précédemment, le peptide de l'ORF1_a est un analogue du peptide ORF1 possédant une sérine en position C-terminale à la place d'une cystéine, dans le but d'éliminer tout risque d'une dimérisation potentielle du peptide via la formation d'un pont disulfure (la sérine ne comprend pas de fonction thiol). Ce peptide a aussi été préparé dans le but de réduire la racémisation qui est souvent induite par une cystéine à cette position lors

de la synthèse peptidique en phase solide (Han *et al.*, 1997), comme démontré à la figure 3.17.

Le peptide ORF1_b a été synthétisé en remplaçant l'acide aminé méthionine situé en position N-terminale par la norleucine pour éliminer l'oxydation potentielle de la méthionine, tout en conservant la cystéine en position C-terminale. Encore une fois, cette substitution conserve la structure, quoique la chaîne latérale de la norleucine soit uniquement composée de carbones et d'hydrogènes sans atomes de soufre qui caractérise la méthionine. Le peptide ORF1_c a été synthétisé pour déterminer l'effet combiné des substitutions des peptides ORF1_a et ORF1_b. Quant au peptide ORF1_d, il a été conçu pour déterminer l'effet de l'omission de la cystéine de la séquence peptidique, alors que le peptide ORF1_e a été synthétisé dans le but d'incorporer un résidu supplémentaire dans la séquence afin que la cystéine ne se retrouve plus en position C-terminale. Une lysine, a été choisie pour être incorporée à la séquence peptidique afin d'obtenir un peptide comprenant une charge positive supplémentaire, pouvant potentiellement favoriser l'effet antibactérien. De plus, le peptide a été synthétisé sans méthionine en position N-terminale puisqu'il est possible que cet acide aminé soit excisé dans le cytoplasme de la bactérie suite à la traduction de la séquence nucléotidique (Frottin *et al.*, 2006). Le peptide ORF1_f correspond à la séquence du peptide de l'ORF1_e mais avec une méthionine en N-terminal. Le peptide l'ORF1_g correspond au peptide ORF1_f qui comprend une norleucine à la place de la méthionine. Et enfin, le peptide ORF1_h a été synthétisé pour déterminer l'effet de l'élimination de la méthionine et de la cystéine de la séquence sur son activité antimicrobienne. Dans le but de favoriser le développement de petits peptides, étant plus convoités (efficacité de synthèse, coût et temps de préparation, etc), certains peptides plus courts mais comprenant une certaine partie du peptide dérivé de l'ORF1 ont été préparés. Par exemple, le peptide dérivé de l'ORF1_i a été synthétisé de façon à éliminer les parties non structurées de part et d'autre du peptide ORF1 (figure 3.15), alors que les peptides ORF1_j et ORF1_k représentent la partie N-terminale et C-terminale du peptide ORF1, respectivement.

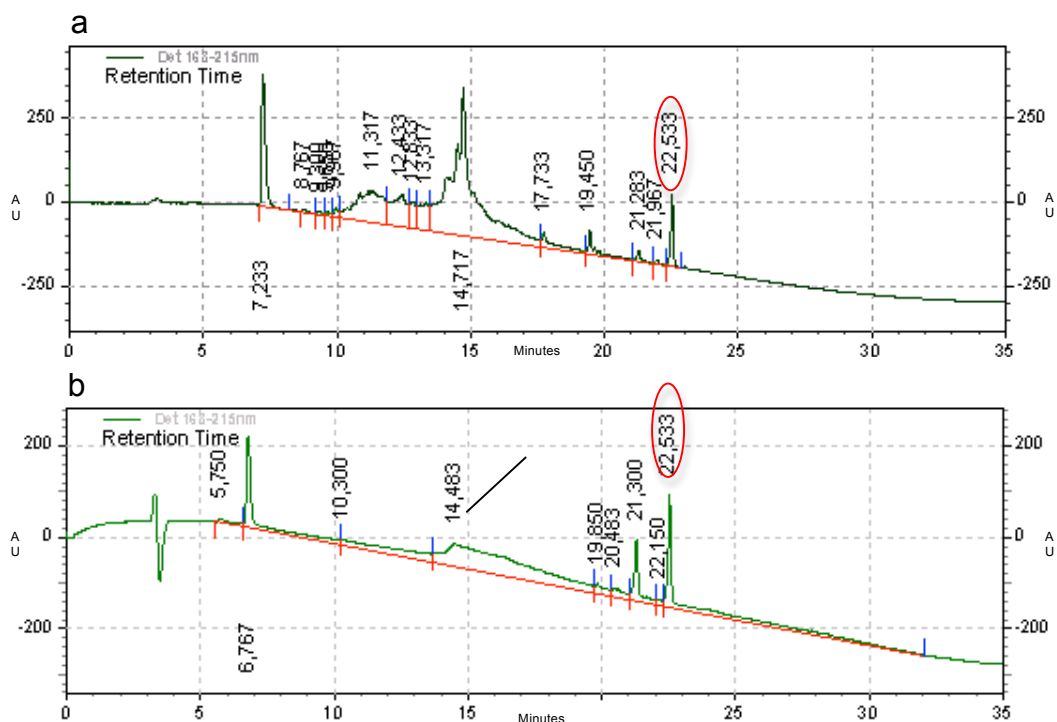


Figure 3.17. Chromatogrammes obtenus par analyse HPLC démontrant la dimérisation du peptide de l'ORF 1. (A) Chromatogramme avant dimérisation complète du peptide et (B) après dimérisation complète du peptide, démontrant la disparition du pic correspondant au peptide dérivé de l'ORF1 (temps de rétention : 14.71 min) et l'augmentation de l'intensité du pic associé au dimère (temps de rétention : 22.53 min).

3.1.3.11. Effet des peptides sur la croissance de SARM

3.1.3.11.1. Effet des dérivés du peptide issu de l'ORF1

Nous avons voulu montrer l'effet du peptide ORF1 sur la croissance de SARM en synthétisant neuf peptides, testés à une concentration de 10 mg/mL (soit 10,000 ppm) dans le milieu de culture MHB. La principale raison de cette forte concentration était de permettre la détection d'un effet quelconque contre cette bactérie cible (faible, modéré ou fort), considérant l'importance de cette bactérie et de sa forte résistance aux antibiotiques. D'après la figure 3.18, en considérant les peptides possédant l'acide aminé méthionine en position N-terminale (1, 1_a, 1_d, 1_f),

on peut remarquer que le peptide ORF1 n'était pas très actif comparativement au peptide l'ORF1_a pour lequel une sérine se trouve à la position de la cystéine. Cette activité augmente en omettant complètement la cystéine de la séquence (peptide ORF1_d). En conservant l'acide aminé cystéine en position C-terminale mais avec une norleucine à la position de la méthionine, soit le peptide ORF1_b, on constate un effet similaire au peptide ORF1. Le fait de remplacer l'acide aminé cystéine par la sérine chez ce peptide pour donner le peptide ORF1_c améliore légèrement l'activité antibactérienne. Cependant, aussitôt qu'une lysine supplémentaire est ajoutée du côté C-terminal (ORF1_e, ORF1_f, et ORF1_g), l'activité du peptide est davantage augmentée même en variant l'acide aminé méthionine. Enfin, le peptide ORF1_h qui ne possédait ni l'acide aminé méthionine au début de la séquence peptidique, ni l'acide aminé cystéine à la fin de la séquence peptidique, s'est avéré être l'analogue le plus actif de cette série de peptide.

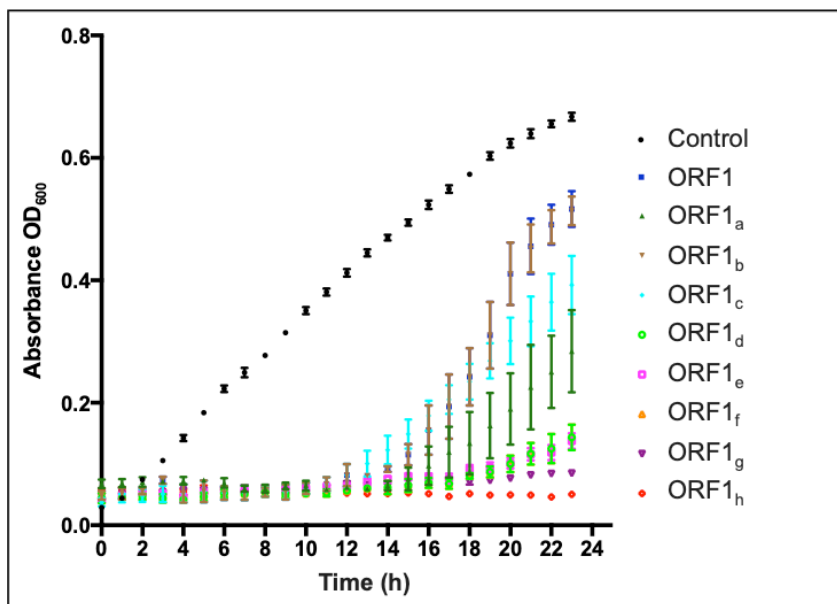


Figure 3.18. Effet du peptide ORF1 et de ses analogues à une concentration de 10 mg/mL sur la croissance de SARM. La croissance en milieu MHB est suivie au bioscreen pendant 23 h à 37°C en mesurant la DO à chaque heure. Les barres d'erreur correspondent à la déviation standard de triplicats.

L'effet de trois autres dérivés du peptide de l'ORF1 (ORF1_i, ORF1_j, ORF1_k) (tableau 4.4) ont aussi été testés contre SARM toujours à la concentration de 10 mg/mL. Le peptide dérivé de l'ORF1_i a été synthétisé de façon à éliminer les parties non-structurées de part et d'autre de ORF1 (figure 3.19), alors que ORF1_j et l'ORF1_k représentent la partie N-terminal et C-terminal du peptide ORF1, respectivement. D'après la figure 3.19, les peptides analogues ont un effet inhibiteur plus long sur la croissance de SARM (inhibition de la croissance pendant 23h) pour le peptide ORF1_h. Il semble donc que les éléments essentiels pour l'activité du peptide se retrouvent dans la portion N-terminale de la séquence.

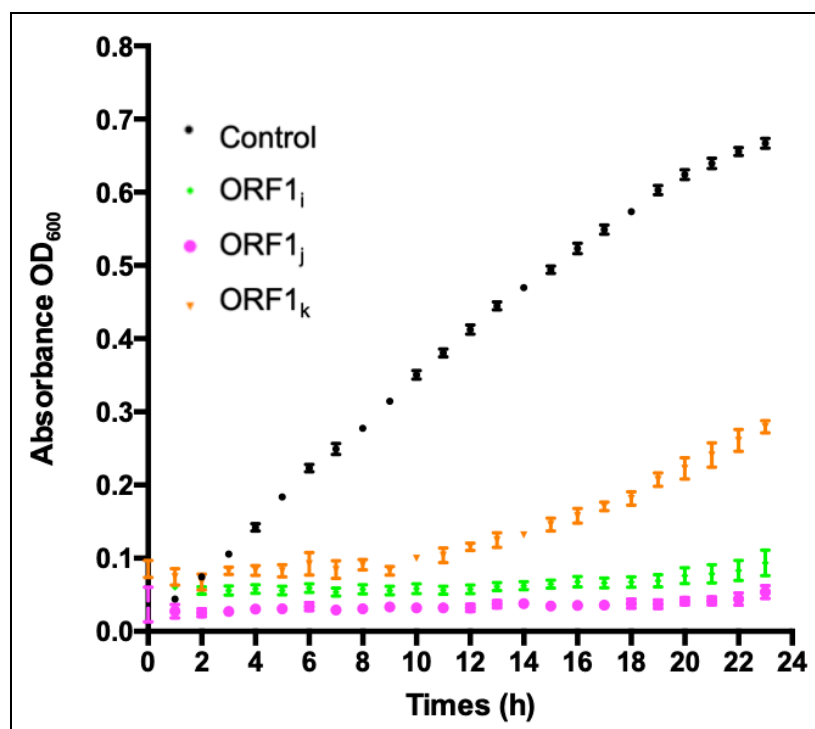


Figure 3.19. Effet des peptides dérivés de l'ORF1 à une concentration 10 mg/mL sur la croissance de SARM. La croissance en milieu MHB est suivie au BioScreen pendant 23 h à 37°C avec des mesures de DO chaque heure. Les barres d'erreur correspondent à la déviation standard de triplicats.

3.1.3.11.2. Effet des peptides issus des ORF2, 5 et 6

Les peptides de ORF2, ORF5 et ORF6 ont aussi été testés contre SARM à une concentration de 10 mg/mL. Ces peptides montrent un effet d'inhibition de croissance sur SARM même au bout de 23 h d'incubation (figure 3.20).

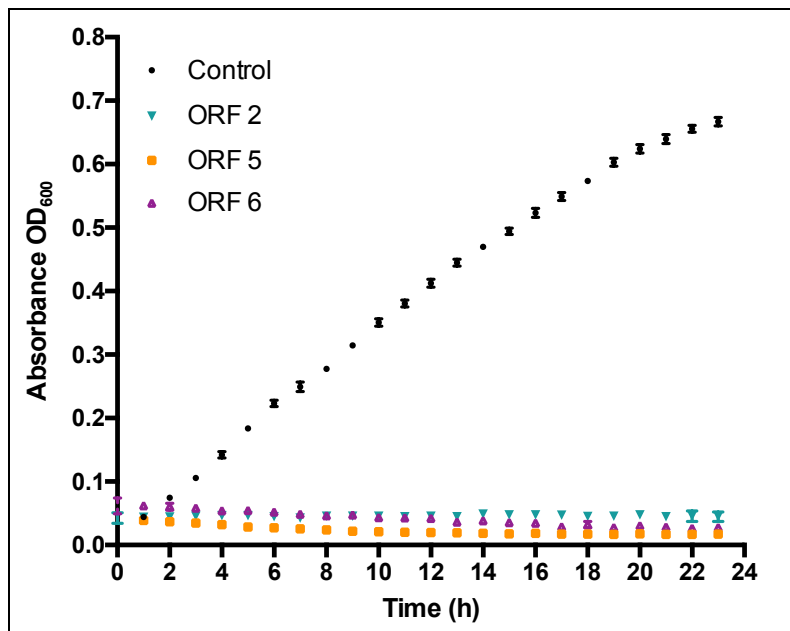


Figure 3.20. Effet des peptides dérivés des ORF2, 5 et 6 à une concentration de 10 mg/mL sur la croissance de SARM. La croissance en milieu MHB est suivie au bioscreen pendant 23 h à 37°C avec des prélèvements de DO à chaque heure. Les barres d'erreur correspondent à la déviation standard de triplicats.

3.1.3.12. Nouvelle synthèse de peptides

En prenant en considération les résultats précédents sur l'effet de certains peptides sur la croissance de SARM, les peptides ORF1_h, ORF1_i, ORF1_j, ORF1_k, ORF2, ORF5 et ORF6 ont été synthétisés à nouveau, dans le but d'être testés à plus faibles concentrations afin de déterminer leur concentration minimale inhibitrice (CMI) de la croissance de SARM. De plus, une bonne solubilité dans le milieu ainsi qu'une quantification plus précise sont essentielles à une juste détermination de la

CMI. Ainsi, l'absence de formation d'agrégats ou de solutions turbides lors de la solubilisation de ces peptides dans l'eau ou dans le milieu MHB a été vérifiée.

Afin d'ajuster la concentration en fonction du contenu peptidique de la masse pesée, une analyse élémentaire CHN (carbone, hydrogène et azote) de ces peptides fut effectuée.

3.1.3.12.1. Analyse de la composition en carbone, hydrogène et azote (CHN) des peptides

Les résultats des analyses élémentaires des peptides ORF1_h, ORF1_i, ORF1_j, ORF1_k, ORF2, ORF5 et ORF6 sont présentés dans le tableau 3.4. Le contenu en peptides de chaque échantillon fut déterminé par analyse élémentaire, en se basant sur le %N comme décrit dans la section 4.1.4.9, chapitre 4. Le contenu en peptide de ces échantillons se situe entre 63 et 86%. Il est à noter que comme chaque charge positive est associée à un contre-ion, plus un peptide est chargé positivement, moins son contenu en peptide est élevé.

Tableau 3.4. Contenu en carbone, hydrogène et azote des peptides ORF1_h, ORF1_i, ORF1_j, ORF1_k, ORF2, ORF5 et ORF6.

Peptide dérivé des ORFs	Formule chimique	Masse moléculaire	%C		%H		%N		% Peptide
			Theo	%C	Theo	%H	Theo	%N	
ORF1 _h	C ₁₃₉ H ₂₃₇ N ₄₇ O ₄₃	3254,70	51,3	44,62	7,34	6,58	20,23	17,00	84
ORF1 _i	C ₈₄ H ₁₄₈ N ₃₀ O ₂₆	1994,00	52,59	47,89	7,91	7,71	19,67	17,00	86
ORF1 _j	C ₇₄ H ₁₃₀ N ₂₄ O ₁₉ S	1692,05	52,53	47,6	7,74	7,11	19,87	16,00	80
ORF1 _k	C ₇₃ H ₁₂₄ N ₂₆ O ₂₆ S	1812,01	48,34	44,47	6,89	6,44	20,08	17,00	85
ORF2	C ₁₃₃ H ₂₂₅ N ₅₇ O ₃₈ S	3263,00	48,96	40,61	6,95	5,16	24,47	17,36	71
ORF2 _a	C ₁₃₃ H ₂₂₅ N ₅₇ O ₃₉	3245,92	49,20	40,29	6,99	5,21	24,59	17,18	70
ORF5	C ₆₇ H ₁₂₄ N ₃₆ O ₂₂	1785,95	45,06	35,41	7,00	5,30	28,23	18,00	63
ORF6	C ₁₅₂ H ₂₄₁ N ₄₅ O ₄₆ S	3466,92	52,66	44,92	7,01	6,28	18,18	14,00	77

3.1.3.12.2. Test d'activité des peptides ORF1_h, ORF1_i, ORF1_j, ORF1_k, ORF2, l'ORF5 et de l'ORF6 sur SARM et d'autres bactéries cibles

À une concentration de 5 mg/mL (en prenant compte l'analyse CHN de tous les peptides) les peptides ORF1_h, ORF1_i, ORF1_j, ORF1_k et ORF5 n'avaient plus aucun effet sur la croissance de SARM. Seuls les peptides ORF2 et ORF6 avaient encore un effet inhibiteur de la croissance de SARM. À 5 mg/mL (Figure 3.21), le peptide ORF6 avait un effet inhibiteur de croissance pendant 8 h avant que la croissance de SARM reprenne graduellement. Par contre le peptide ORF2 inhibait la croissance de SARM pendant 18 h avant que celle-ci reprenne graduellement.

Deux autres concentrations ont été testées : 2,5 mg/mL et 1,25 mg/mL. L'effet de ralentissement de la croissance de SARM est observé de façon plus importante avec le peptide ORF2 puisqu'un plateau de croissance est observé après 12h de croissance.

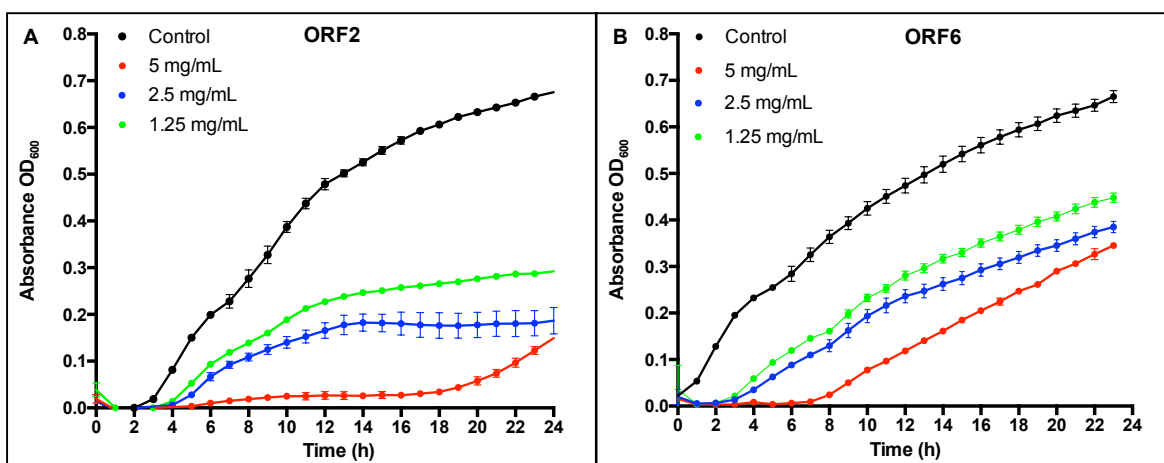


Figure 3.21. Effet des peptides ORF2 (A) et ORF6 (B) sur la croissance de *Staphylococcus aureus* résistante à la méthicilline. Les concentrations 5 mg/mL, 2.25 mg/mL et 1.25 mg/mL ont été testées. La croissance en milieu MHB est suivie au bioscreen pendant 23 h à 37°C avec des prélèvements de DO à chaque heure. Les barres d'erreur correspondent à la déviation standard de triplicats.

Étant donné que la sequence de l'insert indique une identité de 95% avec *Mycobacterium* sp., l'effet de ces deux peptides a aussi été testé contre

Mycobacterium smegmatis MC2 155 pour vérifier si ce peptide pourrait être une bactériocine produite par le genre *Mycobacterium*. Un effet inhibiteur de la croissance de *M. smegmatis* est observé (Figure 3.22) à 2,5 mg/mL, à 1,25 mg/mL et à 0.63 mg/mL après 8 h de croissance seulement avec le peptide ORF6. La croissance rejoint par la suite celle du témoin. Par contre, un effet inhibiteur de croissance est observé avec le peptide ORF2 à 2.5 et 1.25 mg/mL jusqu'à 10 h de croissance. L'effet de ce peptide est bactériostatique, mais un ralentissement de la croissance est quand même observé comparativement au témoin même à 0.63 mg/mL.

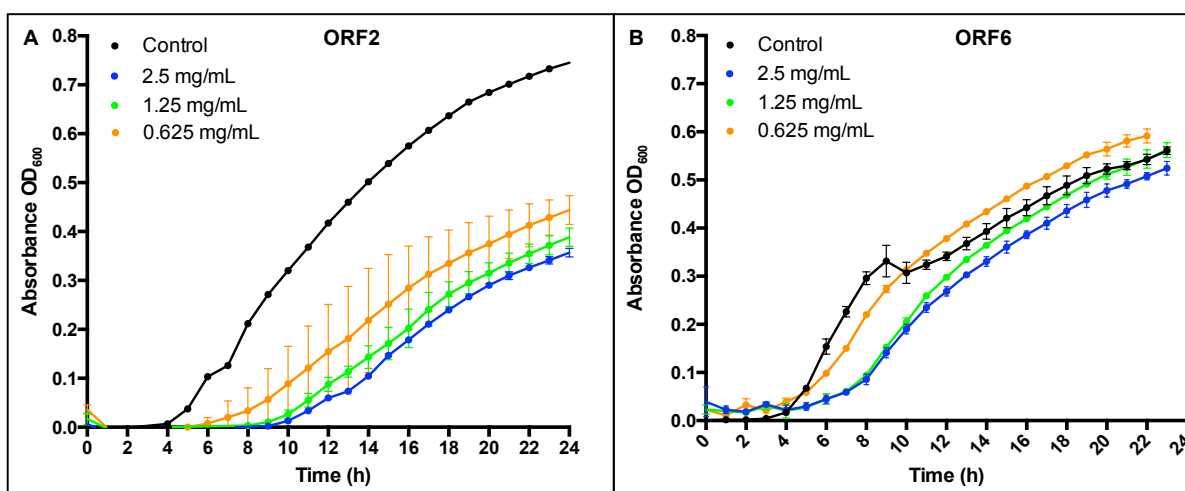


Figure 3.22. Effet des peptides ORF2 (A) et ORF6 (B) sur la croissance de *Mycobacterium smegmatis*. Les concentrations 2.5 mg/mL, 1.25 mg/mL et 0.63 mg/mL ont été testées. La croissance en milieu MHB est suivie au bioscreen pendant 24h à 37°C avec des prélèvements de DO chaque heure. Les prélèvements ont été réalisés en triplicat représentatifs des barres d'erreurs. Les barres d'erreur correspondent à la déviation standard de triplicat.

CHAPITRE 4 : ARTICLE 2

4.1. DISCOVERY OF AN ANTIMICROBIAL PEPTIDE FROM THE FUNCTIONAL METAGENOMIC SCREENING OF A HOSPITAL BIOFILM

Anissa Brahami, Myriam Létourneau, David Chatenet, Steven R. LaPlante, Eric Déziel* and Annie Castonguay*

INRS-Centre Armand-Frappier Santé Biotechnologie, 531 boul. des Prairies, Laval,
Québec, Canada.

Journal: Journal of Natural Products

Submitted: 15th August 2019

Acceptation:

Publication:

Citation:

Contribution:

Acknowledgments: the authors would like to thank Marie-Christine Groleau (INRS) for helpful discussions/technical assistance and Prof. Claude Guertin (INRS) for his help to collect the biofilm samples.

4.1.1. Résumé

La résistance des bactéries aux antibiotiques se répand plus vite que l'introduction de nouveaux antibiotiques en médecine humaine, provoquant une crise de santé publique. Nous sommes actuellement confrontés à une augmentation des infections causées par ces bactéries multirésistantes, ce qui accroît le besoin de découvrir de nouveaux antibiotiques. Étant donné que les microorganismes du sol produisent la plupart des antibiotiques connus et utilisés et que la grande majorité de ces microorganismes restent incultivables en laboratoire, la métagénomique fonctionnelle permet un accès à ce potentiel génétique qui représente une source prometteuse de nouveaux antibiotiques. Dans cette étude, l'ADN de biofilms de différents hôpitaux a été utilisé pour la construction et l'expression hétérologue de l'ADN chez *Escherichia coli* DH10B à l'aide du vecteur pBeloBAC11 suivant les méthodes optimisées. Le criblage fonctionnel par bacteriospray de 150,000 clones contre *Staphylococcus aureus* résistant à la méthicilline (SARM) a permis d'identifier 10 clones actifs. Certains clones avaient des morphologies particulières en étoiles ou formaient de petites colonies, mais un clone en particulier présentait un défaut de croissance en milieu liquide et en milieu solide. Ce clone avait un insert d'ADN d'une taille de 100 pb, 95% identique à une séquence nucléotidique de *Mycobacterium* sp. Une analyse des cadres de lecture ouverts avec le codon ATG comme codon de départ a révélé deux différents peptides qui ont été synthétisés sur support solide afin de tester leur activité. Une prédiction structurale, une analyse par dichroïsme circulaire ainsi qu'une analyse par résonance magnétique nucléaire de ces deux peptides ont montré que le peptide-1 (dérivé de ORF1) adopte une conformation en α -hélice, une structure commune aux peptides antibactériens, tandis que le peptide-2 (dérivé de ORF2) adopte une structure aléatoire. Un résultat inattendu obtenu était que le peptide-2 était faiblement actif contre SARM alors que le peptide-1 ne présentait aucune activité. Des études de structure-activité sont proposées en perspective pour améliorer l'activité biologique du peptide-2.

4.1.2. Abstract

We are now facing a rise in untreatable infections caused by multidrug-resistant bacteria, emphasizing the need to discover novel antibiotics. As soil microorganisms are known to constitute the most important source of antibiotics and considering that the vast majority of existing species remains so far uncultured, functional metagenomic screenings provide access to genetic material representing a very promising source of new bioactive molecules. In an effort to identify some of these compounds, eDNA was used to construct and express metagenomic BAC (Bacterial Artificial Chromosome) vector libraries in an *Escherichia coli* host. With the assumption that environments characterized by multidrug-resistant bacteria competition are more likely to lead to the identification of promising antimicrobials by prompting bacteria to produce metabolites to allow their survival, biofilms were sampled from hospital wastewater drainage pipes. Using the sensitive bacteriospray method we recently developed, the functional screening of 150,000 colonies allowed the identification of 10 clones exhibiting an inhibitory activity against a methicillin-resistant *Staphylococcus aureus* (MRSA) strain. Interestingly, one of the active *E. coli* clones presented a very slow growth phenotype. Analysis of its eDNA insert revealed a short 100 bp sequence identical to a conserved portion of a *Mycobacterium* sp. gene. A blast ORF finder analysis predicted two different peptides from this metagenomic DNA insert. Structure modeling, circular dichroism and nuclear magnetic resonance analyses indicate that peptide-1 (derived from ORF1) likely adopts an α -helical conformation, a common structure among antibacterial peptides, whereas peptide-2 (derived from ORF2) is most likely unstructured. Unexpectedly, peptide-2 could inhibit the growth of MRSA whereas no activity was observed for peptide-1. These results suggest that functional metagenomic screenings performed by exploiting eDNA isolated from an environment with high antimicrobial selective pressure may reveal new metabolites active against antibiotic-resistant bacteria.

4.1.3. Introduction

Antibiotic resistance is becoming a serious human threat (World Health Organization, 2017). Even in the early days of antibiotics development, resistant strains have emerged spontaneously, but with the abusive use of antibiotics nowadays, an increasing number of bacteria acquire such resistance (Demain & Sanchez, 2009). Multidrug-resistant bacteria such as the 'ESKAPE' pathogens (Rice, 2008) are currently responsible for the majority of US hospital infections, as they literally 'escape' the effects of conventional and clinically approved antibacterial drugs (Santajit & Indrawattana, 2016). For instance, methicillin-resistant *Staphylococcus aureus* (MRSA) is the leading cause of morbidity and the mortality associated with hospital infections (DeLeo *et al.*, 2010; Howden *et al.*, 2011). In fact, infections are expected to surpass cancer as one of the most important cause of death in the coming decades (KPMG LLP, 2014). Despite the urgent need for more new antibiotics effective against multidrug-resistant bacteria (World Health Organization, 2017), only few compounds were recently developed (Bierbaum & Sahl, 2014; Cooper & Shlaes, 2011; The Pew Charitable Trusts, 2016).

In recent years, the functional screening of metagenomic libraries made from soil DNA has proven to be a useful approach for the discovery of novel antibiotics (Biver *et al.*, 2013; Craig *et al.*, 2010; Iqbal *et al.*, 2014). Numerous new antimicrobial metabolites and interesting bioactive molecules have been discovered by exploiting the DNA of uncultured bacteria (Cohen *et al.*, 2018; Iqbal *et al.*, 2016; Thies *et al.*, 2016). The choice of the vector used in these studies depends on the quality and length of the DNA after extraction, the targeted genes and the screening strategies (Daniel, 2005). Small-insert libraries are more readily constructed and expressed but can be used mostly for the identification of phenotypes encoded by a single gene or a small operon (Simon & Daniel, 2017), whereas large-insert libraries are required to recover large gene clusters coding most bioactive compounds belonging to polyketide synthase (PKS) and nonribosomal peptide synthase (NRPS) groups

(Huo *et al.*, 2019). Indeed, phenotypic screening approaches are very promising as they are activity-based, enabling the discovery of unknown molecules belonging to completely novel families (Brady *et al.*, 2002; Brady *et al.*, 2004; Craig *et al.*, 2009; Iqbal *et al.*, 2014).

Interestingly, the source of environmental DNA (eDNA) sampling in any metagenomic studies largely influences the properties of the biomolecules to be found (Simon *et al.*, 2009; Thies *et al.*, 2016; Vercammen *et al.*, 2013), likely due to the physiological adaptation of microorganisms to their environment (Brooks *et al.*, 2011). Indeed, based on the ability of bacteria to adapt to extreme conditions (Brooks *et al.*, 2011), it has been hypothesized that environments with a strong antibiotic selection pressure could be a favourable starting point for functional metagenomic studies and therefore the potential identification of new antibiotics (Katouli *et al.*, 2012; Lood *et al.*, 2017). Moreover, hostile environments favour competition between drug-resistant bacteria with the need to develop new ways to defend themselves against each other (Cohen *et al.*, 2018). Among such environments, wastewater materials were previously proven interesting for the discovery of new classes of enzymes (Vercammen *et al.*, 2013) and antibiotic resistance genes (Uyaguari *et al.*, 2011), but were rarely exploited for the discovery of novel antibacterial molecules. For instance, a previous study using biofilms from a slaughterhouse drain and a metagenomic approach led to the discovery of a novel biosurfactant with antibiotic activity against *Bacillus subtilis* (Thies *et al.*, 2016).

In this study, eDNA sampling was performed from hospital biofilms, characterized by a strong antibiotic selection pressure, multidrug-resistant bacteria competition, human feces and microbiota, all associated to a high evolution pressure, encouraging bacteria to produce molecules allowing their survival. Thus, metagenomic libraries were constructed using eDNA from a sanitary sewer near a Laval hospital and a Montreal hospital wastewater pipe. The resulting 150,000 clones were screened and revealed an interesting unstructured antimicrobial peptide that has, to the best of our knowledge, not been so far reported for its activity against methicillin-resistant *Staphylococcus aureus* (MRSA).

4.1.4. Materials and methods

4.1.4.1. Biofilm sampling

Biofilm samples were collected from two distinct sources: biofilm LHW originated from a sanitary sewer located just outside a hospital (where hospital wastewater gets discharged in the municipal sewage system), whereas biofilm MHW was directly collected from the main wastewater drainage pipe of another hospital. Samples were collected in sterile plastic bags and kept on ice while being transferred to the laboratory (on the same day) for DNA extraction and storage at -20°C.

4.1.4.2. DNA extraction

DNA was extracted from biofilms LHW and MHW. Conditions were first optimized to facilitate eDNA release from the biofilm material, rich in extracellular polymeric substances (EPSs) (Fleming *et al.*, 2017). An aqueous enzyme cocktail solution composed of porcine pancreas α -amylase (Sigma A-6880) and fungal cellulase (Sigma C-7502) (0.25 % w/v) was added to 50 mg of the biofilms, respectively, and incubated for 30 min at 37°C. The FastDNA spin kit (MP Biomedicals) was used to extract total genomic DNA. DNA shearing was compared between the bead-beating (FastPrep-24 MPbio) and vortexing (Vortex-Genie 2 - Scientific Industries, Inc) methods for 40 seconds (twice) to recover high-molecular-weight biofilm DNA for library construction. The quality and the size of the extracted DNA were visualized by electrophoresis on 0.8 % agarose gels.

4.1.4.3. Bacterial strains and plasmids

Electrocompetent *Escherichia coli* DH10B was purchased from New England Biolabs (NEB) and was used for library construction and expression. The strain was

grown in Lysogeny Broth (LB) medium at 37°C. The bacterial artificial chromosome pBeloBAC11 (7.5 kb) (Kim *et al.*, 1996), obtained from NEB was maintained in *Escherichia coli* K12 ER2420 with 20 µg/mL chloramphenicol. DH10B transformants were grown in LB medium containing chloramphenicol (40 µg/mL). A clinical MRSA isolate was used as the target bacterium for antimicrobial activity testing.

4.1.4.4. Library construction

DNA extracted from the microbial communities of biofilms LHW and MHW were used to construct metagenomic libraries, following a protocol we recently reported (Brahami *et al.*, 2019). Briefly, to promote DNA insert diversification, purified genomic DNA was partially digested with the *Bam*HI, *Hind*III or *Sph*I restriction enzymes (New England Biolabs, MA, USA), three unique cloning sites present in the *lacZ*α gene of the pBeloBAC11 vector. Resulting digested genomic DNA fragments were ligated overnight to the linearized pBeloBAC11 vector at 16°C using NEB T4 DNA ligase and transferred into *E. coli* DH10B by electroporation.

4.1.4.5. Library screening for antimicrobial activity and phenotype confirmation

The bacteriospray antimicrobial assay was used, as previously described (Brahami *et al.*, 2019), to screen *E. coli* libraries for antibacterial activity against a methicillin-resistant *Staphylococcus aureus* (MRSA) isolate. Briefly, *E. coli* transformants were grown overnight at 37°C in 20 x 20 mm² LB plates containing 20 µg/mL of chloramphenicol and 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for five days at room temperature. A thin film of the MRSA suspension was then sprayed on the surface of the plates with a paintbrush tool, followed by incubation at 30°C for 24 h to 48 h before visual examination for active clones. Colonies surrounded by a zone of MRSA growth inhibition were picked and restreaked for phenotype confirmation. Plasmids from the isolated active clones

were extracted with the Qiagen maxiprep kit. For confirmation, extracted plasmids were retransformed in *E. coli* DH10B followed by the bacteriospray antimicrobial assay, as described above. Plasmids that reproduced the initially observed phenotype were then selected for further analyses.

4.1.4.6. DNA insert amplification and sequencing

Amplification of inserted eDNA from extracted vectors was achieved through PCR using the following universal primers: M13F (5'-GTTTTCCAGTCACG-3') and M13R (5'-CAGGAAACAGCTATG-3'). The PCR was performed in a 60 μ L reaction volume containing 2.5 μ L each of the forward and reverse primers (10 pmol/ μ L), 1 μ L of DNA template (0.5 ng), 1 μ L of 10 μ M dNTPs, 10 μ L of 5X Q5 buffer, 10 μ L of 5X Q5 high GC enhancer and 0.5 μ L of Q5 Taq polymerase (New England Biolabs, MA, USA). Reaction conditions consisted of an initial denaturation by heating the samples at 98°C for 30 s, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 57°C for 30 s, extending at 72°C for 1 min, and final extending at 72°C for 5 min. Amplified products were confirmed by 1% agarose gel electrophoresis. Amplicons were dialyzed using a Millipore mixed cellulose esters membrane filter (0.025 μ M porosity, VSWP01300) in sterile water, before DNA sequencing (Institut de recherches cliniques de Montréal sequencing service).

4.1.4.7. Sequence analysis and structure prediction

Sequence analysis was performed using the CLC sequence viewer (Qiagen, version 7). Since *E. coli* uses ATG as a start codon to initiate approximately 92% of the translations (Villegas & Kropinski, 2008a), inserted sequences were analyzed with ORF (open reading frame) finder from NCBI using standard conditions and considering only ATG as a start codon. The resulting ORF sequences were submitted to the Basic Local Alignment Search Tools (BLASTn) from the NCBI for nucleotide sequence similarity analysis, and then translated into amino acids and peptide sequence similarities using BLASTp. To correlate ORF

sequences with the observed antimicrobial activity, predicted peptides were synthesized for further biological characterization but also structurally analyzed with the PEP_FOLD 3.5 software from *Ressource Parisienne Bioinformatique Structurale* (RPBS) web server.

4.1.4.8. Antimicrobial activity assay with cell culture supernatants

Clone MHW38, selected from the metagenomic library constructed from biofilm MHW, was grown in LB medium supplemented with 20 µg/mL chloramphenicol and incubated at 37°C for 3 hours before 1 mM IPTG was added (when required) for overnight incubation (Grossman *et al.*, 1998). Supernatants were then collected after a 10 min centrifugation at 10,000 x *g* and passed through a 0.22 µm filter. The antibacterial activity was evaluated using the well diffusion method (Holder & Boyce, 1994) by adding 100 µL of the filtrate, sometimes concentrated up to 10 times, in holes bored in a fresh MRSA lawn seeded over Mueller-Hinton (MH, BD Difco) agar plate and concentrated with amicon ultra 10 KDa (sigma aldrich) as recommended by the manufacturer. Plates were incubated overnight at 30°C before the antimicrobial activity was assessed.

