**UNIVERSITY OF LILLE 1**
**SCIENCE AND TECHNOLOGY**

**PhD THESIS**

Thesis presented for the degree of Philosophiae Doctor by

**Qassim Abdullah Ahmed ESMAEEL**

**Major: BIOLOGICAL ENGINEERING**

---

# Deciphering new nonribosomal peptide synthetases and their products from genomic data

---

**No: 42071**
Charles Viollette Institute
Polytech'Lille, France

**8 July 2016**

| | |
|---|---|
| **Monique Royer,** Researcher-HDR **(**CIRAD, Montpellier, France) | Reviewer |
| **Pierre Cornelis**, Professor (University of Brussels, Belgium) | Reviewer |
| **Monica Höfte,** Professor (Université of Ghent, Belgium) | Examinator |
| **Jean-Luc Pernodet**, Professor (CNRS, University of Paris-Sud, France) | Examinator |
| **Maude Pupin**, Assistant Professor -HDR (University of Lille 1, France) | Examinator |
| **Philippe Jacques**, Professor (Université of liège, Belgium) | Co-Supervisor |
| **Valérie Leclère**, Assistant Professor - HDR (University of Lille 1, France) | Supervisor |

**UNIVERSITE DE LILLE 1**
**SCIENCES ET TECHNOLOGIES**

**THESE**

Présentée par
Qassim Abdullah Ahmed **ESMAEEL**

Pour l'obtention du grade de

**DOCTEUR DE L'UNIVERSITE DE LILLE 1**

**Filière: INGENIERIE DES FONCTIONS BIOLOGIQUES**

# Identification de nouvelles synthétases non ribosomiques et de leurs produits à partir de données de génomique

**N° Ordre:** 42071
Préparée à l'Institute Charles Viollette
Polytech'lille, France

**Soutenance le 8 juillet 2016 devant le jury composé de:**

| | |
|---|---|
| **Monique Royer, Directrice de recherche- HDR (CIRAD, Montpellier, France)** | Rapporteur |
| **Pierre Cornelis, Professeur (Université de Bruxelles, Belgique)** | Rapporteur |
| **Monica Höfte, Professeur (Université de Gand, Belgique)** | Examinatrice |
| **Jean-Luc Pernodet, Professeur (CNRS, université paris-sud, France)** | Examinateur |
| **Maude Pupin, Maitre de conférences- HDR (Université de Lille 1, France)** | Examinatrice |
| **Philippe Jacques, Professeur (Université de liège, Belgique)** | Co-directeur |
| **Valérie Leclère, Maitre de conférences- HDR (Université de Lille 1, France)** | Directrice |

*Dedication*

*To my beloved family*

*With love and gratitude*

# Acknowledgement

Words fail me to express how much I am indebted to my supervisor Dr. Valerie Leclère for giving me the opportunity to conduct this research. She provided invaluable support and unflinching encouragement throughout this period. My mere expression of thanks likewise does not suffice for her patience during all the meetings, discussions and continuous interest in supervising me.

I am very grateful to my co supervisor Prof. Philippe Jacques for his relentless scientific guidance, motivation, discussions during my thesis.

I would like to thank Prof. Pascal for allowing me to work in his laboratory and provided me with assistance throughout my studies.

I would like to thank Dr. Monique Royer and Prof. Pierre Cornelis for reviewing my thesis and Prof. Monica Höfte, Dr. Maude Pupin, and Prof. Jean-Luc Pernodet for serving on my examination committee.

Big thanks to Dr. Max Béchet for his unlimited help and cooperation during throughout my studies.

I am indebted to the Sanaa University, Yemen for providing partial funding for my PhD in Lille1 University, France.

My thanks to my friends and colleagues for the great time I had in our group. I enjoyed the atmosphere, their friendship, and their support. My thanks to Yazen, Alaa, Ameen, Oumayma, Sabrine, Amel, Kalim, Lamia, Samiha, Rémi, Juliette, Alexandra, Debarun, Alexandre, Sandy, Justine, Ahmed, Michael, Amirouche, Marwan, Adil, Louis, Mohammed, Antoine, Delphine, Alexandre, Cyril, Maxim, Gabrielle and Delphine Caly for the great collaboration over the years.

Special thanks to Amjed, Suhaib, Mouad, Muaamer, Yahya, and Ameen Berdouz for their support during my study.

It was a pleasure to work with friends and colleagues from laboratory of Phytopathology, university of Ghent, Belgium, to benefit from their knowledge. Especially, I would like to thank Dr. Nam Phuong Kieu for the interesting help and fruitful discussions.

Last, but not least, I would like to thank my great Mum, Father, brothers and sisters for their support and prayers. Of greatest value to me was the tireless help, encouragement and support of my beloved wife Nadya and the patience and cooperation of my lovely daughter Ruba.

I) Publications that were done during this thesis

**1: Qassim Esmaeel**, Mickael Chevalier, Gabrielle Chataigné, Rathinasamy Subashkumar, Philippe Jacques and Valérie Leclère, Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism catalyzes amonabactin synthesis in *Aeromonas*. Appl Microbiol Biotechnol (2016). doi:10.1007/s00253-016-7773-4

**2:Qassim Esmaeel**, Maude Pupin, Nam Phuong Kieu, Gabrielle Chataigné, Max Béchet, Jovana Dervael, François Krier, Monica Höfte, Philippe Jacques and Valérie Leclère, 2016. *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. *Microbiology open*. doi:10.1002/mbo3.347

3: Maude Pupin, **Qassim Esmaeel**, Areski Flissi, Yoann Dufresne, Philippe Jacques, and Valérie Leclère, 2015. Norine: a powerful resource for novel nonribosomal peptide discovery. *Synthetic and Systems Biotechnology*. doi:10.1016/j.synbio.2015.11.001

**4: Qassim Esmaeel**, Philippe Jacques, and Valérie Leclère, Non ribosomal peptides and polyketides of *Burkholderia* as a new source of bioactive compounds potentially implicated in agriculture and pharmaceutical (Review) **(*In preparation*)**

**II) Communications**

Oral Presentation

**1 : Qassim Esmaeel**, Philippe Jacques, and Valérie Leclère. Florine and Norine: powerful tools for *in silico* genome mining leading to discovery of novel nonribosomal peptides. Annual Conference 2016 of Microbiology Society, 21–24 March 2016- Liverpool, UK.

**2: Qassim Esmaeel**, Maude Pupin, Nam Phuong Kieu, Gabrielle Chataigné, Max Béchet, Jovana Deravel, François Krier, Monica Höfte, Philippe Jacques and Valérie Leclère. *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great Potential for novel lipopeptides synthesis. 10th International PGPR Workshop Liège, Belgium, 16-19/06/2015

**3: Qassim Esmaeel**, Maude Pupin, Nam Phuong Kieu, Gabrielle Chataigné, Max Béchet, Jovana Deravel, François Krier, Monica Höfte, Philippe Jacques and Valérie Leclère. *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great Potential for novel lipopeptides synthesis. Journée des doctorants, 21 Novembre 2014, Institut Charles Viollette, Polytech Lille, Lille, France

**III) Poster:**

**1 : Qassim Esmaeel**, Maude Pupin, Nam Phuong Kieu, Gabrielle Chataigné, Max Béchet, Jovana Deravel, François Krier, Monica Höfte, Philippe Jacques and Valérie Leclère, *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great Potential for novel lipopeptides synthesis. The 6th Congress of European Microbiologists (FEMS 2015), 7-11 June 2015, Maastricht, Netherlands.

# Abstract

Microorganisms are considered one of the most important sources of secondary metabolites including ribosomal and non ribosomal peptides. The search of new non ribosomal peptides has been motivated by their wide applications exploited by industries in different area including pharmaceutical and phytosanitary sectors. They are produced through complex synthetases called non ribosomal peptides synthetases. More than 70 % of NRPs have a complex structure that includes one or more cycles and branches. Therefore, development of specific tools dedicated to the screening of these peptides is necessary as they cannot be predicted and analyzed as classical peptides. With the aim to further analyse biosynthesis pathways and to identify new active peptides, I mainly studied two bacterial models: *Burkholderia* and *Aeromonas*. The genome-mining is a very powerful approach for the discovery of new non NRPs. Indeed, among 48 strains of *Burkholderia,* 228 gene clusters containing NRPSs and hybrid NRPS-PKS were found via *in silico* analysis following Florine workflow. The current study lead to the discovery of new peptides in *Burkholderia* including a new siderophore named phymabactin and a cyclic lipopeptide we have called burkhomycin. It also gave new insights on the mechanism of nonribosomal synthetases, exemplified by the detection of dual C/E domains in NRPSs involved in the production of cyclic lipopeptides by *Burkholderia* and the identification of a unique use of domains and modules in the pathway responsible for synthesis of amonabactins in *A. hydrophila*.