4.1.4.9. Peptide synthesis, purification and characterization

The two predicted peptides (see Results and Discussion) were prepared following a standard Fmoc-based solid phase peptide synthesis procedure using a Liberty Blue™ peptide synthesizer. The Fmoc-Rink-amid AM resin (with Nle) used in this study was purchased from Chem-Impex (IL, USA) whereas all other chemical reagents and solvents were obtained from Sigma-Aldrich. Briefly, coupling reactions were performed at room temperature for 0.5-1 h using an excess (3 equivalents) protected amino acid (based on the original substitution of the resin, 0.5 meq/g), *N,N'*-diisopropylcarbodiimide (DIC, 3 equivalents) and Oxyma (1 equivalent) in *N,N*-dimethylformamide (DMF). All amino acids were coupled thrice. Fmoc removal was

achieved in 15 min using piperidine (20% in DMF). After coupling and deprotection of the last amino acid, peptide-linked resins were washed with DMF, dichloromethane and diethyl ether, and dried under vacuum for 15 minutes. Peptides were cleaved from the resins using a trifluoroacetic acid (TFA) solution containing triisopropylsilane (2.5%), ethanedithiol (2.5%) and H₂O (1%) mixture (20 mL/g) for 2 h at room temperature. After cleavage, resins were removed by filtration and washed twice with the previously described TFA-based cleavage cocktail. Filtrates were concentrated under reduced pressure and unprotected peptides were precipitated with diethyl ether. Crude peptides were then recovered by filtration, solubilized in an acetonitrile:H₂O (1:1) solution containing TFA (0.1%), lyophilized and stored at -20°C. Both crude peptides were purified using a preparative reversed-phase HPLC (RP-HPLC) protocol using a 0-100% linear gradient from eluent A (0.1% TFA in H₂O,) to eluent B (0.1% TFA in CH₃CN,) and a C18 column (Kinetex EVO, 5 µm, 250 x 10 mm) at a flow rate of 10 mL/min (40 % B/min), by monitoring the absorbance at 229 nm. The homogeneity of the purified fractions was assessed by analytical RP-HPLC using a 0-60% linear gradient from the same eluents and a C18 column (CSC-Kromasil 100 Å, 5 µm, 250 x 4,6 mm) at a flow rate of 1 mL/min (2% B/min), by monitoring the absorbance at 229 nm. Fractions containing the desired product, as assessed by MALDI-TOF mass spectrometry (Microflex LRF Bruker system) using α-cyano-4-hydroxycinnamic acid (matrix), were pooled and lyophilized. A RP-HPLC analysis revealed the purity of both peptides to be >98%. For each peptide, *m/z* values observed by MALDI-TOF mass spectrometry were in agreement with calculated values: **1**, *m/z*: found ([M+H]⁺) 3254.9 (calc. 3255.7); **2**, *m/z* found ([M+H]⁺) 3263.3 (calc. 3263.6). The peptide content of each sample (**1**, 84%; **2**, 70%) was determined by elemental analysis (CHN), based on the %N (Centre de caractérisation des matériaux, Université de Sherbrooke): **1**, C₁₃₉H₂₃₇O₄₃N₄₇ requires C, 51.30; H, 7.34; N, 20.23. Found (%): C, 44.62; H, 6.58; N, 17.00; **2**, C₁₃₃H₂₂₅O₃₈N₅₇S₁ requires C, 48.96; H, 6.95; N, 24.47. Found (%): C, 40.61; H, 5.16; N, 17.36.

4.1.4.10. Circular dichroism (CD) analysis

Spectra were recorded at room temperature from 200 to 250 nm (0.1 nm steps), using a 1 mm optical path length cuvette with 1 nm step, scanning rate 50 nm/min, 1 nm bandwidth, and an integration time of 4 s on a Jasco J-815 circular dichroism (CD) spectropolarimeter (Easton, MD, USA). Each spectrum represents the mean of three scans corrected for solvent contribution. A digital low-pass filter was used as a smoothing routine. Peptides were dissolved in either water or a trifluoroethanol (TFE): water (50:50) mixture at a final concentration of 200 µg/mL.

4.1.4.11. Nuclear magnetic resonance (NMR) analysis

NMR samples for peptide-1 and peptide-2 were prepared identically. 5 mg of peptide were dissolved in 200 µL of a 1:1 water (10% D₂O) and deuterated trifluoroethanol (TFE-d₃) mixture. Samples were prepared in 3 mm NMR tubes. Experiments were run on a 600 MHz Bruker AV III NMR spectrometer equipped with a helium cryoprobe. The 1D ¹H and 2D ¹H-¹H NOESY NMR spectra involving peptide-1 and peptide-2 were acquired identically. Standard pulse programs were used and included excitation sculpting pulses for water suppression purposes. The ¹H 1D experiments employed 32 scans, and the NOESY experiments were acquired with 32 scans per t₁ point. Both types of experiments employed a relaxation delay of 4 seconds. Data were processed and visualized using Bruker's TOPSPIN software.

4.1.4.12. Antimicrobial activity assay

Assays were performed according to the microdilution method developed by the Clinical Laboratory Standards Institute's committee (Clinical Laboratory Standards Institute, 2012). Briefly, the MRSA used in this study was inoculated and allowed to reach an OD₆₀₀ of 0.08–0.13 prior to dilution to 10⁶ CFU/mL in MHB. Fifty µL of the

bacterial suspension were then added to each well of a 96-well microplate containing serially diluted peptide solutions (50 μL in MHB) and incubated at 37°C under static conditions for 8 h or 24 h. A well containing sterilized medium was used as a sterility control while a well without added peptide served as a growth control. The minimum inhibitory concentration (MIC, expressed in $\mu\text{g}/\text{mL}$) corresponds to the lowest concentration of peptide at which no bacterial growth was observed after the incubation time, as determined by OD_{600} measurements using a Cytation 3 multimode plate reader.

4.1.5. Results and discussion

4.1.5.1. DNA extraction from hospital biofilms

Microbial communities from two hospital biofilms (LHW and MHW) were selected as sources of eDNA for the construction of metagenomic libraries. DNA extraction optimization was first performed using the thicker MHW biofilm. Accessing the microbial DNA from complex environmental biofilms represents a difficult task since bacterial cells are encased in a EPS matrix comprising of various polysaccharides, proteins, and amphiphilic polymers (Jachlewski *et al.*, 2015). Indeed, this physical and chemical association of the bacteria with these EPS complexes can prevent the release of the genomic DNA (Wu & Xi, 2009). Consequently, all DNA extractions were performed following amylase and cellulase pretreatment to degrade the EPS and release the bacteria. Furthermore, in functional metagenomic studies, large DNA fragments appear preferable for generating large-insert metagenomic libraries to identify genes encoding complex biosynthetic pathways, thus increasing the likelihood of uncovering novel, promising molecules (Cohen *et al.*, 2018; Kang & Brady, 2014). Since the physical bead-beating method induces DNA fragmentation even under optimized conditions (Bürgmann *et al.*, 2001), a more gentle vortexing procedure was also attempted for DNA extraction. The average DNA yield obtained

with the bead-beating and the vortexing procedures of three metagenomic DNA extraction replicates (μg per g of biofilm) was $47.6 \pm 10.7 \mu\text{g/g}$ and $37.9 \pm 1.2 \mu\text{g/g}$, respectively. Using both techniques, the purity of the extracted DNA, assessed by determining the A260/A280 ratio, was comprised between 1.8 and 2. As shown in Figure 1, DNA fragments between 4–48 kb were obtained with the bead-beating procedure (Figure 4.1, lane 2) whereas DNA fragments were between 10–30 kb (Figure 4.1, lane 4) were obtained with the vortexing procedure. Following *SphI* digestion, DNA fragments between 0.5 – 15 kb (Figure 4.1, lane 3), were obtained from total DNA extracted using the bead-beating procedure, whereas DNA fragments between 4–15 kb (Figure 4.1, lane 5) were obtained from total DNA extracted using the vortexing procedure. Since the bead-beating procedure caused considerable DNA shearing, the vortexing procedure was selected for all subsequent DNA extractions.

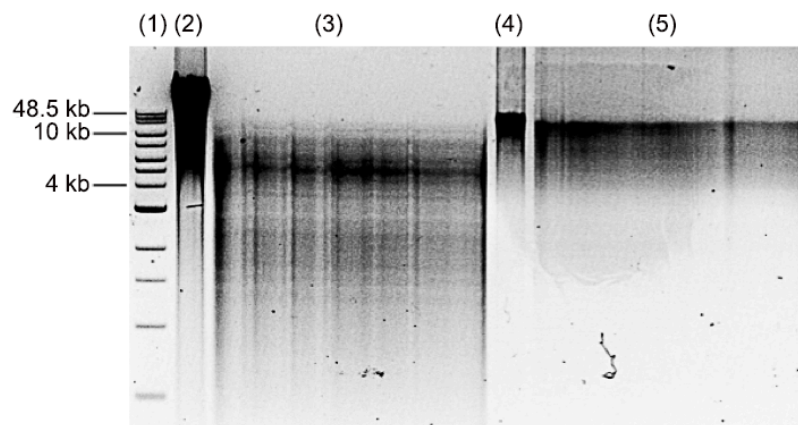


Figure 4.1. DNA extracted from a MHW hospital wastewater drainage pipe biofilm sample. Samples were visualized on a 0.8% agarose gel. Biofilm DNA extraction was performed either through bead-beating (lanes 2 and 3) or vortexing (lanes 4 and 5). Extracted DNA was then respectively submitted to a partial *SphI* digestion (lanes 3 and 5; samples loaded into two fused wells). New England Biolab 1 kb extended DNA ladder (lane 1) was used to determine fragment size.

4.1.5.2. Metagenomic library construction and screening for antimicrobials molecules

DNA samples extracted from the microbial communities of biofilms LHW and MHW were used to construct two distinct metagenomic libraries following a protocol we recently reported (Brahami *et al.*, 2019). The digestion of biofilm DNA with restriction enzymes *BamHI*, *HindIII* or *SphI* produced DNA fragments of a similar size range (data not shown). For this study, the pBeloBAC11 vector was selected increasing the probability of detecting clones expressing antibacterial activity (Gillespie *et al.*, 2002; Rondon *et al.*, 2000a). Moreover, this vector can accommodate large DNA inserts, allowing a reduced number of clones to be tested while increasing the probability of identifying new bioactive molecules from large biosynthetic gene clusters (Sosio *et al.*, 2000).

E. coli DH10B was used as the heterologous host for the expression of the metagenomic libraries construction. *E. coli* represents the most commonly used host, and has previously been exploited for the discovery of bioactive molecules (Brady, 2007; Gillespie *et al.*, 2002; Thies *et al.*, 2016). The transformation efficiency of these newly constructed metagenomic libraries constructed was found to be at least $2.5 \times 10^5 \pm 0.6$ CFU/ μ g, which is comparable to previous reports (Iqbal *et al.*, 2014). From these two libraries, 150,000 transformants were screened for antibacterial activity against a MRSA strain, using the bacteriospray method. Beyond its clinical significance and thus a desirable target for the identification of new antimicrobials, we picked a Gram-positive bacterium as the target of the screening to increase the likelihood of finding active clones, assuming that *E. coli* clones producing antimicrobials that are active against Gram-negative would be less readily isolated.

Ten metagenomic clones, all originating from DNA extracted from the MHW hospital wastewater drainage pipe biofilm, were found to cause a small zone of growth inhibition on the lawn of the MRSA clinical isolate used in this study. After phenotypic confirmation, the DH10B clone named MHW38 presented a very slow growth in LB medium. This clone was selected for further analyses, assuming that

its slow growth could be the result of a produced molecule. Indeed, overnight cultures reached an OD_{600} of 0.33 ± 0.013 as compared to an OD_{600} of 4.67 ± 0.01 for DH10B transformed with the empty pBeloBAC11 vector. Transformation of the pBeloBAC MHW38 vector into a new electrocompetent DH10B confirmed the phenotypes (MRSA inhibition and slow growth of the clone) (Figure 2).

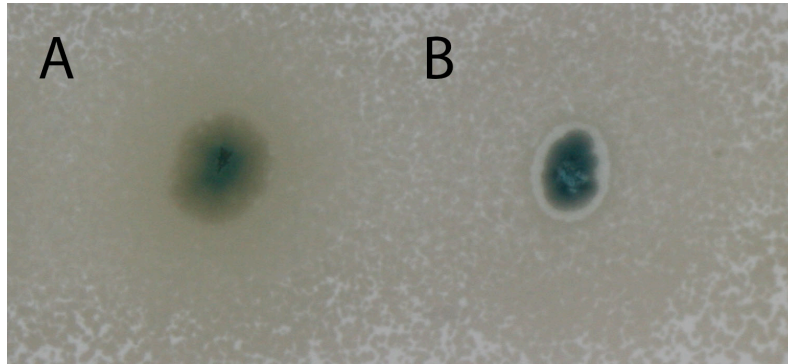


Figure 4.2. Picture of control and inhibitory *E. coli* DH10B clones against MRSA lawn. (A) a spot of 5 μ L *E. coli* DH10 B transformed with the empty pBeloBAC11 vector; (B) a spot of 5 μ L MHW₅₃₈ clone.

4.1.5.2. Selected clone MHW₅₃₈

Attempts to recover the activity from filtered culture supernatants by heterologous expression of the inserted DNA into *E. coli* DH10B failed, suggesting that the molecule responsible for the observed activity might be unstable, too small to be concentrated with a centrifugal device of 10 kDa molecular weight cut-off (since big molecule was expected), or even not secreted (Delmont *et al.*, 2011b)ted at all but rather cell-associated (MacNeil *et al.*, 2001a).

Although a large-molecular-weight DNA insert was expected, the one found to be involved in the antimicrobial activity of the MHW38 clone was less than 200 pb, as shown in the PCR amplification product visualisation (1% agarose gel) presented in Figure 4.3. Indeed the environmental DNA (eDNA) size of the metagenomic DNA

library is dependent on the quality of the extracted DNA (Delmont *et al.*, 2011b; Williamson *et al.*, 2011). In our study, the average gel visible size of DNA after digestion was between 4 and 15 kb (figure 1 lane 5). When ligated (without DNA gel extraction) the small DNA pieces preferentially ligated than larger ones, which is relatively common (Stevenson *et al.*, 2013). Sequence analysis of the PCR product revealed that the eDNA insert is actually only 100 bp long (Figure 4.4), and is identical to a conserved portion of *Mycobacterium* sp coding region (accession number CP011022.1). A conserved domain search on NCBI indicated that the inserted DNA is part of gene sequence predicted to contain the cobaltochelatase subunit (CobN) of *Mycobacterium* sp (accession number WP_019512951.1). This family of proteins catalyzes the insertion of cobalt into the corrin ring of hydrogenobyrinic acid a,c-diamide. This aerobic branch of corrin ring synthesis is part of the adenosylcobalamin biosynthetic pathway (Debusche *et al.*, 1992). Due to the size of the identified DNA insert, the expression of the product likely requires both the promoter and the RBS of the pBeloBAC11 vector (Gabor *et al.*, 2004). Moreover, in order to produce a zone of growth inhibition, the active molecule had to be released from the bacteria. This clone could be identified thanks to the use of a low copy number vector, which likely prevented excessive toxicity of the insert-derived product to the host, and to the use of the bacteriospray screening method, which allowed the detection of the smallest zone of growth inhibition. The observed zone of growth inhibition could be the result of the release of the antimicrobial following *E. coli* lysis.

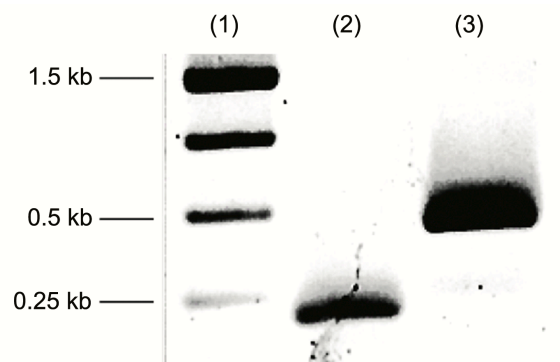


Figure 4.3. Visualisation of PCR amplification products on the eDNA fragment inserted into clone HMW₅₃₈. Visualization on a 1% agarose gel. Lane 1: 1 kb

DNA Ladder, lane 2: PCR amplification product of the cloning site of the empty pBeloBAC11 vector, lane 3: PCR amplification product of clone MHW 3. pBeloBAC11 from M13 forward and reverse size: 158 bp empty vs 168 bp with the inserted DNA.

A BLAST ORF finder analysis of the inserted DNA, restrained to ATG as a start codon (Villegas & Kropinski, 2008a), revealed two ORFs giving rise to two theoretical peptides of 33 (peptide-1 from ORF1) and 32 amino acids (peptide-2 from ORF2), respectively (Figure 4). Since the ORF 2 sequence is in the same direction of the *T7* promoter, it could be the one expressed and produced. Using the Antimicrobial Peptide Database (<http://aps.unmc.edu/AP/main.php>), the peptide-1 sequence alignment revealed some similarities (around 30%) with known peptides: a helical peptide from alligator plasma with a weak activity against *Staphylococcus aureus* (Bishop *et al.*, 2015), bacteriocins produced *Lactobacillus salivarius* effective against the sensitive *Listeria innocua* (Barrett *et al.*, 2007; Flynn *et al.*, 2002), and dermatoxins (Amiche *et al.*, 2000; Chen *et al.*, 2003). The peptide-2 sequence alignment, also showed some similarities (around 30%) with known peptides: bacteriocin BacFL31 produced by *Enterococcus faecium* active against *Listeria monocytogenes* (Chakchouk-Mtibaa *et al.*, 2014), a platypus defensin-like peptide with a small activity against *Staphylococcus aureus* and *Pseudomonas aeruginosa* (Torres *et al.*, 2014), the antibacterial peptide A3-APO effective against *Escherichia coli* infections in mouse models (Szabo *et al.*, 2010), an arasin-like peptide active against some aquatic Gram-positives and Gram-negatives bacteria (Imjongjirak *et al.*, 2011) and nemuri antimicrobial peptide active against *Serratia marcescens* and *Escherichia coli* (Toda *et al.*, 2019). Note that peptide-2 is on the same coding direction as the CobN subunit whereas peptide-1 is the reverse complement sequence.

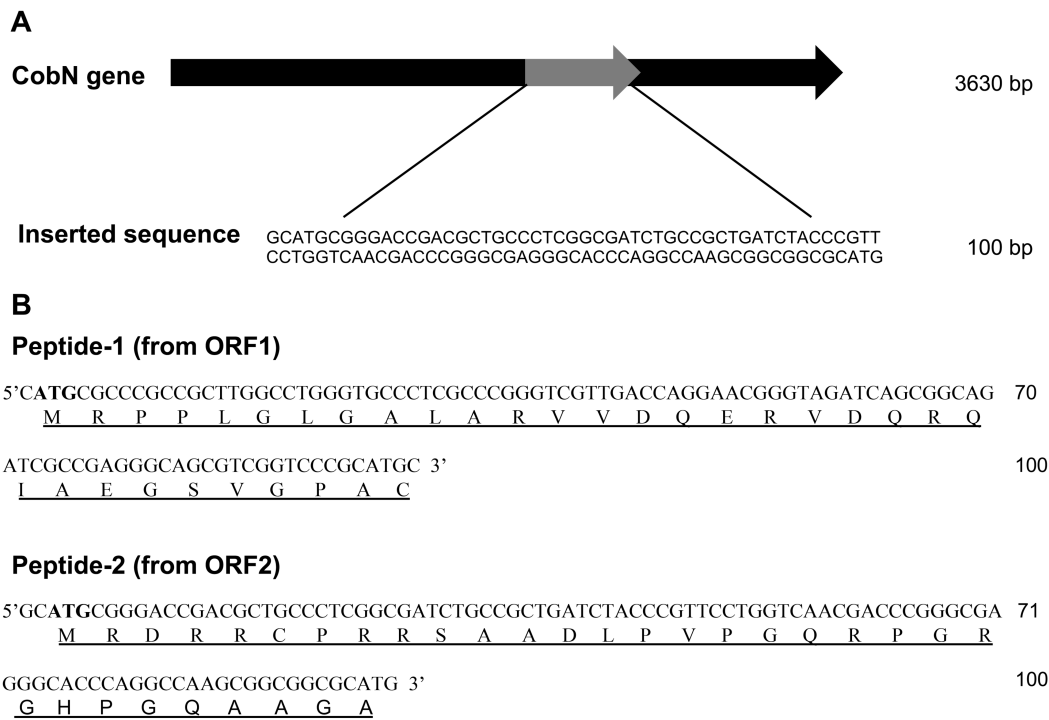


Figure 4.4. Representation of the clone WHW38 inserted DNA, (A) in whole cobaltochelatae subunit CobN gene of *Mycobacterium* sp. The inserted sequence is on the same coding direction as the subunit CobN gene, but not on the same reading frame (n+2) as the peptide-2 nucleotide sequence. (B) Sequence of peptide-1 and peptide-2 translated from ORF1 and ORF2 (ORF2 is the reverse complement of ORF1) suggested by ORF finder (NCBI) restrained to *E. coli* ATG (Villegas and Kropinski 2008) as the start codon.

4.1.5.3. Synthesis, activity assessment test and characterisation of ORF1- and ORF2-derived peptides

The negative selection in the host conferred by the expression of this eDNA fragment prompted us to directly synthesize both predicted peptides, as an alternative to heterologously expressing them from a multiple copy vector. In fact, Courtois et al., 2003 have already observed a negative selection phenomenon and their active clone was lost before the nature of the active molecule could be identified (Courtois *et al.*, 2003). The peptide synthesis strategy also had the added advantage of bypassing any laborious downstream processing of the peptide,

allowing the work with a pure product. Hence, peptide-1 and peptide-2 were synthesized based on the MHW38 plasmid sequence and in accordance with the *E. coli* codon preference (Kudla *et al.*, 2009) because of the heterologous expression system that was used for library screening. However, for peptides-1 the C-terminal cysteine residue was omitted in order to prevent the formation of an intermolecular disulfide bridge during synthesis (Yang *et al.*, 2016).

As depicted in Figure 4.5, no effect was noted when a MRSA clinical isolate was exposed to peptide-1 (up to 1.53 mM) whereas complete growth inhibition was observed when the same strain was exposed to a similar concentration of peptide-2 (1.54 mM), up to eighteen hours. It is noteworthy that although to a lesser extent, peptide-2 was also found to be active at 0.38 mM.

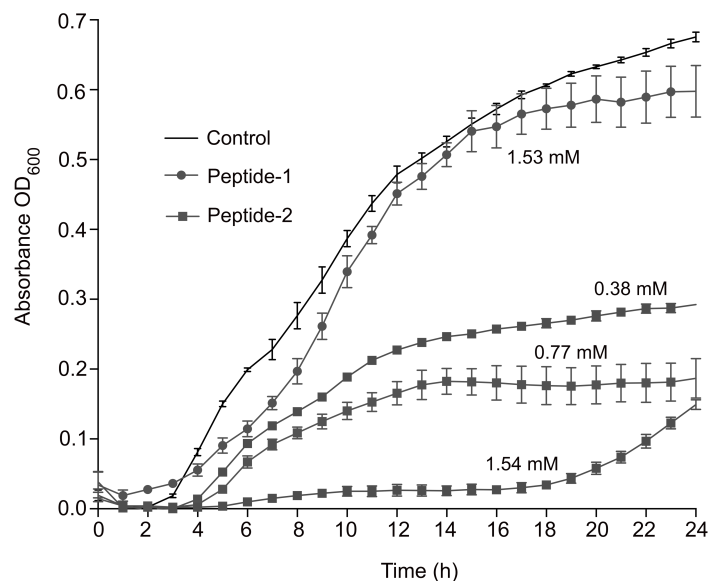


Figure 4.5. Methicillin-resistant *Staphylococcus aureus* growth after exposure to peptides 1 and 2. Bacteria were placed in wells of 96-well plates containing various concentrations of peptides in Mueller-Hinton broth and incubated at 37°C under agitation conditions. OD₆₀₀ measurements were performed at given times using a Bioscreen plate reader.

Structural analysis of both peptides revealed distinct secondary structures. As shown in Figure 4.6, peptide-1, including four positively charged residues, should adopt an α -helix conformation whereas peptide-2, presenting seven positively charged residues, should be mostly unstructured.

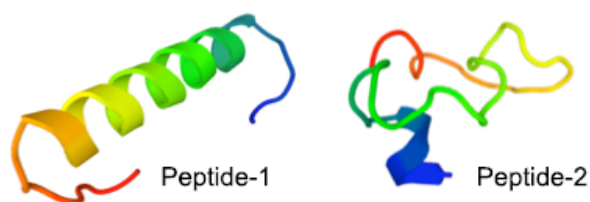


Figure 4.6. Peptide structure models of peptide-1 and peptide-2 as predicted from their amino acid sequence using the PEP-FOLD 3.5 software.

Since peptide conformation is dependent of its local environment, the secondary structure of peptide-1 and peptide-2 was investigated by circular dichroism (CD) and in a 2,2,2-trifluoroethanol (TFE)/water mixture. In water, both peptides appeared unstructured, as shown in Figure 4.7. However, in a media (50% TFE/water) that mimics the hydrophobic environment similar to that of the cell membrane, peptide-1 exhibited two negative minima at 208 and 222 nm characteristic of an α -helix structure. This helical behaviour in the presence of TFE was not observed for peptide-2 in agreement with the previously discussed predicted models. To further explore these observations, both peptides were placed in the same TFE:water mixture and NOESY NMR spectra were acquired. Figure 4.8 displays an overlay of the NOESY spectrum of peptide-1 (blue spectrum) with that of peptide-2 (red spectrum) which clearly show distinct solution behaviors. The NOESY data of peptide-1 has large NH-NH crosspeaks that report the predominance of helical secondary structure. In contrast, the NOESY spectrum of peptide-2 lacks this signature peaks. These data are clearly in agreement with the CD data in Figure 4.7 and the 3D structures shown in Figure 4.6. Taken together, these results suggest that the primary sequence rather than the secondary structure alone is responsible for the observed biological activity of peptide-2.

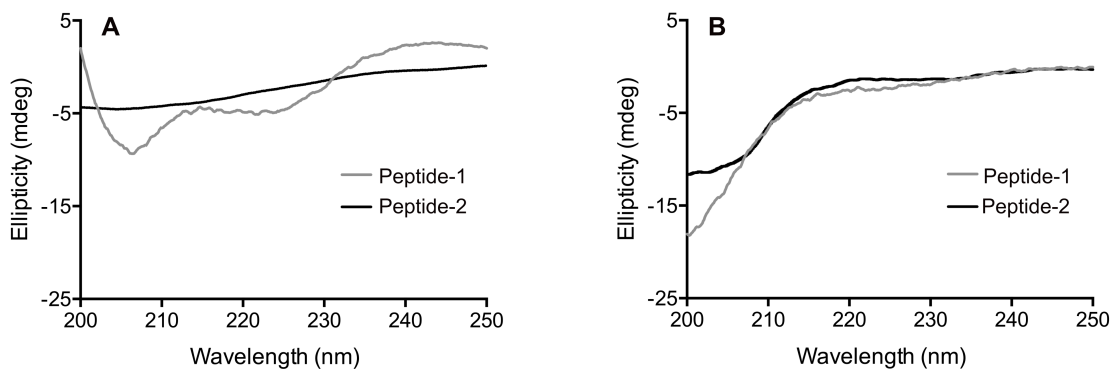


Figure 4.7. Circular dichroism analysis of peptide-1 and peptide-2. Peptides were dissolved (A) in TFE/water (1:1) mixture and (B) in water, at a concentration of 200 $\mu\text{g}/\text{mL}$. Each curve represents the mean of three scans corrected for solvent contribution.

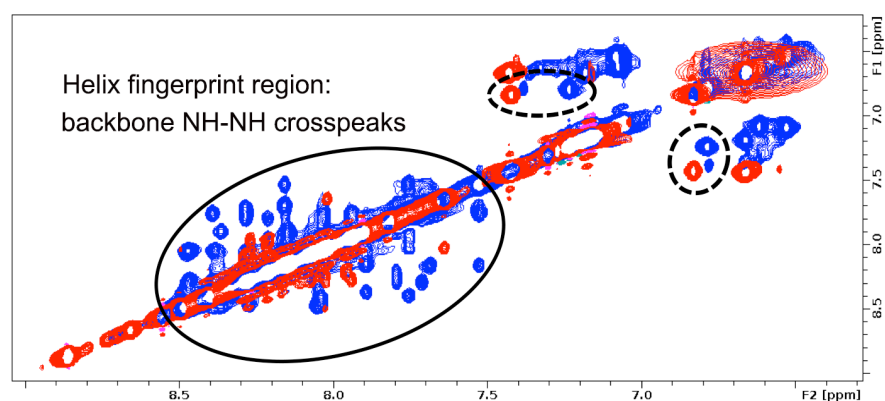


Figure 4.8. Overlay of NMR NOESY spectra of peptide-1 (blue) and peptide-2 (red). Shown are subregions of the ^1H homonuclear NOESY spectra that are considered as the fingerprint region for monitoring helices (Laplante *et al.*, 1990). Multiple blue crosspeaks for peptide-1 are observed within the non-dashed circle, which is consistent with the existence of helical secondary structure. On the other hand, only few red crosspeaks are observed for peptide-2, which is consistent with the lack of helical secondary structure. Sample preparation and experimental conditions were identical for both datasets involving peptide-1 and peptide-2, thus allowing one to normalize crosspeak intensities using crosspeaks from fixed-distance side-chain hydrogens of arginine (see crosspeaks within the dashed circles).

The unexpected lack of activity of peptide-1, despite its α -helical conformation, could be associated with a lack of amphipathic character. Using the antimicrobial database for sequence analysis, peptide-1 was characterized as a hydrophobic (42%) glycine-rich peptide (12%) with a net charge of 0. The absence of a positively charged environment/face, might explain its lack of activity. However, the possibility that post-translational modifications could have occurred in the *E. coli* host cannot be ruled out. Indeed, prokaryote post-translational modifications are less common than in eukaryotes but recent studies successfully identified such processes in bacteria, particularly under stress conditions (Brown *et al.*, 2017).

Using a similar approach, we observed that peptide-2 is a hydrophobic (25%) proline- (16%) and arginine- (22%) rich molecule with a net charge of +5. Interestingly, arasin-1, an antimicrobial proline- and arginine-rich peptide isolated from spider crab hemocytes (Stensvag *et al.*, 2008) is also completely unstructured, as observed by circular dichroism (Paulsen *et al.*, 2014). Peptide-2 could therefore be a novel member of the proline/arginine-rich group of AMPs based on its low similarity with other known AMPs, and be classified as a proline/arginine-rich peptide because of its high content in these amino acids (nearly 20% each). Interestingly, these proline-rich AMPs use a mode of action that does not involve the lysis of bacterial membranes but entails penetration into susceptible cells where they can act intracellularly (Scocchi *et al.*, 2011). In particular, proline-rich AMPs target protein synthesis (Graf *et al.*, 2017), emphasizing their favourable cell penetration properties rather than lytic activities. It is therefore possible that the antimicrobial activity of peptide-2 requires a relatively high concentration due to its inability to efficiently reach the intracellular compartment of assayed target bacteria, which could be associated to its lack of lipophilicity. For instance, peptide-2 displays a lower hydrophobicity than humimycins, antimicrobial peptides predicted from the genomes of *Rhodococcus* spp. from the human microbiome that were found to be active against MRSA (MIC <150 mg/mL). These peptides are composed of uncharged amino acids and are N-acylated with β -hydroxymyristic acid (Chu *et al.*,

2016). On the other hand, it cannot be ruled out that peptide-2 might also readily undergo degradation, as a consequence of the intracellular presence of various enzymes including proteases/peptidases. However, it is noteworthy that the entire natural molecule, presumably naturally encoded by a *Mycobacterium* sp., might not have been picked up by our screening. Although out of the scope of the current study, further experiments could allow the improvement of the antimicrobial activity of peptide-2 through strategic modifications aiming at increasing hydrophobicity or stability, and/or with structure-activity studies using truncations, deletions or substitutions. For instance, peptide cyclisation or introduction of D-amino acids have been used to decrease bacterial protease degradation as well as to improve peptide activity (Carmona *et al.*, 2013; Joo, 2012; Park *et al.*, 2000; Pavlova *et al.*, 2008).

4.1.6. Conclusion

The functional metagenomic approach used in this study, based on eDNA sampled from the biofilms growing in wastewater drainage pipes from hospital settings, allowed the discovery of a new antimicrobial peptide, i.e. peptide-2, with an activity against MRSA that could not be predicted by bioinformatics tools. This peptide could represent a new molecular template for the elaboration of therapeutically relevant MRSA antibiotics. The high concentration required to affect the growth of MRSA could be explained by the fact that the full molecule normally produced by *Mycobacterium* sp strain is not the one synthesized in this study, and/or that it is aiming at a different target than MRSA. Finally, it has to be pointed out that highly potent antimicrobial molecules could have precluded the growth of clones containing their DNA sequence, thus preventing their identification; an obvious limitation of any functional metagenomic screenings. The use of other species for heterologous expression, including other bacteria, fungi or plants, is currently envisioned to allow the identification of new and possibly more potent molecules (Parachin *et al.*, 2012).

CHAPITRE 5 : DISCUSSION GÉNÉRALE

5.1.1. Préambule

Les molécules naturelles sont une source majeure de composés pharmaceutiques pour le traitement des maladies humaines. Le processus classique de ces découvertes consiste en l'isolement de l'organisme producteur, le dépistage de son activité ainsi que la purification approfondie des molécules actives. Ces molécules représentent souvent un point de départ intéressant pour le développement de nouveaux médicaments. Par exemple, les microorganismes produisent des molécules leur permettant de compétitionner avec d'autres organismes en produisant des antibiotiques. La grande majorité des antibiotiques utilisés en médecine humaine, en médecine vétérinaire et en agriculture sont directement produits ou dérivés de molécules naturelles produites par des microorganismes. Depuis l'avancement du séquençage à haut débit, les chercheurs se sont aperçus qu'une vaste majorité de la diversité microbienne est encore inconnue et que cette richesse inexploitée pourrait être un avantage au développement de nouveaux médicaments. Pour exploiter cette diversité, nous avons utilisé la métagénomique fonctionnelle. Cette approche n'est pas nouvelle et a déjà été employée par certains chercheurs pour la découverte de nouvelles molécules antibiotiques (Brady, 2007), enzymes (Leis *et al.*, 2015), gènes de résistance aux antibiotiques (Uyaguari *et al.*, 2011), etc. Cette technique de choix est donc un excellent allié pour notre objectif de travail. Sa particularité est qu'en absence de culture en laboratoire, le lieu d'échantillonnage de l'ADN métagénomique conditionne le type de molécules recherchées. De ce fait, nous avons choisi d'échantillonner l'ADN métagénomique à partir de biofilms d'hôpitaux riches en BMR.

L'ensemble de nos résultats nous permet de tirer plusieurs observations. Tel qu'attendu, avec la construction des banques d'ADN métagénomiques, plusieurs

problèmes techniques ont été rencontrés: d'une part, de faibles efficacités de transformation nous ont mené à se questionner sur les facteurs qui pourraient être en cause, entraînant le développement d'une méthode permettant d'améliorer et de maximiser cette efficacité. D'autre part, une lacune majeure en ce qui a trait au criblage fonctionnel nous a incité à développer une nouvelle méthode facilitant cette étape. Ce n'est que par la suite que le projet de construction de banques d'ADN métagénomiques et celui du criblage fonctionnel ont pu être entamés. Pour terminer, nous avons découvert un nouveau peptide antimicrobien actif contre SARM à une concentration supérieure aux peptides antimicrobiens existant dans la littérature. Ce résultat est un début de recherche très intéressant sur le développement pharmacologique de ce peptide, en particulier parce qu'il ne présente aucune homologie suffisante pour avoir été prédit par les outils disponibles. Donc, ce modèle n'ayant jamais été découvert auparavant, nous pouvons émettre l'hypothèse que l'optimisation des résidus d'acides aminés responsables de la structure/activité de ce peptide puisse encore largement augmenter son activité.

Dans la suite de ce document, nous mettrons nos résultats en perspectives par rapport à la littérature globale sur la métagénomique fonctionnelle et nous discuterons de certaines améliorations à effectuer pour encourager l'utilisation de la métagénomique fonctionnelle dans plusieurs niveaux d'études.

5.1.2. Choix de *Burkholderia thailandensis* pour les optimisations de construction de banques d'ADN métagénomiques et de criblage fonctionnel

Dans cette étude, une optimisation de construction de banques d'ADN métagénomiques ainsi que le développement d'une méthode de criblage ont été réalisés. Étant donné que l'ADN des biofilms tiré des effluents hospitaliers était

difficile à obtenir, l'utilisation de l'ADN de *Burkholderia thailandensis* semblait un choix stratégique pour cette preuve de concept. En effet, son génome héberge au moins 21 groupes de gènes de biosynthèse de métabolites secondaires prédits par l'analyse antiSMASH, ce qui indique la grande capacité de cette bactérie à produire des métabolites secondaires (Liu & Cheng, 2014). Ces groupes de gènes représentent potentiellement des voies de production de molécules antibiotiques intéressantes. Même si plusieurs d'entre eux ont déjà été identifiés, tels que thailandépsines/burkholdacs A et B (Biggins, Gleber et al. 2011), capistrain (Knappe, Linne et al. 2008), malleilactone (Biggins, Ternei et al. 2012), acide burkholderique (Franke, Ishida et al. 2012), thailandamide (Ishida, Lincke et al. 2010), bactobolin (Seyedsayamdost, Chandler et al. 2010), la majorité de ces BGC demeurent silencieux (Liu et Cheng 2014). D'autre part, même la réidentification de certains de ces BGC aurait également pu contribuer à prouver la validité de la méthode développée.

Découpler ces opérons de leurs promoteurs natifs et les placer sous contrôle d'un promoteur inductible n'a jamais encore été réalisé avec l'ADN de *B. thailandensis* afin d'étudier ses opérons silencieux. Généralement, dans de nombreux cas, la variation des conditions de laboratoire de culture (Gupta, Bedre et al. 2017), l'ajout d'inducteurs (Seyedsayamdost, 2014) ou l'inactivation d'un répresseur (Mao *et al.*, 2017) sont les méthodes utilisées pour déclencher et activer ces voies métaboliques.

Au cours de notre étude (chapitre 2), nous avons identifié 6 clones inhibant la croissance de la souche de laboratoire *S. aureus* Newman par expression hétérologue de l'ADN de *B. thailandensis*. Le choix de cette bactérie cible a été effectué dans le but d'augmenter la probabilité d'identifier des clones positifs, comme preuve de concept du fonctionnement de la méthode de construction de banques d'ADN métagénomiques et de celle du criblage fonctionnel.

5.1.3. Choix des biofilms d'effluents hospitaliers pour l'étude métagénomique fonctionnelle

Le choix du lieu d'échantillonnage de l'ADN dans les études métagénomiques fonctionnelles détermine les propriétés des molécules recherchées. Cela est dû aux mécanismes d'évolution des microorganismes aux changements environnementaux, qui sélectionnent des traits génétiques leur conférant un avantage face aux conditions de l'environnement. Ces adaptations se manifestent à divers niveaux hiérarchiques: traitement de l'information génétique, récepteurs pour la détection des facteurs environnementaux, niveau transcriptionnel, post-transcriptionnel, et métabolique (Brooks, Turkarslan et al. 2011).

Dans notre étude principale, les inserts des clones actifs sélectionnés proviennent tous de l'ADN du biofilm échantillonné directement d'un conduit d'égout d'un hôpital de la grande région de Montréal. Même si l'analyse de la composition en microorganismes du biofilm n'a pas pu encore être réalisée, les biofilms d'hôpitaux sont généralement composés d'un cocktail de bactéries multirésistantes (Katouli *et al.*, 2012), d'une forte concentrations d'antibiotiques créant une pression de sélection, de gènes de résistance aux antibiotiques et d'éléments mobiles (Le *et al.*, 2016; Sabri *et al.*, 2018; Wang *et al.*, 2018). Ces biofilms contiennent aussi de la matière fécale et de l'urine de patients hospitalisés, colonisés par des bactéries multirésistantes (Buelow *et al.*, 2017). Ce microbiote humain est aussi connu pour produire des bactériocines (Cohen, Han et al. 2018), des antimicrobiens (Chu, Vila-Farres et al. 2016) et des peptides cycliques (Zipperer, Konnerth et al. 2016).

Si l'on se penche sur la littérature en métagénomique fonctionnelle (tableau 1.1 chapitre 1), pour la majorité de ces études de métagénomique fonctionnelles, le lieu d'isolement de l'ADN sélectionné fut les sols forestiers, donc sans pression d'antibiose particulière. Le prélèvement d'ADN à partir d'un biofilm présent dans un système d'évacuation des effluents d'hôpitaux n'avait jamais encore été rapporté.

Évidemment, notre étude nous ne permet pas affirmer que les environnements tels les effluent d'hôpitaux nous conduiront systématiquement à la découverte de nouvelles molécules antibiotiques actives contre les bactéries multirésistantes. Toutefois, il est important de noter que cette hypothèse a été récemment supportée par Peek *et al.*, 2018 qui postulent que la compétition entre les microorganismes de l'environnement pourrait accélérer l'évolution de certains antibiotiques (comme par exemple la rifamycine) étant capables de contourner les mécanismes de résistance aux antibiotiques actuels pour le traitement de certaines bactéries. Cette hypothèse les a menés à la découverte de la kanglemycine, une molécule active contre *Mycobacterium tuberculosis* résistante à la rifampicine (Mosaei *et al.*, 2018; Peek *et al.*, 2018).