# Résumé

Les micro-organismes sont considérés comme l'une des sources les plus importantes de métabolites secondaires, y compris les peptides ribosomiques et non ribosomiques. La recherche de nouveaux peptides non ribosomiques a été motivée par leurs larges applications dans diverses industries telles que les secteurs pharmaceutiques et phytosanitaires. Ils sont produits par complexes enzymatiques appelés NRPSs. Plus de 70% des peptides non ribosomiques ont une structure complexe qui comprend un ou plusieurs cycles et des ramifications. Par conséquent, le développement d'outils spécifiques dédiés à la prédiction de ces peptides est nécessaire car ils ne peuvent pas être prédits et analysés comme les peptides classiques. Dans le but d'analyser les voies de biosynthèse et d'identifier de nouveaux peptides actifs, j'ai étudié principalement deux modèles bactériens: *Burkholderia* et *Aeromonas*. Le genome-mining est un approche très performante pour la découverte de nouveaux NRPs. En effet, pour 48 souches de *Burkholderia* analysées *in silico* à l'aide du workflow Florine, 228 clusters de gènes contenant des gènes de NRPS et hybrides NRPS-PKS ont été trouvés. Cette étude a permis de mettre en évidence de nouveaux peptides produits par des *Burkholderia*, incluant la phymabactin, un nouveau siderophore, et un lipopeptide cyclique que nous avons appelé burkhomycin. La présente étude a d'autre part permis d'éclaircir le mécanisme de fonctionnement des synthétases non ribosomiques, illustrés par la détection des domaines C/E dans des synthétases de lipopeptides cycliques et une utilisation originale des domaines et modules dans les NRPS impliquées dans la biosynthèse des amonabactines chez *A. hydrophila*.

# *List of contents*

# List of abbreviations

| | |
|---|---|
| 4'PP | 4'-Phosphopantetheine |
| AA | Amino Acid |
| ACP | Acyl carrier protein |
| A-domain | Adenylation domain |
| AMMD | An isolate of *Burkholderia ambifaria* |
| Amp | Ampicillin |
| ATCC | American Type Culture Collection |
| ATP | Adenosine triphosphate |
| Bamb | *B. ambifaria* genes |
| Bcc | *Burkholderia cepacia* complex bacteria |
| BCCM/LMG | Belgian coordinated collections of microorganisms |
| BGCs | Biosynthetic gene clusters |
| BLAST | Basic Local Alignment Search Tool |
| Bp | Base pair |
| C-domain | Condensation domain |
| CFU | Colony Forming Unit |
| CIP | La collection de l'Institut Pasteur |
| CoA | Coenzyme A |
| DSMZ | German collection of microorganisms and cell cultures |
| E-domain | Epimerization domain |
| EDTA | Ethylene di-amine tetra- acetate |
| GC% | Guanine cytosine content |
| Gm | Gentamicin |
| HEPES | 4-(2- hydroxylethyl)-1- piperazineethane sulfonic acid |
| HPLC | High performance liquid chromatography |
| IPTG | Isopropyl β-D- thiogalactopyranoside |
| Kb | Kilo bases |
| KB | Kings' B |
| Km | Kanamycin |

| | |
|---|---|
| LB | Luria Bertani |
| LC | Liquid chromatography |
| LC-MS | Liquid chromatography-Mass spectrometry |
| Mb | Mega bases (1Mb = 1,000,000 bases) |
| MeOH | Methanol |
| MHA | Muller Hinton Agar medium |
| MOPS | 3-(N-morpholino)- propane sulfonic acid |
| MS | Mass Spectrometry |
| MW | Molecular weight |
| NCBI | National center for biotechnology information |
| NCTC | National Collection of Type Culture |
| ND | Not detected |
| NIS | NRPS-independent siderophores |
| NRPS | Non ribosomal peptide synthetase |
| NRPs | Non ribosomal peptides |
| OD | Optical density |
| ORF | Open Reading Frame |
| PCP | Peptidyl carrier protein |
| PDA | Potato Dextrose Agar |
| PEG | Polyethylene glycol |
| PKS | Polyketide synthetase |
| Pks | Polyketides |
| SM | Secondary metabolites |
| TE | Tris-EDTA |
| Te | Thioesterase domain |
| TSB, TSA | Tryptone Soya broth/ agar medium |
| UV | Ultraviolet light |
| W/V | Weight per volume |
| X-Gal | Bromo-4- chloro-3- indolyl β-D- galactopyranoside |
| YE | Yeast extract |
| YPD | Yeast extract-peptone dextrose |