En plus de cette hypothèse, nous supportons que les bactéries multirésistantes aux antibiotiques puissent aussi produire des molécules antibiotiques pour compétitionner avec les bactéries étant également multirésistantes aux antibiotiques. Étant donné que ces bactéries sont difficiles à étudier pour déterminer l'arsenal de molécules bioactives qu'elles produisent du fait de leur multirésistance, étudier leurs ADN par métagénomique fonctionnelle est une alternative valable.

Un des exemples les plus acceptés pour justifier le lieu d'isolement d'un ADN, en fonction de la molécule recherchée est la recherche d'enzymes. En effet, le besoin d'enzymes pour le développement de procédés industriels a poussé Angelov et collègues à réaliser un criblage fonctionnel à partir d'une banque d'ADN métagénomique d'un sol de volcan. Dans leur étude, ils ont découvert une enzyme, β -glucanase, qui s'est avérée active à 90°C avec un profil d'activité différent des glucanases connues, et qui possède une structure modulaire actuellement non caractérisée (Angelov *et al.*, 2017). Un autre exemple est la découverte d'une ADN polymérase active à 18°C à partir d'un criblage d'un ADN provenant d'un glacier, dans le but de développer de nouvelles méthodes en biologie moléculaire (Simon, Herath et al. 2009).

Le microbiome humain a aussi suscité beaucoup d'intérêt récemment pour découvrir de nouvelles molécules bioactives. En effet, l'un des meilleur exemples est la découverte de la lugdunine, un peptide antimicrobien produit simplement en cultivant la souche nasale *Staphylococcus lugdunensis*; (Zipperer *et al.*, 2016),

5.1.4. Rôle des facteurs influençant l'efficacité de transformation

Les problèmes d'efficacité de transformation rencontrés au cours des constructions de banques d'ADN métagénomiques ne sont pas ouvertement discutés dans la littérature. Pourtant, plusieurs étapes se succèdent pour la construction des banques d'ADN métagénomiques, détaillées dans la figure 3.8 (chapitre 3). Si l'une de ces étapes échoue, il est probable qu'aucun clone ou qu'un nombre très limiter de clones ne résulte de la transformation, reflétant une faible efficacité de transformation. Nous avons alors décidé d'entreprendre l'étude des facteurs les plus influants sur l'efficacité de transformation.

L'un des facteurs les plus limitants dans l'efficacité de transformation que nous avons pu observer (chapitre 2 section 2.1.5.1.) est l'étape de dialyse du produit de ligation avant la transformation. En effet, entre un produit de ligation dialysé et un autre non dialysé, une efficacité de transformation 1,000 fois supérieure a pu être observée. De plus, durant l'électroporation, la dialyse du produit de ligation permet une faible conductivité, ce qui évite les arcs électriques lors de l'électroporation (2.0 kv). Saraswat et collègues ont constaté que le dessalage du produit de ligation par purification en microcolonnes résultait en des efficacités de transformation moindres à comparer à lorsque le dessalage était effectué par membrane de nitrocellulose (Saraswat *et al.*, 2013).

Bien déphosphorylé le vecteur est essentiel pour maximiser l'étape de ligation (Brady, 2007). Nous avons d'ailleurs pu constater qu'une omission de cette étape diminuait l'efficacité de transformation de 1,000 fois.

Contre toutes attentes, augmenter le ratio d'ADN génomique (BAC/ADNe) n'a pas eu un grand effet sur l'efficacité de transformation. Afin de bien constater cet effet sans que les autres paramètres influencent l'efficacité de transformation, la déphosphorylation du vecteur et la dialyse du produit de ligation ont d'abord été réalisés. Nous avons remarqué que l'efficacité de transformation augmentait avec l'augmentation du ratio d'ADN 1:3 jusqu'à 1:9, mais qu'au-delà de ce ratio, un effet négatif sur l'efficacité de transformation était observé. Donc, étant donné que le vecteur pBeloBAC11 est saturé en ADN digéré, une quantité considérable d'ADN non ligué au vecteur demeure dans le produit de ligation. En transformant ce produit de ligation chez *E. coli* DH10B, les fragments d'ADN non ligués sont plus faciles à transformer que le vecteur ayant un insert. Ces fragments d'ADN n'étant pas liés au vecteur ne procurèrent pas à *E. coli* le gène de résistance à la chloramphénicol nécessaire à la croissance sur gélose de sélection après transformation.

Cette stratégie nous a permis de maximiser l'efficacité de transformation dans le but de maximiser la diversité des inserts d'ADN des clones transformants pour le criblage fonctionnel. Toutefois, d'autres facteurs auraient pu être étudiés. Par exemple, avant d'exclure directement les cellules compétentes préparées au laboratoire, la concentration de CaCl_2 , le temps d'incubation au bain chaud, le temps de récupération sur glace et la variation de certains ions auraient pu être également optimisés (Liu *et al.*, 2014).

La transformation du produit de ligation par électroporation est la méthode la plus utilisée en métagénomique fonctionnelle pour la construction de banques d'ADN métagénomiques et c'est pour cette raison que nous l'avons utilisée (Biver *et al.*, 2013; Cohen *et al.*, 2018; Craig *et al.*, 2009; Gillespie *et al.*, 2002; Iqbal *et al.*,

2014; MacNeil *et al.*, 2001a; Rondon *et al.*, 1999; Simon & Daniel, 2010; Thies *et al.*, 2016). Il est à noter que d'autres chercheurs utilisent la méthode de transduction où l'ADN inséré au vecteur est emballé dans un phage qui infecte directement la bactérie hôte (Brady *et al.*, 2004; Brady & Clardy, 2004; Courtois *et al.*, 2003). Cette méthode permet de directement transformer le produit de ligation sans passer par la dialyse du produit de ligation, qui peut dans certains cas causer une perte d'ADN.

5.1.5. Intervention des gènes de la bactérie hôte dans l'expression hétérologue

L'importance de l'hôte d'expression hétérologue dans une étude métagénomique fonctionnelle est largement rapportée dans la littérature. Pour que la molécule antibiotique produite soit attribuée à l'ADN inséré, il est préférable que l'hôte ne produise pas de molécules antibiotiques efficaces envers la cible choisie (Handelsman, 2007). Dans le cadre de notre étude, *E. coli* est le seul hôte d'expression hétérologue utilisé car cela facilite les conditions de croissance. Avec cet hôte, nous avons quand même pu constater une morphologie particulière en étoile chez le clone MHW₂178 et plusieurs clones positifs.

Idéalement, il aurait été préférable de réaliser cette étude métagénomique chez au moins un autre hôte d'expression hétérologue tel qu'une bactérie à Gram + (*Bacillus subtilis* ou *Streptomyces lividans*). La principale raison est que différents phénotypes peuvent être exprimés d'un hôte d'expression hétérologue à un autre. Par exemple, dans l'étude de Craig et collègues, le criblage fonctionnel chez plusieurs hôtes d'expression hétérologue montre aucun clone positif ni aucune morphologie particulière chez l'hôte *E. coli* contrairement à une colonie à morphologie particulière chez l'hôte *Agrobacterium tumefaciens* (Craig *et al.*, 2010) et une colonie pigmentée chez les hôtes *Pseudomonas putida*, *Agrobacterium tumefaciens* et *Ralstonia metallidurans*. De même aucun clone positif n'a été

observé dans l'étude de Biver *et al.*, 2013, en utilisant l'hôte d'expression hétérologue *E. coli* contre un clone positif en transférant la même banque d'ADN chez *B. subtilis*. Néanmoins, de nombreuses molécules antibiotiques ont quand même été découvertes en utilisant seulement *E. coli* comme hôte d'expression hétérologue: la violaceine (Brady *et al.*, 2001b), la turbomycine (Gillespie *et al.*, 2002), les longues chaînes N-acyl-tyrosine (Brady *et al.*, 2002), l'indirubine et l'indigo (MacNeil *et al.*, 2001a).

La dépendance de l'expression de l'ADN inséré sur l'hôte hétérologue ne nous a pas considérablement affecté dans notre étude sur le criblage fonctionnel de la banque d'ADN génomique de *B. thailandensis* (chapitre 2), puisque *B. thailandensis* et *E. coli* sont toutes deux des proteobactéries. L'expression hétérologue dépend beaucoup de la machinerie de la bactérie hôte et de ses préférences de codon. De plus, pour que la molécule produite soit active, un repliement adéquat et une excrétion à l'extérieur de la bactérie hôte sont nécessaires (Kudla *et al.*, 2009). En revanche, l'expression hétérologue pour notre étude métagénomique fonctionnelle à partir d'ADN de biofilms (chapitre 3 et 4) a sans doute été affectée par le biais de l'utilisation de codons de *E. coli* étant donnée la plus grande diversité microbienne présente dans des biofilms (Wang *et al.*, 2018).

Cette préférence de codons chez cet hôte a été étudiée par Kudla *et al.*, 2009 en créant une librairie de 154 gènes codant pour la protéine de fluorescence GFP, avec une mutation silencieuse à la troisième position du triplet codon. Le niveau d'expression de la protéine GFP différait de 250 fois selon ces mutations, ainsi que le niveau de l'ARNm, de son mode de dégradation et du taux de croissance bactériens variait également. La variation de ce niveau d'expression est expliquée par la variation du taux d'initiation de la traduction, liée aux différents repliements de l'ARNm. Quant au biais de codon, il influence directement l'efficacité de la traduction et la séquence (donc la forme) de la protéine finale (Kudla *et al.*, 2009).

De plus, les microorganismes ont une préférence pour certains codons d'initiation de la traduction (Saier, 1995). Dans le cas de *E. coli*, le codon d'initiation de la traduction AUG est utilisé à 92%, GUG est utilisé à 6,7%, et UUG est utilisé à 1,2%. Donc les codons d'initiation non AUG peuvent ne pas être reconnus efficacement et entraîner un problème de traduction dans ce modèle (Villegas & Kropinski, 2008a). C'est en suivant cette stratégie et en vérifiant les codons préférentiels de *E. coli* que le peptide antimicrobien dérivé de ORF2 a été découvert dans notre étude (chapitre 3). La recherche d'ORF par ORF Finder s'est concentrée seulement sur les ORFs avec un codon de départ AUG, donnant seulement deux options de peptides possibles qui pourraient être à l'origine de l'activité observée. Évidemment, il est possible que la molécule produite par expression hétérologue de cet insert chez DH10B soit totalement différente des prédictions effectuées. La molécule hypothétique pourrait adopter une toute autre conformation suite à des modifications post-transcriptionnelles par des enzymes de modification de protéines propres à la bactérie hôte *E. coli*, pouvant alors modifier son activité biologique (Gabor *et al.*, 2004).

Une autre raison qui justifie l'utilisation d'au moins deux hôtes d'expression hétérologue est que la molécule produite par l'hôte peut être toxique chez l'hôte et qu'elle ne sera donc jamais représentée (Rondon *et al.*, 2000a). Un phénomène similaire a été observé chez le clone MHW₅38 qui avait des difficultés de croissance en milieu liquide et solide. Sa croissance était probablement affectée par la production d'une molécule toxique. C'est l'une des raisons qui nous a poussé à exploiter directement la séquence de l'insert comme base de synthèse chimique de la molécule en fonction des préférences codons de *E. coli*.

5.1.6. Importance du nombre de clones à cribler

Le criblage d'un nombre élevé de clones est nécessaire afin d'augmenter la probabilité d'identifier des clones positifs, actifs contre la bactérie cible testée. Dans

le cadre de notre étude, le taux de succès varie de 1/10,000 pour la banque d'ADN génomique construite à partir de l'ADN de *B. thailandensis*, et à 1/15,000 pour la banque d'ADN métagénomique construite à partir de l'ADN de biofilms. Bien sûr, le nombre de clones actifs dépend considérablement de la bactérie cible utilisée. Dans l'étude effectuée à partir de l'ADN génomique de *B. thailandensis*, qui consistait en une preuve de concept, nous avons choisi *S. aureus* Newman comme bactérie cible afin d'augmenter la probabilité d'identifier des clones actifs. Pour notre étude métagénomique à partir de l'ADN de biofilms, nous étions plutôt intéressés de tenter d'identifier une molécule active contre SARM. Par exemple d'après la littérature, le taux de succès de ce type d'étude varie considérablement, étant par exemple de 1/3,000 pour l'étude de MacNeil *et al.*, 2001, où les bactérie très sensibles *B. subtilis* et *S. aureus* ont été utilisée comme bactéries cibles (MacNeil *et al.*, 2001a), de 1/5,000 pour l'étude de Courtois et collègues où la bactérie cible utilisée était *B. subtilis* (Courtois *et al.*, 2003), et de 1/25,000 pour l'étude de Biver *et al.*, 2013 où la bactérie cible utilisée était *B. cereus*. La seule étude en métagénomique fonctionnelle qui a rapporté l'utilisation d'une bactérie cible résistante à un antibiotique est l'étude de Iqbal *et al.* (Iqbal *et al.*, 2014) pour laquelle la bactérie cible *B. subtilis* résistante à la tétracycline a été utilisée comme bactérie cible. Le taux de succès était alors d'environ de 1/40,000

5.1.7. Analyse des inserts des clones sélectionnés

Parmi tous les clones positifs sélectionnés, c'est-à-dire six clones issus du criblage fonctionnel de la banque d'ADN génomique de *B. thailandensis* E264, et dix clones du criblage fonctionnel de la banque d'ADN métagénomique du biofilm de l'hôpital de Montréal, la taille des inserts se situait entre 99 pb et 4019 pb et ce, malgré les précautions ayant été prises pour construire des banques d'ADN à large insert. En effet, comme il était envisagé de construire des banques d'ADN dont la taille minimum dépassait 10,000 pb, le vecteur pBeloBAC11 qui peut théoriquement intégrer des fragments d'ADN allant jusqu'à 100 kb (Daniel, 2004) a été

sélectionné. En plus, une méthode d'extraction d'ADN qui limitait les cassures d'ADN lors de l'extraction d'ADN a été utilisée, le bead-beating de l'ADN durant l'extraction a été remplacé par le vortex, et la trousse d'extraction d'ADN qui permettait d'obtenir les fragments d'ADN les plus larges a été choisie. Il est à noter qu'une digestion partielle de cette ADN s'est révélée suffisante pour conserver la longueur des fragments d'ADN. Cependant, lors de l'extraction sur gel des fragments dont la taille était supérieure à 10 kb, le rendement d'ADN après extraction sur gel était trop faible, aucun fragment d'ADN de 10 kb et plus ont été obtenus. Pour remédier au problème de rendement, l'alternative choisie fut en premier lieu de concentrer tous les produits de l'extraction sur gel en un petit volume. Comme cette technique ne semblait pas suffisante pour obtenir de longs fragments d'ADN, l'étape d'extraction sur gel n'a simplement pas été réalisée pour la construction des banques d'ADN métagénomiques en chapitre 3 et 4. L'une des stratégies alternatives pouvant être employée pour récupérer des grands fragments d'ADN est l'électroélution de l'ADN après digestion, afin de remplacer l'étape d'extraction sur gel qui était l'étape limitante (Zarzosa-Alvarez *et al.*, 2010).

Peu d'études métagénomiques pour la découverte de nouvelles molécules antibiotiques ont utilisé des vecteurs plasmidiques (tableau 1.1, chapitre 1). Ceci s'explique par le fait qu'il soit plus avantageux de construire des banques d'ADN à larges inserts, sachant que la majorité des voies de synthèse de métabolites secondaires tels que les antibiotiques de types polyketides synthase (PKS) et les peptides non ribosomiques (NRPS) sont regroupées en opérons ou «clusters» et requiert plusieurs gènes. La probabilité de découvrir une nouvelle molécule bioactive se voit alors augmentée (Rondon *et al.*, 2000b; Simon & Daniel, 2011). De plus, un métagénome du sol requiert théoriquement approximativement 10^6 vecteurs BAC avec un insert d'ADN de 100 kb (Daniel, 2004), ce qui réduit largement le nombre de clones à tester pour une étude donnée. Mais, en raison de certains problèmes techniques rencontrés en fonction des différentes méthodes d'extraction d'ADN citées plus haut, il est très difficile d'obtenir des fragments

d'ADN d'une taille supérieur à 100 kb. En pratique, la taille de l'ADN inséré dans les études métagénomiques se limite à 40 kb, et cette taille ne permet pas de cloner des groupes de gènes polycistronique de type PKS et NRPS (Kang & Brady, 2014).

De plus, retrouver le gène responsable de la production de la molécule bioactive reste un défi majeur pour les banques d'ADN à larges inserts. Des prédictions d'ORF par certaines bases de données comme antiSMASH peuvent contribuer à l'identification des groupes de gènes responsables du phénotype observé. Une mutagenèse dirigée ou aléatoire peu alors apporter une première confirmation du phénotype, suivie d'un sous-clonage. Par exemple, dans l'étude de Rondon *et al.* un BAC-clone avec un insert de 25 kb ayant une activité antibactérienne a été identifié. Après l'identification du locus par inactivation du gène, le sous-clonage et l'induction de ce locus seul a mené à l'expression d'une petite protéine de 55 kDa qui inhibait la croissance de la bactérie hôte. Il est à noter qu'une fois purifiée, cette protéine ne démontrait plus d'activité (Rondon *et al.*, 2000a).

La construction et le criblage de banques d'ADN à petits inserts est une bonne alternative pour l'étude des petits gènes exprimés par les promoteurs inductibles du vecteur utilisé (MacNeil *et al.*, 2001a). Par exemple, dans notre étude, l'insert du clone actif sélectionné était seulement de 100 pb. Mais, la prédiction d'ORFs possibles à partir de tous types de codons de départ nous a suggéré 9 possibilités de peptides antimicrobiens. Plus la taille de l'insert augmente et plus il y a de possibilités de prédictions de gènes responsables du phénotype observé. Il serait intéressant de sous-cloner cet insert dans un plasmide inductible à multicopies, et de changer d'hôte d'expression hétérologue étant donné que *E. coli* est sensible à la molécule produite. En fait, de choisir un hôte qui permet l'excrétion de ces molécules dans le surnageant de culture afin d'induire l'expression hétérologue de cette molécule, de la purifier à partir du surnageant et de comparer son profil aux peptides antimicrobiens prédit à partir des ORFs de notre étude pourrait être avantageux.

5.1.8. Effet des peptides antimicrobiens découverts à partir de la séquence de l'insert du clone MHW₅₃₈ sur la croissance de *S. aureus* résistant à la méthicilline.

Les peptides antimicrobiens sont très étudiés de nos jours, la plupart sont cationiques, ont une structure en hélice et ciblent directement la membrane des cellules bactériennes, typiquement en déstabilisant la bicouche lipidique (Bahar & Ren, 2013). Ils représentent une alternative intéressante pour contrer le problème de résistance aux antibiotiques. C'est dans cette optique que la séquence de l'insert du clone MHW₅₃₈, a été traduite en acides aminés pour étudier les caractéristiques des molécules résultantes. Vu le caractère en hélice du peptide ORF1, il fut le premier candidat pour lequel une synthèse chimique a été considérée. De plus, suivant l'hypothèse que ces peptides en hélice sont connus pour avoir une activité antimicrobienne, les dérivés de ce peptide ont également initialement été synthétisés.

En constatant leur très faible activité antimicrobienne, une deuxième série de peptides dérivés des autres ORFs prédits en fonction des codons de départ aléatoires ont aussi été considérées en priorisant les peptides ayant une structure en hélice (ORF5 et ORF 6) ainsi que les peptides ayant un fort caractère cationique (ORF 2). Ce dernier s'est avéré le plus actif de tous ceux synthétisés, un ralentissement de la croissance de SARM ayant été observé à 1,53 mM. En scrutant de plus près l'origine des peptides dérivés des ORF2 et ORF6, on note que qu'ils sont prédits qu'à partir d'un décalage de cadre de lecture de deux nucléotides seulement (Figure 5.1). Le peptide dérivé de l'ORF6 consiste en une séquence 100% identique (même brin codon et même cadre de lecture) à la séquence de la cobaltochélatase 3 de *Mycobacterium* sp (figure 3.13 chapitre 3). Quant à la séquence du peptide dérivé de l'ORF2, aucune homologie de séquence n'est prédite sur NCBI.

Séquence de l'ORF2

5'GC**ATG**CGGGACCGACGCTGCCCTCGGGCGATCTGCCGCTGATCTACCCGTTCCCTGGTC
AACGACCCGGGCGAGGGCACCCAGGCCAAGCGGCGGGCGGCA 3'

Séquence de l'ORF6

5'**GCA**TGCGGGACCGACGCTGCCCTCGGGCGATCTGCCGCTGATCTACCCGTTCCCTGGTC
AACGACCCGGGCGAGGGCACCCAGGCCAAGCGGCGGGCGGCAT 3'

Figure 5.1. Décalage du cadre de lecture de deux nucléotides entre la séquence codante du peptide ORF2 avec ATG comme codon d'initiation et le peptide ORF6 avec GCA comme codon d'initiation.

Il est important de rappeler que les membranes bactériennes sont chargées négativement en raison de leur teneur élevée en phospholipides anioniques (acides lipotéichoïques et lipopolysaccharides retrouvés chez les bactéries à Gram + et à Gram -). Lors du contact des peptides antimicrobiens cationiques avec la membrane bactérienne, un repliement du peptide en structure secondaire est induit (*eg.* en conformation en hélice alpha) ce qui déstabilise la membrane bactérienne (Bahnsen *et al.*, 2013). Bien que le peptide dérivé de l'ORF2 soit chargé positivement, aucune conformation en hélice n'était prédite par PEP-FOLD, ni même par analyse dichroïsme circulaire et par l'analyse RMN dans le solvant TFE (trifluoroéthanol), qui pourtant favorise l'adoption de structure secondaires (Oman & van der Donk, 2010). Cette observation suggère que l'action du peptide dérivé de l'ORF2 puisse ne pas se situer au niveau de la membrane bactérienne, comme il est généralement noté chez les peptides antimicrobiens. En effet, dans ce cas, cet effet serait plutôt noté une fois que le peptide se retrouverait à l'intérieur de la bactérie. Il s'agit probablement de la raison pour laquelle une concentration saturante de ce peptide soit nécessaire pour permettre sa pénétration. Toutefois, des expériences supplémentaires sont nécessaires pour démontrer cela. Des études de relation structure-activité (RSA) seront nécessaires afin d'améliorer et de comprendre l'activité de ce peptide (Huang *et al.*, 2010). La production de peptides homologues au peptide-2 ayant une meilleure stabilité et capacité de pénétration à l'intérieure de la bactérie cible (Brogden & Brogden, 2011) pourrait grandement améliorer son activité (Kim *et al.*, 2014; Park *et al.*, 2000).

5.1.9. Autres hypothèses expliquant l'effet du peptide-2 à forte concentration

Le peptide-2 provient de l'ORF2, une séquence d'ADN 95% identique à la séquence du brin codon de la cobaltochélatase de *Mycobacterium* sp. mais dans un cadre de lecture différent de n+2 (figure 4.4). Afin d'essayer de comprendre les raisons de la faible activité de ce peptide, certaines hypothèses ont été émises.

*i. Le peptide-2 n'est pas le peptide réellement codé par le genre *Mycobacterium* sp.*

En comparant la séquence nucléotidique du peptide ORF6 (portion identique à celle de la sous-unité CobN en figure 3.14), à la séquence nucléotidique de la sous-unité CobN, certaines, cinq nucléotides étaient différents en position (18, 21, 37, 69, 93) (Figure 5.2). Mais en considérant le cadre de lecture de la sous-unité CobN, la même séquence en acides aminés était retrouvée (séquence de l'ORF6 tableau 3.2).

En position 18 : **GCG** au lieu de **GCT** : alanine

En position 21 : **GCG** au lieu de **GCC** : alanine

En position 37 : **TTG** au lieu de **CTG** : leucine

En position 69 : **CCT** au lieu de **CCG** : proline

En position 93 : **CGA** au lieu de **CGG** : arginine

Sequence ID: Query_122795 Length: 99 Number of Matches: 1				
Range 1: 1 to 99 Graphics ▼ Next Match ▲ Previous Match				
Score	Expect	Identities	Gaps	Strand
156 bits(84)	1e-43	94/99(95%)	0/99(0%)	Plus/Plus
Query 4		GCATGCGGGACCGACGCGGCGCTCGGCGATCTGCCGTTGATCTACCCGTTCTGGTCAAC		63
Sbjct 1		GCATGCGGGACCGACGCTGCCCTCGGCGATCTGCCGCTGATCTACCCGTTCTGGTCAAC		60
Query 64		GACCCTGGCGAGGGCACCCAGGCCAAGCGACGGGCGCAT	102	
Sbjct 61		GACCCGGGCGAGGGCACCCAGGCCAAGCGGCGGGCGCAT	99	

Figure 5.2. Aligement de la séquence nucléotidique ORF6 avec la partie similaire de la séquence de CobN chez *Mycobacterium* sp. NRRL B-3805. ‘Query’ correspond à la séquence de l’insert du clone MHW₅₃₈ tandis que ‘Subject’ correspond à la séquence de *Mycobacterium* sp.

En revanche, si l’on change de cadre de lecture pour obtenir la séquence du peptide-2, certains acides aminés sont différents: la cystéine en position 6 est remplacée par une glycine, la proline en position 7 est remplacée par une alanine, l’alanine en position 12 est remplacée par une valine, la glycine en position 22 est remplacé par une tryptophane et finalement l’alanine en position 30 est substituée par une thréonine. Toutefois, une fois l’analyse PEP-FOLD réalisée, le peptide ORFM est tout de même non structuré, tout comme le peptide-2.

ORF2	MRDRRC P RRSAADLPVPGQRP G RGHPGQAAGA
ORFM	MRDRRG A RRSAVDLPVPGQRP W RGHPGQATGA

Figure 5.3. Comparaison entre la séquence peptidique de ORF2 et celle de *Mycobacterium* sp. ORFM : séquence peptidique de *Mycobacterium* sp. de la région CobN.

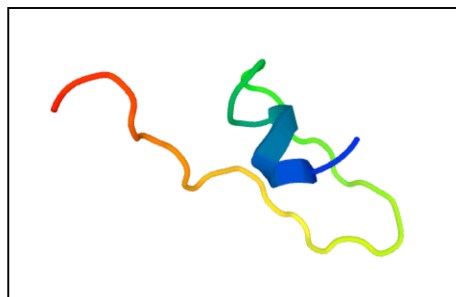


Figure 5.4. Analyse structurale prédictive de la séquence ORFM par le serveur PEP-FOLD. Dégradé de couleurs allant du N-terminal (bleu) au C-terminal (rouge).

ii. La séquence du peptide ORF2 n'est qu'une partie de la séquence d'une molécule réellement produit par Mycobacterium sp.

En analysant la séquence nucléotidique de la sous-unité CobN, et en considérant les codons stops en amont et en aval de la séquence insérée, une protéine de 160 acides aminés en résulte (Figure 5.5). En considérant la méthionine comme codon de départ, une analyse structurale prédictive des 50 premiers acides aminés de cette séquence a été réalisée par PEP-FOLD (50 acides aminés étant le maximum pour le serveur PEP-FOLD). Le résultat n'indique aucune possibilité de structure de cette protéine.

Stop DAGDEQRM**RD**RR**GARR**SAVDLPVPGQRPWRGHPGQATGACHPGRPS
DPADGARRNVWRHRASGTAARRALQCVRAGPGQAARHPPADLDPDARGED
GSRPGSGGSARRGLLRRHAAARRRLAVRDQGRPDSRRSAHPGPRARGGLR
TGPGAGDPAGPPAVRW Stop

Figure 5.5. Séquence en acides aminés de la partie en amont et en aval de la séquence de du peptide ORF2. La sequence de couleur représente la séquence du peptide-2.

iii. Le peptide-2 est actif contre une autre cible que celles testés.

Jusqu'à date, la production d'une quelconque molécule naturelle antibactérienne par le genre *Mycobacterium* sp n'a jamais été rapportée dans la littérature. Par contre, l'analyse de ce génome indique la présence de nombreuses protéines à fonctions inconnues (Amoutzias *et al.*, 2016; Muller *et al.*, 2015; Quadri, 2014). Il est fortement possible que le peptide-2 soit actif contre une bactérie cible spécifique qui n'a pas encore été testée. En effet, certains peptides ont démontré une sélectivité à l'égard de certaines cibles spécifiques (Debbabi *et al.*, 2018).

CHAPITRE 6 : CONCLUSION ET PERSPECTIVES

6.1. CONCLUSION

La capacité d'exploiter les génomes des microorganismes sans les cultiver au préalable a apporté une nouvelle ère dans la recherche d'antibiotiques. Dans cette optique, cette thèse avait pour objectif de rechercher par métagénomique fonctionnelle un nouvel antibiotique ou une nouvelle classe d'antibiotiques actifs contre la BMR SARM. Les résultats obtenus soutiennent l'hypothèse que l'isolement de l'ADN à partir d'un environnement riche en compétition entre les microorganismes sélectionnerait l'évolution de microorganismes capables de produire des d'antibiotiques naturels susceptibles de contourner les mécanismes de résistance aux antibiotiques actuel.

Au cours de cette étude, nous nous sommes aperçus que, malgré l'existence de méthodes établies pour la construction des banques d'ADN métagénomiques (Brady, 2007; Simon & Daniel, 2010), de nombreux facteurs influençaient cette étape cruciale des études de métagénomique fonctionnelle. Les facteurs limitants ont été identifiés : l'importance de la déphosphorylation de l'ADN du vecteur, de la déphosphorylation de l'ADN génomique, l'inactivation de la ligase après ligation, du dessalage du produit de ligation et du rapport ADN vecteur/ADN génomique optimal sur l'efficacité de transformation.

De plus, au cours du criblage fonctionnel, nous nous sommes aperçus du défi que représentait cette étape dans les études métagénomiques, et ainsi de la nécessité de développer une nouvelle méthode plus efficace que la méthode de double couche d'agar conventionnelle. De ce fait, une méthode nommée *bacteriospray* permettant le criblage simultané de 5 à 10 fois plus de clones que la

méthode de double couche d'agar a été développée. En effet, une couche homogène de bactéries cibles détectent une plus petite zone d'inhibition produite par les clones transformants.

Ce n'est que par la suite que l'étude de recherche de nouvelles molécules antibiotiques à partir d'ADN de biofilms a pu être débutée, en exploitant nos méthodes optimisées. Notre revendication, que le lieu d'isolement de l'ADN métagénomique détermine le type de molécules recherchées est supportée par d'autres études récentes (Peek *et al.*, 2018; Thies *et al.*, 2016). Néanmoins, nous ne pouvons pas encore conclure que ces lieux favorisent la découverte de nouveaux antibiotiques, bien que nos résultats soutiennent fortement cette hypothèse.

De plus, une stratégie de métagénomique fonctionnelle basée sur la prédiction de la séquence de l'insert en acides aminés en vue d'une synthèse chimique de la molécule prédite a été entreprise. Généralement, dans les études de métagénomiques fonctionnelles, la molécule bioactive est plutôt identifiée par expression hétérologue de l'insert, mais notre approche permet de contourner tous les obstacles liés à l'expression hétérologue tels que ceux de l'excrétion et de la dégradation de la molécule, en plus des problèmes de perte d'activité. Cette stratégie a été nommée par Culligan *et al.*, 2014 "métagénomique synthétique" où la synthèse chimique du gène d'intérêt est priorisée à l'expression hétérologue. La différence entre leur proposition et la nôtre est que dans leur cas, les séquences à étudier sont directement exploitées à partir des bases de données, alors que dans notre approche, les séquences proviennent des inserts de clones ayant déjà un phénotype observé.

Les résultats des criblages fonctionnels réalisés dans cette étude nous ont permis d'une part de se rallier à la littérature sur l'apport de *B. thailandensis* en terme de molécules bioactives (Gupta *et al.*, 2017). Toutefois, nous sommes les

premiers à avoir recherché de nouvelles molécules bioactives à partir de cette bactérie par génomique fonctionnelle. En effet, les analyses des clones actifs à partir de cette banque indiquent la présence de plusieurs protéines hypothétiques qui pourraient être à l'origine de l'activité observée. D'autre part, le criblage fonctionnel à partir de l'ADN de biofilm, a aussi permis d'identifier de nombreux clones intéressants, et de nouveaux modèles de peptides dont certains sont antimicrobiens. Il s'agit du cas du peptide nommé 'peptide-2' présentant un effet négatif sur la croissance de SARM. Néanmoins, certaines modifications de ce peptide seront sans doute nécessaires pour améliorer son activité.

De plus, nous nous sommes aperçus que de déterminer le gène responsable de la molécule bioactive à partir d'un ADN métagénomique très diversifié représentait un autre défi majeur dans les études métagénomiques.

En somme, notre étude nous a permis d'améliorer notre compréhension des études métagénomiques fonctionnelles spécifiques à la recherche de nouvelles molécules antibiotiques.

6.2. DIFFICULTÉS INHÉRENTES À CE PROJET ET PERSPECTIVES

Les BMR pourraient éventuellement tuer plus de personnes que le cancer. C'est ainsi qu'un nouveau rapport sur les dangers des superbactéries résistantes aux antibiotiques stipule que si des mesures importantes ne sont pas prises pour lutter contre ce problème, la résistance aux antibiotiques dont les infections causées par la tuberculose et le SARM pourrait tuer 10 millions de personnes dans le monde, soit un mort toutes les trois secondes en 2050 (World Bank, 2017). Si les antibiotiques perdent de leur efficacité, les chirurgies de routine présenteront des risques plus élevés de mort notamment chez les patients avec un système immunitaire affaibli. Plusieurs stratégies sont employées pour remédier à ce problème de résistance aux antibiotiques. Dans cette thèse, nous avons contribué à

ces avancements en recherchant de nouvelles molécules antibiotiques par métagénomique fonctionnelle.

Il y a plusieurs défis à l'utilisation de l'expression hétérologue d'un ADN exogène en métagénomique fonctionnelle, et ces limites freinent l'utilisation de cette technique très prometteuse. Par exemple, Iqbal, (2016) du laboratoire spécialisé en métagénomique fonctionnelle de Sean Brady, a dédié sa thèse entière à l'identification du meilleur hôte d'expression hétérologue avec des capacités innés à exprimer des groupes de gènes de biosynthèse exogènes.

De nombreux obstacles ont aussi été rencontrés durant cette thèse. Malgré le fait que nous en avons résolu plusieurs, il est important de soulever d'autres problèmes persistants en perspectives pour ceux qui poursuivront des études de métagénomiques fonctionnelles.

Tout d'abord, nous continuons à soutenir l'hypothèse que l'ADN provenant de lieux riches en BMR soit prometteur pour identifier de nouveaux groupes de gènes de biosynthèse de métabolites secondaires. Une analyse de la composition de l'ADN du biofilm de l'hôpital de Montréal (MHW) devrait être effectuée pour différencier l'ADN bactérien de l'ADN humain présent dans l'échantillon. Une étude métagénomique par séquençage serait intéressante à exécuter pour répertorier les différentes BMR, et les gènes de résistance présents dans l'échantillon de ce biofilm. De plus, isoler les BMR de ce biofilm pour évaluer leur niveau de résistance aux antibiotiques serait intéressant à effectuer. Aussi, pour confirmer l'impact de ce choix de milieu pour les études métagénomiques fonctionnelles, d'autres biofilms d'hôpitaux devraient servir de base pour la construction et le criblage de banques d'ADN métagénomiques.

Ensuite, cette étude a montré que la méthode d'extraction d'ADN était la première étape limitante pour obtenir des fragments d'ADN de grande taille et par la suite, que la méthode d'extraction sur gel n'était pas adaptée pour la récupération

des grands fragments d'ADN. Pour une prochaine étude, il serait intéressant de considérer plusieurs méthodes d'extraction d'ADN parallèles pour favoriser une plus grande diversité. De plus, il serait suggéré de construire deux types de banques d'ADN métagénomiques : à petits inserts (insert de 10 kb et moins), pour faciliter l'identification du gène responsable de la production de la molécule bioactive, et à larges inserts (inserts de plus de 10 kb), pour rechercher les plus gros groupes de gènes de type PKS et NRPS. Pour récupérer les fragments d'ADN à larges inserts après digestion, il paraît impératif d'utiliser l'électroélution de l'ADN au lieu de l'extraction sur gel.

Ainsi, nous avons longuement discuté de la dépendance de l'expression hétérologue, de l'hôte d'expression utilisé. Dans cette étude, *E. coli* était le seul hôte d'expression hétérologue. Étant donné que le vecteur pBeloBAC11 est spécifique à *E. coli*, les banques déjà construites ne peuvent être transférées chez d'autres hôtes d'expression hétérologue. En revanche pour une prochaine étude métagénomique fonctionnelle, il serait fortement suggéré d'utiliser un vecteur de type navette idéalement *E. coli-Streptomyces* pour exprimer la même banque chez les deux hôtes d'expression hétérologue pour distinguer les différences d'expression et maximiser la probabilité de découvrir de nouvelles molécules.

Dans un second temps, développer la méthode de criblage fonctionnel par *bacteriospray* était une sérieuse avancée étant donné qu'il semble que ce soit la première fois qu'un aérographe soit utilisé pour vaporiser des bactéries. Plusieurs optimisations ont dû être réalisées (chapitre 2) pour déterminer les conditions optimales de vaporisation. Toutefois, cette méthode a seulement été testée pour la vaporisation de *S. aureus* et de SARM. Il sera intéressant de la développer pour la vaporisation d'autres bactéries cibles telle que *Pseudomonas aeruginosa* qui est aussi une bactérie pour laquelle le développement de nouvelles molécules antibiotiques est urgent.

Dans le contexte des résultats obtenus, il serait intéressant d'étudier en profondeur les clones actifs sélectionnés B1, B2 et B5, issus de la banque génomique de *B. thailandensis*, et le clone MHW₃128 issu de la banque d'ADN du biofilm de l'hôpital de Montréal, notamment par restriction sous clonage de l'insert dans un plasmide inductible à multicopies au lieu d'utiliser le vecteur pBeloBAC11 originalement choisi pour exprimer les grands inserts.

Tester préalablement l'effet de ces clones transformants contre d'autres bactéries cibles permettra aussi de déterminer le potentiel de la molécule bioactive afin de sélectionner ceux dont la molécule agit à plus large spectre pour des études plus poussées.

Pour les clones MHW₁23, MHW₁64 et MHW₂105, MHW₂300, dont l'activité n'a pas pu être confirmée par retransformation du vecteur chez une autre souche de *E. coli* DH10B, avant de les exclure, d'autres contextes d'expression hétérologues seraient à considérer, comme la restriction sous clonage dans un autre hôte d'expression hétérologue. Pour le clone MHW₂178 qui ne possède pas d'insert mais dont l'activité est reproduite après la première étape de confirmation de l'activité, le séquençage complet de la souche *E. coli* DH10B du clone nous renseignera sur la modification possible de cette souche.

Finalement, selon l'ensemble des observations réalisées concernant le peptide-2, ayant un faible effet sur la croissance de SARM et de *Mycobacterium smegmatis*, nous proposons un ensemble d'améliorations pour augmenter son efficacité. Mais avant cela, l'expression hétérologue des deux séquences ORF1 et ORF2 devrait être réalisée à l'aide d'un promoteur plus faible que le promoteur T7 (Wang *et al.*, 2019; Wei *et al.*, 2005).

Par la suite, le peptide ORFM devrait être synthétisé et testé, pour évaluer l'impact de la substitution des cinq acides aminés identifiés. Par la suite le pourcentage élevé en prolines-arginines de ce peptide-2 et la confirmation de l'absence de structure de ce peptide par RMN nous laisse croire que ce dernier appartiendrait à la famille des prolines-arginines riche tout comme le peptide non structuré arasine-

1. En effet, étant donné que l'activité de cette famille de peptides n'est pas basée sur la lyse de la paroi bactérienne mais plutôt au niveau intracellulaire, ceci pourrait expliquer l'action du peptide-2 à forte concentration. La difficulté de pénétration de ce peptide pourrait être associée à sa faible hydrophobicité. Pour confirmer cette hypothèse, certaines modifications peuvent être envisagées pour accroître le pouvoir hydrophobe ou la stabilité de ce peptide. Par exemple, cycliser ce peptide pourrait réduire la dégradation de ce dernier par les protéases bactériennes et ainsi améliorer son activité. Par la suite, une étude de relation structure-activité, telle que des délétions successives en C- et N-terminal, permettra d'identifier les séquences minimales nécessaires reproduisant l'activité sur la croissance de SARM afin de diminuer les coûts de production de ce peptide. De plus, des bibliothèques d'analogues du peptide-2 pourront être construites, par la méthodologie *Ala-scan* et *D-scan* afin d'identifier les déterminants responsables de l'amélioration des propriétés et de l'efficacité de ce peptide. Et pour finir, si les peptides dérivés ont un potentiel de développement pharmacologique, des études de stabilité du/des peptides dérivés de l'optimisation du 'peptide-2' devront être réalisées pour établir leur(s) propriété(s) pharmacologique(s) par marquage fluorescent et radioactif *in vivo*.