# GENERAL INTRODUCTION

Microorganisms including bacteria and fungi live in different conditions, and to survive under such conditions, they synthesize a wide range of secondary metabolites with varied metabolic origin and exotic chemical structure. Many of them are biologically active peptides of nonribosomal origin that represent a family of fantastic compounds with complex structure and various biological activities, such as antibiotics (Bacitracin), biosurfactants (Surfactin), siderophores (Ornibactin), antitumors (Bleomycin), immunosuppressive agents (Cyclosporin), toxins (Microcystin) and lipopeptides (Fengycin). These metabolites find use as natural products in pharmaceutics, agriculture and manufacturing.

The first study of non ribosomal peptides synthesis was reported in the 1960s by biochemical studies with cell free extracts of *Brevibacillus brevis*. These experiments showed that synthesis of gramicidin S was observed even in the presence of RNAses or inhibitors of the ribosomal biosynthesis (Gevers et al. 1968). Lipmann and coworkers also reported that the synthesis of antibiotics gramicidin S and tyrocidine in the bacterium *Bacillus brevis* was resistant to ribonucleases, and appeared to occur through enzymes capable of catalyzing peptide bond formation from thiol linked intermediates in a way resembling that of fatty acid synthesis (Lipmann 1971) .

The biosynthetic mechanisms and regulation system of non ribosomal peptides have been extensively studied over the past few years. The biosynthesis of these compounds is carried out by nonribosomal peptide synthetases (NRPSs), which are large multifunctional enzymes organized in sets of domains which constitute modules containing the information needed to complete an elongation step in an original peptide biosynthesis. Genes encoding the enzymes incorporating monomers or modifying the different residues are, most of the time, organized in clusters. Different pathways of NRPS mechanisms have been discussed including linear, iterative, and non linear biosynthesis. However, the biosynthesis of some peptides is not totally understood. Therefore, information about these clusters, pathways and metabolites is needed to facilitate the process of natural product discovery. This will help to increase knowledge about the mechanisms of synthesis and facilitate the screening of new non ribosomal peptides.

The advancement in the bioinformatics tools and comprehensive knowledge in nonribosomal peptide synthetases (NRPSs) has been instrumental in the identification of new biosynthetic gene clusters from genomes. Thus, this approach called “genome mining” can be applied for the discovery of natural products useful for natural product development. The

availability of whole bacterial genome sequencing combined with bioinformatics approaches facilitated the understanding of secondary metabolites biosynthesis and gave insight into bacterial metabolism. By combining technological development in genomics, bioinformatics and structure biology, translated gene sequence data can be used to rapidly derive structural elements encoded by secondary metabolic gene clusters from microorganisms. Once the biosynthesis gene cluster for the novel compound is clarified, the structure of predicted molecules can be described on the biochemical and genetic levels. Theses exploration gives insight into the mechanisms of synthesis for non ribosomal peptides.

As many natural products are synthesized by NRPSs, it is very essential to study their mechanisms of synthesis and develop a protocol to facilitate the screening of natural products including non ribosomal peptides. Two interesting models of bacteria were used in this study to investigate the mechanisms of synthesis of non ribosomal peptides and their discovery by following a strategy we have developed with these objectives.

The genus *Aeromonas* is ubiquitous bacteria found in a variety of aquatic environments worldwide, including well, bottled, flood and heavily polluted water (Janda and Abbott 2010). Due to their pathogenicity and persistent threat to the aquaculture sector, different strains of *Aeromonas* received very high economical attention in the last two decades. It has been shown that under conditions of low- iron concentrations, most isolates of *Aeromonas* species produce iron transporting siderophore called amonabactins (Barghouthi et al. 1989a; Barghouthi et al. 1989b). The production of amonabactins by *A. hydrophila* has been reported since more than twenty five years. Although the gene cluster of amonabactins is known, their mechanism of synthesis is not totally understood. Therefore, the first part of the thesis was to:

I: Investigate in details the biosynthetic gene cluster of amonabactins in *Aeromonas hydrophila*

II: Study the relationship between the domains organization of the amonabactin gene cluster and the structures of the amonabactins and to confirm the hypothetical non-ribosomal biosynthesis using mutants.

III: Decipher the distribution of the NRPS genes including siderophores that are probably associated with the pathogenicity of these bacteria in all sequenced genomes of *Aeromonas* strains available in NCBI.

This will help to increase knowledge about the mechanisms of nonribosomal synthetases and guide to understand the production of virulence factor including siderophores in *Aeromonas* strains.

The genus *Burkholderia* consists of over sixty species, and live in various ecological niches including soil, water, plant as well as animal and human (Mahenthiralingam et al. 2008a). Due to their wide geographical distribution and versatile ecological properties including their significant role as pathogenic, plant promoting growth, and biocontrol, *Burkholderia* have received many scientific attentions in recent years. This makes them interesting model organisms to study, and for putative applications, a good understanding of their occurrence, ecological potential in the environment and their secondary metabolites is necessary. Members of *Burkholderia* genus are well known for their significant role in the environments through the excretion of a wide variety of extracellular products including non ribosomal peptides and polyketides. These metabolites represent a widely distributed biocontrol and biomedically important class of peptidic natural products including antibiotics and anticancers as well as biopesticides that are considered as a novel source that will be used to defend ecological niche from competitors and promote growth of plant. However, up to now only a few NRPSs have been characterized for a given members of *Burkholderia* due to the restricted conditions used that are conductive to the production of NRPSs. Thus, the majority of *Burkholderia* biosynthetic gene cluster are silent under standard laboratory conditions.

The focus of the second part of the study was to:

I: Develop a global strategy to screen all the potential NRPSs gene clusters of 48 gapless genomes of *Burkholderia* species.

II: Design the organization of domains including the C-domain sub-types and predict the most probable produced peptides for all detected NRPSs gene clusters.

III: Investigate in details all NRPSs gene clusters potentially involved in biocontrol by structure and genetic approaches to support the predictions obtained *in silico*.

# CHAPTER 1- STATE OF THE ART REVIEW

## A) Biodiversity of non ribosomal peptides

Microorganisms are a prolific source of natural products that being extremely valuable in drug discovery and agriculture researches. Many of these compounds that have been isolated belong to a class of compounds called Polyketides (PKs) and nonribosomal peptides (NRPs) which represent a promising basis for the development of substances that attract a great interest and applications in medicine and agriculture. The ability to produce nonribosomal bioactive peptides is widely spread among microorganisms including bacteria and fungi. They represent a diversity of natural products with a broad range of biological activities and pharmacological properties (Cane et al. 1998) and complex chemical structures (Kopp and Marahiel 2007). They are mostly cyclic or branched cyclic compounds containing not only the 20 common proteinogenic L-amino acids but also non-proteinogenic amino acids, small heterocyclic rings and other unusual modifications in the peptide backbone (Grünewald and Marahiel 2006). They are synthesized through an alternative pathway that allows production of polypeptides other than through the traditional translation mechanism. The peptides are created by multimodular enzymes called nonribosomal peptide synthetases (NRPSs), from simple building blocks, leading to the construction of compounds containing 2 to 26 residues. Non ribosomal peptides are classified based on their functions as antibiotics, siderophores, surfactants, immunosuppressors, toxins, etc (Fig. 1) or based on their chemical structure as peptides, lipopeptides, glycolipopeptides, etc (Finking and Marahiel 2004).