Pour finir, les autres ORF non testés ORF3, ORF4, ORF7 et ORF8 devraient être synthétisés et leur activité devrait être évaluée.

En résumé, ces études d'amélioration proposées pour le peptide-2 pourront mener à la découverte d'un nouveau peptide antimicrobien actif contre SARM. Étant donné l'effet de ce peptide sur la croissance de *Mycobacterium smegmatis*, des tests d'activité notamment contre *Mycobacterium tuberculosis* pourront aussi être envisagés car peu de molécules de source naturelle sont connues pour agir spécifiquement contre cet agent pathogène. Toutefois, pour atteindre l'objectif ultime qui consiste en la conception d'un nouveau peptide antimicrobien, diverses phases d'évaluation devront être réalisées pour identifier la molécule répondant aux exigences de l'industrie pharmaceutique, notamment une évaluation de sa toxicité contre les cellules humaines.

6.3. PERSPECTIVES D'AVENIR EN MÉTAGÉNOMIQUE FONCTIONNELLE

La métagénomique fonctionnelle permet d'accéder aux molécules bioactives produites par les microorganismes sans devoir les cultiver, grâce à l'expression hétérologue d'un ADN environnemental, ce qui évite de devoir recourir aux méthodes classiques de culture au laboratoire qui ne sont pas représentatives de la communauté microbienne. Néanmoins, en analysant la littérature sur l'utilisation de cette technique, la découverte de nouvelles molécules antibiotiques reste lente comparativement à la découverte de nouvelles enzymes (Cheng *et al.*, 2017; Silva-Portela *et al.*, 2016; Zhang *et al.*, 2017), car un grand nombre de clones doivent être testés afin d'obtenir un seul clone actif, c'est-à-dire présentant une zone d'inhibition détectable. La nouvelle méthode de criblage fonctionnel par *bacteriospray* développée au cours de cette thèse a le potentiel d'apporter une plus grande utilisation de cette approche de criblage. Avec cette technique, un plus grand nombre de clones peut être testé, en plus de la possibilité de détecter de plus petites zones d'inhibition. Mais, à la base, l'une des raisons pour laquelle le nombre de clones à tester est important réside dans le fait que les groupes de gènes de biosynthèse de type PKS et NRPS ne peuvent être insérés dans un seul clone. En effet, Feng *et al.*, 2010 ont pu reconstituer un groupe de gènes de type PKS à partir de deux clones sélectionnés d'un criblage de clones de banques d'ADN métagénomiques avec des amorces de type PKS. Cela leur a permis de découvrir le nouvel antibiotique fluostatine (Feng *et al.*, 2010). Dans cette optique, l'antibiotique fassamycine A et B actif contre SARM et *E. faecalis* résistant à la vancomycine a été identifié en réassemblant le groupe de gènes complets à partir de plusieurs clones (Feng *et al.*, 2012; Feng *et al.*, 2011). Dans ce contexte, sélectionner ces groupes de gènes à l'aide d'amorces spécifiques, directement à partir d'un échantillon d'ADN, permettrait de cibler directement ces gènes aux activités potentiellement nouvelles pour la construction de banques d'ADN métagénomiques. Cette stratégie a déjà été employée et a permis de découvrir de

nombreuses nouvelles molécules : l'hydroxisporine et la réductasporine (Chang *et al.*, 2015), la malacidine A B (Hover *et al.*, 2018), et la kanglemycin A (Mosaei *et al.*, 2018).

On peut donc constater que les études métagénomiques fonctionnelles seules se développent plus lentement que les études métagénomiques basées sur la recherche de séquences d'intérêts. Néanmoins, déterminer la fonction de gènes d'intérêt par expression hétérologue et criblage fonctionnel demeure toujours d'actualité. Il est donc impératif de développer des méthodes performantes pour améliorer l'expression hétérologue.

CHAPITRE 7 : BIBLIOGRAPHIE

Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Volker U, Ellingsen TE & Valla S (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiol. Lett.* 296(2):149-158.

Alduina R & Gallo G (2012a) Artificial Chromosomes to Explore and to Exploit Biosynthetic Capabilities of Actinomycetes. *J. Biomed. Biotechnol.* 2012:1-10.

Alduina R & Gallo G (2012b) Artificial chromosomes to explore and to exploit biosynthetic capabilities of actinomycetes. *J. Biomed. Biotechnol.* 2012:462049.

Ali MM, Aburowes AH, Albakush AM, Rzeg MM, Alrtail A & Ghenghesh KS (2014) Identification of multidrug-resistant bacteria and *Bacillus cereus* from healthcare workers and environmental surfaces in a hospital. *Libyan J. Med.* 9:25794.

Allen HK, Moe LA, Rodbumrer J, Gaarder A & Handelsman J (2009) Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *The ISME journal* 3(2):243-251.

Amiche M, Seon AA, Wroblewski H & Nicolas P (2000) Isolation of dermatoxin from frog skin, an antibacterial peptide encoded by a novel member of the dermaseptin genes family. *Eur. J. Biochem.* 267(14):4583-4592.

Andrushchenko VV, Vogel HJ & Prenner EJ (2006) Solvent-dependent structure of two tryptophan-rich antimicrobial peptides and their analogs studied by FTIR and CD spectroscopy. *Biochim. Biophys. Acta* 1758(10):1596-1608.

Angelov A, Pham VTT, Ubelacker M, Brady S, Leis B, Pill N, Brolle J, Mechelke M, Moerch M, Henrissat B & Liebl W (2017) A metagenome-derived thermostable beta-glucanase with an unusual module architecture which defines the new glycoside hydrolase family GH148. *Sci. Rep.* 7(1):17306.

Arvani S, Markert A, Loeschcke A, Jaeger KE & Drepper T (2012) A T7 RNA polymerase-based toolkit for the concerted expression of clustered genes. *J. Biotechnol.* 159(3):162-171.

Bahar AA & Ren D (2013) Antimicrobial peptides. *Pharmaceuticals (Basel)* 6(12):1543-1575.

Bahnsen JS, Franzyk H, Sandberg-Schaal A & Nielsen HM (2013) Antimicrobial and cell-penetrating properties of penetratin analogs: effect of sequence and secondary structure. *Biochim. Biophys. Acta* 1828(2):223-232.

Banik JJ & Brady SF (2008) Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc. Natl. Acad. Sci. U. S. A.* 105(45):17273-17277.

Banik JJ & Brady SF (2010) Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opin. Microbiol.* 13(5):603-609.

Barrett E, Hayes M, O'Connor P, Gardiner G, Fitzgerald GF, Stanton C, Ross RP & Hill C (2007) Salivaricin P, one of a family of two-component antilisterial bacteriocins produced by intestinal isolates of *Lactobacillus salivarius*. *Appl. Environ. Microbiol.* 73(11):3719-3723.

Bentley S, Chater K, Cerdeno-Tarraga A-M, Challis G, Thomson N, James K, Harris D, Quail M, Kieser H & Harper D (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* 417(6885):141-147.

Bierbaum G & Sahl HG (2014) The search for new anti-infective drugs: untapped resources and strategies. *Int. J. Med. Microbiol.* 304(1):1-2.

Bisant D & Maizel J (1995) Identification of ribosome binding sites in *Escherichia coli* using neural network models. *Nucleic Acids Res.* 23(9):1632-1639.

Bishop BM, Juba ML, Devine MC, Barksdale SM, Rodriguez CA, Chung MC, Russo PS, Vliet KA, Schnur JM & van Hoek ML (2015) Bioprospecting the American alligator (*Alligator mississippiensis*) host defense peptidome. *PLoS One* 10(2):e0117394.

Biver S, Steels S, Portetelle D & Vandenberg M (2013) *Bacillus subtilis* as a tool for screening soil metagenomic libraries for antimicrobial activities. *J. Microbiol. Biotechnol.* 23(6):850-855.

Blaskovich MA, Butler MS & Cooper MA (2017) Polishing the tarnished silver bullet: the quest for new antibiotics. *Essays Biochem.* 61(1):103-114.

Bleich R, Watrous JD, Dorrestein PC, Bowers AA & Shank EA (2015) Thiopeptide antibiotics stimulate biofilm formation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* 112(10):3086-3091.

Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B & Bartlett J (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* 48(1):1-12.

Brady SF (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat. Protoc.* 2(5):1297-1305.

Brady SF, Chao CJ & Clardy J (2002) New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* 124(34):9968-9969.

Brady SF, Chao CJ & Clardy J (2004) Long-chain N-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* 70(11):6865-6870.

Brady SF, Chao CJ, Handelsman J & Clardy J (2001a) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* 3(13):1981-1984.

Brady SF, Chao CJ, Handelsman J & Clardy J (2001b) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org Lett* 3(13):1981-1984.

Brady SF & Clardy J (2000) Long-Chain N-Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental DNA. *J. Am. Chem. Soc.* 122(51):12903-12904.

Brady SF & Clardy J (2004) Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J. Nat. Prod.* 67(8):1283-1286.

Brady SF & Clardy J (2005a) Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. *Angew. Chem.* 117(43):7225-7227.

Brady SF & Clardy J (2005b) N-Acyl Derivatives of Arginine and Tryptophan Isolated from Environmental DNA Expressed in *Escherichia coli*. *Org. Lett.* 7(17):3613-3616.

Brady SF, Wright SA, Lee JC, Sutton AE, Zumoff CH, Wodzinski RS, Beer SV & Clardy J (1999) Pantocin B, an antibiotic from *Erwinia herbicola* discovered by heterologous expression of cloned genes. *J. Am. Chem. Soc.* 121(50):11912-11913.

Brahami A, Castonguay A & Déziel É (2019) Novel 'Bacteriospray' Method Facilitates the Functional Screening of Metagenomic Libraries for Antimicrobial Activity. *Methods and Protocols* 2(1):4.

Brogden NK & Brogden KA (2011) Will new generations of modified antimicrobial peptides improve their potential as pharmaceuticals? *Int. J. Antimicrob. Agents* 38(3):217-225.

Brooks AN, Turkarslan S, Beer KD, Lo FY & Baliga NS (2011) Adaptation of cells to new environments. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3(5):544-561.

Brown S, Santa Maria JP, Jr. & Walker S (2013) Wall teichoic acids of gram-positive bacteria. *Annu. Rev. Microbiol.* 67:313-336.

Buelow E, Bayjanov JR, Willems RJL, Bonten MJM, Schmitt H & Van Schaik W (2017) The microbiome and resistome of hospital sewage during passage through the community sewer system 10.1101/216242.

Bürgmann H, Pesaro M, Widmer F & Zeyer J (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. *J. Microbiol. Methods* 45(1):7-20.

Cardenas E & Tiedje JM (2008) New tools for discovering and characterizing microbial diversity. *Curr. Opin. Biotechnol.* 19(6):544-549.

Carmona G, Rodriguez A, Juarez D, Corzo G & Villegas E (2013) Improved protease stability of the antimicrobial peptide Pin2 substituted with D-amino acids. *Protein J.* 32(6):456-466.

Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Dore J, Henrissat B, Remaud-Simeon M, Monsan P & Potocki-Veronese G (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* 8(9):72766.

Chakchouk-Mtibaa A, Elleuch L, Smaoui S, Najah S, Sellem I, Abdelkafi S & Mellouli L (2014) An antilisterial bacteriocin BacFL31 produced by *Enterococcus faecium* FL31 with a novel structure containing hydroxyproline residues. *Anaerobe* 27:1-6.

Chang AY, Chau V, Landas JA & Pang Y (2017) Preparation of calcium competent *Escherichia coli* and heat-shock transformation. *JEMI Methods* 1:22-25.

Chang F-Y & Brady SF (2013a) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proceedings of the National Academy of Sciences* 110(7):2478-2483.

Chang FY & Brady SF (2013b) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc. Natl. Acad. Sci. U. S. A.* 110(7):2478-2483.

Chang FY, Ternei MA, Calle PY & Brady SF (2013) Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment. *J. Am. Chem. Soc.* 135(47):17906-17912.

Chang FY, Ternei MA, Calle PY & Brady SF (2015) Targeted metagenomics: finding rare tryptophan dimer natural products in the environment. *J. Am. Chem. Soc.* 137(18):6044-6052.

Chen T, Tang L & Shaw C (2003) Identification of three novel *Phyllomedusa sauvagei* dermaseptins (sVI–sVIII) by cloning from a skin secretion-derived cDNA library. *Regulatory Peptides* 116(1-3):139-146.

Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD & Charles TC (2017) Functional metagenomics reveals novel beta-galactosidases not predictable from gene sequences. *PLoS One* 12(3):e0172545.

Chu J, Vila-Farres X, Inoyama D, Ternei M, Cohen LJ, Gordon EA, Reddy BVB, Charlop-Powers Z, Zebroski HA & Gallardo-Macias R (2016) Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat. Chem. Biol.* 12(12):1004-1006.

Chu X, He H, Guo C & Sun B (2008) Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl. Microbiol. Biotechnol.* 80(4):615-625.

Clardy J, Fischbach MA & Walsh CT (2006) New antibiotics from bacterial natural products. *Nat. Biotechnol.* 24(12):1541-1550.

Clinical Laboratory Standards Institute (2012) Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically : approved standard - ninth edition M07-A9. (CLSI, Wayne, Pa.), p 68.

Cohen LJ, Han S, Huang YH & Brady SF (2018) Identification of the Colicin V Bacteriocin Gene Cluster by Functional Screening of a Human Microbiome Metagenomic Library. *ACS Infect Dis* 4(1):27-32.

Colman A & Burley JC (2001) A legal and ethical tightrope. *EMBO reports* 2(1):2-5.

Conrad B, Savchenko RS, Breves R & Hofemeister J (1996) A T7 promoter-specific, inducible protein expression system for *Bacillus subtilis*. *Mol. Gen. Genet.* 250(2):230-236.

Control CfD & Prevention (2013) *Antibiotic resistance threats in the United States, 2013*. Centres for Disease Control and Prevention, US Dep. of Health and Human Services.

Cooper MA & Shlaes D (2011) Fix the antibiotics pipeline. *Nature* 472(7341):32.

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osborne MS, August PR, Nalin R, Guerineau M, Jeannin P, Simonet P & Pernodet JL (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* 69(1):49-55.

Craig JW, Chang FY & Brady SF (2009) Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chem. Biol.* 4(1):23-28.

Craig JW, Chang FY, Kim JH, Obiajulu SC & Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* 76(5):1633-1641.

Culligan EP, Sleator RD, Marchesi JR & Hill C (2014) Metagenomics and novel gene discovery: promise and potential for novel therapeutics. *Virulence* 5(3):399-412.

D'Onofrio A, Crawford JM, Stewart EJ, Witt K, Gavrish E, Epstein S, Clardy J & Lewis K (2010) Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17(3):254-264.

Daegelen P, Studier FW, Lenski RE, Cure S & Kim JF (2009) Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394(4):634-643.

Daniel R (2004) The soil metagenome a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15(3):199-204.

Davies J & Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74(3):417-433.

De Jong A, van Hijum SA, Bijlsma JJ, Kok J & Kuipers OP (2006) BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res.* 34(Web Server issue):W273-279.

Delbruck M (1945) Interference Between Bacterial Viruses: III. The Mutual Exclusion Effect and the Depressor Effect. *J. Bacteriol.* 50(2):151-170.

DeLeo FR, Otto M, Kreiswirth BN & Chambers HF (2010) Community-associated meticillin-resistant *Staphylococcus aureus*. *Lancet* 375(9725):1557-1568.

Delmont TO, Robe P, Clark I, Simonet P & Vogel TM (2011) Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J. Microbiol. Methods* 86(3):397-400.

Demain AL & Sanchez S (2009) Microbial drug discovery: 80 years of progress. *J. Antibiot. (Tokyo)* 62(1):5-16.

Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M & Mullany P (2003) Novel Tetracycline Resistance Determinant from the Oral Metagenome. *Antimicrob. Agents Chemother.* 47(4):1430-1432.

Feng Z, Chakraborty D, Dewell SB, Reddy BV & Brady SF (2012) Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis. *J. Am. Chem. Soc.* 134(6):2981-2987.

Feng Z, Kallifidas D & Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. U. S. A.* 108(31):12629-12634.

Feng Z, Kim JH & Brady SF (2010) Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J. Am. Chem. Soc.* 132(34):11902-11903.

Ferrand J, Patron K, Legrand-Frossi C, Fripiat JP, Merlin C, Alauzet C & Lozniewski A (2014) Comparison of seven methods for extraction of bacterial DNA from fecal and cecal samples of mice. *J. Microbiol. Methods* 105:180-185.

Fleming D, Chahin L & Rumbaugh K (2017) Glycoside Hydrolases Degrade Polymicrobial Bacterial Biofilms in Wounds. *Antimicrob. Agents Chemother.* 61(2).

Florez AB, Vazquez L & Mayo B (2017) A Functional Metagenomic Analysis of Tetracycline Resistance in Cheese Bacteria. *Front. Microbiol.* 8:907.

Flynn S, van Sinderen D, Thornton GM, Holo H, Nes IF & Collins JK (2002) Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. *salivarius* UCC118. *Microbiology* 148(4):973-984.

Gabor EM, Alkema WB & Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* 6(9):879-886.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM & Handelsman J (2002) Isolation of Antibiotics Turbomycin A and B from a Metagenomic Library of Soil Microbial DNA. *Appl. Environ. Microbiol.* 68(9):4301-4306.

Gillings M (2015) How antibiotic pollution of waterways creates superbugs. in *The Conversation USA*.

Graf M, Mardirossian M, Nguyen F, Seefeldt AC, Guichard G, Scocchi M, Innis CA & Wilson DN (2017) Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* 34(7):702-711.

Grossman TH, Kawasaki ES, Punreddy SR & Osburne MS (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* 209(1):95-103.

Gupta A, Bedre R, Thapa SS, Sabrin A, Wang G, Dassanayake M & Grove A (2017) Global Awakening of Cryptic Biosynthetic Gene Clusters in *Burkholderia thailandensis*. *ACS Chem. Biol.* 12(12):3012-3021.

Handelsman J (2007) Metagenomics and microbial communities. *eLS*.

Handelsman J, Liles M, Mann D, Riesenfeld C & Goodman RM (2002) 13 Cloning the Metagenome: Culture-independent Access to the Diversity and Functions of the Uncultivated Microbial World. *Methods in Microbiology* (3) p 241-255.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998a) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5(10):R245-R249.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998b) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology* 5(10):R245-249.

Harris G (2014) Superbugs kill India's babies and pose an overseas threat. *The New York Times* (New York), World-Asian Pacific, p 1.

Holder I & Boyce S (1994) Agar well diffusion assay testing of bacterial susceptibility to various antimicrobials in concentrations non-toxic for human cells in culture. *Burns* 20(5):426-429.

Hover BM, Kim SH, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB, Molina H, Park S, Perlin DS & Brady SF (2018) Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol* 10.1038/41564-018-0110-1.

Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, Robins-Browne R, Davies JK, Seemann T

& Stinear TP (2011) Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Pathog.* 7(11):1002359.

Huang Y, Huang J & Chen Y (2010) Alpha-helical cationic antimicrobial peptides: relationships of structure and function. *Protein & cell* 1(2):143-152.

Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 3(2):reviews0003. 0001.

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M & Omura S (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* 21(5):526-531.

Imjongjirak C, Amparyup P & Tassanakajon A (2011) Two novel antimicrobial peptides, arasin-likeSp and GRPSp, from the mud crab *Scylla paramamosain*, exhibit the activity against some crustacean pathogenic bacteria. *Fish Shellfish Immunol.* 30(2):706-712.

Imjongjirak C, Amparyup P, Tassanakajon A & Sittipraneed S (2009) Molecular cloning and characterization of crustin from mud crab *Scylla paramamosain*. *Mol. Biol. Rep.* 36(5):841-850.

Iqbal H (2016) Natural Products From Functional Screening of Soil Metagenomic Libraries. *J. Am. Chem. Soc.* 138, 9341–9344.

Iqbal HA, Craig JW & Brady SF (2014) Antibacterial enzymes from the functional screening of metagenomic libraries hosted in *Ralstonia metallidurans*. *FEMS Microbiol. Lett.* 354(1):19-26.

Iqbal HA, Low-Beinart L, Obiajulu JU & Brady SF (2016) Natural Product Discovery through Improved Functional Metagenomics in *Streptomyces*. *J. Am. Chem. Soc.* 138(30):9341-9344.

Jachlewski S, Jachlewski WD, Linne U, Brasen C, Wingender J & Siebers B (2015) Isolation of Extracellular Polymeric Substances from Biofilms of the Thermoacidophilic Archaeon *Sulfolobus acidocaldarius*. *Front Bioeng Biotechnol* 3:123.

Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Segurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P & Kim JF (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394(4):644-652.

Jiang L, Wang L, Zhang J, Liu H, Hong B, Tan H & Niu G (2015) Identification of novel mureidomycin analogues via rational activation of a cryptic gene cluster in *Streptomyces roseosporus* NRRL 15998. *Sci. Rep.* 5:14111.

Joo SH (2012) Cyclic peptides as therapeutic agents and biochemical tools. *Biomol. Ther. (Seoul)* 20(1):19-26.

Kaeberlein T, Lewis K & Epstein SS (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296(5570):1127-1129.

Kakirde KS, Parsley LC & Liles MR (2010) Size Does Matter: Application-driven Approaches for Soil Metagenomics. *Soil Biol. Biochem.* 42(11):1911-1923.

Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W & Liles MR (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene* 475(2):57-62.

Kang HS & Brady SF (2014) Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. *ACS Chem. Biol.* 9(6):1267-1272.

Katouli M, Thompson JM, Gündoğdu A & Stratton HM (2012) Antibiotic Resistant Bacteria in Hospital Wastewaters and Sewage Treatment Plants. in *Science Forum and Stakeholder Engagement: Building Linkages, Collaboration and Science Quality (UWSRA 4th Science Forum)*, Édité Begbie D.K. KSJ, Biermann S.M. And Wakem S.L. (Urban Water Security Research Alliance, Brisbane, Australia), p 223-229.

Kim H, Jang JH, Kim SC & Cho JH (2014) De novo generation of short antimicrobial peptides with enhanced stability and cell specificity. *J. Antimicrob. Chemother.* 69(1):121-132.

Kim SJ, Chang J & Singh M (2015) Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR. *Biochim. Biophys. Acta* 1848(1 Pt B):350-362.

Kim U-J, Birren BW, Slepak T, Mancino V, Boysen C, Kang H-L, Simon MI & Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* 34(2):213-218.

King RW, Bauer JD & Brady SF (2009) An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. *Angew. Chem. Int. Ed. Engl.* 48(34):6257-6261.

Kodama T, Endo K, Ara K, Ozaki K, Kakeshita H, Yamane K & Sekiguchi J (2007) Effect of *Bacillus subtilis* spo0A mutation on cell wall lytic enzymes and extracellular proteases, and prevention of cell lysis. *J. Biosci. Bioeng.* 103(1):13-21.

KPMG LLP (2014) The global economic impact of anti-microbial resistance. *KPMG International, UK.* p 41.

Krsek M & Wellington EMH (1999) Comparison of different methods for the isolation and purification of total community DNA from soil. *J. Microbiol. Methods* 39(1):1-16.

Kudla G, Murray AW, Tollervey D & Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324(5924):255-258.

Lam KN, Cheng J, Engel K, Neufeld JD & Charles TC (2015) Current and future resources for functional metagenomics. *Front. Microbiol.* 6.

Lazo N & Downing DT (1997) Circular dichroism of model peptides emulating the amphipathic α -helical regions of intermediate filaments. *Biochemistry* 36(9):2559-2565.

Le TH, Ng C, Chen H, Yi XZ, Koh TH, Barkham TM, Zhou Z & Gin KY (2016) Occurrences and Characterization of Antibiotic-Resistant Bacteria and Genetic Determinants of Hospital Wastewater in a Tropical Country. *Antimicrob. Agents Chemother.* 60(12):7449-7456.

Leis B, Angelov A, Mientus M, Li H, Pham VT, Lauinger B, Bongen P, Pietruszka J, Goncalves LG, Santos H & Liebl W (2015) Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front. Microbiol.* 6:275.

Lewis K (2012) Antibiotics: recover the lost art of drug discovery. *Nature* 485(7399):439.

Lewis K (2013) Platforms for antibiotic discovery. *Nature reviews Drug discovery* 12(5):371.

Li JW-H & Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? *Science* 325(5937):161-165.

Li JX, Zhao LM, Wu RJ, Zheng ZJ & Zhang RJ (2013) High-level overproduction of Thermobifida enzyme in *Streptomyces lividans* using a novel expression vector. *Int. J. Mol. Sci.* 14(9):18629-18639.

Liles MR, Manske BF, Bintrim SB, Handelsman J & Goodman RM (2003) A Census of rRNA Genes and Linked Genomic Sequences within a Soil Metagenomic Library. *Appl. Environ. Microbiol.* 69(5):2684-2691.

Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY & Lee SW (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.* 71(12):7768-7777.

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schaberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C & Lewis K (2015) A new antibiotic kills pathogens without detectable resistance. *Nature* 10.1038/14098.

Liu X & Cheng YQ (2014) Genome-guided discovery of diverse natural products from Burkholderia sp. *J. Ind. Microbiol. Biotechnol.* 41(2):275-284.

Liu X, Liu L, Wang Y, Wang X, Ma Y & Li Y (2014) The study on the factors affecting transformation efficiency of *E. coli* competent cells. *Cell* 5:106.

Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu L-F, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu J-H & Shen J (2016) Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 16(2):161-168.

Lood R, Erturk G & Mattiasson B (2017) Revisiting Antibiotic Resistance Spreading in Wastewater Treatment Plants - Bacteriophages as a Much Neglected Potential Transmission Vehicle. *Front. Microbiol.* 8:2298.

Lopez PJ, Marchand I, Joyce SA & Dreyfus M (1999) The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol. Microbiol.* 33(1):188-199.

Lussier FX, Chambenoit O, Cote A, Hupe JF, Denis F, Juteau P, Beaudet R & Shareck F (2011) Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector. *J. Ind. Microbiol. Biotechnol.* 38(9):1321-1328.

Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 25(6):1203-1210.

MacNeil I, Tiong C, Minor C, August P, Grossman T, Loiacono K, Lynch B, Phillips T, Narula S & Sundaramoorthi R (2001a) Expression and isolation of

antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* 3(2):301-308.

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D & Osburne MS (2001b) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* 3(2):301-308.

Mahlapuu M, Hakansson J, Ringstad L & Bjorn C (2016) Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Front Cell Infect Microbiol* 6:194.

Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA & Osburne MS (2004) Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Appl. Environ. Microbiol.* 70(4):2452-2463.

Martiny AC, Martiny JB, Weihe C, Field A & Ellis JC (2011) Functional metagenomics reveals previously unrecognized diversity of antibiotic resistance genes in gulls. *Front. Microbiol.* 2:238.

Mary RZa & Betlach C (1999) Recombinant polyketide synthesis in *Streptomyces*: engineering of improved host strains. *BioTechniques* 26:106-110.

McKenzie NL, Thaker M, Koteva K, Hughes DW, Wright GD & Nodwell JR (2010) Induction of antimicrobial activities in heterologous streptomycetes using alleles of the *Streptomyces coelicolor* gene *absA1*. *J. Antibiot. (Tokyo)* 63(4):177-182.

Mocali S & Benedetti A (2010) Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res. Microbiol.* 161(6):497-505.

Mosaei H, Molodtsov V, Kepplinger B, Harbottle J, Moon CW, Jeeves RE, Ceccaroni L, Shin Y, Morton-Laing S, Marrs ECL, Wills C, Clegg W, Yuzenkova Y, Perry JD, Bacon J, Errington J, Allenby NEE, Hall MJ, Murakami KS & Zenkin N (2018) Mode of Action of Kanglemycin A, an Ansamycin Natural Product that Is Active against Rifampicin-Resistant *Mycobacterium tuberculosis*. *Mol. Cell* 72(2):263-274 e265.

Mullany P (2014) Functional metagenomics for the investigation of antibiotic resistance. *Virulence* 5(3):443-447.

Oman TJ & van der Donk WA (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* 6(1):9-18.

Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH & Brady SF (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proceedings of the National Academy of Sciences* 110(29):11797-11802.

Park CB, Yi KS, Matsuzaki K, Kim MS & Kim SC (2000) Structure-activity analysis of buforin II, a histone H2A-derived antimicrobial peptide: the proline hinge is responsible for the cell-penetrating ability of buforin II. *Proc. Natl. Acad. Sci. U. S. A.* 97(15):8245-8250.

Pavlova O, Mukhopadhyay J, Sineva E, Ebright RH & Severinov K (2008) Systematic structure-activity analysis of microcin J25. *J. Biol. Chem.* 283(37):25589-25595.

Peek J, Lilic M, Montiel D, Milshteyn A, Woodworth I, Biggins JB, Ternei MA, Calle PY, Danziger M, Warriar T, Saito K, Braffman N, Fay A, Glickman MS, Darst SA, Campbell EA & Brady SF (2018) Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nature communications* 9(1):4147.

Poirel L, Kieffer N, Liassine N, Thanh D & Nordmann P (2016) Plasmid-mediated carbapenem and colistin resistance in a clinical isolate of *Escherichia coli*. *The Lancet Infectious Diseases* 16(3).

Rabausch U, Juergensen J, Ilmberger N, Bohnke S, Fischer S, Schubach B, Schulte M & Streit WR (2013) Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl. Environ. Microbiol.* 79(15):4551-4563.

Ranjan R, Grover A, Kapardar RK & Sharma R (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochem. Biophys. Res. Commun.* 335(1):57-65.

Rice LB (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J. Infect. Dis.* 197(8):1079-1081.

Riesenfeld CS, Goodman RM & Handelsman J (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* 6(9):981-989.

Robe P, Nalin R, Capellano C, Vogel TM & Simonet P (2003) Extraction of DNA from soil. *Eur. J. Soil Biol.* 39(4):183-190.

Roberts RR, Hota B, Ahmad I, Scott RD, Foster SD, Abbasi F, Schabowski S, Kampe LM, Ciavarella GG & Supino M (2009) Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin. Infect. Dis.* 49(8):1175-1184.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA & Minor C (2000a) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66(6):2541-2547.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J & Goodman RM (2000b) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66(6):2541-2547.

Rondon MR, Raffel SJ, Goodman RM & Handelsman J (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *Proceedings of the National Academy of Sciences* 96(11):6451-6455.

Sabree ZL, Rondon MR & Handelsman J (2009) Metagenomics. *Encyclopedia of Microbiology (Third Edition)*, Schaechter M (Édit.) Academic Press, Oxford <http://dx.doi.org/10.1016/B978-012373944-5.00034-1>. p 622-632.

Sabri NA, Schmitt H, Van der Zaan B, Gerritsen HW, Zuidema T, Rijnaarts HHM & Langenhoff AAM (2018) Prevalence of antibiotics and antibiotic resistance genes in a wastewater effluent-receiving river in the Netherlands. *Journal of Environmental Chemical Engineering* 10.1016.

Saier MH (1995) Differential codon usage: a safeguard against inappropriate expression of specialized genes? *FEBS Lett.* 362(1):1-4.

Saraswat M, Grand RS & Patrick WM (2013) Desalting DNA by drop dialysis increases library size upon transformation. *Biosci. Biotechnol. Biochem.* 77(2):402-404.

Schloss PD & Handelsman J (2004) Status of the microbial census. *Microbiol. Mol. Biol. Rev.* 68(4):686-691.

Schneider D, Duperchy E, Depeyrot J, Coursange E, Lenski R & Blot M (2002) Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiol.* 2:18.

Seyedsayamdost MR (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* 111(20):7266-7271.

Silva-Portela RC, Carvalho FM, Pereira CP, de Souza-Pinto NC, Modesti M, Fuchs RP & Agnez-Lima LF (2016) ExoMeg1: a new exonuclease from metagenomic library. *Sci. Rep.* 6:19712.

Simon C & Daniel R (2010) Construction of small-insert and large-insert metagenomic libraries. *Methods Mol. Biol.* 668:39-50.

Simon C & Daniel R (2011) Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* 77(4):1153-1161.

Simon C, Herath J, Rockstroh S & Daniel R (2009) Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl. Environ. Microbiol.* 75(9):2964-2968.

Sosio M, Giusino F, Cappellano C, Bossi E, Puglia AM & Donadio S (2000) Artificial chromosomes for antibiotic-producing actinomycetes. *Nat. Biotechnol.* 18(3):343-345.

Sperstad SV, Haug T, Vasskog T & Stensvag K (2009) Hyastatin, a glycine-rich multi-domain antimicrobial peptide isolated from the spider crab (*Hyas araneus*) hemocytes. *Mol. Immunol.* 46(13):2604-2612.

Steffensky M, Muhlenweg A, Wang ZX, Li SM & Heide L (2000) Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. *Antimicrob. Agents Chemother.* 44(5):1214-1222.

Stensvag K, Haug T, Sperstad SV, Rekdal O, Indrevoll B & Styrvold OB (2008) Arasin 1, a proline-arginine-rich antimicrobial peptide isolated from the spider crab, *Hyas araneus*. *Dev. Comp. Immunol.* 32(3):275-285.

Studier FW, Daegelen P, Lenski RE, Maslov S & Kim JF (2009) Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes. *J. Mol. Biol.* 394(4):653-680.

Studier FW & Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189(1):113-130.

Szabo D, Ostorhazi E, Binas A, Rozgonyi F, Kocsis B, Cassone M, Wade JD, Nolte O & Otvos L, Jr. (2010) The designer proline-rich antibacterial peptide A3-APO is effective against systemic *Escherichia coli* infections in different mouse models. *Int. J. Antimicrob. Agents* 35(4):357-361.

Tamaki H, Sekiguchi Y, Hanada S, Nakamura K, Nomura N, Matsumura M & Kamagata Y (2005) Comparative analysis of bacterial diversity in freshwater

sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. *Appl. Environ. Microbiol.* 71(4):2162-2169.

Tao Q & Zhang HB (1998) Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. *Nucleic Acids Res.* 26(21):4901-4909.

Tatum EL & Lederberg J (1947) Gene Recombination in the Bacterium *Escherichia coli*. *J. Bacteriol.* 53(6):673-684.

Thies S, Rausch SC, Kovacic F, Schmidt-Thaler A, Wilhelm S, Rosenau F, Daniel R, Streit W, Pietruszka J & Jaeger KE (2016) Metagenomic discovery of novel enzymes and biosurfactants in a slaughterhouse biofilm microbial community. *Sci. Rep.* 6:27035.

Toda H, Williams JA, Gullledge M & Sehgal A (2019) A sleep-inducing gene, *nemuri*, links sleep and immune function in *Drosophila*. *Science* 363(6426):509-515.

Torres AM, Bansal P, Koh JM, Pages G, Wu MJ & Kuchel PW (2014) Structure and antimicrobial activity of platypus 'intermediate' defensin-like peptide. *FEBS Lett.* 588(9):1821-1826.

Troeschel SC, Thies S, Link O, Real CI, Knops K, Wilhelm S, Rosenau F & Jaeger KE (2012) Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *J. Biotechnol.* 161(2):71-79.

Trusts PC (2016) A Scientific Roadmap for Antibiotic Discovery: A sustained and robust pipeline of new antibacterial drugs and therapies is critical to preserve public health. Philadelphia), p 42.

Uchiyama T & Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* 20(6):616-622.

Uyaguari MI, Fichot EB, Scott GI & Norman RS (2011) Characterization and quantitation of a novel beta-lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. *Appl. Environ. Microbiol.* 77(23):8226-8233.

Uyterhoeven ET, Butler CH, Ko D & Elmore DE (2008) Investigating the nucleic acid interactions and antimicrobial mechanism of buforin II. *FEBS Lett.* 582(12):1715-1718.

Vercammen K, Garcia-Armisen T, Goeders N, Van Melderens L, Bodilis J & Cornelis P (2013) Identification of a metagenomic gene cluster containing a new

class A beta-lactamase and toxin-antitoxin systems. *Microbiologyopen* 2(4):674-683.

Villegas A & Kropinski AM (2008a) An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. *Microbiology* 154(Pt 9):2559-2661.

Villegas A & Kropinski AM (2008b) An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. *Microbiology* 154(9):2559-2661.

Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ & Davies J (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* 2(16):2401-2404.

Wang L, Hatem A, Catalyurek UV, Morrison M & Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS One* 8(11):78507.

Wang Q, Wang P & Yang Q (2018) Occurrence and diversity of antibiotic resistance in untreated hospital wastewater. *Sci. Total Environ.* 621:990-999.

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Muller R, Wohlleben W, Breitling R, Takano E & Medema MH (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43(W1):W237-243.

Westers L, Westers H & Quax WJ (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta* 1694(1-3):299-310.

Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences* 95(12):6578-6583.

Williamson KE, Kan J, Polson SW & Williamson SJ (2011) Optimizing the indirect extraction of prokaryotic DNA from soils. *Soil Biol. Biochem.* 43(4):736-748.

Wong SL (1995) Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. *Curr. Opin. Biotechnol.* 6(5):517-522.

Wooley JC, Godzik A & Friedberg I (2010) A primer on metagenomics. *PLoS Comput. Biol.* 6(2):1000667.

World Bank (2017) Drug Resistant Infections: A Threat to Our Economic Future. (World Bank, Washington, DC), p 141.

World Health Organization (2017) Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. (WHO, s.l.), p 7.

World Health Organization. (2014) *Antimicrobial resistance : global report on surveillance.(In IRIS)*. World Health Organization, Geneva. p 232.

Wright GD (2017) Opportunities for natural products in 21st century antibiotic discovery. *Nat. Prod. Rep.* 10.1039/7-00019.

Wu SC, Yeung JC, Duan Y, Ye R, Szarka SJ, Habibi HR & Wong SL (2002) Functional Production and Characterization of a Fibrin-Specific Single-Chain Antibody Fragment from *Bacillus subtilis*: Effects of Molecular Chaperones and a Wall-Bound Protease on Antibody Fragment Production. *Appl. Environ. Microbiol.* 68(7):3261-3269.

Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC & Moore BS (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* 111(5):1957-1962.

Zarzosa-Alvarez AL, Sandoval-Cabrera A, Torres-Huerta AL & Bermudez-Cruz RM (2010) Electroeluting DNA fragments. *Journal of visualized experiments : JoVE* 10.3791/2136(43).

Zhang L, Fu Q, Li W, Wang B, Yin X, Liu S, Xu Z & Niu Q (2017) Identification and characterization of a novel beta-glucosidase via metagenomic analysis of *Bursaphelenchus xylophilus* and its microbial flora. *Sci. Rep.* 7(1):14850.

Zhu H, Swierstra J, Wu C, Girard G, Choi YH, van Wamel W, Sandiford SK & van Wezel GP (2014) Eliciting antibiotics active against the ESKAPE pathogens in a collection of actinomycetes isolated from mountain soils. *Microbiology* 160(8):1714-1725.

Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, Burian M, Schilling NA, Slavetinsky C, Marschal M, Willmann M, Kalbacher H, Schittek B, Brötz-Oesterhelt H, Grond S, Peschel A & Krismer B (2016) Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* 535(7613):511-516.