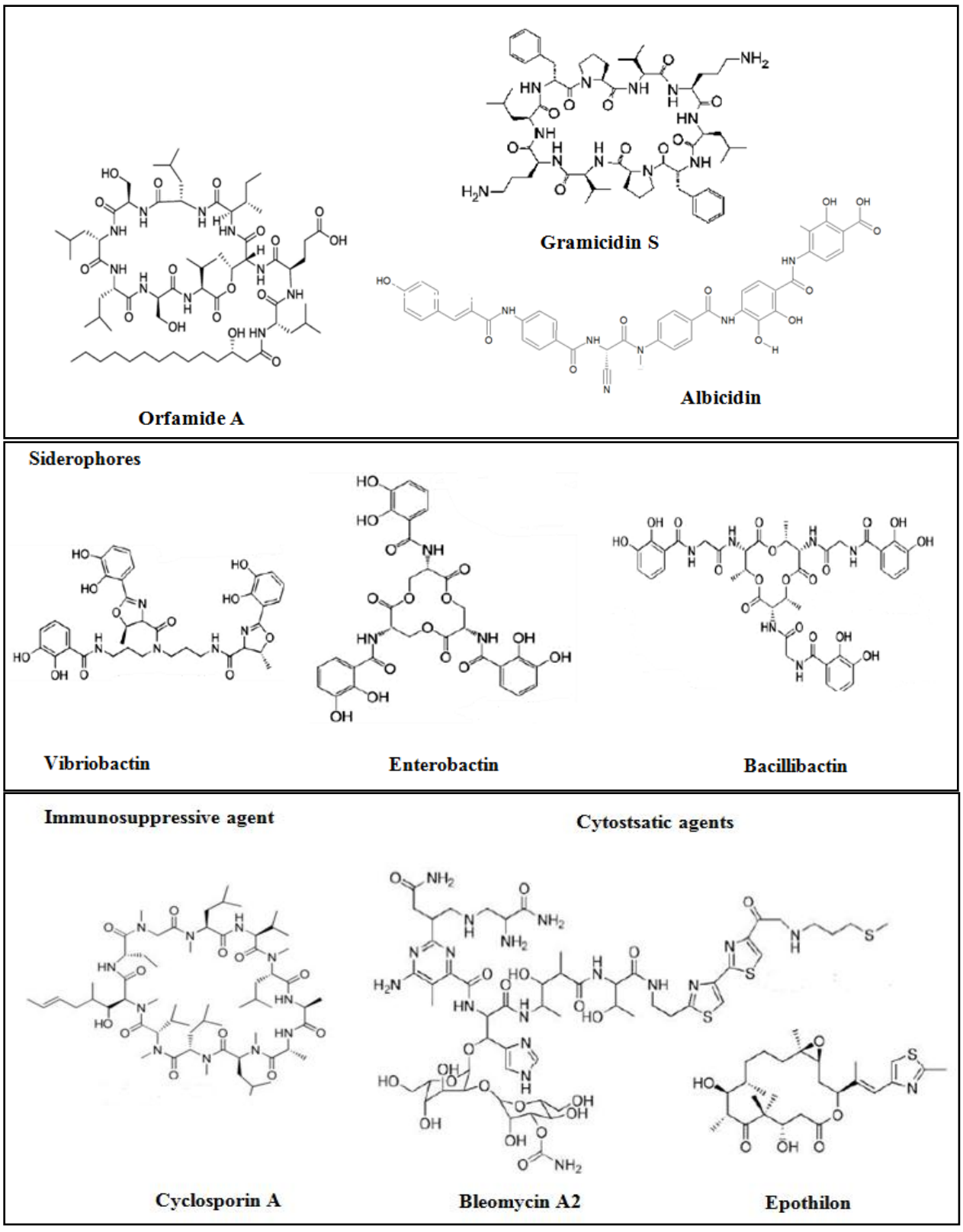



Figure 1: Examples of the structural diversity of bioactive compounds of nonribosomal origin (Jang et al. 2013; Schwarzer et al. 2003).

### B) Structures of nonribosomal peptides

Non ribosomal peptides are different from ribosomal peptides based on their primary structure as linear, cyclic (partially cyclic, branched or double cyclic) containing small heterocyclic rings and other unusual modifications in the peptide backbone and the diversity of different building blocks (More than 500 different monomers). Among them are D-amino acids, N-terminally attached fatty acids, methylated, hydroxylated and acetylated amino acids, as well as various phosphorylated and glycosylated residues (Caboche et al. 2010). Based on their compositions, non ribosomally peptides (NRPs) are divided into six classes including peptides, lipopeptides, glycopeptides, chromopeptides, peptaibols, and PK-NRPs (Fig. 2) (Caboche et al. 2008). Each class will be detailed in the following sections:

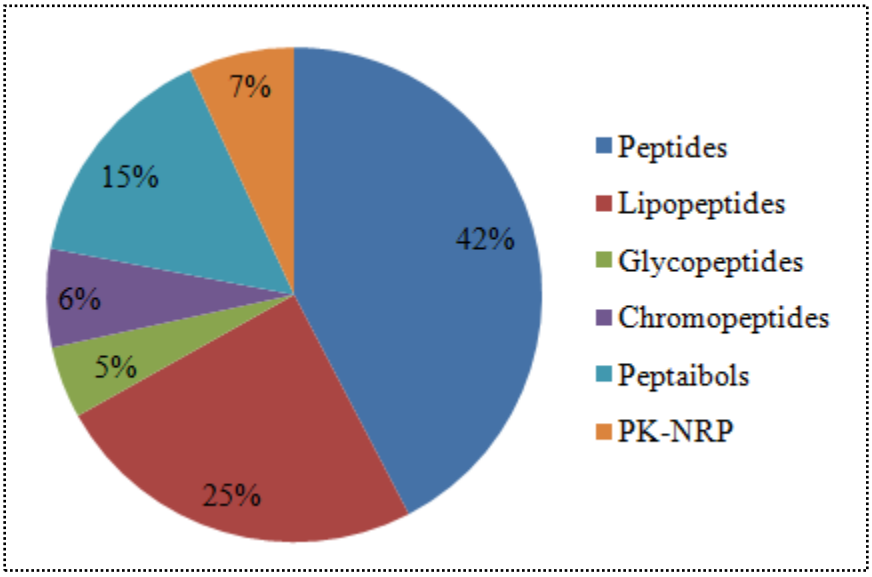



Figure 2: Schematic representation of nonribosomal peptide classes found in Norine (1172 peptides) (Caboche et al. 2008; Flissi et al. 2016).

### B.1. Peptides

The peptides are only composed of monomers including the 20 common proteinogenic L-amino acids, non-proteinogenic amino acids and different building blocks (Caboche et al. 2008). For example, the ACV, the famous precursor of penicillin and cephalosporin, is a linear tripeptide composed of L-delta-(alpha-aminoadipoyl), L-cysteinyl, and D-valine (Fig. 3).

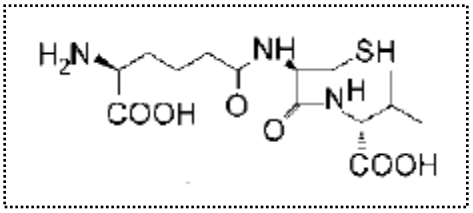



Figure 3: Structure of ACV as an example of peptides (Mootz et al. 2002b).

### B.2. Lipopeptides

Lipopeptides have linear or cyclic structure of a peptide moiety, linked to a fatty acid via ester or amide bonds or both. The most studied lipopeptides are the lipopeptides produced by *Bacillus* as fengycin and surfactin, and *Pseudomonas* as arthrofactin (Lange et al. 2012) (Fig.4) and tolaasine. Many of these lipopeptides have surfactant, antibacterial, and antifungal properties that have attracted interest from industry.

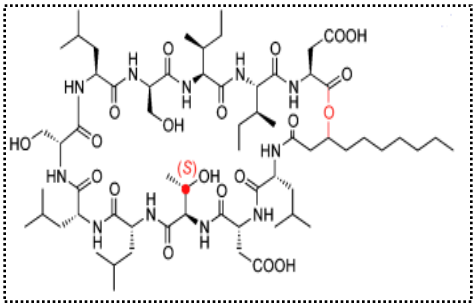



Figure 4: Structure of arthrofactin as an example of lipopeptides (Lange et al. 2012).

### B.3. Glycolipopeptides

The glycolipopeptides are cyclic or polycyclic nonribosomal peptides that contain carbohydrates moieties (glycans) covalently attached to the side chains of the amino acid residues that constitute the peptide. They are considered as essential for the control of infectious diseases caused by pathogenic gram positive bacteria. For example, Vancomycin (Fig.5) is a glycopeptide antibiotic produced by *Amycolatopsis orientalis* (Hubbard and Walsh 2003).

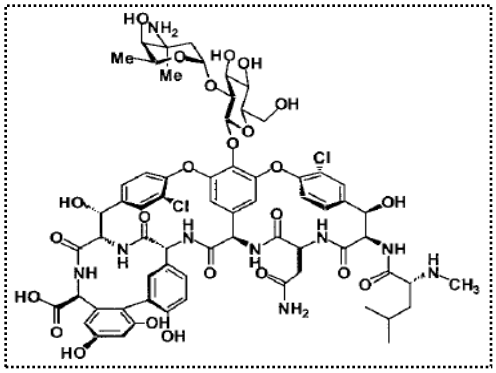


Figure 5: Vancomycin as an example of glycopeptides (Hubbard and Walsh 2003).