CHAPITRE 8 : ANNEXES

8.1. ANNEXE A: article revue à soumettre prochainement

Functional metagenomics for the discovery of new antibiotics

Anissa Brahami, Éric Déziel* and Annie Castonguay*

INRS-Institut Armand-Frappier, Laval, Quebec, H7V 1B7, Canada;

Journal:

Submitted:

Acceptation:

Publication:

Citation:

Contribution:

Acknowledgments:

8.1.1. Abstract

New bacterial natural antibiotics are urgently needed in human medicine to counter the emergence of multidrug-resistant pathogens. Since most commercial antibiotics are derived from cultured soil microorganisms, many successful approaches have been developed to improve the bacterial culturability for drug discovery. IChip and co-culturing techniques are good examples, but still not enough in front of the uncultured microbial diversity. On the other hand, functional metagenomic tools have proven their potential to allow the discovery of new antibiotics. However, although being a very promising technique, only a few active compounds were so far identified from functional metagenomic screenings when compared to the plethora of natural product biosynthetic gene clusters reported. This review provides an overview of the functional metagenomic reports that led to the discovery of new antimicrobials. It summarises some of the problems and limitations encountered by using this technique, which might have mainly contributed to the so far low success of functional studies. In addition, it provides a discussion that highlights how sequence-based screenings can assist functional-based screenings in drug discovery. Importantly, as the chosen habitat for eDNA sampling determines the properties of the biomolecules to be found, a survey of the functional metagenomic studies reported so far according to their selected source of eDNA is presented. From the studies reported in this review, it is hypothesized that polluted environments with strong antibiotic selection pressure, drug-resistant bacteria competition and human feces, such as hospital-derived biofilms, would be advantageous to further exploit in order to yield more novel bioactive antibiotics in a near future.

Keywords

Drug discovery. Uncultured bacteria. Functional metagenomic. Wastewater eDNA sampling.

8.1.2. Introduction

The overuse of antibiotics is the main cause of antibiotic resistance development worldwide. Importantly, around 50% of all antibiotics prescribed in human medicine are not necessary nor effective in the prescribed manner (Control & Prevention, 2013). Notably, the use of antibiotics during viral infections exposes bacteria to sub-lethal antibiotic concentrations that promote the growth of a sub-population of resistant strains (Blaskovich *et al.*, 2017). Antibiotics are also used excessively in food-producing animals to prevent, control and treat diseases, promoting animal growth.

The development of bacterial resistance has been widely documented (Davies & Davies, 2010) since the discovery of the first antibiotics (Demain & Sanchez, 2009). Indeed, when Fleming discovered penicillin in 1945, penicillin-resistant bacteria were also observed (World Health Organization., 2014). For example, the US Infectious Disease Society reports 2 million people getting a serious MDR infection (Control & Prevention, 2013). Many of these multidrug-resistant bacteria are continually evolving and spreading, especially in hospitals where patients die from bacterial infections after a surgery, or from previously treatable bacterial infections (Roberts *et al.*, 2009). Of a particular interest are the ESKAPE pathogens comprising *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and/or *Enterobacter cloacae* (Santajit & Indrawattana, 2016; Zhu *et al.*, 2014). This scourge is particularly problematic because of the lack of effective antibiotics, especially for the treatment of infections caused by multidrug-resistant Gram-negative bacteria (Blaskovich *et al.*, 2017). More worryingly, some studies even indicate the emergence of strains resistant to colistin, a last-resort antibiotic (Liu *et al.*, 2016; Poirel *et al.*, 2016).

This review examines important aspects of the functional metagenomic studies reported to date for the discovery of antibiotics able to eradicate multidrug-resistant bacteria. More specifically, a description of the antimicrobial screening approaches

used from the age of bacterial culturing to the culture-free concept is surveyed, along with a discussion of the functional metagenomic approaches favored in terms of heterologously expressed environmental DNA (eDNA), with a special focus on details such as the average DNA insert size, the choice of plasmid, the selection of host and the challenges associated with these approaches and the choice of the eDNA. Finally, we suggest the combination of functional and sequence-based metagenomic approaches as a promising strategy to promote the discovery of new antimicrobials.

8.1.3. Bacteria as a source of antibiotic

Most natural products of economic value such as antibiotics or other pharmaceuticals are derived from cultured soil microorganisms (Daniel, 2005). These soil microorganisms have been exploited based on the cultivation, referred to the Waksman platform, by growing bacteria in the laboratory and identifying natural antagonist products using bioactivity assays (Waksman & Woodruff, 1940). This approach allowed the discovery of more than 28 000 compounds and most of them (like Streptomycin, Fosfomycin, Tetracycline, etc) were discovered by screening soil-derived actinomycetes (Wright, 2017). Thereby, the golden age of antibiotic discovery was from the 1940s to 1960s (Lewis, 2013). Since then, the number of novel natural product discovered is declining (The Pew Charitable Trusts, 2016), mainly due to the traditional laboratory culturing method that led to high rates of molecule rediscovery due to the heavy reliance on cultured bacteria (Lewis, 2013) known as 'the great plate anomaly' (Lewis *et al.*, 2010).

The microbial diversity in terrestrial biosphere is estimated at up to 10^{30} of different microorganism species (Whitman *et al.*, 1998) and contains approximately 10^7 bacterial species per gram of soil (Colman & Burley, 2001). The diversity is very scarcely studied because only 1% of it remains culturable by conventional *in vitro* methods (Handelsman *et al.*, 1998a). This theory is further supported by the fact that one-third of the bacterial diversity does not have cultured representative

species, and are known only from 16S rDNA sequencing, suggesting that the number of antibiotics to be discovered from these sources are two to three orders of magnitude higher than those from cultured sources (Clardy *et al.*, 2006).

To isolate new environmental species, an adapted culture condition closely related to the original bacterial natural environment is needed (Hugenholtz, 2002). The manipulation of growth laboratory media often had limited success and the unculturability of many species remains a major problem (Rappe & Giovannoni, 2003). The fact that most microorganisms never grow in culture is a clear indication of the high importance of some components from their natural environment (D'Onofrio *et al.*, 2010). For instance, growing unculturable microorganisms in semi-diffusible chambers with nutrients and growth factors from their natural environment was previously found successful. Briefly, bacteria from a specific environment are mixed with agar, sandwiched between two semi-permeable membranes and placed back in their initial environment (Kaeberlein *et al.*, 2002). This approach improved their growth rate from 1% to 40% (Bollmann *et al.*, 2007; Lewis *et al.*, 2010) and has recently led to the discovery of the new antibiotic Teixobactin (Ling *et al.*, 2015).

8.1.4. Functional screening of metagenomic libraries

Metagenomics provides tools to study the majority of unculturable microorganisms. The direct cloning of an environmental DNA (eDNA) into a host was first proposed in 1985 by Pace *et al.*, for the purpose of cloning DNA from picoplankton in a phage vector for subsequent 16S rRNA gene sequence analyses (Schmidt *et al.*, 1991). The first metagenomic libraries screening was proposed by Jo Handelsman (Handelsman *et al.*, 1998a), and the first successful function-driven screening for antibiotic discovery was conducted in 2000 by Rondon *et al.* The construction of metagenomic libraries is a critical step in DNA-based metagenomic projects aimed at the discovery of new antimicrobials. Notably, high quality environmental DNA (eDNA) isolated and purified is needed for cloning a suitable vector, heterologously expressing the eDNA-inserted vector in a surrogate host

(usually *Escherichia coli*) and analysing the transformants by phenotypic-based approaches (Handelsman, 2007). The choice of vector depends essentially on the DNA size, targeted genes, and screening strategy. Small-insert libraries can be employed for the identification of novel molecules encoded by a single gene or small operon, whereas large-insert libraries are required to recover large gene clusters coding complex pathways (Wang *et al.*, 2000).

Several antimicrobials were discovered using functional metagenomics and details about the studies that led to the identification of these molecules are presented in Tableau 8.1. Functional metagenomic studies are often associated to a number of significant challenges and some of them are discussed in the next section.

Tableau 8.1. eDNA library construction for the identification of antimicrobials

Antibiotic function	eDNA origin	Average insert size (kb)	Vector type	Host	Hit rate	Biological activity	Reference
55 KDA protein	Agricultural soil	27	BAC pBeloBAC11	<i>E. coli DH10B</i>	1/28,224	<i>B. subtilis</i> , <i>S. aureus</i> (weak)	(Rondon <i>et al.</i> , 2000b)
Terragin A, B, C, D, E	Soil	-	Vector	<i>S. lividans</i>	2/1,020	<i>E. coli</i> , <i>B. subtilis</i>	(Wang <i>et al.</i> , 2000)
Indigo & Indirubin	Soil from New England	37	BAC pBTP2	<i>E. coli DH10B</i>	4/12,000	<i>B. subtilis</i> , <i>S. aureus</i>	(MacNeil <i>et al.</i> , 2001b)
Violacein	Soil from Ithaca, New York	-	Cosmid SuperCosI	<i>E. coli</i>	-	<i>B. subtilis</i> , <i>S. aureus</i> , <i>Streptococcus</i> <i>us</i> sp.	(Brady <i>et al.</i> , 2001a)*
Turbomycin	Soil	44.5	BAC pBeloBAC11	<i>E. coli DH10B</i>	1/24,546	<i>E. coli</i> , <i>B. subtilis</i> , <i>S. aureus</i> , ...	(Gillespie <i>et al.</i> , 2002)
N-acyl tyrosine	Soil from Ithaca, New	-	Cosmid	<i>E. coli</i>	-	-	(Brady <i>et al.</i> , 2002)*

	York						
Antibacterial molecule	Organic Arable field from France	50	Shuttle cosmid pOS700I	<i>E. coli</i> DH10B, <i>S. lividans</i> TK24	1/5,000	<i>B. subtilis</i>	(Courtois et al., 2003)
Palmitoylputrescine	Bromeliad tank water from Costa Rica	-	Cosmid pCSLF16	<i>E. coli</i>	-	<i>B. subtilis</i>	(Brady & Clardy, 2004)*
Long-chain <i>N</i> -acyl tyrosine	Soil from Boston Ithaca & Costa Rica	-	Cosmid	<i>E. coli</i>	-	<i>B. subtilis</i>	(Brady et al., 2004)*
Indigo & indirubin	Soil from Jindong Valley, Corea	35	Fosmid pEpiFOS-5	<i>E. coli</i>	2/113,700	<i>B. subtilis</i>	(Lim et al., 2005)
Isocyanide	Soil from Boston	-	Cosmid pCSLG18	<i>E. coli</i> ,	-	<i>B. subtilis</i>	(Brady & Clardy, 2005a)*
Long-chain <i>N</i> -acyl tryptophan and arinine	Soil	-	Cosmid pSuperCosI	<i>E. coli</i>	-	<i>B. subtilis</i>	(Brady & Clardy, 2005b)
Antibacterial molecule	Environmental soil	-	Cosmid pWEB, pWEB::TNC	<i>E. coli</i> XL-1Blue MR	1/20,000	<i>B. subtilis</i>	(Brady, 2007)
Type III Polyketide synthase	Soil from Oregon and Pennsylvania	-	Shuttle cosmid pTR101	<i>E. coli</i> , <i>R. metallidurans</i>	ND	<i>B. subtilis</i>	(Craig et al., 2009)
Antibacterial molecule	Soil from Pennsylvania and Oregon	-	Shuttle cosmid pJWC1	<i>E. coli</i> , <i>A. tumefaciens</i> , <i>B. graminis</i> , <i>C. vibrioides</i> , <i>E. coli</i> , <i>P. putida</i> , <i>R. metallidurans</i>	8/-	<i>B. subtilis</i>	(Craig et al., 2010)
Antimicrobial molecule	Forest from Belgium	12	Shuttle plasmid PHT01	<i>E. coli</i> , <i>B. subtilis</i>	1/25,000	<i>B. cereus</i>	(Biver et al., 2013)

Antibacterial molecule	Soil from Arizona	-	Cosmid pJWC1, pJSS	<i>E. coli</i> , <i>Ralstonia metallidurans</i>	19/700,000	Tetracyclin-resistant <i>B. subtilis</i>	(Iqbal <i>et al.</i> , 2014)
N-acyltyrosine	Slaughterhouse drain biofilm	2-10	Shuttle vector pEBP18	<i>E. coli</i> DH10B	-	<i>B. subtilis</i>	(Thies <i>et al.</i> , 2016)
Colicin V	Human microbiome	35	Shuttle cosmid pJWC1	<i>E. coli</i> EC100	12/350,000	<i>E. coli</i>	(Cohen <i>et al.</i> , 2018)
Antimicrobial peptide	Hospital biofilm (wastewater)	0.001	BAC pBeloBAC11	<i>E. coli</i> DH10B	1/150,000	MRSA	(Our work, 2019)

* Related studies

8.1.5. Functional metagenomics : challenges and alternatives

Source of eDNA

An important aspect that is clearly neglected in metagenomic studies is the careful selection of the source of eDNA to be exploited. The choice of the habitat for eDNA sampling in metagenomic studies determines the properties of the bioactive molecules to be found. This is due to the physiological adaptation of the microorganisms to the environmental conditions (Brooks *et al.*, 2011). A reflective choice of the sampling site depends on the phenotype screening and increases the chances of discovering new genes (table II) while preventing the rediscovery of bioactive molecules. For instance, a highly thermostable enzyme, β -glucanase (active at 90°C), was identified from a volcanic soil (Angelov *et al.*, 2017). A related study allowed the discovery of a novel DNA polymerase from Glacial Ice (Simon *et al.*, 2009), which was found to display an activity at 18°C. Polluted environments are also interesting eDNA sources for the discovery of new molecules (tableau 8.2). Antibiotics released in the environment from agriculture and medicine have allowed the selection of antibiotic-resistant bacteria. These antibiotic resistance genes can be moved between bacteria via horizontal transfer of mobile genetic elements

(Martinez *et al.*, 2009). It is now well-accepted that the antibiotic resistance gene can be used for identification of secondary metabolites in functional metagenomic library screenings (Kealey *et al.*, 2017; Wright, 2017). This new approach was developed by Thaker *et al.*, (2013), on exploiting actinomycetes vancomycin-resistance traits to identify glycopeptide antibacterial producers (Thaker *et al.*, 2014; Thaker *et al.*, 2013). Polluted environments such as wastewater, especially from hospitals, have a strong antibiotic selection pressure (Katouli *et al.*, 2012; Lood *et al.*, 2017) in addition to drug-resistant bacteria competition, human feces and microbiota that have been found to produce bacteriocins (Cohen *et al.*, 2018), antimicrobials (Chu *et al.*, 2016) and cyclic peptides (Zipperer *et al.*, 2016). The competition between these microorganisms might have selected for the evolution of antibiotics congeners that could be actives against the current resistant bacteria. Furthermore, these resistant bacteria could themselves produce new molecules to meet the strong selection pressure. This was recently demonstrated from the discovery of a new antimicrobial peptide (ref, Tableau 8.1) for which eDNA was sampled from a hospital biofilm. We believe that this type of polluted environment should be explored further.

Tableau 8.2. Unusual eDNA sampling in functional-based metagenomics studies

Type of molecule isolated	eDNA sampling source	References
Novel β -glucosidase	<i>Bursaphelenchus xylophilus</i> microbial flora	(Zhang <i>et al.</i> , 2017)
Novel β -galactosidase	Corn field soil	(Cheng <i>et al.</i> , 2017)
Thermostable β -glucanase	Volcano soil	(Angelov <i>et al.</i> , 2017)
Exonuclease (ExoMeg1)	Semiarid Biome in Brazil	(Silva-Portela <i>et al.</i> , 2016)
Enzymes		
Esterase	Hot environments	(Leis <i>et al.</i> , 2015)
Amylases, cellulases	Rumen of cows	(Wang <i>et al.</i> , 2013)
Novel Flavonoid-Modifying Enzymes	Elephant feces	(Rabausch <i>et al.</i> , 2013)
Hydrolytic activity	Human Gut Bacteria	(Cecchini <i>et al.</i> , 2013)
DNA polymerases	Glacial Ice	(Simon <i>et al.</i> , 2009)

	Tetracyclin	Cheese bacteria	(Florez <i>et al.</i> , 2017)
	New class of A beta-lactamase	Polluted environments,	(Vercammen <i>et al.</i> , 2013)
Resistance genes	Novel β -lactamases	Wastewater Treatment	(Uyaguari <i>et al.</i> , 2011)
	Ampicillin, amoxicillin, p \acute{e} nicillin	Gull feces	(Martiny <i>et al.</i> , 2011)
	β -lactamases	Alaskan soil	(Allen <i>et al.</i> , 2009)
	Tetracyclin	Oral metagenome	(Diaz-Torres <i>et al.</i> , 2003)

Antimicrobial discovery and rediscovery

An important question arising from functional metagenomic studies is whether or not the various newly found molecules were previously identified. For instance, although Indirubin was identified in 1985, it was used as antileukemic agents in Chinese medicine (Wu *et al.*, 1985) and its antibacterial activity was first reported from functional metagenomic studies conducted by MacNeil *et al.*, 2001. This molecule has been reported in the drug bank (<https://www.drugbank.ca/>) as an organoheterocyclic antibiotic and antineoplastic compound with an action as a protein kinase inhibitor. Some derivatives of this molecule were also developed to improve its biological activity (Ndolo *et al.*, 2017; Vougianniopoulou *et al.*, 2008), which is now undergoing clinical trials (Park *et al.*, 2016). Violacein, produced by *Chromobacterium violaceum*, has initially been reported in 1945 (Lichstein & Van De Sand, 1945) and rediscovered in 2001 (Brady *et al.*, 2001b; Lichstein & Van De Sand, 1945). The activity of this molecule against cancer cells (Hashimi *et al.*, 2015) and membrane-permeabilizing Meticillin-resistant *S. aureus* (Aruldass *et al.*, 2018) was recently reported. Turbomycin A has been characterized in 1972 from a fungal source *Saccharomyces cerevisiae* (Budzikiewicz *et al.*, 1972), but was never been reported as a bacterial metabolite whereas Turbomycin B was not previously reported (Gillespie *et al.*, 2002). Colicins were first discovered in 1925 by Gratia who showed that one strain of *E. coli* released a toxic substance (Gratia, 1925). The bacteriocin colicin V was recently found to be produced by the human microbiome using functional metagenomics (Cohen *et al.*, 2018). Landomycin A-D were

discovered in 1990 after being isolated from *Streptomyces* sp strains (Henkel *et al.*, 1990) whereas landomycin E was discovered by functional metagenomic studies (Feng *et al.*, 2011).

eDNA size for library construction

Large-insert library construction depends on the quality of the eDNA extracted and purified (Delmont *et al.*, 2011b; Williamson *et al.*, 2011). eDNA BAC capturing is generally limited to DNA inserts of about 40 kb (table I). This size does not allow the cloning of polycistronic gene clusters of PKS (polyketide synthase) and NRPS (non ribosomal peptide synthase) types with a length of approximately 100 kb (Kang & Brady, 2014). Recently, multiple different protocols for metagenomic libraries construction were developed, resulting in different average insert-size of the library (Brady, 2007; Liles *et al.*, 2008).

Heterologous hosts

One of the most important limitation of functional metagenomic is the dependence of the metagenomic DNA expression in the heterologous expression host used. *E. coli* is the most widely used host in functional metagenomic studies due to its ease of culture and manipulation (table I). Importantly, the use of *E. coli* as a single host limits the expression of metagenomic DNA. Indeed, it has been shown using *in silico* analysis that from 32 complete prokaryotic genome sequences only 40% of the gene could be expressed in *E. coli*. And even when the gene is expressed, the level of heterologous expression and detection can be low (Gabor *et al.*, 2004). However, this does not mean that the expression of heterologous genes in *E. coli* is limited to DNA from Proteobacteria. The expression of the *B. cereus* genome integrated into a BAC vector maintained in *E. coli* has shown a significant number of such activities detected in the library (Rondon *et al.*, 1999). Usually, libraries are constructed in shuttle vectors that could be maintained in *E. coli* and then transferred by conjugation to other hosts to detect other phenotypes that are not expressed using *E. coli* as a single host (Biver *et al.*, 2013; Craig *et al.*, 2009; Craig

et al., 2010). Therefore, using different hosts for expressing the same metagenomic eDNA library is more suitable because the antibiotic molecule produced may be toxic to the host selected (Rondon *et al.*, 2000b). Considering that *Streptomyces* strains are great prolific producers of various known secondary metabolites (Barka *et al.*, 2016), they are considered as convoluted heterologous expression hosts (Courtois *et al.*, 2003; Wang *et al.*, 2000). An interesting study conducted by Iqbal *et al.*, (2016) investigated the native prerequisite heterologous expression of 39 *Streptomyces* strains that would be more productive hosts for large metagenomic studies and high transformation frequency. 97 eDNA cosmid clones containing minimal type II polyketide synthase (PKS) were expressed in these *Streptomyces* strains to investigate the colored phenotype. The best hit was observed with *Streptomyces albus*. Craig *et al.*, (2010) also reported the construction of the same cosmid library in different hosts for antibacterial activity phenotype as such as altered pigmentation or altered colony morphology. Interestingly, no *E. coli*-hosted eDNA conferring this phenotype were found in *P. putida*-hosted eDNA, *A. tumefaciens*-hosted eDNA, *R. metallidurans*-hosted eDNA and wrinkled colonies found in *A. tumefaciens*-hosted eDNA. Similarly, no antibacterial activity was found from a *E. coli* library reported by Biver *et al.*, (2013), whereas some activity was observed with the use of *B. subtilis* as the heterologous host. More recently, an interesting enzyme with cell wall lytic activity was identified from the functional screening of a metagenomic library hosted in *R. metallidurans* (Iqbal *et al.*, 2014).

Gene expression.

The inability to express a targeted gene is mostly due to a plethora of factors such as codon usage, improper promoter recognition, the lack of initiation factors, improper protein folding, absence of essential co-factors, toxicity of the gene product, or the inability of the host to excrete the gene expression product (Kudla *et al.*, 2009). As a result, the production of the molecule is considerably reduced, or even skewed and the final product would not be the one expected (Kallifidas *et al.*, 2012). For example, microorganisms have a particular preference for some

translation initiation codons (Saier, 1995). *E. coli* uses AUG as a start codon to initiate approximately 92% of translations, GUG in 6.7% of translations and UUG for only 1.2%, therefore these non-AUG start codons may not be efficiently recognized and can lead to translation problems (Villegas & Kropinski, 2008a). The codon usage preference varies between species (referred to as codon usage bias, CUB) (Goodarzi *et al.*, 2008). Kudla *et al.*, (2009) have synthesized a library of 154 genes encoding green fluorescent protein (GFP) that randomly have a silent mutation in the third-base position. Resulting expression levels varied 250-fold across all variants, illustrating the dramatic effect of CUB on gene expression.

Screening techniques

Since the sequence information is not required for function-based screenings, it is the only strategy that allows the identification of an entirely novel class of gene-encoding novel functions (Daniel, 2005). To date, the screening of metagenomic libraries through antibiotic phenotypic-based approaches conferred by the inserted DNA have resulted in the identification of several novel antibacterial molecules (table 1). The screening of morphological traits such as colored colonies have often been associated with the production of secondary metabolite pigmented antibiotics (Brady, 2007). Therefore, color can be used as a visual screening characteristic to identify eDNA clones containing biosynthetic gene clusters. This very simple filter to identify potential natural product-producing clones was previously found successfully to identify new antibiotics, as for instance the purple-colored Violaceine (Brady *et al.*, 2001b), the orange-colored Turbomycin A, the red-colored Turbomycin B (Gillespie *et al.*, 2002), the blue-colored Indigo and the pink-colored Indirubin (MacNeil *et al.*, 2001a). For antibiotic activity screenings using the top agar overlay assay, metagenomic libraries are usually overlaid with a thin layer of top agar containing an assay bacteria strain. The observation of a zone of growth inhibition around the colonies is used to identify the clones producing antibacterial molecules (Biver *et al.* 2013, Brady, 2007, Chang & Brady, 2013b, Courtois *et al.*, 2003b, Craig *et al.*, 2010, Martinez *et al.*, 2004, Rondon *et al.*, 2000b). A more

sensitive antibacterial assay called *bacteriospray* was recently developed for a more suitable screening of large-scale functional metagenomic libraries for antibacterial activity (Brahmi *et al.*, 2019).

Clone absence/loss

Attempts to heterologously express eDNA have taken a variety of forms. In this report, most studies attempted the identification of antibacterial phenotypes. In some functional studies, the convoted antibacterial phenotypes are absent or lost. This is the case for Terragine A for which no *in vitro* inhibition of *E. coli* or *B. subtilis* was observed (Wang *et al.*, 2000), despite the fact that it is associated with metal-chelating hydroxamic acids which often exhibit an antibacterial activity. In 2000, Rondon *et al.* identified the antibacterial activity of a 55 kb protein against *B. subtilis*. Interestingly, in the course of their investigation, they were unable to extract or concentrate the molecule from liquid or agar culture, even after using a variety of growth conditions, filtration steps, organic extractions and pH variations, suggesting that the activity is due to some aspect of the protein itself, or its effect upon the host cell on the inhibitory activity (Rondon *et al.*, 2000a). In 2003, Courtois *et al.* identified an antibacterial active clone which was lost before its analysis could be made possible, due to the strong negative selection of the host used, resulting from the expression of this molecule (Courtois *et al.*, 2003).

8.1.7. Future prospects in functional metagenomics

Functional metagenomic led to the remarkable exploitation of gene function of uncultured microorganisms for the screening and development of new antimicrobial therapies. However, since only few metagenomic library clones will contain complete biosynthetic gene clusters of secondary metabolites, directly targeting these clones using specific primers of conserved secondary metabolites regions could be a successful approach to identify uncharacterized natural product gene clusters from metagenomic libraries. Thereby, even if the gene targeting

involved polymerase chain reaction (PCR) amplification of conserved region similar to already known genes, this method has been successfully applied for the detection of novel biocatalysts (Owen *et al.*, 2013). In Banik *et al.*, (2008) study, degenerate primers of an oxidative coupling enzyme, OxyC were used to screen eDNA clone libraries. This led to identification new glycopeptide gene clusters. A related study used degenerate primers based on conserved polyketide synthase (PKS) regions followed by heterologous expression in *S. albus* led to the discovery of Erdacin a novel pentacyclic type II polyketide pathway (King *et al.*, 2009), Utahmycins A and B, two azaanthraquinones with biological activity against drug-resistant *Staphylococcus aureus*, (Bauer *et al.*, 2010) and metatricycloene (Iqbal *et al.*, 2016).

Furthermore, the large size of biosynthetic gene clusters cannot be captured on a single cosmid clone and must be recovered in a series of overlapping clones. These clones are reassembled into a bacterial artificial chromosome using transformation associated recombination (TAR) (Kallifidas & Brady, 2012). Thereby, PKS positives clones obtained by PCR screen reassembled and heterologous expressed yielded for example to the novel antibiotic fluostatin (Feng *et al.*, 2010). Another successful sequence-based identification and heterologous expression of reassembled type II PKS led to identification of the well-known antibiotic landomycin E, an uncharacterized pentacyclic ring, and a unique KB-3346-5 metabolites with biological activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faecalis* (Feng *et al.*, 2011). The KB-3346-5 metabolite have been given the name of fasamycin A and B (Feng *et al.*, 2012).

Interestingly, some gene clusters remain silent even after an attempt of heterologous expression, this was observed by the study of Kallifidas *et al.*, (2012) in the discovery of tetarimycin A, an MRSA active antibiotic molecule. The entire *tam* gene cluster was reconstructed from AZ60 an AZ919 clones that did not produce tetarimycin unless SARP (Streptomyces antibiotic regulator protein) was conjugated (Kallifidas *et al.*, 2012) concluding that certain transcriptional factor(s)

may be important for the successful expression.

The prescreening of gene clusters directly from eDNA samples, followed by library construction could be a way to directly access to biosynthetic gene clusters with potentially novel activities. This strategy was employed to target highly conserved biosynthetic genes as starting point for the library construction (Owen *et al.*, 2013). New members of relevant natural products were discovered by targeting the unique tryptophan dimerization biosynthetic gene cluster that led to the discovery of the indolotryptoline antiproliferative agent Borregomycin A, B and D (Chang & Brady, 2013a), erdasporines cytotoxines with novel carboxy-indolocarbazole thryptophane dimer (TD) (Chang *et al.*, 2013), hydroxisporine and reductasporine (Chang *et al.*, 2015). Also, PCR searching in eDNA sequence of a KS_{β} gene a phylogenetic markers of Type II PKS gene cluster led to the identification of arixanthomycins A-C, a pentangular phenol with antiproliferative properties (Kang & Brady, 2014).

A more recent study of sequence-guided metagenomics using degenerate PCR primer screening, that encode calcium-binding motifs representative of known calcium dependent antibiotics biosynthesized by non-ribosomal peptide synthase (NRPS), led to the discovery of malacidin A-B active against methicillin-resistant *S. aureus* (SARM) (Hover *et al.*, 2018).

8.1.8. Conclusion

Functional metagenomics provides access to bioactive molecules produced by uncultured environmental bacteria through the direct cloning of eDNA, bypassing the requirement for laboratory culturing, whereas sequence-based metagenomics relies on the sequencing and bioinformatics analysis to identify biosynthetic gene clusters of interest. The functional approach develops at a much slower pace than the sequence-based approach probably since the latter is akin to high-throughput screening efforts.

To accelerate the development this functional approach, innovative methods are needed, new environmental sampling sources have to be investigated, considering the dimension of roughly two to three orders of magnitude deeper biosynthetic diversity potential. It becomes clear from the studies reported in this survey that the combination of both functional and sequence-based metagenomic strategies would be highly beneficial and could greatly enhance the discovery of highly needed novel antibiotics.

8.1.9 Bibliography

Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Volker U, Ellingsen TE & Valla S (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiol. Lett.* 296(2):149-158.

Alduina R & Gallo G (2012a) Artificial chromosomes to explore and to exploit biosynthetic capabilities of actinomycetes. *J. Biomed. Biotechnol.* 2012:462049.

Alduina R & Gallo G (2012b) Artificial Chromosomes to Explore and to Exploit Biosynthetic Capabilities of Actinomycetes. *J. Biomed. Biotechnol.* 2012:1-10.

Allen HK, Moe LA, Rodbumrer J, Gaarder A & Handelsman J (2009) Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *The ISME journal* 3(2):243-251.

Amiche M, Seon AA, Wroblewski H & Nicolas P (2000) Isolation of dermatoxin from frog skin, an antibacterial peptide encoded by a novel member of the dermaseptin genes family. *Eur. J. Biochem.* 267(14):4583-4592.

Amoutzias GD, Chaliotis A & Mossialos D (2016) Discovery Strategies of Bioactive Compounds Synthesized by Nonribosomal Peptide Synthetases and Type-I Polyketide Synthases Derived from Marine Microbiomes. *Mar. Drugs* 14(4).

Angelov A, Pham VTT, Ubelacker M, Brady S, Leis B, Pill N, Brolle J, Mechelke M, Moerch M, Henrissat B & Liebl W (2017) A metagenome-derived thermostable beta-glucanase with an unusual module architecture which defines the new glycoside hydrolase family GH148. *Sci. Rep.* 7(1):17306.

Aruldass CA, Masalamany SRL, Venil CK & Ahmad WA (2018) Antibacterial mode of action of violacein from *Chromobacterium violaceum* UTM5 against *Staphylococcus aureus* and methicillin-resistant *Staphylococcus aureus* (MRSA). *Environ. Sci. Pollut. Res. Int.* 25(6):5164-5180.

Arvani S, Markert A, Loeschcke A, Jaeger KE & Drepper T (2012) A T7 RNA polymerase-based toolkit for the concerted expression of clustered genes. *J. Biotechnol.* 159(3):162-171.

Bahar AA & Ren D (2013) Antimicrobial peptides. *Pharmaceuticals (Basel)* 6(12):1543-1575.

Bahnsen JS, Franzyk H, Sandberg-Schaal A & Nielsen HM (2013) Antimicrobial and cell-penetrating properties of penetratin analogs: effect of sequence and secondary structure. *Biochim. Biophys. Acta* 1828(2):223-232.

Banik JJ & Brady SF (2008) Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc. Natl. Acad. Sci. U. S. A.* 105(45):17273-17277.

Banik JJ & Brady SF (2010) Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opin. Microbiol.* 13(5):603-609.

Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, Klenk HP, Clement C, Ouhdouch Y & van Wezel GP (2016) Taxonomy, Physiology, and Natural Products of *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* 80(1):1-43.

Barrett E, Hayes M, O'Connor P, Gardiner G, Fitzgerald GF, Stanton C, Ross RP & Hill C (2007) Salivaricin P, one of a family of two-component antilisterial bacteriocins produced by intestinal isolates of *Lactobacillus salivarius*. *Appl. Environ. Microbiol.* 73(11):3719-3723.

Bauer JD, King RW & Brady SF (2010) Utahmycins a and B, azaquinones produced by an environmental DNA clone. *J. Nat. Prod.* 73(5):976-979.

Bentley S, Chater K, Cerdeno-Tarraga A-M, Challis G, Thomson N, James K, Harris D, Quail M, Kieser H & Harper D (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* 417(6885):141-147.

Bierbaum G & Sahl HG (2014) The search for new anti-infective drugs: untapped resources and strategies. *Int. J. Med. Microbiol.* 304(1):1-2.

Bisant D & Maizel J (1995) Identification of ribosome binding sites in *Escherichia coli* using neural network models. *Nucleic Acids Res.* 23(9):1632-1639.

Bishop BM, Juba ML, Devine MC, Barksdale SM, Rodriguez CA, Chung MC, Russo PS, Vliet KA, Schnur JM & van Hoek ML (2015) Bioprospecting the American alligator (*Alligator mississippiensis*) host defense peptidome. *PLoS One* 10(2):e0117394.

Biver S, Steels S, Portetelle D & Vandenberg M (2013) *Bacillus subtilis* as a tool for screening soil metagenomic libraries for antimicrobial activities. *J. Microbiol. Biotechnol.* 23(6):850-855.

Blaskovich MA, Butler MS & Cooper MA (2017) Polishing the tarnished silver bullet: the quest for new antibiotics. *Essays Biochem.* 61(1):103-114.

Bleich R, Watrous JD, Dorrestein PC, Bowers AA & Shank EA (2015) Thiopeptide antibiotics stimulate biofilm formation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* 112(10):3086-3091.

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T & Medema MH (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45(W1):W36-W41.

Bollmann A, Lewis K & Epstein SS (2007) Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates. *Appl. Environ. Microbiol.* 73(20):6386-6390.

Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B & Bartlett J (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* 48(1):1-12.

Brady SF (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat. Protoc.* 2(5):1297-1305.

Brady SF, Chao CJ & Clardy J (2002) New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* 124(34):9968-9969.

Brady SF, Chao CJ & Clardy J (2004) Long-chain N-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* 70(11):6865-6870.

Brady SF, Chao CJ, Handelsman J & Clardy J (2001a) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org Lett* 3(13):1981-1984.

Brady SF, Chao CJ, Handelsman J & Clardy J (2001b) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* 3(13):1981-1984.

Brady SF & Clardy J (2000) Long-Chain N-Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental DNA. *J. Am. Chem. Soc.* 122(51):12903-12904.

Brady SF & Clardy J (2004) Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J. Nat. Prod.* 67(8):1283-1286.

Brady SF & Clardy J (2005a) Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. *Angew. Chem.* 117(43):7225-7227.

Brady SF & Clardy J (2005b) N-Acyl Derivatives of Arginine and Tryptophan Isolated from Environmental DNA Expressed in *Escherichia coli*. *Org. Lett.* 7(17):3613-3616.

Brady SF, Wright SA, Lee JC, Sutton AE, Zumoff CH, Wodzinski RS, Beer SV & Clardy J (1999) Pantocin B, an antibiotic from *Erwinia herbicola* discovered by heterologous expression of cloned genes. *J. Am. Chem. Soc.* 121(50):11912-11913.

Brahami A, Castonguay A & Déziel É (2019) Novel 'Bacteriospray' Method Facilitates the Functional Screening of Metagenomic Libraries for Antimicrobial Activity. *Methods and Protocols* 2(1):4.

Brogden NK & Brogden KA (2011) Will new generations of modified antimicrobial peptides improve their potential as pharmaceuticals? *Int. J. Antimicrob. Agents* 38(3):217-225.

Brooks AN, Turkarslan S, Beer KD, Lo FY & Baliga NS (2011) Adaptation of cells to new environments. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3(5):544-561.

Brown CW, Sridhara V, Boutz DR, Person MD, Marcotte EM, Barrick JE & Wilke CO (2017) Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. *BMC Genomics* 18(1):301.

Budzikiewicz H, Eckau H & Ehrenberg M (1972) Bis (3-indolyl)-3H-indolyldenmethan, ein von *Saccharomyces cerevisiae* Gebildeter Farbstoff. *Tetrahedron Lett.* 13(36):3807-3810.

Buelow E, Bayjanov JR, Willems RJL, Bonten MJM, Schmitt H & Van Schaik W (2017) The microbiome and resistome of hospital sewage during passage through the community sewer system 10.1101/216242.

Bürgmann H, Pesaro M, Widmer F & Zeyer J (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. *J. Microbiol. Methods* 45(1):7-20.

Cardenas E & Tiedje JM (2008) New tools for discovering and characterizing microbial diversity. *Curr. Opin. Biotechnol.* 19(6):544-549.

Carmona G, Rodriguez A, Juarez D, Corzo G & Villegas E (2013) Improved protease stability of the antimicrobial peptide Pin2 substituted with D-amino acids. *Protein J.* 32(6):456-466.

Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Dore J, Henrissat B, Remaud-Simeon M, Monsan P & Potocki-Veronese G (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* 8(9):e72766.

Chakchouk-Mtibaa A, Elleuch L, Smaoui S, Najah S, Sellem I, Abdelkafi S & Mellouli L (2014) An antilisterial bacteriocin BacFL31 produced by *Enterococcus faecium* FL31 with a novel structure containing hydroxyproline residues. *Anaerobe* 27:1-6.

Chang AY, Chau V, Landas JA & Pang Y (2017) Preparation of calcium competent *Escherichia coli* and heat-shock transformation. *JEMI Methods* 1:22-25.

Chang F-Y & Brady SF (2013a) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proceedings of the National Academy of Sciences* 110(7):2478-2483.

Chang FY & Brady SF (2013b) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc. Natl. Acad. Sci. U. S. A.* 110(7):2478-2483.

Chang FY, Ternei MA, Calle PY & Brady SF (2013) Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment. *J. Am. Chem. Soc.* 135(47):17906-17912.

Chang FY, Ternei MA, Calle PY & Brady SF (2015) Targeted metagenomics: finding rare tryptophan dimer natural products in the environment. *J. Am. Chem. Soc.* 137(18):6044-6052.

Chen T, Tang L & Shaw C (2003) Identification of three novel *Phyllomedusa sauvagei* dermaseptins (sVI–sVIII) by cloning from a skin secretion-derived cDNA library. *Regulatory Peptides* 116(1-3):139-146.

Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD & Charles TC (2017) Functional metagenomics reveals novel beta-galactosidases not predictable from gene sequences. *PLoS One* 12(3):e0172545.

Chu J, Vila-Farres X, Inoyama D, Ternei M, Cohen LJ, Gordon EA, Reddy BVB, Charlop-Powers Z, Zebroski HA & Gallardo-Macias R (2016) Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat. Chem. Biol.* 12(12):1004-1006.

Chu X, He H, Guo C & Sun B (2008) Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl. Microbiol. Biotechnol.* 80(4):615-625.

Clardy J, Fischbach MA & Walsh CT (2006) New antibiotics from bacterial natural products. *Nat. Biotechnol.* 24(12):1541-1550.

Clinical Laboratory Standards Institute (2012) Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically : approved standard - ninth edition M07-A9. (CLSI, Wayne, Pa.), p 68.

Cohen LJ, Han S, Huang YH & Brady SF (2018) Identification of the Colicin V Bacteriocin Gene Cluster by Functional Screening of a Human Microbiome Metagenomic Library. *ACS Infect Dis* 4(1):27-32.

Colman A & Burley JC (2001) A legal and ethical tightrope. *EMBO reports* 2(1):2-5.

Conrad B, Savchenko RS, Breves R & Hofemeister J (1996) A T7 promoter-specific, inducible protein expression system for *Bacillus subtilis*. *Mol. Gen. Genet.* 250(2):230-236.

Control CfD & Prevention (2013) Antibiotic resistance threats in the United States, 2013. *Centres for Disease Control and Prevention, US Dep. Health and Human Services.*

Cooper MA & Shlaes D (2011) Fix the antibiotics pipeline. *Nature* 472(7341):32.

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS, August PR, Nalin R, Guerineau M, Jeannin P, Simonet P & Pernodet JL (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* 69(1):49-55.

Craig JW, Chang FY & Brady SF (2009) Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chem. Biol.* 4(1):23-28.

Craig JW, Chang FY, Kim JH, Obiajulu SC & Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* 76(5):1633-1641.

Culligan EP, Sleator RD, Marchesi JR & Hill C (2014) Metagenomics and novel gene discovery: promise and potential for novel therapeutics. *Virulence* 5(3):399-412.

D'Onofrio A, Crawford JM, Stewart EJ, Witt K, Gavrish E, Epstein S, Clardy J & Lewis K (2010) Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17(3):254-264.

Daegelen P, Studier FW, Lenski RE, Cure S & Kim JF (2009) Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394(4):634-643.

Daniel R (2004) The soil metagenome--a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15(3):199-204.

Daniel R (2005) The metagenomics of soil. *Nat. Rev. Microbiol.* 3(6):470-478.

Davies J & Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74(3):417-433.

De Jong A, van Hijum SA, Bijlsma JJ, Kok J & Kuipers OP (2006) BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res.* 34:W273-279.

Debbabi S, Groleau MC, Letourneau M, Narayanan C, Gosselin LL, Iddir M, Gagnon J, Doucet N, Deziel E & Chatenet D (2018) Antibacterial properties of the pituitary adenylate cyclase-activating polypeptide: A new human antimicrobial peptide. *PLoS One* 13(11):e0207366.

Debussche L, Couder M, Thibaut D, Cameron B, Crouzet J & Blanche F (1992) Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a, c-diamide during coenzyme B12 biosynthesis in *Pseudomonas denitrificans*. *J. Bacteriol.* 174(22):7445-7451.

Delbruck M (1945) Interference Between Bacterial Viruses: III. The Mutual Exclusion Effect and the Depressor Effect. *J. Bacteriol.* 50(2):151-170.

DeLeo FR, Otto M, Kreiswirth BN & Chambers HF (2010) Community-associated methicillin-resistant *Staphylococcus aureus*. *Lancet* 375(9725):1557-1568.

Delmont TO, Robe P, Clark I, Simonet P & Vogel TM (2011a) Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J. Microbiol. Methods* 86(3):397-400.

Delmont TO, Robe P, Clark I, Simonet P & Vogel TM (2011b) Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J. Microbiol. Methods* 86(3):397-400.

Demain AL & Sanchez S (2009) Microbial drug discovery: 80 years of progress. *J. Antibiot. (Tokyo)* 62(1):5-16.

Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M & Mullany P (2003) Novel Tetracycline Resistance Determinant from the Oral Metagenome. *Antimicrob. Agents Chemother.* 47(4):1430-1432.

Feng Z, Chakraborty D, Dewell SB, Reddy BV & Brady SF (2012) Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis. *J. Am. Chem. Soc.* 134(6):2981-2987.

Feng Z, Kallifidas D & Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. U. S. A.* 108(31):12629-12634.

Feng Z, Kim JH & Brady SF (2010) Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J. Am. Chem. Soc.* 132(34):11902-11903.

Fleming D, Chahin L & Rumbaugh K (2017) Glycoside Hydrolases Degrade Polymicrobial Bacterial Biofilms in Wounds. *Antimicrob. Agents Chemother.* 61(2).

Florez AB, Vazquez L & Mayo B (2017) A Functional Metagenomic Analysis of Tetracycline Resistance in Cheese Bacteria. *Front. Microbiol.* 8:907.

Flynn S, van Sinderen D, Thornton GM, Holo H, Nes IF & Collins JK (2002) Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. *salivarius* UCC118. *Microbiology* 148(4):973-984.

Food U & Administration D (2014) New animal drugs and new animal drug combination products, administered in or on medicated feed or drinking water of food-producing animals: recommendations for drug sponsors for voluntarily aligning product use conditions with GFI# 209.).

Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, Giglione C & Meinel T (2006) The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics* 5(12):2336-2349.

Gabor EM, Alkema WB & Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* 6(9):879-886.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM & Handelsman J (2002) Isolation of Antibiotics Turbomycin

A and B from a Metagenomic Library of Soil Microbial DNA. *Appl. Environ. Microbiol.* 68(9):4301-4306.

Goodarzi H, Torabi N, Najafabadi HS & Archetti M (2008) Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons. *Gene* 407(1-2):30-41.

Graf M, Mardirossian M, Nguyen F, Seefeldt AC, Guichard G, Scocchi M, Innis CA & Wilson DN (2017) Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* 34(7):702-711.

Gratia A (1925) Sur un remarquable exemple d'antagonisme entre deux souches de coillbacille. *CR Seances Soc. Biol. Fil.* 93:1040-1041.

Grossman TH, Kawasaki ES, Punreddy SR & Osburne MS (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* 209(1):95-103.

Gupta A, Bedre R, Thapa SS, Sabrin A, Wang G, Dassanayake M & Grove A (2017) Global Awakening of Cryptic Biosynthetic Gene Clusters in *Burkholderia thailandensis*. *ACS Chem. Biol.* 12(12):3012-3021.

Han Y, Albericio F & Barany G (1997) Occurrence and Minimization of Cysteine Racemization during Stepwise Solid-Phase Peptide Synthesis^{1, 2}. *The Journal of organic chemistry* 62(13):4307-4312.

Handelsman J (2007) Metagenomics and microbial communities. *eLS*.

Handelsman J, Liles M, Mann D, Riesenfeld C & Goodman RM (2002) 13 Cloning the Metagenome: Culture-independent Access to the Diversity and Functions of the Uncultivated Microbial World. *Methods in Microbiology* (3) p 241-255.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998a) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5(10):R245-R249.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998b) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology* 5(10):R245-249.

Harris G (2014) Superbugs kill India's babies and pose an overseas threat. *The New York Times* (New York), World-Asian Pacific, p A1.

Hashimi SM, Xu T & Wei MQ (2015) Violacein anticancer activity is enhanced under hypoxia. *Oncol. Rep.* 33(4):1731-1736.

Henkel T, Rohr J, Beale JM & Schwenen L (1990) Landomycins, new angucycline antibiotics from *Streptomyces* sp. I. Structural studies on landomycins A-D. *J. Antibiot. (Tokyo)* 43(5):492-503.

Holder I & Boyce S (1994) Agar well diffusion assay testing of bacterial susceptibility to various antimicrobials in concentrations non-toxic for human cells in culture. *Burns* 20(5):426-429.

Hover BM, Kim SH, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB, Molina H, Park S, Perlin DS & Brady SF (2018) Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol* 10.1038/s41564-018-0110-1.

Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, Robins-Browne R, Davies JK, Seemann T & Steinar TP (2011) Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Pathog.* 7(11):e1002359.

Huang Y, Huang J & Chen Y (2010) Alpha-helical cationic antimicrobial peptides: relationships of structure and function. *Protein & cell* 1(2):143-152.

Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 3(2):reviews0003. 0001.

Huo L, Hug JJ, Fu C, Bian X, Zhang Y & Muller R (2019) Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* 10.1039/8-00091.

Imjongjirak C, Amparyup P & Tassanakajon A (2011) Two novel antimicrobial peptides, arasin-likeSp and GRPSp, from the mud crab *Scylla paramamosain*, exhibit the activity against some crustacean pathogenic bacteria. *Fish Shellfish Immunol.* 30(2):706-712.

Iqbal HA, Craig JW & Brady SF (2014) Antibacterial enzymes from the functional screening of metagenomic libraries hosted in *Ralstonia metallidurans*. *FEMS Microbiol. Lett.* 354(1):19-26.

Iqbal HA, Low-Beinart L, Obiajulu JU & Brady SF (2016) Natural Product Discovery through Improved Functional Metagenomics in *Streptomyces*. *J. Am. Chem. Soc.* 138(30):9341-9344.

Jachlewski S, Jachlewski WD, Linne U, Brasen C, Wingender J & Siebers B (2015) Isolation of Extracellular Polymeric Substances from Biofilms of the Thermoacidophilic Archaeon *Sulfolobus acidocaldarius*. *Frontiers in bioengineering and biotechnology* 3:123.

Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Segurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P & Kim JF (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394(4):644-652.

Jiang L, Wang L, Zhang J, Liu H, Hong B, Tan H & Niu G (2015) Identification of novel mureidomycin analogues via rational activation of a cryptic gene cluster in *Streptomyces roseosporus* NRRL 15998. *Sci. Rep.* 5:14111.

Joo SH (2012) Cyclic peptides as therapeutic agents and biochemical tools. *Biomol. Ther. (Seoul)* 20(1):19-26.

Kaeberlein T, Lewis K & Epstein SS (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296(5570):1127-1129.

Kakirde KS, Parsley LC & Liles MR (2010) Size Does Matter: Application-driven Approaches for Soil Metagenomics. *Soil Biol. Biochem.* 42(11):1911-1923.

Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W & Liles MR (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene* 475(2):57-62.

Kallifidas D & Brady SF (2012) Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. *Methods Enzymol.* 517:225-239.

Kallifidas D, Kang HS & Brady SF (2012) Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. *J. Am. Chem. Soc.* 134(48):19552-19555.

Kang HS & Brady SF (2014) Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. *ACS Chem. Biol.* 9(6):1267-1272.

Katouli M, Thompson JM, Gündoğdu A & Stratton HM (2012) Antibiotic Resistant Bacteria in Hospital Wastewaters and Sewage Treatment Plants. in *Science Forum and Stakeholder Engagement*. p 223-229.

Kealey C, Creaven CA, Murphy CD & Brady CB (2017) New approaches to antibiotic discovery. *Biotechnol. Lett* 39(6):805-817.

Kim H, Jang JH, Kim SC & Cho JH (2014) De novo generation of short antimicrobial peptides with enhanced stability and cell specificity. *J. Antimicrob. Chemother.* 69(1):121-132.

Kim U-J, Birren BW, Slepak T, Mancino V, Boysen C, Kang H-L, Simon MI & Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* 34(2):213-218.

King RW, Bauer JD & Brady SF (2009) An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. *Angew. Chem. Int. Ed. Engl.* 48(34):6257-6261.

Klevens RM, Edwards JR, Tenover FC, McDonald LC, Horan T, Gaynes R & System NNIS (2006) Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* in intensive care units in US hospitals, 1992–2003. *Clin. Infect. Dis.* 42(3):389-391.

Kodama T, Endo K, Ara K, Ozaki K, Kakeshita H, Yamane K & Sekiguchi J (2007) Effect of *Bacillus subtilis* spo0A mutation on cell wall lytic enzymes and extracellular proteases, and prevention of cell lysis. *J. Biosci. Bioeng.* 103(1):13-21.

KPMG LLP (2014) The global economic impact of anti-microbial resistance. (KPMG International, UK), p 41.

Krsek M & Wellington EMH (1999) Comparison of different methods for the isolation and purification of total community DNA from soil. *J. Microbiol. Methods* 39(1):1-16.

Kudla G, Murray AW, Tollervey D & Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324(5924):255-258.

Lam KN, Cheng J, Engel K, Neufeld JD & Charles TC (2015) Current and future resources for functional metagenomics. *Front. Microbiol.* 6.

Laplante SR, Mikou A, Robin M, Guittet E, Delsuc MA, Charpentier I & Lallemand JY (1990) Rapid determination and NMR assignments of antiparallel sheets and helices of a scorpion and a cobra toxin. *Int. J. Pept. Protein Res.* 36(3):227-230.

Le TH, Ng C, Chen H, Yi XZ, Koh TH, Barkham TM, Zhou Z & Gin KY (2016) Occurrences and Characterization of Antibiotic-Resistant Bacteria and Genetic Determinants of Hospital Wastewater in a Tropical Country. *Antimicrob. Agents Chemother.* 60(12):7449-7456.

Leis B, Angelov A, Mientus M, Li H, Pham VT, Lauinger B, Bongen P, Pietruszka J, Goncalves LG, Santos H & Liebl W (2015) Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front. Microbiol.* 6:275.

Lewis K (2012) Antibiotics: recover the lost art of drug discovery. *Nature* 485(7399):439.

Lewis K (2013) Platforms for antibiotic discovery. *Nature reviews Drug discovery* 12(5):371.

Lewis K, Epstein S, D'Onofrio A & Ling LL (2010) Uncultured microorganisms as a source of secondary metabolites. *J. Antibiot. (Tokyo)* 63(8):468-476.

Li JW-H & Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? *Science* 325(5937):161-165.

Li JX, Zhao LM, Wu RJ, Zheng ZJ & Zhang RJ (2013) High-level overproduction of Thermobifida enzyme in *Streptomyces lividans* using a novel expression vector. *Int. J. Mol. Sci.* 14(9):18629-18639.

Lichstein HC & Van De Sand VF (1945) Violacein, an Antibiotic Pigment Produced by *Chromobacterium violaceum*. *The Journal of Infectious Diseases* 76(1):47-51.

Liles MR, Manske BF, Bintrim SB, Handelsman J & Goodman RM (2003) A Census of rRNA Genes and Linked Genomic Sequences within a Soil Metagenomic Library. *Appl. Environ. Microbiol.* 69(5):2684-2691.

Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Goodman RM & Handelsman J (2008) Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Appl. Environ. Microbiol.* 74(10):3302-3305.

Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY & Lee SW (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.* 71(12):7768-7777.

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schaberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C & Lewis K (2015) A new antibiotic kills pathogens without detectable resistance. *Nature* 10.1038/14098.

Liu X & Cheng YQ (2014) Genome-guided discovery of diverse natural products from Burkholderia sp. *J. Ind. Microbiol. Biotechnol.* 41(2):275-284.

Liu X, Liu L, Wang Y, Wang X, Ma Y & Li Y (2014) The study on the factors affecting transformation efficiency of *E. coli* competent cells. *Cell* 5:106.

Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu L-F, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu J-H & Shen J (2016) Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 16(2):161-168.

Lood R, Erturk G & Mattiasson B (2017) Revisiting Antibiotic Resistance Spreading in Wastewater Treatment Plants - Bacteriophages as a Much Neglected Potential Transmission Vehicle. *Front. Microbiol.* 8:2298.

Lopez PJ, Marchand I, Joyce SA & Dreyfus M (1999) The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol. Microbiol.* 33(1):188-199.

Lussier FX (2011) *Développement de vecteurs et d'hôtes bactériens pour l'expression de protéines recombinantes et l'analyse métagénomique*. Québec. p 178.

Lussier FX, Chambenoit O, Cote A, Hupe JF, Denis F, Juteau P, Beaudet R & Shareck F (2011) Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector. *J. Ind. Microbiol. Biotechnol.* 38(9):1321-1328.

Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 25(6):1203-1210.

MacNeil I, Tiong C, Minor C, August P, Grossman T, Loiacono K, Lynch B, Phillips T, Narula S & Sundaramoorthi R (2001a) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* 3(2):301-308.

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D & Osburne MS (2001b) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* 3(2):301-308.

Mao D, Bushin LB, Moon K, Wu Y & Seyedsayamdost MR (2017) Discovery of scmR as a global regulator of secondary metabolism and virulence in *Burkholderia thailandensis* E264. *Proc. Natl. Acad. Sci. U. S. A.* 114(14):E2920-E2928.

Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA & Osburne MS (2004) Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Appl. Environ. Microbiol.* 70(4):2452-2463.

Martinez JL, Fajardo A, Garmendia L, Hernandez A, Linares JF, Martinez-Solano L & Sanchez MB (2009) A global view of antibiotic resistance. *FEMS Microbiol. Rev.* 33(1):44-65.

Martiny AC, Martiny JB, Weihe C, Field A & Ellis JC (2011) Functional metagenomics reveals previously unrecognized diversity of antibiotic resistance genes in gulls. *Front. Microbiol.* 2:238.

Mary RZa & Betlach C (1999) Recombinant polyketide synthesis in *Streptomyces*: engineering of improved host strains. *BioTechniques* 26:106-110.

McKenzie NL, Thaker M, Koteva K, Hughes DW, Wright GD & Nodwell JR (2010) Induction of antimicrobial activities in heterologous streptomycetes using alleles of the *Streptomyces coelicolor* gene *absA1*. *J. Antibiot. (Tokyo)* 63(4):177-182.

Mosaei H, Molodtsov V, Kepplinger B, Harbottle J, Moon CW, Jeeves RE, Ceccaroni L, Shin Y, Morton-Laing S, Marrs ECL, Wills C, Clegg W, Yuzenkova Y, Perry JD, Bacon J, Errington J, Allenby NEE, Hall MJ, Murakami KS & Zenkin N (2018) Mode of Action of Kanglemycin A, an Ansamycin Natural Product that Is Active against Rifampicin-Resistant *Mycobacterium tuberculosis*. *Mol. Cell* 72(2):263-274 e265.

Mullany P (2014) Functional metagenomics for the investigation of antibiotic resistance. *Virulence* 5(3):443-447.

Muller CA, Oberauner-Wappis L, Peyman A, Amos GC, Wellington EM & Berg G (2015) Mining for Nonribosomal Peptide Synthetase and Polyketide Synthase Genes Revealed a High Level of Diversity in the Sphagnum Bog Metagenome. *Appl. Environ. Microbiol.* 81(15):5064-5072.

Ndolo KM, Park KR, Lee HJ, Yoon KB, Kim YC & Han SY (2017) Characterization of the Indirubin Derivative LDD970 as a Small Molecule Aurora Kinase A Inhibitor in Human Colorectal Cancer Cells. *Immune Netw.* 17(2):110-115.

Oman TJ & van der Donk WA (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* 6(1):9-18.

Overmann J, Abt B & Sikorski J (2017) Present and future of culturing bacteria. *Annu. Rev. Microbiol.* 71:711-730.

Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH & Brady SF (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proceedings of the National Academy of Sciences* 110(29):11797-11802.

Parachin NS, Mulder KC, Viana AAB, Dias SC & Franco OL (2012) Expression systems for heterologous production of antimicrobial peptides. *Peptides* 38(2):446-456.

Park CB, Yi KS, Matsuzaki K, Kim MS & Kim SC (2000) Structure-activity analysis of buforin II, a histone H2A-derived antimicrobial peptide: the proline hinge is responsible for the cell-penetrating ability of buforin II. *Proc. Natl. Acad. Sci. U. S. A.* 97(15):8245-8250.

Park MH, Lee YY, Cho KH, La S, Lee HJ, Yim DS, Ban S, Park MY, Kim YC, Kim YG & Shin YG (2016) Validation of a liquid chromatography-triple quadrupole mass spectrometric method for the determination of 5-nitro-5'-hydroxy-indirubin-3'-oxime (AGM-130) in human plasma and its application to microdose clinical trial. *Biomed. Chromatogr.* 30(3):323-329.

Paulsen JS, Long JD, Ross CA, Harrington DL, Erwin CJ, Williams JK, Westervelt HJ, Johnson HJ, Aylward EH, Zhang Y, Bockholt HJ & Barker RA (2014) Prediction of manifest Huntington's disease with clinical and imaging measures: a prospective observational study. *The Lancet Neurology* 13(12):1193-1201.

Pavlova O, Mukhopadhyay J, Sineva E, Ebright RH & Severinov K (2008) Systematic structure-activity analysis of microcin J25. *J. Biol. Chem.* 283(37):25589-25595.

Peek J, Lilic M, Montiel D, Milshteyn A, Woodworth I, Biggins JB, Ternei MA, Calle PY, Danziger M, Warriar T, Saito K, Braffman N, Fay A, Glickman MS, Darst SA, Campbell EA & Brady SF (2018) Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nature communications* 9(1):4147.

Poirel L, Kieffer N, Liassine N, Thanh D & Nordmann P (2016) Plasmid-mediated carbapenem and colistin resistance in a clinical isolate of *Escherichia coli*. *The Lancet Infectious Diseases* 16(3).

Quadri LE (2014) Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. *Crit. Rev. Biochem. Mol. Biol.* 49(3):179-211.

Rabausch U, Juergensen J, Ilmberger N, Bohnke S, Fischer S, Schubach B, Schulte M & Streit WR (2013) Functional screening of metagenome and genome

libraries for detection of novel flavonoid-modifying enzymes. *Appl. Environ. Microbiol.* 79(15):4551-4563.

Ranjan R, Grover A, Kapardar RK & Sharma R (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochem. Biophys. Res. Commun.* 335(1):57-65.

Rappe MS & Giovannoni SJ (2003) The uncultured microbial majority. *Annu. Rev. Microbiol.* 57:369-394.

Rice LB (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J. Infect. Dis.* 197(8):1079-1081.

Riesenfeld CS, Goodman RM & Handelsman J (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* 6(9):981-989.

Robe P, Nalin R, Capellano C, Vogel TM & Simonet P (2003) Extraction of DNA from soil. *Eur. J. Soil Biol.* 39(4):183-190.

Roberts RR, Hota B, Ahmad I, Scott RD, Foster SD, Abbasi F, Schabowski S, Kampe LM, Ciavarella GG & Supino M (2009) Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin. Infect. Dis.* 49(8):1175-1184.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA & Minor C (2000a) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66(6):2541-2547.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J & Goodman RM (2000b) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66(6):2541-2547.

Rondon MR, Raffel SJ, Goodman RM & Handelsman J (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *Proceedings of the National Academy of Sciences* 96(11):6451-6455.

Sabree ZL, Rondon MR & Handelsman J (2009) Metagenomics. *Encyclopedia of Microbiology (Third Edition)*, Schaechter M (Édit.) Academic Press, Oxford p 622-632.

Sabri NA, Schmitt H, Van der Zaan B, Gerritsen HW, Zuidema T, Rijnaarts HHM & Langenhoff AAM (2018) Prevalence of antibiotics and antibiotic resistance genes in a wastewater effluent-receiving river in the Netherlands. *Journal of Environmental Chemical Engineering* 10.1016/j.jece.2018.03.004.

Saier MH (1995) Differential codon usage: a safeguard against inappropriate expression of specialized genes? *FEBS Lett.* 362(1):1-4.

Santajit S & Indrawattana N (2016) Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed research international* 2016:2475067.

Saraswat M, Grand RS & Patrick WM (2013) Desalting DNA by drop dialysis increases library size upon transformation. *Biosci. Biotechnol. Biochem.* 77(2):402-404.

Schmidt TM, DeLong EF & Pace NR (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* 173(14):4371-4378.

Schneider D, Duperchy E, Depeyrot J, Coursange E, Lenski R & Blot M (2002) Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiol.* 2:18.

Scocchi M, Tossi A & Gennaro R (2011) Proline-rich antimicrobial peptides: converging to a non-lytic mechanism of action. *Cell. Mol. Life Sci.* 68(13):2317-2330.

Seyedsayamdost MR (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* 111(20):7266-7271.

Shimamura H, Gouda H, Nagai K, Hirose T, Ichioka M, Furuya Y, Kobayashi Y, Hirono S, Sunazuka T & Omura S (2009) Structure determination and total synthesis of bottromycin A2: a potent antibiotic against MRSA and VRE. *Angew. Chem. Int. Ed. Engl.* 48(5):914-917.

Silva-Portela RC, Carvalho FM, Pereira CP, de Souza-Pinto NC, Modesti M, Fuchs RP & Agnez-Lima LF (2016) ExoMeg1: a new exonuclease from metagenomic library. *Sci. Rep.* 6:19712.

Simon C & Daniel R (2010) Construction of small-insert and large-insert metagenomic libraries. *Methods Mol. Biol.* 668:39-50.

Simon C & Daniel R (2011) Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* 77(4):1153-1161.

Simon C & Daniel R (2017) Construction of Small-Insert and Large-Insert Metagenomic Libraries. *Metagenomics: Methods and Protocols*, Streit WR & Daniel R (Édit.) Springer New York, New York, NY10.1007/978-1-4939-6691-2_1. p 1-12.

Simon C, Herath J, Rockstroh S & Daniel R (2009) Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl. Environ. Microbiol.* 75(9):2964-2968.

Sosio M, Giusino F, Cappellano C, Bossi E, Puglia AM & Donadio S (2000) Artificial chromosomes for antibiotic-producing actinomycetes. *Nat. Biotechnol.* 18(3):343-345.

Steffensky M, Muhlenweg A, Wang ZX, Li SM & Heide L (2000) Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. *Antimicrob. Agents Chemother.* 44(5):1214-1222.

Stensvag K, Haug T, Sperstad SV, Rekdal O, Indrevoll B & Styrvold OB (2008) Arasin 1, a proline-arginine-rich antimicrobial peptide isolated from the spider crab, *Hyas araneus*. *Dev. Comp. Immunol.* 32(3):275-285.

Stevenson J, Krycer JR, Phan L & Brown AJ (2013) A practical comparison of ligation-independent cloning techniques. *PLoS One* 8(12):e83888.

Stewart EJ (2012) Growing unculturable bacteria. *J. Bacteriol.* 194(16):4151-4160.

Studier FW, Daegelen P, Lenski RE, Maslov S & Kim JF (2009) Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J. Mol. Biol.* 394(4):653-680.

Studier FW & Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189(1):113-130.

Szabo D, Ostorhazi E, Binas A, Rozgonyi F, Kocsis B, Cassone M, Wade JD, Nolte O & Otvos L, Jr. (2010) The designer proline-rich antibacterial peptide A3-APO is effective against systemic *Escherichia coli* infections in different mouse models. *Int. J. Antimicrob. Agents* 35(4):357-361.

Tamaki H, Sekiguchi Y, Hanada S, Nakamura K, Nomura N, Matsumura M & Kamagata Y (2005) Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. *Appl. Environ. Microbiol.* 71(4):2162-2169.

Tao Q & Zhang HB (1998) Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. *Nucleic Acids Res.* 26(21):4901-4909.

Tatum EL & Lederberg J (1947) Gene Recombination in the Bacterium *Escherichia coli*. *J. Bacteriol.* 53(6):673-684.

Thaker MN, Waglechner N & Wright GD (2014) Antibiotic resistance-mediated isolation of scaffold-specific natural product producers. *Nat. Protoc.* 9(6):1469-1479.

Thaker MN, Wang W, Spanogiannopoulos P, Waglechner N, King AM, Medina R & Wright GD (2013) Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol.* 31(10):922-927.

The Pew Charitable Trusts (2016) A Scientific Roadmap for Antibiotic Discovery: A Sustained and Robust Pipeline of New Antibacterial Drugs and Therapies is Critical to Preserve Public Health. (The Pew Charitable Trusts, Philadelphia, PA.), p 42.

Thies S, Rausch SC, Kovacic F, Schmidt-Thaler A, Wilhelm S, Rosenau F, Daniel R, Streit W, Pietruszka J & Jaeger KE (2016) Metagenomic discovery of novel enzymes and biosurfactants in a slaughterhouse biofilm microbial community. *Sci. Rep.* 6:27035.

Toda H, Williams JA, Gullledge M & Sehgal A (2019) A sleep-inducing gene, *nemuri*, links sleep and immune function in *Drosophila*. *Science* 363(6426):509-515.

Torres AM, Bansal P, Koh JM, Pages G, Wu MJ & Kuchel PW (2014) Structure and antimicrobial activity of platypus 'intermediate' defensin-like peptide. *FEBS Lett.* 588(9):1821-1826.

Troeschel SC, Thies S, Link O, Real CI, Knops K, Wilhelm S, Rosenau F & Jaeger KE (2012) Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *J. Biotechnol.* 161(2):71-79.

Uchiyama T & Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* 20(6):616-622.

Uyaguari MI, Fichot EB, Scott GI & Norman RS (2011) Characterization and quantitation of a novel beta-lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. *Appl. Environ. Microbiol.* 77(23):8226-8233.

Vercammen K, Garcia-Armisen T, Goeders N, Van Melderren L, Bodilis J & Cornelis P (2013) Identification of a metagenomic gene cluster containing a new

class A beta-lactamase and toxin-antitoxin systems. *Microbiologyopen* 2(4):674-683.

Villegas A & Kropinski AM (2008a) An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. *Microbiology* 154(9):2559-2661.

Villegas A & Kropinski AM (2008b) An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. *Microbiology* 154(Pt 9):2559-2661.

Vougogiannopoulou K, Ferandin Y, Bettayeb K, Myriantopoulos V, Lozach O, Fan Y, Johnson CH, Magiatis P, Skaltsounis A-L & Mikros E (2008) Soluble 3', 6-substituted indirubins with enhanced selectivity toward glycogen synthase kinase-3 alter circadian period. *J. Med. Chem.* 51(20):6421-6431.

Waisvisz J, Van der Hoeven M & Nijenhuis BT (1957) The structure of the sulfur-containing moiety of bottromycin. *J. Am. Chem. Soc.* 79(16):4524-4527.

Waksman SA & Woodruff HB (1940) The soil as a source of microorganisms antagonistic to disease-producing bacteria. *J. Bacteriol.* 40(4):581.

Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ & Davies J (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* 2(16):2401-2404.

Wang L, Hatem A, Catalyurek UV, Morrison M & Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS One* 8(11):e78507.

Wang M, Lin J, Sun Q, Zheng K, Ma Y & Wang J (2019) Design, expression, and characterization of a novel cecropin A-derived peptide with high antibacterial activity. *Appl. Microbiol. Biotechnol.* 103(4):1765-1775.

Wang Q, Wang P & Yang Q (2018) Occurrence and diversity of antibiotic resistance in untreated hospital wastewater. *Sci. Total Environ.* 621:990-999.

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Muller R, Wohlleben W, Breitling R, Takano E & Medema MH (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43(W1):W237-243.

Wei Q, Kim YS, Seo JH, Jang WS, Lee IH & Cha HJ (2005) Facilitation of expression and purification of an antimicrobial peptide by fusion with baculoviral polyhedrin in *Escherichia coli*. *Appl. Environ. Microbiol.* 71(9):5038-5043.

Westers L, Westers H & Quax WJ (2004) *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta* 1694(1-3):299-310.

Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences* 95(12):6578-6583.

Williamson KE, Kan J, Polson SW & Williamson SJ (2011) Optimizing the indirect extraction of prokaryotic DNA from soils. *Soil Biol. Biochem.* 43(4):736-748.

Wong SL (1995) Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. *Curr. Opin. Biotechnol.* 6(5):517-522.

Wooley JC, Godzik A & Friedberg I (2010) A primer on metagenomics. *PLoS Comput. Biol.* 6(2):e1000667.

World Bank (2017) Drug Resistant Infections: A Threat to Our Economic Future. (World Bank, Washington, DC), p 141.

World Health Organization (2017) Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. p 7.

World Health Organization. (2014) *Antimicrobial resistance : global report on surveillance.*(In IRIS). World Health Organization, Geneva. p 232.

Wright GD (2017) Opportunities for natural products in 21st century antibiotic discovery. *Nat. Prod. Rep.* 10.1039/7-00019.

Wu J & Xi C (2009) Evaluation of different methods for extracting extracellular DNA from the biofilm matrix. *Appl. Environ. Microbiol.* 75(16):5390-5395.

Wu K-M, Zhang M-Y, Fang Z & Huang L (1985) Potential antileukemic agents, synthesis of derivatives of indirubin, indigo, and isoindigotin. *Acta Pharm. Sin* 20:821-826.

Wu SC, Yeung JC, Duan Y, Ye R, Szarka SJ, Habibi HR & Wong SL (2002) Functional Production and Characterization of a Fibrin-Specific Single-Chain Antibody Fragment from *Bacillus subtilis*: Effects of Molecular Chaperones and a Wall-Bound Protease on Antibody Fragment Production. *Appl. Environ. Microbiol.* 68(7):3261-3269.

Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC & Moore BS (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* 111(5):1957-1962.

Yang Y, Ren Y & Shi Y (2016) Intermolecular disulfide bond in the dimerization of S-periaxin mediated by Cys88 and Cys139. *Acta biochimica et biophysica Sinica* 48(4):326-333.

Zarzosa-Alvarez AL, Sandoval-Cabrera A, Torres-Huerta AL & Bermudez-Cruz RM (2010) Electroeluting DNA fragments. *Journal of visualized experiments : JoVE* 10.3791/2136(43).

Zhang L, Fu Q, Li W, Wang B, Yin X, Liu S, Xu Z & Niu Q (2017) Identification and characterization of a novel beta-glucosidase via metagenomic analysis of *Bursaphelenchus xylophilus* and its microbial flora. *Sci. Rep.* 7(1):14850.

Zhu H, Swierstra J, Wu C, Girard G, Choi YH, van Wamel W, Sandiford SK & van Wezel GP (2014) Eliciting antibiotics active against the *ESKAPE* pathogens in a collection of actinomycetes isolated from mountain soils. *Microbiology* 160(Pt 8):1714-1725.

Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, Burian M, Schilling NA, Slavetinsky C, Marschal M, Willmann M, Kalbacher H, Schitteck B, Brötz-Oesterhelt H, Grond S, Peschel A & Krismer B (2016) Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* 535(7613):511-516.

8.2. ANNEXE B: séquences nucléotidique des inserts B1, B2, B3, B4, B5, B6 issus du criblage de la banque d'ADN de *Burkholderia thailandensis* actifs contre *Staphylococcus aureus* souche Newman

8.2.1. Séquence du clone B1

La région du chromosome II insérée s'étend de 270994 pb à 271913 pb soit 919 pb.

```
CATGCCGAATGCGAAGCGAAAACGATTCGGCGAATTCGCCGGCCCGGTAATG
CAAACGGTTCGAATGGCGGCGCTTTTGCCGGATTGTCTCAAGCATTCTCTAATA
CTCACGCCCGATGGAACCGATCACGCAAGCATTTCGCGCAGCAATTCAGCCGG
GAATGGATCGATGCATGGAACGCGCACGATCTCGACGCGATCCTGTGCGCACT
ACGCGGACGGATTTCGAAATGTCGTCGCCGATGATCGTGCGAGATCGCCGGCGA
GCCGAGCGGGCGCCTGCGCGGCAAGGAGCAGGTTCGGCGCATACTGGCGCG
AGGCGTTGCGGATGATTCCGGATCTGCACTTCGAATGGATCGCGACGCTCGC
GGGCGTCGACAGCGTCGCGATTCACTACCGTGGCGCGAAAGGGCGGGCTCGC
GCTCGAAGTATTTCACTTCGGACCGGATCGGCGTGTCGTGAAAGCGCTGGCT
CACTACGCCGGATAATGGAATAACGAGATCGCGGGACATCGGCGCGGCTGTA
TTTTCTACCGTCTCCGAAACGAACTGACCGGCGGGTTTTACCGAATCGAGCAA
TGCGGCATCGATCGAAATTCGTTAGTACATTTCGATTGGATGCCGAATGCGCCG
GCGGCCGGTTTATCGAGTGAACGGCGTCGACGCGAAGCGCTTCGTCAACCGG
CAGCGGGAACGCGGAGCGGGCGCAGGCGCGGAACGAGCGCGATGCGGCAG
CGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
TGCGACGGTGCGTTCCGCGGACTCACCTATGTCATAACCGGCGGCGGCTATATT
GTCGACGGGGCAGATCGGGAGGGCATCACTCATGGTCAAGAAACGGATTTCG
GCCATGTTCCGCGTCGCGCTCGTCCGGGCTGGCCGCGCATG
```

8.2.2. Séquence du clone B2

La région du chromosome II insérée s'étend de 295870 pb à 299889 pb soit 4019 pb.

```
GCATGCGCTGCGCGATGCAGCTCTACATGGCGTCGGCATCGCGCAACTGCCG
TGCATGGTTGTCGAGGATGACTTGCGCGACGGCACCCCTTGTCGACGTCCTGC
CGGGCTGGGTGCCCAAGGGAGGCGTGATCCATGCGGTATTCCTCCGTCGCCGC
GCGGCCTTCTGCCGCGTGTGCGCTTGCTGATCGACTTTCTGGCGGCCCATAT
CCAGAAGGACTGACGTGGGCGATTCCCGCGGGCCTGCCTCATCGCTTCGCA
```

TTCCGAGAGCCGGTTCAAAGGGACGTCGCTCGATGCCGGCCGGCTCCGACGA
GGCTGCCCTCGGGACAGCGAAGCAGACCGAAGGACGATGACGTGGCGGCCG
GCGAGCGGTGCTTGCCGGCCAGGCGCCGTCATTGACTCAGATGAACTGTT
GACATCCAGGAGGCCGCTTTGCTCCTCGAGGCGTGCGCGCAAGAGGGA
GACGATCGCGTCATTCTCTTTCCGTATCGAAAGGCATCGGCTCAAGCCGCG
GCGCGCGCGGGCCGCGTGTGTGCCTCGCCGCATAATCGTCCGATTAG
CCCAGCCAGTCGACGAACATCTTGACCCGCGCATCTTCCGGCTGCGCGCA
AATCGCCTTGTATTGCGCGGGAGCGGGCACGCGACATCCGAGCACGTTGACG
AGCCGGCCGCGCGCGATTTGTCGTTGACGAGCGCTTCCCGGCACACCACCA
CGCCTTGCCCGCAACCGCCGCTCCGATCGCGAGACTCGCCAGATTGAACTC
CGGCCCCCGCAACTCCCCGCATAACCCGGCACGACGGCATCGAGCCACGT
CCGCCATTCGACGTACCGCGGGCGCATCCGGCCACGCCGATGCATCGTGCAGC
GGCACGACGCTCGATATCCATTTGCGCGTGCAGCGCGGGCGGATGCCGGCCG
AGATACGCGGGCGCGGGCCGCGACGAGATACTCGTGAACAGCGGCTCG
ACGTGCAGCGCGCGGTAGTCGAACGGGTCCCAGCGAATCGCGACGTCGATGT
CGTCCGCGCGCATCGTCTCGCGGTGATCGCGTGAACTCCGCTCGCAGCCG
CACCGGCGTCTGCGGATGCAGCCGATGAAATTCGCTCAGCCTCGGCATCAGC
CATTGCAGCGCGAACGACGGCAGGCAGCTCACCTGCAACGGCGCGTCCGGC
GCGCTCAGTTGCCGACGCTCCGGCCGAGCGAACCGAACGCCGTCGCGGTC
GTTTCGTACAGCACGCTGCCCTTCGGCGTCAGCGCGAGCCCGCGCGCATGG
CGCACGAACAGCGGGTAGCCGAGGCGCGCTTCGAGCTGGCGGATCTGCTGG
CTGATCGCGCTTTGCGTCAGATGCAGCGCGCCGGCCGCTTTTCGTGAAGCTGC
CGAGCCGGCCCGCCGCTCGAAGCAGCGCAGCGCGCCGAGCAGCGGGCGCA
TCGAGCTTGATATCCATTAGCGTGAATTATGCATTTGCAAGAATTGATCGCTTC
ATCGTGCCGCTCGTGGGTGTCGACGATAAGCGCTTCTTTATCGCATTTTATTTCA
TGCTCATGCACGACGTCCGACCGCTTCCGCTAGATCCCAATCTATTAGCCATA
CTTCAGGCTGGCTCGCCGACGCTGTGGCTCAATCCGCATCAGGGCGAGCCGC
TGCCGGAATTCGCGCCGACCGCCGCGGATCTCGCCGAAGCCGGCGCGCGCC
TGCACCGCTGCGCGGGCTTGTGGCCGAGCTGTTTCCCGAACTGCGGCCGTT
GGGCGGGCGCATCGCGTCCCGTTGCAGCCGGCCGAGCCGTTGAAGCGCGC
CGATCACGCGCAAGCGGGCGCGTGGTTCGTCAAGCGCGACGACGCGTTGCC
GGTTGCGGGGTGATCAAGGCGCGCGGGCGGTTTTTCATGAGGTGCTCGCGCTC
GCCGAGTCGATCGCCGAACGGCATGGCCTCGTCAGCGCCGGCGCCGACCGG
CGCGCGCTCGCGAGCGGGCAAGCGCACGCGCTGTTTCGCGCGTCATACGGTG
ATGGTCGGCAGCACCGGCAATCTCGGCTTGAGCATCGGGATGCTCGCGTCGG
CGCTCGGGTTCCGGACCGTCGTGCACATGTCCGCCGATGCGAAGGCGTGGAA
GAGGGCGCGCTGCGCACGCGCGGGCGTGGACGTGGTTCGAACACGCGGGCG
ACTACGCAAAGGCGGTTCGACGCCGGGCGGGCGTTCAGGCCGCCGGGATGCCGC
GCTGTCAATTCGTCGACGACGAAGGCTCGCGGATGCTGTTTCTCGGCTACGC
GACGGCCGCCGCGAACTCGCCGCGCAGCTCGCGCAGGCGGGCCGCCCGG
TCGACGCGCGCCATCCGCTGTTTCGTCCACCTTCATGCGGCGTCCGGCGCGC
GCCGGGCGGCATCGCGTACGGGCTGAAAGCGCGGTACGGCGAGCACGTGCA
TGTGTTTCGTGGCCGAGCCCACCGCGTCCGCTTGCCTGCTCGTGCAGCTCGCG
AGCGGCGGGCGCGCATCCGGTTTTCCGCTACGACGTCGGCCTCGACAATCGGA
CCGAGGCGGACGGGCTGGCGGTGCGCGAGGCGTGCATCTCGCCGGTCCGC

TGTTGCGCGCGCAAGCGGGCGGGCGTGTTCACGGTCGACGATCGTCAGTTGTT
CGCGCATCTGCTCGACGCGCGCGAGCGGCTGGGCATCGATCTCGAGCCGTC
CGCGGCGGGCGGCGTTCGGCGGGCCCGGCGTGGCTCGCGGGATCGGAGGCGG
GGCGCGCCTATCTGCGCGGCCGCGGAATCGTTCCGGAGGCGGCGACGCACG
TGATCTGGGCGACGGGCGGCTCGCTCGTGCCGGCGGAAGAGCATCGGCGCT
TCCAGGCGCGGGCGTGCAGCGAGCGGGAGAGCGGCGCGGGCGCGTGA
GGGCGCACGAGGATGCGCGAAGGTGAGCGAGGAGTTTCGACGGCGTGTGAA
GTCGCGAGGGCGATGGCGGGCGCGACGCCTGCCGGCAGGTTCGATCCGGC
GATCCGCGCCGACTCCGGCTCCGGCTCCGGCTCCGACTCCGGCTCCGGCTC
CGGCTCCGGCTCCGGCTCCGGCTCCGACTCCGACTCCGACTCCGACTCCGG
CGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGAGT
GCGTGCGCTACGCGCCGTCGCGCCGCGCCTCGCGGATCAGCCAGGCTCGG
TTTTCGCGATGTCGTCCGGACGCCGCGGCTCGTTCCGGATAGACGATGTAATAG
CCGATCGACAGCGGCAGCGCGAAATCGAAGAGCCGGCGCAGCCGTCGGCC
TCGATGTCGCGCGCGACGAGCTGATCGCTCGCGAGCGCAACGCCGTGCCCG
GCGATCGCCGCGTCGATCGCGAGCGACGACTGATTGAAGCGCGGCCCTTG
GCCGGATCGACGGCGGCGCGCTCGGGCAGCGCGGGCAGAAACTCGGGCCA
CAGGTCGTGCGCGTTCGTGCAGCAGCACCTGTCCGGCGAGATCGCGCGCGCG
CGCGCCGGACGGCGCGAGCAGGGCGGGGCTGCAAACCGCATGGATGTGAC
CGGAAACAGCAAGCGCGCGACGAGCCCCTTGCCGAACGGCGGCTTGCCGTA
GCGAATCGCGAGATCGACGCCGTGCCCTGAAAGGTCGCGAATTGCGAATCG
GCGATCACGCGGATGTGACGTCCGGATGCTCGTCGCCGAAGCTCGCGAGC
CGCGGAATCAGCCAGCGCGACGCGAACGACGGCGTTCGTGCTGATCGTGACG
CTCGAACGGCGCTTCGCGAGCCGGTCCGTTCGCGTTCGCCGATCAGGTTTCAGC
GCGCGGCGCACGTCCGGCGAAATAGGCGAGCCCCTCGGCCGTTCAGCGCGAGG
CCGCGCGGCAGGCGTTCGAACAGTCTCACGCCAGCGTGTCTCGAGATGGC
GCACCTGTTGCGCGACCGCGCCTTGCGTGACGCCGATCTCGCCGGCCGCGG
CGCGGAAGTTCAGATGCCGCGCCGACACCTCGAACGCACGAAGCGCATTTCAG
CGGCGGCAGGC

8.2.3. Séquence du clone B3 et B4

La région du chromosome II insérée s'étend de 237640 pb à 239649 pb soit 2009 pb.

ATCGTGTGCCGCCTGAGCGGCGTTCGACGCGGACCCGGAAGCGGGCGTCCGGC
CAGATCGAGGCGATGCAGTCGCAACTGGCGAACGAGTGGCCGCAATTGCTGA
GCGCGAAGACGTCGTACCGGGAACCTCGTGTGTTTCGAGGCGATCGATGCGCG
CGCGGTTTCGCGACGGCGCGCTCGCGACGCACGCCGGCAAGCGGCTCGCGG
CGCTGGAGGGCACGTTGCGCAAGGTGCGGACGCAGCCGTTTCGCGCTCGCGC
TCGGCATCGCGCTCGACGGAATCGACGGCTTCGAGCGCTCGTACCAGTGCGC
GCTCGCGACGCTTCGCGTTCGGCGCGGGCGCGCACGCCGACGAGGCCACCTA
TTCGTAACGACTTCGTGCTGCCCGTTCTGCTGTTCGGCGCTGTTCGCAAGGCT

GGCAGGCGCGGCAGTTGCGGCTGCCGCTCACGCGGCTGCTCGCGCGCGAGC
GGCGCTCCGGCGCGCTGCTCGACACGTTGAAAGCCTGGTACGCGAACGATG
GCCACCCGGGCGCGACGGCGCAAGCGCTGGGCATCCACCGCAACACGCTCG
ACTACCGGCTGCAGCAGATCCGCGACGCGACGGGCCTCGATCTCGGCGCGA
TCGACGATCGGCTCTGGCTCTATATCGCGCTGCAGACGATGGCGGGCAGCCG
CGACGGCGACGAGAGCCGCGACGGCGACGCCGACGACCGCCATCCGGCGA
GCTGATCGGCGCACGCGGGAAGCGGCAGGCGCCGCACGCGACGGCGGATTC
TTTCATGAAGAAAGGCGGCGGCCACGCGCGAGGCCACGCGCGGGATGAGCC
GCGTGCGCCGTTTCGTGTGGCGATGCGTGAAAGCGATGCCGATGCGCAATGG
GCATTGGCATCTGTCTGCTCGGCATCCGCGTCCGCCGCGGTGCGGCTGCGCC
GGCCTCGTCGCATCGCGCCCGATCCGCTCCCCCTCTCAATCGAAGCGAACGG
CCCCCGCCTCGGCCGATCGATTTCCGAAACATACGTAGAAAATACGGATGTCC
ATTGTGAAATATTAGGTTTCGATAAATGTCGCTAACATGTCGTCCGGATCGGATA
TGAATCGCCAAATAAAATCCAATGATTTTCGATGGTGTGAATGGATTGCGGAATT
GATTTGTTTGAATCGATTTTCATAAAAAACACCGCACTTTTTTACAAATTCCT
CGGGTCGATCGAGTTTAAAATTCATGTGTTTTGTCGTAGACTGCGTTCCGTTTT
AATCGTCCGGACTGGTACGGCGACATCGTCGCAGAAGCCCGATAGCTTGCAC
TGAAGCGCCGACCGTTGCCCGTTTTCGCTTCAATGCTTGCCGCTTGAATCGGC
GCCGGCGCAAAGACGGCGCCGGAATGCCGATGATCGGCCGATTTCCGGACG
GCGTTCCGCCGGTTCGTCCGGATTTCGGTTGATTGCATGAGATCGATGAATCGCA
TCGATCATTGGCGCATGGAAAACGATCGAATCATTTCGTCATCTTTCATTGCTGC
AATCGGCGGATTATTTCAAATCCAGTTCGTTCGCCCGGTGCGATTTCGATATGA
AATCGAGCATTTCGGATCGCCTCGGCTCACGTTGTTCGGGACCGGCTCGAACA
ATCCCGTTTTTTGCCGGGCAAGCGGCCGGTTCGTCTCACGCTGTCGAGTGTCA
TTCGGCCCGCGATTTCGCGTTCGCAGGGATCGGCG
GCCAGTCAGGTAGGGATTACCCCGCAATCGTCAATCGATGCGTCCGCGCGGA
CACGCCCGGACGTTTGTTCCTTCCACTTGCACCATCAGCATGGAGTGCCTGCACCA
TGACGCCATCCACTCTCGATTTCGCCGCGGTTTTCCGAACACGAGTCGCGCGC
CTCTTCGCCGCAAACATCGTCGACCTGCTGTTGCGGGCCGCACGGCTGCAT
CCGCATACGGGCGTTCGCTTCATCGCCGCGCAAGCCGAGGAAAAGGGCGCC
TTGTCACGTATCCGGAGCTGCTCGACGAGGCGCGCCGCA TCCTGGGCGG

8.2.4. Séquence du clone B5

La région du chromosome II insérée s'étend de 2739816 à 2742854 pb soit 3038 pb

CTGCTCGCGATGCCGGTGAGCCGCCCGATCGAATGCACGAAGCCGAGCGCG
CGGCTGCGGATTCGGGTGGGGAACACCTCGGACTGGTACGCGGTTCGATTG
CTCGCAAGCACTGTGTTGCACAGCGTGACGAGCACGCCGAACGCGATCACCG
CGGCGGGAGCGCGGACATCGAAAGCCGGTGCCGAACAGCGCGACGCCGAA
CGCGGCCGCGACGATCAGCCATTTGCGCTCGATCCGATCCGCGAGCGCTGC
GGCGGCAAACGGCGAAACCGGATACGCGAACGCGATCACGAACGCATACCAG

AGGCTCTTCATCACCGACGCGCCCTGCGCGGGCAGCAGCGTCGGCAGCCATT
GGCTGAAGCCGAAGAAGCCGATGCTCAGGAATGCGTTGAATACGATCAGCAT
CGCCGTGCGGCCGCGGTTGTGCGGGTCCCACATCGACGCGCGACGCGGCGA
TGCAACCGCATGGCGCGGCGTGTGACGGGCGGCAGCGGCCGCGCCGCT
CGCGCTCGACGGCCGCTTCGAGCCGGCTCAGGATCGCATCGGCCTGCGCGG
CTCGGCCGCGCTCCGCAAGCCAGCGCGGCGACTCGGGCAGGCAGCGACTGCA
GCCACCAGACGACGAGCGCGCCGCTGCCGCCGACGCCACCACCCAGCGCCA
TCCCGATATGCCGAGCGGCGCGCGTGGCACCCACAGCCACGACACGAACGC
GAGCACCGGCATCGCGAGATAGCCGATCGCGAAGCACACCCGCGAACGCGCG
GCCGCGCACGGCTTTCGGCACGAGCTCCGACAGGAACGCGCCGATCGTGAT
CAGCTCCGCGCCGACGCCGCGAGCTCGCGACGAATCGGCACGCGAGAACGCC
GAGCGCGCTCGACTGCACGCACATCGCGACGCTCGCGGCCGTATAGAGCAG
CAGCGAGCCCGTGAACAACGCGCGGGCGCCCGAAGCGGTGCGCGAGGCGCG
ACACGAACATCTCGCCGATGAACAGGCCCGCGAACGTCACCGCGCCGAGCGC
GCCCTGATCCGAGATGCCGAACGCGCCGCGCGCGCCCGCGTCAAAGATGCC
GTCGCGGATCATGCCGGCAGGCAGATAAGTCATCTGGAACAGGTCGTACACC
TCGAAGAAGCCGCCGATGCTCAGCACCCAGACGAGCTTCCAGATCGTCGCGG
TCGCGGGCAGGCGATCGATGCGCGCCGCGAGCGATGCAGCTTGGACATCCG
ACATGGCGTCTCCTCCATGATCCGGCGTTCGTGAGGCGCGCCGGCTATTCGAA
ATTGAGTTGGAATGCTGACGATTCAAGCCGATGGCCGATGCCTCGAAGTCAA
CCGTGCGACCGGATTGACGGGCGAGCCCACCAGCCCGCGAATGTGCAGCG
GCGGGCGCCGTGAGCAGAAACCGGCTGCGCCCTTCCCTTCGCAGCCACTGCG
CCAGCGGCGTCAGCCTCCACAGCTCGCCGAGATGTATGCCAGCTTGAACAG
ACACAGCTCATGCAACGGCATCAGCGCACCCGGCCTCTCCCTCGACTTCAACA
CCCCAGGCCGCTCCTCCACCGCGTGATTGTCCGCCGCCAGCACCCGACAGCCC
GCTCTCGTCCACCCACTTCAGCAGCCGCCCGTCTCTCCGTCCAGCACGCAG
CCCCCGTCAACGACCCAGCCTCCTCTCCAGCAATCGCTCCGCGAACCCCG
TGTGCACGCACACCATGTCCCCACGCTCCACTTCCACGCGATCCGCCTCCAG
CACCCGCATCAGCGCCTCGTAGCCACCTTGCGACGTTTCGTCTCCGTAGTGG
TGCCACAAGTCGATCAGCACCCCGCGACCCTGCACCCCGTCTGCGCCATCA
CCTCTATCCCCAGCGCGTGCAGCGCCTCCCTCGCTGTCCCCCTCCTTCGGCAC
CACCATGTGCTCGCCCAACCGGTAGCCGTTGTAGAACACCACCTCCGCATGA
CGTCGTGTCGCGGTCGAACACCCCGCCACGTGCGACAGCGCATCCCCTG
CGTCGAGTATTGCGAGTGCACGCAGAACGCGTTCGTGCGCACACCAGTCCGTC
GCGTTGCGCGCCGACTCGTCCGCCCGGTAGCCGAAATACGGCCGGCCGCC
AGTTGCGCCGCCATCACCCGCGGCGGCTGCCGCCGAGCGTTCAGCCACCC
CCACGCGGAACATCCAGCGGCAGGCTCAGCGAAAACACCTTCCCCGCCTTCA
CCTCCGCCAGCCCCTCCAGAACCTTGCGCTCCGTCAGCCAGTTCAGCCTGCC
CTTCTCGTCTCATCCCCGAACTCGCCCCAGTTCGACCCCGCCGGGCGCCGT
CTCCACCGTTCGTTCCATCGATGCGTCTCCTCGATCGCGCTGCATCGCGCCG
GTGCTCGGCGCGATGCGCGACTGGCGGCCCTCACTCGGGCCGAATCACGCCG
TTGCGGGCGAGCCGCGCGTACTCGTGCAGCGCCGAGCAACGGCAGCAGAATC
TCGCGGTTGTGCTCGCCAAGCCTCGGCGCGGCACGCCTGAATCCGCCCGGC
GTGCCGGACAGGCGCGGCACGACGCCGTGCATCGCAAGCTGCCAACGTCAT
CGTCGGGCAGCTCGACGAGCGCCTCGCGTTCAATCACGTAGCGATCCTGAAC

GATCTGCGCGATGTCGAGTATCGGCCCGACCGTCACGCCGCTCTTCTCGAAG
AACGCGAGGTTGTCGGCGAGATCGCGTTCGGCGACGAACGCGCCGACGATC
GCATCGAGCTCGCCCGCGTGCTGCACGCGCTGCACGTTTCGTTCGCGTAGCGC
GGATCGTCGATCAGGTCCGCGCGGCCGATCGCGCGAAACAGCCGCTCGGCC
ATGCCCTGCGTCGAGCTCGACAGACACAGCCACTTGTTCGTTCGCGCGTGC
ACGCGTTGCGCGGCCGCGGTGTTTCGACGAACGGCTGCCGGTGCCTCCTTCA
CGCGCCCCGTCAGCACGTAGTTTCGCGGCGGCCCGGCCGAGAATCGAGAACA
GCGGCTCGAA CAGCGATACGTTCGATCACCTGCCCGTTCGCGCCGTCGCG

8.2.5. Séquence du clone B6

La région du chromosome II insérée s'étend de 446857 à 447934 pb soit 1077 pb

CGGCTCGCCCCGCTGCTCGAGAACATATCCCATTCCGCGCACCGTGTGGATC
AGCTTGGGCTCGTAGTCCTGATCGACCTTCGCCCGCACGCGCTTGATCGCCG
CATCGACGACGTTTGAATCGCTGTCTGAAGTTGATGTCCAGATTTGCGAGGGC
ATCAGCGGCGCGGCAGGATCTCGCCCTGCCGGCGCATCAGCAGCCACAGCA
GCAGGAATTCCTTCGAGGTGAGCAGAATCGTGCTGCCCTGGCGGGCGCGCCTT
GCGCTGGACGAGATCGAGCTCGAGATTCGCGACGCGCAGCACCGGCGCGTC
GCTTGATTGCCCGCGGCGGAGCAGCGAGCGCACGCGTGCAGCAACTCGGC
GAAATCGAACGGCTTCGCGAGATAGTCGTTCGCCCGGAGTTTCGAGGCCTTTC
ACGCGATCGTCGACGTGATCGCGCGCGGTGAGGAACAGCACCGGCGTGC
TGCGTTTCTGCGCAGCTCGCGCACACCAGTCCATCCGTCGCGAGCCGGGAGCA
TCACGTTCGAGGATCGCGAGATCGTAGTTTCTCCGTCTGCAACTGGTGCAGACC
GCTGATGCCGTCTGTGTCCAGTCGCGAGACGAAACCCGCCTCCGTAAGCCCA
CGCTTCAAATACATGCCCGTCTTCGGCTCGTCTTCTACGATCAGGACCTTCATT
TCAAACATTAAGAACACGTTAAGTTTTCTTTGAGAATGCGTGTCGTTACAG
TTCAACATTTCCGGTGTCCATGATTAACGTCTGATTCGAAATTCGTTCTTTGCC
GGTGAATGCCGCGATAAAACGTTCGCATCCGAGAACCAATGGACGGCACGAAC
GTGCAACGCGTGGAAGCGGACGCGCGCGAGCGATGAAGCATGAGACGAGCG
CCCGTGTTCGATCGAACATGACAAGAATGTCATGCGGTCGCGACGACAATCGAT
CCGATGCGTTCGATGCCGATGTGCCGTCTCGCGACGTGAGGGAGCGAAAACGA
GAATCGCGGCCGCGATCGGGAAAATCGAAGAAATGCGATGTGCATGTACGCG
ATGAAAAGTCTCGATCGTCACGAGCGGCGCATGC

8.3. ANNEXE C : données supplémentaire relatives au chapitre 3

8.3.1. Séquence nucléotidique des clones 300 et 128

8.3.1.1. Clone 300

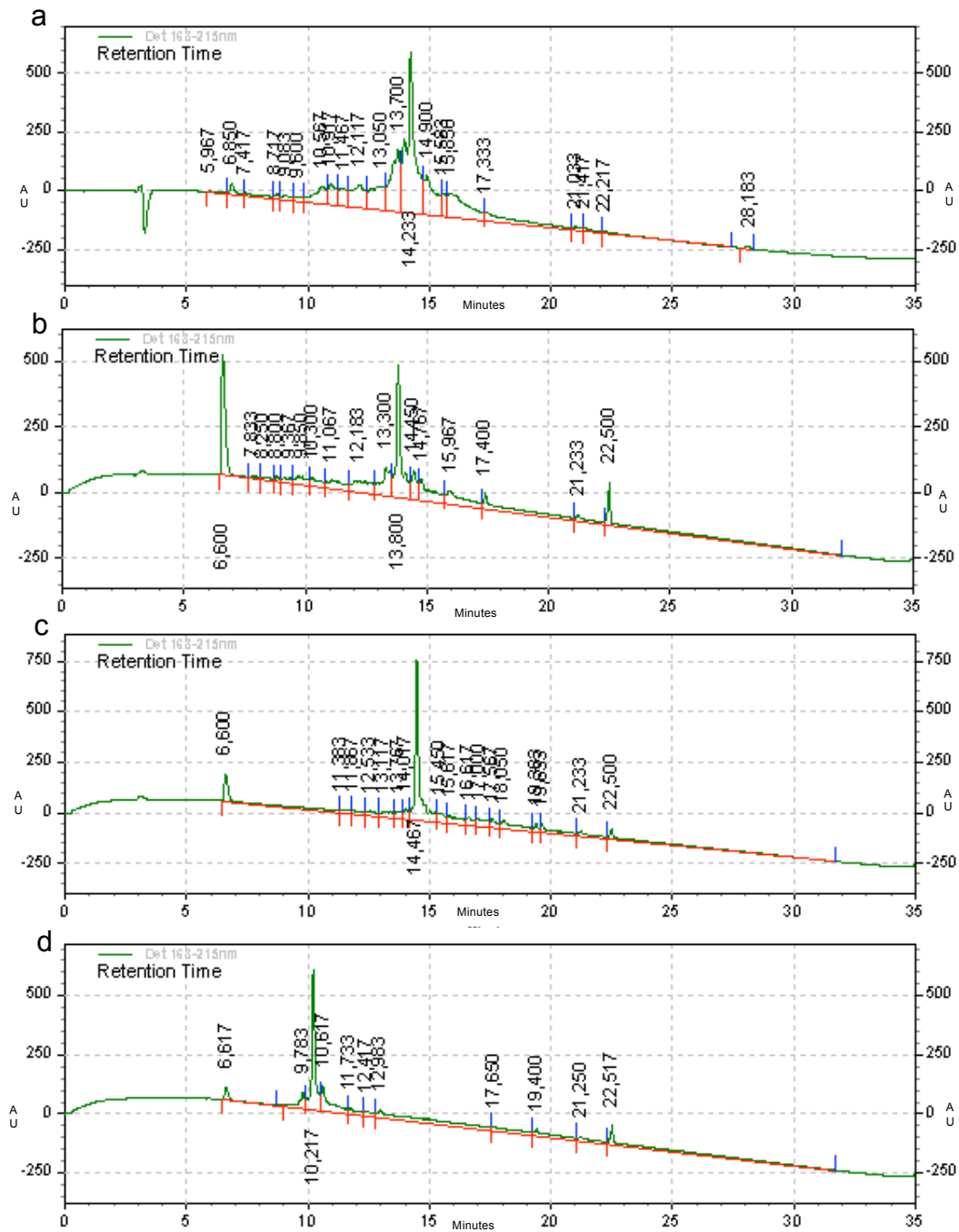
CTGGAGCTGACCGAAGAGTTGGGCAATCTGGCCAAGCAGCACGGCGCCTGG
GTGCGTTTGCACACTACGTCTACCCATATCCCACGGTGGACAACATCCTGCCGCT
GATGGCCGAAGGCAAGATCCTGCCGTACCTGGATGTGCCGCTGCAGCACAGC
CACCCCGATGTCTTGAAGCGTATGAAGCGCCCCGCATCGGGCGAAAAGAATC
TGGACCGCATCCGCGAATGGCGTGCCATTTGCCCCGAGTTGGTGATCCGCTC
TACCTTTATCGCTGGCTTCCCTGGCGAGACCGAGGAAGAGTTTGAGCATCTGC
TCAACTTCATACGCGAAGCCGAGATCGACCGCGCCGGTTGCTTTGCCTATAGC
CCGGTGGAAGGCGCCACCGCCAACGACCTGCCCGGCATG

8.3.1.2. Clone 128

CGCAGGTGCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGCCCTATA
GTGAGTCGTATTACAATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGRAA
AACGAAGGGAGTTTTTTGTTTAGACTCATTGGGGCGAATTCGAGCACGGKGCC
GGCCATCCTCKASAATCGACCTGCGGGCATG

8.3.2. Spectres HPLC et MALDI-TOF de certains peptides

8.3.2.1. Analyse des brutes de clivage par HPLC analytique



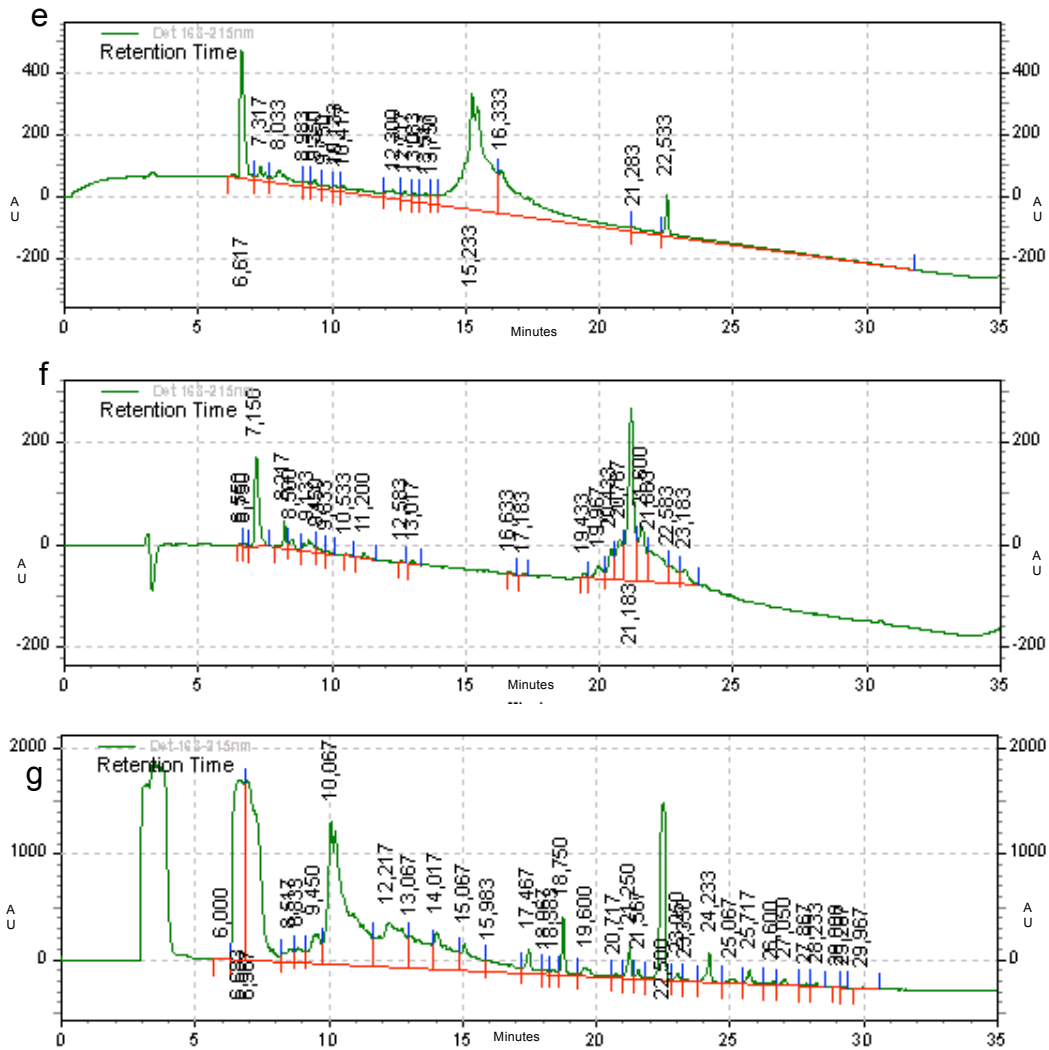
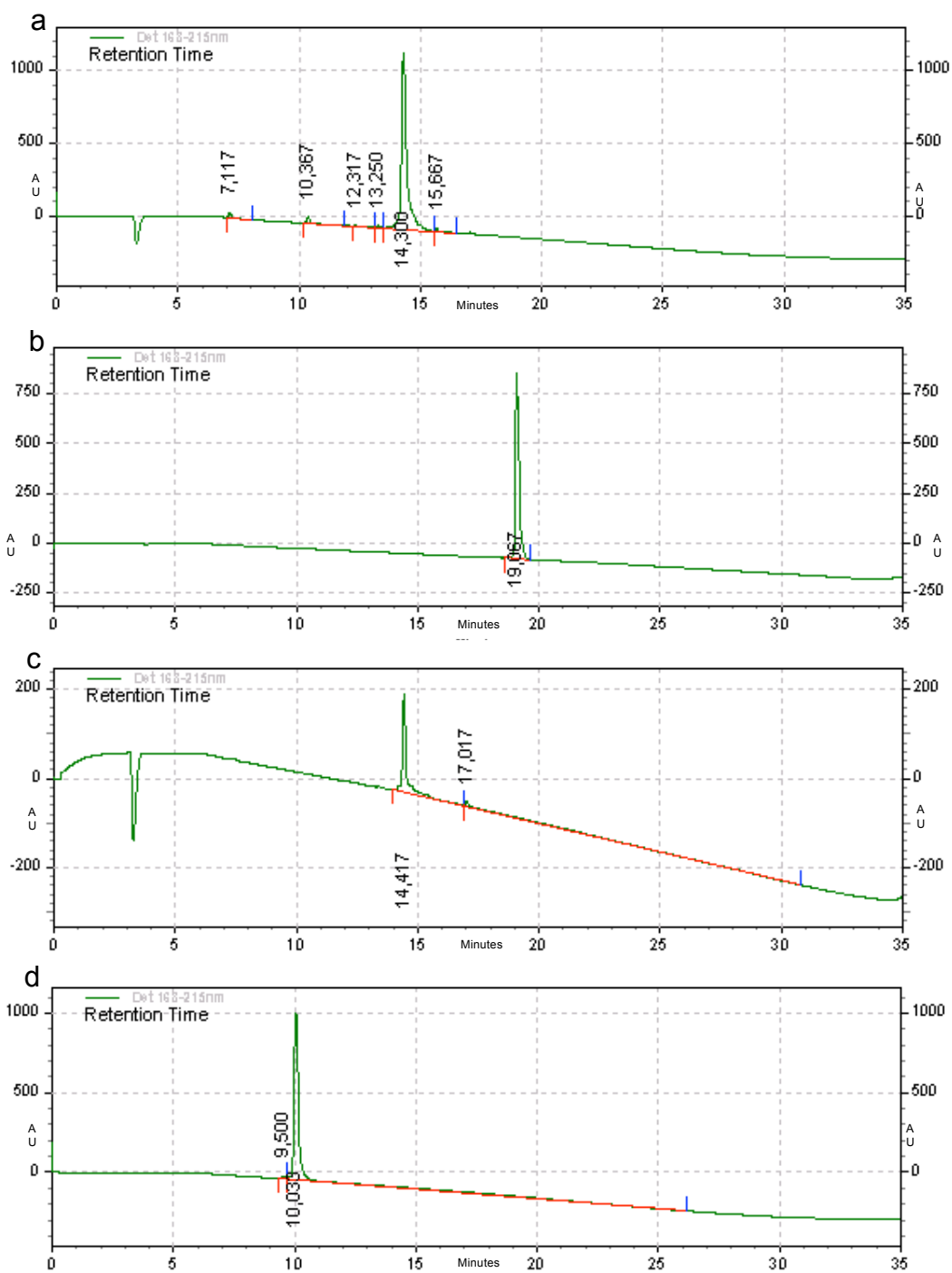


Figure 8.1. HPLC analytique du brut de clivage des peptides sur un gradient linéaire de 0 à 100% acétonitrile/H₂O en 30 minutes sur colonne C18 CSC-Kromasil 100 A/ODS 5 µm, (a) brute de clivage du peptide dérivé de l'ORF 1_h qui a un temps de rétention de 14.23 min, (b) brute de clivage du peptide dérivé de l'ORF 1_i qui a un temps de rétention de 13.80 min, (c) brute de clivage du peptide dérivé de l'ORF 1_j qui a un temps de rétention de 14.46 min, (d) brute de clivage du peptide dérivé de l'ORF 1_k qui a un temps de rétention de 10.21 min, (e) brute de clivage du peptide dérivé de l'ORF 5 qui a un temps de rétention de 15.23 min, (f) brute de clivage du peptide dérivé de l'ORF 6 qui a un temps de rétention de 21.18 min, (g) brute de clivage du peptide dérivé de l'ORF 9 qui a un temps de rétention de 10.60 min.

8.3.2.2. HPLC des peptides après purification sur colonne



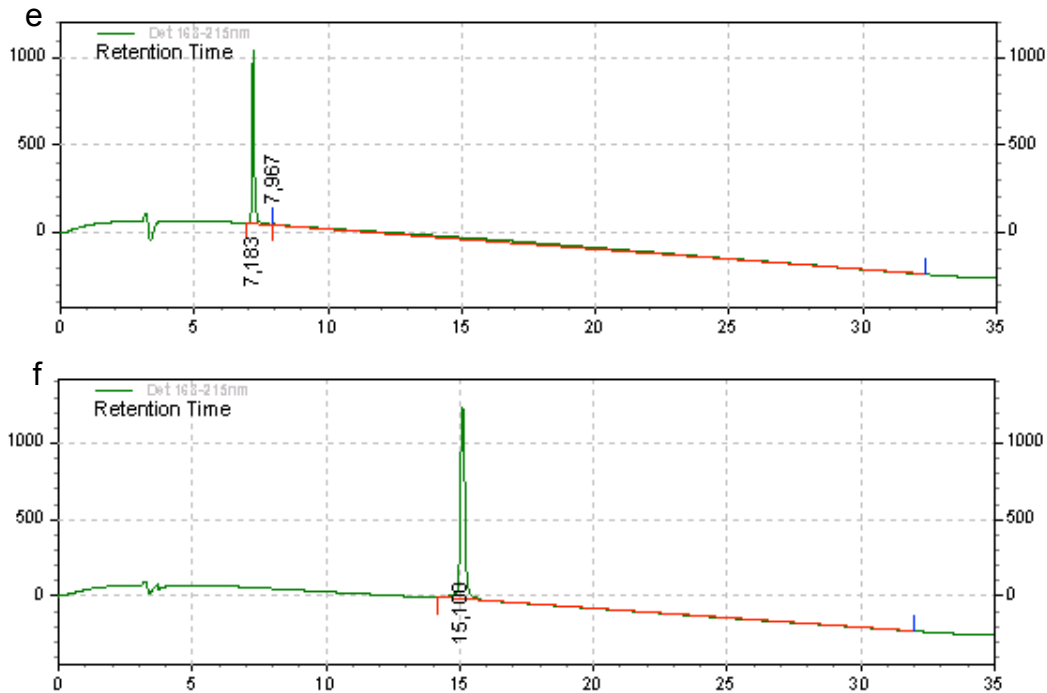


Figure 8.2. HPLC analytique des peptides sur un gradient linéaire de 0 à 100% acétonitrile/H₂O en 30 minutes sur colonne C18 CSC-Kromasil 100 A/ODS 5 μ m, les peptides sont purs à 98%, (a) peptide dérivé de l'ORF 1_h qui a un temps de rétention de 14.30 min, (b) peptide dérivé de l'ORF 1_i qui a un temps de rétention de 19.06 min sur un gradient linéaire de 0 à 60% acétonitrile/H₂O en 30 minutes, (c) peptide dérivé de l'ORF 1_j qui a un temps de rétention de 14.41 min, (d) peptide dérivé de l'ORF 1_k qui a un temps de rétention de 10.03 min, (e) peptide dérivé de l'ORF 5 qui a un temps de rétention de 7.83 min, (f) peptide dérivé de l'ORF 6 qui a un temps de rétention de 15.10 min.

8.3.2.3. BLASTn et BLASTx de la séquence de l'insert MHW₅₃₈

[BLAST®](#) » [blastn suite](#) » RID-XUE3B0V5014

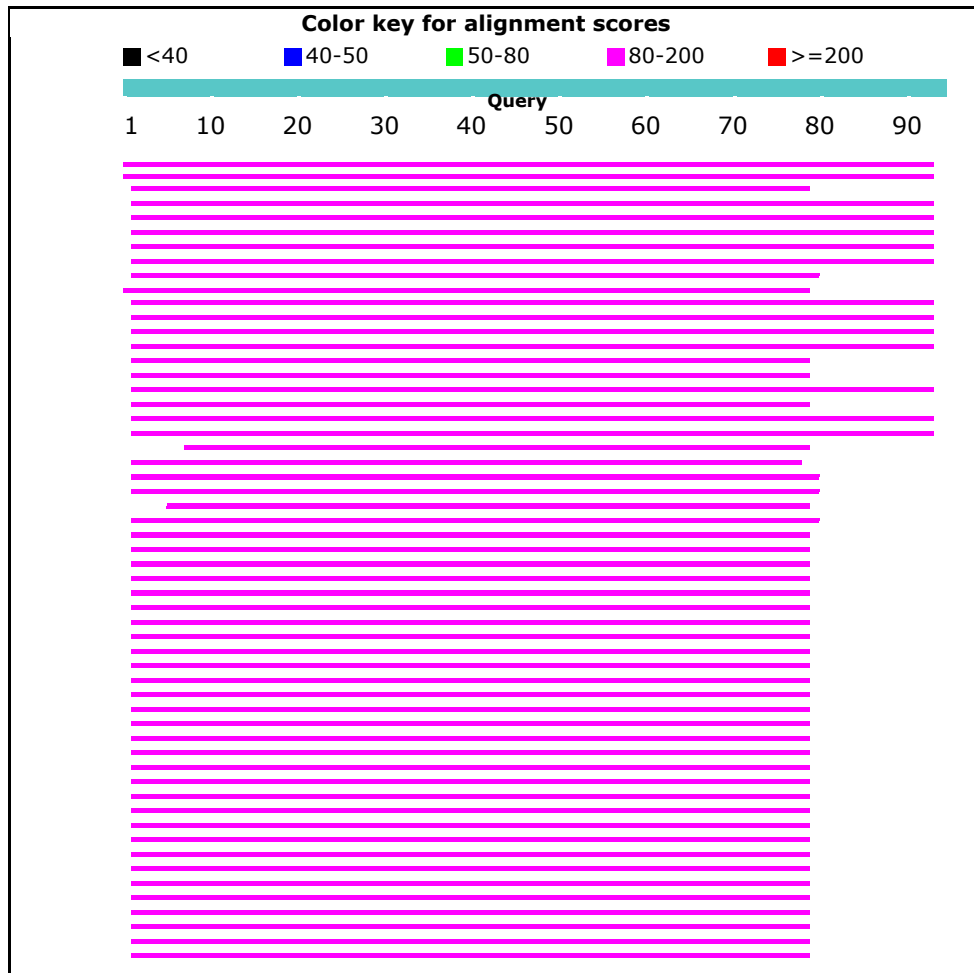
BLAST Results

Job title: Nucleotide Sequence (91 letters)

RID	XUE3B0V5014 (Expires on 10-12 05:15 am)	Database Name	nr
Query ID	lcl Query_114597	Description	Nucleotide collection (nt)
Description	None	Program	BLASTN 2.7.0+
Molecule type	nucleic acid		
Query Length	91		

Graphic Summary

Distribution of the top 100 Blast Hits on 100 subject sequences



Descriptions

Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium sp. NRRL B-3805, complete genome	127	127	100%	4e-26	93%	CP011022.1
Mycobacterium neoaurum VKM Ac-1815D, complete genome	127	127	100%	4e-26	93%	CP006936.2
Saccharopolyspora erythraea NRRL2338 complete genome	118	118	83%	2e-23	95%	AM420293.1
Mycobacterium dioxanotrophicus strain PH-06, complete genome	116	116	98%	6e-23	90%	CP020809.1
Actinoplanes derwentensis strain DSM 43941 genome assembly, chromosome: I	116	116	98%	6e-23	90%	LT629758.1
Rhodococcus sp. PBTS1, complete genome	116	116	98%	6e-23	90%	CP015219.1
Mycobacterium sp. EPa45, complete genome	116	116	98%	6e-23	90%	CP011773.1
Nocardia brasiliensis ATCC 700358, complete genome	116	116	98%	6e-23	90%	CP003876.1
Streptomyces clavuligerus strain F613-1, complete genome	114	114	84%	2e-22	94%	CP016559.1
Mycobacterium canettii CIPT 140070017 complete genome	114	114	84%	2e-22	94%	FO203510.1
Pseudonocardia sp. HH130630-07, complete genome	113	113	98%	8e-22	89%	CP013854.1
Mycobacterium fortuitum strain CT6, complete genome	113	113	98%	8e-22	89%	CP011269.1
Actinoplanes friuliensis DSM 7358, complete genome	113	113	98%	8e-22	89%	CP006272.1
Actinoplanes sp. N902-109, complete genome	113	113	98%	8e-22	89%	CP005929.1
Mycobacterium canettii CIPT 140070010 complete genome	113	113	83%	8e-22	94%	FO203509.1
Modestobacter marinus str. BC501 chromosome, complete genome	113	113	83%	8e-22	94%	FO203431.1
Rhodococcus equi 103S chromosome	113	113	98%	8e-22	89%	FN563149.1
Thermobispora bispora DSM 43833, complete genome	113	113	83%	8e-22	94%	CP001874.1
Mycobacterium sp. KMS, complete genome	113	113	98%	8e-22	89%	CP000518.1
Mycobacterium sp. MCS, complete genome	113	113	98%	8e-22	89%	CP000384.1
Streptomyces autolyticus strain CGMCC0516, complete genome	111	111	76%	3e-21	96%	CP019458.1
Nonomuraea sp. ATCC 39727 isolate nono1 genome assembly, chromosome: I	111	111	82%	3e-21	93%	LT559118.1
Streptomyces sp. TLI_053 genome assembly, chromosome: I	111	111	84%	3e-21	92%	LT629775.1
Micromonospora echinospora strain DSM 43816 genome assembly, chromosome: I	111	111	84%	3e-21	92%	LT607413.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
Streptomyces violaceusniger Tu 4113, complete genome	111	111	79%	3e-21	95%	CP002994.1
Frankia alni str. ACN14A chromosome, complete sequence	111	111	84%	3e-21	92%	CT573213.2
Mycobacterium tuberculosis strain TBV4952 chromosome, complete genome	109	109	83%	1e-20	92%	CP023640.1
Mycobacterium tuberculosis strain TBV4768 chromosome, complete genome	109	109	83%	1e-20	92%	CP023639.1
Mycobacterium tuberculosis strain TBV4766 chromosome, complete genome	109	109	83%	1e-20	92%	CP023638.1
Mycobacterium tuberculosis strain TBDM2717 chromosome, complete genome	109	109	83%	1e-20	92%	CP023637.1
Mycobacterium tuberculosis strain TBDM2699 chromosome, complete genome	109	109	83%	1e-20	92%	CP023636.1
Mycobacterium tuberculosis strain TBDM2489 chromosome, complete genome	109	109	83%	1e-20	92%	CP023635.1
Mycobacterium tuberculosis strain TBDM2487 chromosome, complete genome	109	109	83%	1e-20	92%	CP023634.1
Mycobacterium tuberculosis strain TBDM2444 chromosome, complete genome	109	109	83%	1e-20	92%	CP023633.1
Mycobacterium tuberculosis strain TBDM2189 chromosome, complete genome	109	109	83%	1e-20	92%	CP023632.1
Mycobacterium tuberculosis strain TBDM1506 chromosome, complete genome	109	109	83%	1e-20	92%	CP023631.1
Mycobacterium tuberculosis strain MDRMA2441 chromosome, complete genome	109	109	83%	1e-20	92%	CP023630.1
Mycobacterium tuberculosis strain MDRMA2260 chromosome, complete genome	109	109	83%	1e-20	92%	CP023629.1
Mycobacterium tuberculosis strain MDRMA2082 chromosome, complete genome	109	109	83%	1e-20	92%	CP023628.1
Mycobacterium tuberculosis strain MDRMA2019 chromosome, complete genome	109	109	83%	1e-20	92%	CP023627.1
Mycobacterium tuberculosis strain MDRMA1565 chromosome, complete genome	109	109	83%	1e-20	92%	CP023626.1
Mycobacterium tuberculosis strain MDRMA863 chromosome, complete genome	109	109	83%	1e-20	92%	CP023625.1
Mycobacterium tuberculosis strain MDRMA701 chromosome, complete genome	109	109	83%	1e-20	92%	CP023624.1
Mycobacterium tuberculosis strain MDRMA203 chromosome, complete genome	109	109	83%	1e-20	92%	CP023623.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium tuberculosis strain MDRDM827 chromosome, complete genome	109	109	83%	1e-20	92%	CP023622.1
Mycobacterium tuberculosis strain LN2900 chromosome, complete genome	109	109	83%	1e-20	92%	CP023621.1
Mycobacterium tuberculosis strain LN1856 chromosome, complete genome	109	109	83%	1e-20	92%	CP023620.1
Mycobacterium tuberculosis strain LN1100 chromosome, complete genome	109	109	83%	1e-20	92%	CP023619.1
Mycobacterium tuberculosis strain LN3695 chromosome, complete genome	109	109	83%	1e-20	92%	CP023618.1
Mycobacterium tuberculosis strain LN3672 chromosome, complete genome	109	109	83%	1e-20	92%	CP023617.1
Mycobacterium tuberculosis strain LN3668 chromosome, complete genome	109	109	83%	1e-20	92%	CP023616.1
Mycobacterium tuberculosis strain LN3589 chromosome, complete genome	109	109	83%	1e-20	92%	CP023615.1
Mycobacterium tuberculosis strain LN3588 chromosome, complete genome	109	109	83%	1e-20	92%	CP023614.1
Mycobacterium tuberculosis strain LN3584 chromosome, complete genome	109	109	83%	1e-20	92%	CP023613.1
Mycobacterium tuberculosis strain LN2978 chromosome, complete genome	109	109	83%	1e-20	92%	CP023612.1
Mycobacterium tuberculosis strain LN763 chromosome, complete genome	109	109	83%	1e-20	92%	CP023611.1
Mycobacterium tuberculosis strain LN317 chromosome, complete genome	109	109	83%	1e-20	92%	CP023610.1
Mycobacterium tuberculosis strain LN55 chromosome, complete genome	109	109	83%	1e-20	92%	CP023609.1
Mycobacterium tuberculosis strain LE410 chromosome, complete genome	109	109	83%	1e-20	92%	CP023608.1
Mycobacterium tuberculosis strain LE371 chromosome, complete genome	109	109	83%	1e-20	92%	CP023607.1
Mycobacterium tuberculosis strain LE103 chromosome, complete genome	109	109	83%	1e-20	92%	CP023606.1
Mycobacterium tuberculosis strain LE79 chromosome, complete genome	109	109	83%	1e-20	92%	CP023605.1
Mycobacterium tuberculosis strain LE76 chromosome, complete genome	109	109	83%	1e-20	92%	CP023604.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium tuberculosis strain LE63 chromosome, complete genome	109	109	83%	1e-20	92%	CP023603.1
Mycobacterium tuberculosis strain LE13 chromosome, complete genome	109	109	83%	1e-20	92%	CP023602.1
Mycobacterium tuberculosis strain CSV9577 chromosome, complete genome	109	109	83%	1e-20	92%	CP023601.1
Mycobacterium tuberculosis strain CSV3611 chromosome, complete genome	109	109	83%	1e-20	92%	CP023600.1
Mycobacterium tuberculosis strain CSV383 chromosome, complete genome	109	109	83%	1e-20	92%	CP023599.1
Mycobacterium tuberculosis strain SLM100 chromosome, complete genome	109	109	83%	1e-20	92%	CP023598.1
Mycobacterium tuberculosis strain SLM088 chromosome, complete genome	109	109	83%	1e-20	92%	CP023597.1
Mycobacterium tuberculosis strain SLM063 chromosome, complete genome	109	109	83%	1e-20	92%	CP023596.1
Mycobacterium tuberculosis strain SLM060 chromosome, complete genome	109	109	83%	1e-20	92%	CP023595.1
Mycobacterium tuberculosis strain SLM056 chromosome, complete genome	109	109	83%	1e-20	92%	CP023594.1
Mycobacterium tuberculosis strain SLM040 chromosome, complete genome	109	109	83%	1e-20	92%	CP023593.1
Mycobacterium tuberculosis strain SLM036 chromosome, complete genome	109	109	83%	1e-20	92%	CP023592.1
Mycobacterium tuberculosis strain TBV5365 chromosome, complete genome	109	109	83%	1e-20	92%	CP023591.1
Mycobacterium tuberculosis strain TBV5362 chromosome, complete genome	109	109	83%	1e-20	92%	CP023590.1
Mycobacterium tuberculosis strain TBV5000 chromosome, complete genome	109	109	83%	1e-20	92%	CP023589.1
Mycobacterium tuberculosis strain TBDM425 chromosome, complete genome	109	109	83%	1e-20	92%	CP023588.1
Mycobacterium tuberculosis strain ME1473 chromosome, complete genome	109	109	83%	1e-20	92%	CP023587.1
Mycobacterium tuberculosis strain MDRMA2491 chromosome, complete genome	109	109	83%	1e-20	92%	CP023586.1
Mycobacterium tuberculosis strain MDRDM1098 chromosome, complete genome	109	109	83%	1e-20	92%	CP023585.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium tuberculosis strain MDRDM627 chromosome, complete genome	109	109	83%	1e-20	92%	CP023584.1
Mycobacterium tuberculosis strain MDRDM260 chromosome, complete genome	109	109	83%	1e-20	92%	CP023583.1
Mycobacterium tuberculosis strain LN3756 chromosome, complete genome	109	109	83%	1e-20	92%	CP023582.1
Mycobacterium tuberculosis strain LN2358 chromosome, complete genome	109	109	83%	1e-20	92%	CP023581.1
Mycobacterium tuberculosis strain LN180 chromosome, complete genome	109	109	83%	1e-20	92%	CP023580.1
Mycobacterium tuberculosis strain LE492 chromosome, complete genome	109	109	83%	1e-20	92%	CP023579.1
Mycobacterium tuberculosis strain LE486 chromosome, complete genome	109	109	83%	1e-20	92%	CP023578.1
Mycobacterium tuberculosis strain CSV11678 chromosome, complete genome	109	109	83%	1e-20	92%	CP023577.1
Mycobacterium tuberculosis strain CSV10399 chromosome, complete genome	109	109	83%	1e-20	92%	CP023576.1
Mycobacterium tuberculosis strain CSV5769 chromosome, complete genome	109	109	83%	1e-20	92%	CP023575.1
Mycobacterium tuberculosis strain CSV4644 chromosome, complete genome	109	109	83%	1e-20	92%	CP023574.1
Mycobacterium tuberculosis strain CSV4519 chromosome, complete genome	109	109	83%	1e-20	92%	CP023573.1
Actinosynnema pretiosum strain X47 chromosome, complete genome	109	109	83%	1e-20	92%	CP023445.1
Mycobacterium tuberculosis strain C3 chromosome	109	109	83%	1e-20	92%	CP023170.1
Mycobacterium tuberculosis strain S3 chromosome	109	109	83%	1e-20	92%	CP023169.1
Mycobacterium avium subsp. paratuberculosis strain JII-1961 genome	109	109	83%	1e-20	92%	CP022105.1
Mycobacterium avium subsp. paratuberculosis strain FDAARGOS_305, complete genome	109	109	83%	1e-20	92%	CP022095.1
Mycobacterium tuberculosis strain MTB2, complete genome	109	109	83%	1e-20	92%	CP022014.1

Alignments

Mycobacterium sp. NRRL B-3805, complete genome

Sequence ID: **CP011022.1** Length: 5421338 Number of Matches: 1

Range 1: 3033704 to 3033797

Score	Expect	Identities	Gaps	Strand	Frame
127 bits(140)	4e-26()	87/94(93%)	3/94(3%)	Plus/Minus	

Features:

cobalamin biosynthesis protein CobN

Query	1	ATGCGCCCCGCGCTTGGCCTGGGTGCCCTCGCCCCGGT-GTTGACCAGGAACGGGTAGAT	59
Sbjct	3033797	ATGCGCCCCGCTCGCTTGGCCTGGGTGCCCTCGCCAGGGTCGTTGACCAGGAACGGGTAGAT	3033738
Query	60	CAGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91
Sbjct	3033737	CAACGGCAGATCGCCGAGCGCCGCTCGGTCCCG	3033704

Mycobacterium neoaurum VKM Ac-1815D, complete genome

Sequence ID: **CP006936.2** Length: 5421267 Number of Matches: 1

Range 1: 3033705 to 3033798

Score	Expect	Identities	Gaps	Strand	Frame
127 bits(140)	4e-26()	87/94(93%)	3/94(3%)	Plus/Minus	

Features:

cobalamin biosynthesis protein CobN

Query	1	ATGCGCCCCGCGCTTGGCCTGGGTGCCCTCGCCCCGGT-GTTGACCAGGAACGGGTAGAT	59
Sbjct	3033798	ATGCGCCCCGCTCGCTTGGCCTGGGTGCCCTCGCCAGGGTCGTTGACCAGGAACGGGTAGAT	3033739
Query	60	CAGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91
Sbjct	3033738	CAACGGCAGATCGCCGAGCGCCGCTCGGTCCCG	3033705

Saccharopolyspora erythraea NRRL2338 complete genome

Sequence ID: **AM420293.1** Length: 8212805 Number of Matches: 1

Range 1: 6683671 to 6683747

Score	Expect	Identities	Gaps	Strand	Frame
118 bits(130)	2e-23()	73/77(95%)	1/77(1%)	Plus/Minus	

Features:

cobalamin biosynthesis protein N

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCCGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	6683747	TGCACCCGGCGCTTGGCCTGGGTGCCCTCGCCCGGGTCGTTGACCAGGAACGGGTAGATC	6683688
Query	61	AGCGGCAGATCGCCGAG	77
Sbjct	6683687	AGCGGCAGGTCGCCGAG	6683671

Mycobacterium dioxanotrophicus strain PH-06, complete genome

Sequence ID: **CP020809.1** Length: 7595921 Number of Matches: 1

Range 1: 2953226 to 2953318

Score	Expect	Identities	Gaps	Strand	Frame
116 bits(128)	6e-23()	84/93(90%)	3/93(3%)	Plus/Minus	

Features:

cobaltochelataase subunit CobN

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCCGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	2953318	TGGGCCCGGCGTTTGGCCTGGGTGCCCTCCCCGGGTCGTTGACCAGGAACGGGTAGATC	2953259
Query	61	AGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91
Sbjct	2953258	AGCGGCAGATCGCCGAGTGCAGCGTCCGTGCCG	2953226

Actinoplanes derwentensis strain DSM 43941 genome assembly, chromosome: I

Sequence ID: **LT629758.1** Length: 10654181 Number of Matches: 1
 Range 1: 9282086 to 9282178

Score	Expect	Identities	Gaps	Strand	Frame
116 bits(128)	6e-23()	84/93(90%)	3/93(3%)	Plus/Minus	

Features:
cobaltochelatase CobN subunit

Query	2	TGCGCCC	CGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	9282178	TGCGCCC	CGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT	9282119
Query	61	AGCGGCAGATCGCCGA-GGCAGC-TCGGTCCCG	91	
Sbjct	9282118	AGCGGCAGGTACCGAGGGCCGCGTCGGTGCCG	9282086	

Rhodococcus sp. PBTS1, complete genome
 Sequence ID: **CP015219.1** Length: 4251687 Number of Matches: 1
 Range 1: 1893406 to 1893498

Score	Expect	Identities	Gaps	Strand	Frame
116 bits(128)	6e-23()	84/93(90%)	3/93(3%)	Plus/Minus	

Features:

Query	2	TGCGCCC	CGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	1893498	TGGGCC	CGGCGCTTGGCCTGGGTCCCTCACCGGGT	1893439
Query	61	AGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91	
Sbjct	1893438	AGCGGCAGATCGCCGAGCGGGCGTCGGTCCCG	1893406	

Mycobacterium sp. EPa45, complete genome
 Sequence ID: **CP011773.1** Length: 6177406 Number of Matches: 1
 Range 1: 3108474 to 3108566

Score	Expect	Identities	Gaps	Strand	Frame
116 bits(128)	6e-23()	84/93(90%)	3/93(3%)	Plus/Minus	

Features:
cobalamin biosynthesis protein CobN

Query	2	TGCGCCC	CGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	3108566	TGCGCGC	GGCGCTTGGCCTGGGTGCCCTCCCCGGGT	3108507
Query	61	AGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91	
Sbjct	3108506	AGCGGCAGGTGCGCCGAGCGGGCGTCGGTCCCG	3108474	

Nocardia brasiliensis ATCC 700358, complete genome
 Sequence ID: **CP003876.1** Length: 9436348 Number of Matches: 1
 Range 1: 6081252 to 6081344

Score	Expect	Identities	Gaps	Strand	Frame
116 bits(128)	6e-23()	84/93(90%)	3/93(3%)	Plus/Plus	

Features:
cobaltochelatase subunit CobN

Query	2	TGCGCCC	CGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	6081252	TGCGCCC	GGCGCTTGGCCTGGGTGCCCTCGCCGGGATCGTTCACCAGGAACGGGTAGATC	6081311
Query	61	AGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91	
Sbjct	6081312	AGCGGCAGATCGCCGAGCGGGCGTCGGTGCCG	6081344	

Streptomyces clavuligerus strain F613-1, complete genome
 Sequence ID: **CP016559.1** Length: 6883702 Number of Matches: 1
 Range 1: 3844546 to 3844623

Score	Expect	Identities	Gaps	Strand	Frame
114 bits(126)	2e-22()	73/78(94%)	1/78(1%)	Plus/Minus	

Features:
cobaltochelataase subunit CobN

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	3844623	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGTTCGTTACCAGGAACGGGTAGACG	3844564
Query	61	AGCGGCAGATCGCCGAGG	78
Sbjct	3844563	AGCGGCAGATCGCCGAGG	3844546

Mycobacterium canettii CIPT 140070017 complete genome
 Sequence ID: **FO203510.1** Length: 4524466 Number of Matches: 1
 Range 1: 2409427 to 2409504

Score	Expect	Identities	Gaps	Strand	Frame
114 bits(126)	2e-22()	73/78(94%)	1/78(1%)	Plus/Plus	

Features:

Query	1	ATGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGAT	59
Sbjct	2409427	ATGCGCGCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGTTCGTTACCAGAAACGGGTAGAT	2409486
Query	60	CAGCGGCAGATCGCCGAG	77
Sbjct	2409487	CAGCGGCAGATCGCCGAG	2409504

Pseudonocardia sp. HH130630-07, complete genome
 Sequence ID: **CP013854.1** Length: 6186048 Number of Matches: 1
 Range 1: 4876365 to 4876457

Score	Expect	Identities	Gaps	Strand	Frame
113 bits(124)	8e-22()	83/93(89%)	3/93(3%)	Plus/Minus	

Features:
3-hydroxy-2-methylbutyryl-CoA dehydrogenasecobalamin biosynthesis protein CobN

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	4876457	TGCGCGCGCCGCTTGGCCTGGGTGCCCTCCCCCGGGTTCGTTACCAGGAACGGGTAGATC	4876398
Query	61	AGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91
Sbjct	4876397	AGCGGCAGGTCGCCGAGCACGGCGTCGGTCCCG	4876365

Mycobacterium fortuitum strain CT6, complete genome
 Sequence ID: **CP011269.1** Length: 6254616 Number of Matches: 1
 Range 1: 3644236 to 3644328

Score	Expect	Identities	Gaps	Strand	Frame
113 bits(124)	8e-22()	83/93(89%)	3/93(3%)	Plus/Plus	

Features:
CobN component of cobalt chelatase

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	3644236	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCCCCCGGATCGTTACCAGGAACGGGTAGATC	3644295
Query	61	AGCGGCAGATCGCCGA-GGCAGC-TCGGTCCCG	91
Sbjct	3644296	AGCGGAAGATCGCCGAGGGCGGCGTCGGTGCCG	3644328

Actinoplanes friuliensis DSM 7358, complete genome
 Sequence ID: **CP006272.1** Length: 9376071 Number of Matches: 1
 Range 1: 7410528 to 7410620

Score	Expect	Identities	Gaps	Strand	Frame
113 bits(124)	8e-22()	83/93(89%)	3/93(3%)	Plus/Plus	

Features:
cobaltochelatase subunit CobN

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	7410528	TGGGCGCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGTCTGTTGACCAGGAACGGGTAGATC	7410587
Query	61	AGCGGCAGATCGCCGA-GGCAGC-TCGGTCCCG	91
Sbjct	7410588	AGCGGCAGGTCACCGAGGGCCGCTCGGTGCCG	7410620

Actinoplanes sp. N902-109, complete genome
 Sequence ID: **CP005929.1** Length: 9228054 Number of Matches: 1
 Range 1: 7480589 to 7480681

Score	Expect	Identities	Gaps	Strand	Frame
113 bits(124)	8e-22()	83/93(89%)	3/93(3%)	Plus/Plus	

Features:
cobaltochelatase

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	7480589	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCACCCGGTCTGTTGACCAGGAACGGGTAGATC	7480648
Query	61	AGCGGCAGATCGCCGAG-GCAGC-TCGGTCCCG	91
Sbjct	7480649	AGCGGCAGGTTGCCACGCGCGCTCGGTGCCG	7480681

Mycobacterium canettii CIPT 140070010 complete genome
 Sequence ID: **FO203509.1** Length: 4525948 Number of Matches: 1
 Range 1: 2438387 to 2438463

Score	Expect	Identities	Gaps	Strand	Frame
113 bits(124)	8e-22()	72/77(94%)	1/77(1%)	Plus/Plus	

Features:

Query	2	TGCGCCCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGT-GTTGACCAGGAACGGGTAGATC	60
Sbjct	2438387	TGCGCGCGCCGCTTGGCCTGGGTGCCCTCGCCCGGGTCTGTTGACCAGAAACGGGTAGATC	2438446
Query	61	AGCGGCAGATCGCCGAG	77
Sbjct	2438447	AGCGGCAGATCGCCAG	2438463

BLAST is a registered trademark of the National Library of Medicine



[Support center Mailing list](#) [YouTube](#)



• [National Library Of Medicine](#)



• [National Institutes Of Health](#)



- [U.S. Department of Health & Human Services](#)



- [USA.gov](#)

NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA
[Policies and Guidelines](#) | [Contact](#)

BLAST® » blastx » RID-XU3G0MBR014

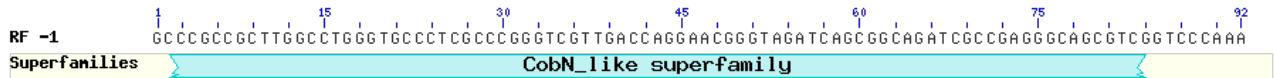
BLAST Results

Job title: Nucleotide Sequence (89 letters)

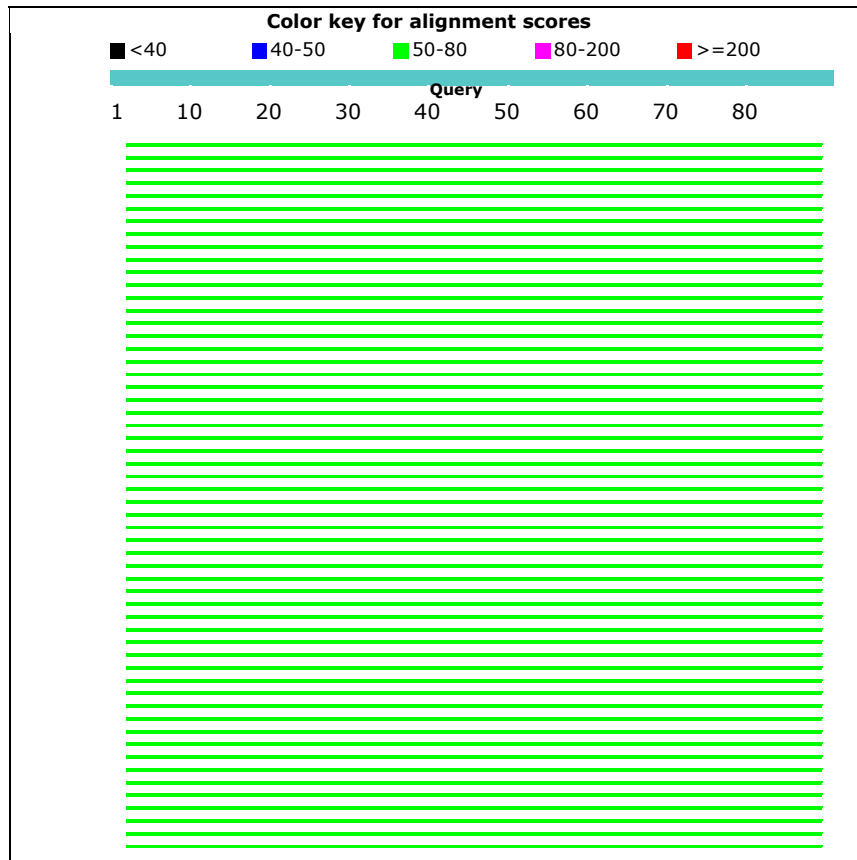
RID	XU3G0MBR014 (Expires on 10-12 02:14 am)	Database Name	nr
Query ID	Id Query_367019	Description	All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Description	None	Program	BLASTX 2.7.0+
Molecule type	nucleic acid		
Query Length	89		

Graphic Summary

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of the top 100 Blast Hits on 100 subject sequences



Descriptions

Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
cobaltochelatae subunit CobN [Rhodococcus fascians]	63.2	63.2	97%	9e-11	100%	WP_037181271.1
cobaltochelatae subunit CobN [Rhodococcus sp. 114MFTsu3.1]	63.2	63.2	97%	9e-11	100%	WP_020110056.1
cobaltochelatae subunit CobN [Rhodococcus sp. 06-621-2]	63.2	63.2	97%	9e-11	100%	WP_094724882.1
cobaltochelatae subunit CobN [Rhodococcus sp. RS1C4]	63.2	63.2	97%	9e-11	100%	WP_094728646.1
cobaltochelatae subunit CobN [Rhodococcus sp. 14-2470-1b]	63.2	63.2	97%	9e-11	100%	WP_094721623.1
MULTISPECIES: cobaltochelatae subunit CobN [Rhodococcus]	63.2	63.2	97%	9e-11	100%	WP_033189721.1
cobaltochelatae subunit CobN [Rhodococcus sp. 06-418-1B]	63.2	63.2	97%	9e-11	100%	WP_094712849.1
cobaltochelatae subunit CobN [Rhodococcus sp. 15-1154-1]	63.2	63.2	97%	1e-10	100%	WP_094732627.1
cobaltochelatae subunit CobN [Rhodococcus sp. 06-1059B-a]	63.2	63.2	97%	1e-10	100%	WP_094628196.1
cobaltochelatae subunit CobN [Nocardia crassostreae]	63.2	63.2	97%	1e-10	100%	WP_067538688.1
cobaltochelatae subunit CobN [Rhodococcus sp. 14-2470-1a]	63.2	63.2	97%	1e-10	100%	WP_094655786.1
cobaltochelatae subunit CobN [Rhodococcus sp. 29MFTsu3.1]	63.2	63.2	97%	1e-10	100%	WP_019666247.1
cobaltochelatae subunit CobN [Rhodococcus fascians]	63.2	63.2	97%	1e-10	100%	WP_094652804.1
cobaltochelatae subunit CobN [Rhodococcus sp. 14-1411-2a]	63.2	63.2	97%	1e-10	100%	WP_094688650.1
cobaltochelatae subunit CobN [Rhodococcus fascians]	63.2	63.2	97%	1e-10	100%	WP_037163206.1
cobaltochelatae subunit CobN [Rhodococcus sp. 15-649-1-2]	63.2	63.2	97%	1e-10	100%	WP_094711009.1
cobaltochelatae subunit CobN [Rhodococcus fascians]	63.2	63.2	97%	1e-10	100%	WP_037142244.1
MULTISPECIES: cobaltochelatae subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_037113312.1
cobaltochelatae subunit CobN [Rhodococcus sp. 14-2483-1-2]	63.2	63.2	97%	1e-10	100%	WP_094741942.1
cobaltochelatae subunit CobN [Rhodococcus sp. 14-2496-1d]	63.2	63.2	97%	1e-10	100%	WP_094685769.1
cobaltochelatae subunit CobN [Rhodococcus sp. 14-2483-1-1]	63.2	63.2	97%	1e-10	100%	WP_094739074.1
cobaltochelatae subunit CobN [Rhodococcus sp. 1163]	63.2	63.2	97%	1e-10	100%	WP_084348911.1
cobaltochelatae subunit CobN [Nocardia soli]	63.2	63.2	97%	1e-10	100%	WP_071927686.1
cobaltochelatae subunit CobN [Rhodococcus qingshengii]	63.2	63.2	97%	1e-10	100%	WP_064688188.1
cobaltochelatae subunit CobN [Rhodococcus sp. 1168]	63.2	63.2	97%	1e-10	100%	WP_088943079.1
cobaltochelatae subunit CobN [Rhodococcus yunnanensis]	63.2	63.2	97%	1e-10	100%	WP_072801801.1
MULTISPECIES: cobaltochelatae subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_094641795.1
cobaltochelatae subunit CobN [Rhodococcus qingshengii]	63.2	63.2	97%	1e-10	100%	WP_083057138.1
MULTISPECIES: cobaltochelatae subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_064074789.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_063316391.1
cobaltochelatase subunit CobN [Rhodococcus sp. KB6]	63.2	63.2	97%	1e-10	100%	WP_058038798.1
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_030536358.1
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_042919753.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_003944890.1
cobaltochelatase subunit CobN [Rhodococcus sp. ADH]	63.2	63.2	97%	1e-10	100%	WP_054188251.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_042450955.1
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_021344299.1
cobaltochelatase subunit CobN [Rhodococcus qingshengii]	63.2	63.2	97%	1e-10	100%	WP_007729837.1
cobalamin biosynthesis protein CobN [Rhodococcus hoagii]	63.2	63.2	97%	1e-10	100%	OCC19942.1
cobaltochelatase CobN subunit [Rhodococcus sp. OK302]	63.2	63.2	97%	1e-10	100%	OYD70286.1
putative cobaltochelatase CobN subunit [Rhodococcus erythropolis PR4]	63.2	63.2	97%	1e-10	100%	BAH33849.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	OXM19713.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_084325976.1
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_070797025.1
cobaltochelatase subunit CobN [Rhodococcus sp. 164Chir2E]	63.2	63.2	97%	1e-10	100%	WP_092763989.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_060938577.1
cobaltochelatase subunit CobN [Rhodococcus sp. ARP2]	63.2	63.2	97%	1e-10	100%	WP_047888159.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_046378567.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_029538136.1
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_020969649.1
MULTISPECIES: cobaltochelatase subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_056074858.1
cobaltochelatase subunit CobN [Rhodococcus sp. MEB064]	63.2	63.2	97%	1e-10	100%	WP_042575704.1
cobaltochelatase subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_095970114.1
cobaltochelatase subunit CobN [Rhodococcus sp. BH4]	63.2	63.2	97%	1e-10	100%	WP_081557886.1
cobaltochelatase subunit CobN [Rhodococcus sp. D-1]	63.2	63.2	97%	1e-10	100%	WP_076948441.1
cobaltochelatase subunit CobN [Rhodococcus sp. YH3-3]	63.2	63.2	97%	1e-10	100%	WP_064444609.1
cobaltochelatase subunit CobN [Rhodococcus sp. 311R]	63.2	63.2	97%	1e-10	100%	WP_050655273.1
cobaltochelatase subunit CobN [Mycobacterium peregrinum]	63.2	63.2	97%	1e-10	100%	WP_064934535.1
cobaltochelatase subunit CobN [Rhodococcus kyotonensis]	63.2	63.2	97%	1e-10	100%	WP_068424109.1
cobaltochelatase subunit CobN [Rhodococcus sp. OK302]	63.2	63.2	97%	1e-10	100%	WP_094275033.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
cobaltochelatae subunit CobN [Rhodococcus sp. 4J2A2]	63.2	63.2	97%	1e-10	100%	WP_092807773.1
cobaltochelatae subunit CobN [Nocardia alba]	63.2	63.2	97%	1e-10	100%	WP_067453550.1
cobaltochelatae subunit CobN [Rhodococcus sp. UNC23MFCrub1.1]	63.2	63.2	97%	1e-10	100%	WP_027505379.1
cobaltochelatae subunit CobN [Rhodococcus erythropolis]	63.2	63.2	97%	1e-10	100%	WP_029255378.1
cobaltochelatae subunit CobN [Gordonia amarae]	63.2	63.2	97%	1e-10	100%	WP_005181911.1
cobaltochelatae subunit CobN [Rhodococcus sp. AD45]	63.2	63.2	97%	1e-10	100%	WP_045067816.1
cobaltochelatae subunit CobN [Nocardia caishijiensis]	63.2	63.2	97%	1e-10	100%	WP_084458329.1
MULTISPECIES: cobaltochelatae subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_033231975.1
MULTISPECIES: cobaltochelatae subunit CobN [Rhodococcus]	63.2	63.2	97%	1e-10	100%	WP_056444543.1
MULTISPECIES: cobaltochelatae subunit CobN [Actinobacteria]	63.2	63.2	97%	1e-10	100%	WP_019047469.1
cobaltochelatae subunit CobN [Gordonia malaquae]	63.2	63.2	97%	1e-10	100%	WP_008376255.1
cobaltochelatae subunit CobN [Mycobacterium peregrinum]	63.2	63.2	97%	1e-10	100%	WP_064881082.1
cobaltochelatae subunit CobN [Rhodococcus sp. ACPA4]	63.2	63.2	97%	1e-10	100%	WP_095886348.1
cobaltochelatae subunit CobN [Rhodococcus marinonascens]	62.8	62.8	97%	1e-10	100%	WP_072689956.1
MULTISPECIES: cobaltochelatae subunit CobN [Mycobacterium]	62.8	62.8	97%	1e-10	100%	WP_090431022.1
cobaltochelatae subunit CobN [Mycobacterium peregrinum]	62.8	62.8	97%	1e-10	100%	WP_055114836.1
cobaltochelatae subunit CobN [Nocardia soli]	62.8	62.8	97%	1e-10	100%	WP_063055124.1
cobaltochelatae subunit CobN [Nocardia concava]	62.8	62.8	97%	1e-10	100%	WP_040808888.1
cobaltochelatae subunit CobN [Nocardia salmonicida]	62.8	62.8	97%	1e-10	100%	WP_062983447.1
cobaltochelatae subunit CobN [Pseudonocardia autotrophica]	62.8	62.8	97%	1e-10	100%	WP_051737945.1
cobaltochelatae subunit CobN [Nocardia cummidelens]	62.8	62.8	97%	1e-10	100%	WP_063006465.1
MULTISPECIES: cobaltochelatae subunit CobN [Mycobacterium]	62.8	62.8	97%	1e-10	100%	WP_036460715.1
cobaltochelatae subunit CobN [Pseudonocardia oroxyli]	62.8	62.8	97%	1e-10	100%	WP_093088747.1
cobaltochelatae subunit CobN [Mycobacterium septicum]	62.8	62.8	97%	1e-10	100%	WP_044523977.1
cobaltochelatae [Rhodococcus jostii RHA1]	62.8	62.8	97%	1e-10	100%	ABG92007.1
cobaltochelatae subunit CobN [Nocardia jejuensis]	62.8	62.8	97%	1e-10	100%	WP_067686530.1
cobaltochelatae subunit CobN [Nocardia ignorata]	62.8	62.8	97%	1e-10	100%	WP_067492125.1
cobaltochelatae subunit CobN [Mycobacterium sp. GA-1841]	62.8	62.8	97%	1e-10	100%	WP_076198706.1
cobaltochelatae subunit CobN [Mycobacterium sp. URHD0025]	62.8	62.8	97%	1e-10	100%	WP_029108886.1
cobaltochelatae subunit CobN [Gordonia westfalica]	62.8	62.8	97%	1e-10	100%	WP_074848915.1
cobaltochelatae subunit CobN [Mycobacterium sp. 88mf]	62.8	62.8	97%	1e-10	100%	SER28471.1
cobaltochelatae subunit CobN [Gordonia amicalis]	62.8	62.8	97%	1e-10	100%	WP_006437927.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
cobaltochelata subunit CobN [Mycobacterium boenickei]	62.8	62.8	97%	1e-10	100%	WP_077741033.1
cobaltochelata subunit CobN [Nocardia coubleae]	62.8	62.8	97%	1e-10	100%	WP_067635658.1
MULTISPECIES: cobaltochelata subunit CobN [Mycobacterium]	62.8	62.8	97%	1e-10	100%	WP_070946436.1
cobaltochelata subunit CobN [Nocardia arizonensis]	62.8	62.8	97%	1e-10	100%	WP_054813130.1
cobaltochelata subunit CobN [Gordonia soli]	62.8	62.8	97%	1e-10	100%	WP_007624158.1
cobaltochelata subunit CobN [Mycobacterium neoaurum]	62.8	62.8	97%	1e-10	100%	WP_030135476.1
MULTISPECIES: cobaltochelata subunit CobN [Gordonia]	62.8	62.8	97%	1e-10	100%	WP_024498042.1
cobaltochelata subunit CobN [Nocardia altamirensis]	62.8	62.8	97%	1e-10	100%	WP_069162551.1

Alignments

cobaltochelata subunit CobN [Rhodococcus fascians]

Sequence ID: **WP_037181271.1** Length: 1224 Number of Matches: 1
Range 1: 567 to 595

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	9e-11()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

```
Query 89 GTDAALGDLPLIYPFLVNDPGEQTAKRR 3
          GTDAALGDLPLIYPFLVNDPGEQTAKRR
Sbjct 567 GTDAALGDLPLIYPFLVNDPGEQTAKRR 595
```

cobaltochelata subunit CobN [Rhodococcus sp. 114MFTsu3.1]

Sequence ID: **WP_020110056.1** Length: 1224 Number of Matches: 1
Range 1: 567 to 595

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	9e-11()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

```
Query 89 GTDAALGDLPLIYPFLVNDPGEQTAKRR 3
          GTDAALGDLPLIYPFLVNDPGEQTAKRR
Sbjct 567 GTDAALGDLPLIYPFLVNDPGEQTAKRR 595
```

cobaltochelata subunit CobN [Rhodococcus sp. 06-621-2]

Sequence ID: **WP_094724882.1** Length: 1224 Number of Matches: 1

See 1 more title(s)

Range 1: 567 to 595

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	9e-11()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

```
Query 89 GTDAALGDLPLIYPFLVNDPGEQTAKRR 3
          GTDAALGDLPLIYPFLVNDPGEQTAKRR
Sbjct 567 GTDAALGDLPLIYPFLVNDPGEQTAKRR 595
```

cobaltochelata subunit CobN [Rhodococcus sp. RS1C4]

Sequence ID: **WP_094728646.1** Length: 1213 Number of Matches: 1

See 1 more title(s)

Range 1: 556 to 584

Score	Expect	Method	Identities	Positives	Gaps	Frame
-------	--------	--------	------------	-----------	------	-------

63.2 bits(152) 9e-11() Compositional matrix adjust. 29/29(100%) 29/29(100%) 0/29(0%) -1

Features:

Query 89 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 3
 GTDAALGDLPLIYPFLVNDPGEGTQAKRR
 Sbjct 556 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 584

cobaltochelataase subunit CobN [Rhodococcus sp. 14-2470-1b]
 Sequence ID: **WP_094721623.1** Length: 1224 Number of Matches: 1

See 1 more title(s)
 Range 1: 567 to 595

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	9e-11()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

Query 89 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 3
 GTDAALGDLPLIYPFLVNDPGEGTQAKRR
 Sbjct 567 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 595

MULTISPECIES: cobaltochelataase subunit CobN [Rhodococcus]
 Sequence ID: **WP_033189721.1** Length: 1224 Number of Matches: 1

See 6 more title(s)
 Range 1: 567 to 595

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	9e-11()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

Query 89 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 3
 GTDAALGDLPLIYPFLVNDPGEGTQAKRR
 Sbjct 567 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 595

cobaltochelataase subunit CobN [Rhodococcus sp. 06-418-1B]
 Sequence ID: **WP_094712849.1** Length: 1213 Number of Matches: 1

See 1 more title(s)
 Range 1: 556 to 584

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	9e-11()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

Query 89 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 3
 GTDAALGDLPLIYPFLVNDPGEGTQAKRR
 Sbjct 556 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 584

cobaltochelataase subunit CobN [Rhodococcus sp. 15-1154-1]
 Sequence ID: **WP_094732627.1** Length: 1213 Number of Matches: 1

See 1 more title(s)
 Range 1: 556 to 584

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	1e-10()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

Query 89 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 3
 GTDAALGDLPLIYPFLVNDPGEGTQAKRR
 Sbjct 556 GTDAALGDLPLIYPFLVNDPGEGTQAKRR 584

cobaltochelataase subunit CobN [Rhodococcus sp. 06-1059B-a]
 Sequence ID: **WP_094628196.1** Length: 1213 Number of Matches: 1

See 1 more title(s)
 Range 1: 556 to 584

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	1e-10()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

Query	89	GTDAALGDLPLIYPFLVNDPGEGTQAKRR	3
Sbjct	556	GTDAALGDLPLIYPFLVNDPGEGTQAKRR	584

cobaltochelatae subunit CobN [Nocardia crassostreae]

Sequence ID: **WP_067538688.1** Length: 1297 Number of Matches: 1

Range 1: 551 to 579

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.2 bits(152)	1e-10()	Compositional matrix adjust.	29/29(100%)	29/29(100%)	0/29(0%)	-1

Features:

Query	89	GTDAALGDLPLIYPFLVNDPGEGTQAKRR	3
Sbjct	551	GTDAALGDLPLIYPFLVNDPGEGTQAKRR	579

BLAST is a registered trademark of the National Library of Medicine

[Support center Mailing list](#)[YouTube](#)

- [National Library Of Medicine](#)



- [National Institutes Of Health](#)



- [U.S. Department of Health & Human Services](#)



- [USA.gov](#)

NCBI

[National Center for Biotechnology Information](#), [U.S. National Library of Medicine](#) 8600 Rockville Pike, Bethesda MD, 20894 USA
[Policies and Guidelines](#) | [Contact](#)