### B.4. Chromopeptides

Chromopeptides are compounds that consist of cyclic or linear peptide chain and aromatic or heteroaromatic rings or ring systems called chromophores. The chromophore parts of chromopeptides absorb light in the far-ultraviolet range. They are usually important for their biological activities such as acting as pharmacophors. The most famous examples of chromopeptides are pyoverdines (Fig.6), siderophores with yellow- green fluorescent colour excreted by different species of *Pseudomonas* (Cornelis and Matthijs 2002).

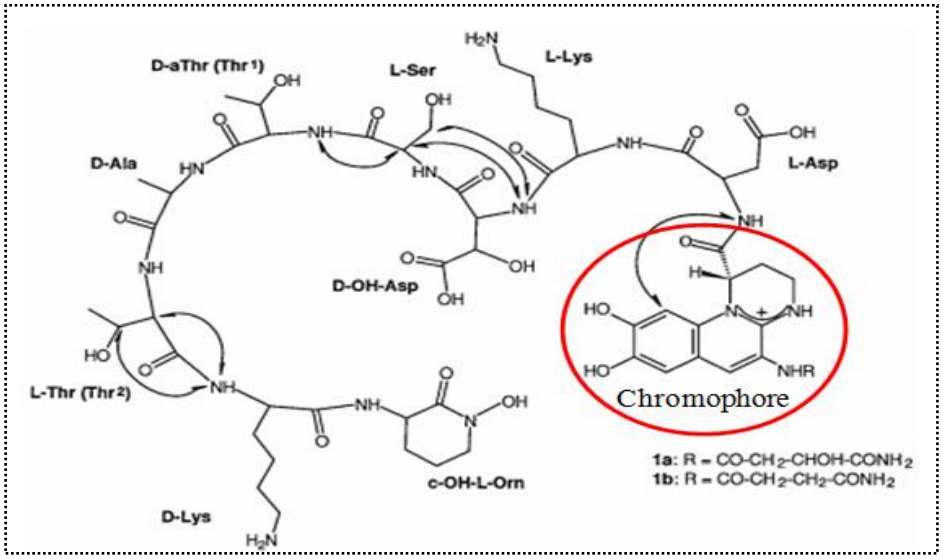



Figure 6: Structure of pyoverdine 2461 as an example of chromopeptides (Beiderbeck et al. 1999).

### B.5. Peptaibols

Peptaibols are linear peptides containing 5 to 20 amino acids; most of them are non-proteinogenic amino acids (Fig.7). They are characterized by containing α-aminoisobutyric acid (Aib) in the peptide chain and ending by an alcohol as the C-terminal amino acid is hydroxylated to an acid alcohol. They are produced by some fungi and have biological properties as antibiotics.

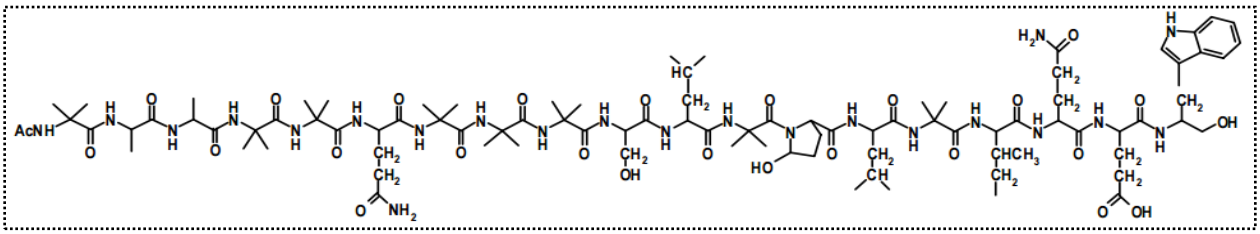


Figure 7: Structure of trichorzianine TB IIa as an example of peptaibols (Schuhmacher et al. 2007).

## C) Biological activities of non ribosomal peptides

Based on their original way of synthesis and compositions including the non proteinogenic monomers, cycles and branching, nonribosomal peptides represent a wide range of biological activities and pharmaceutical applications including antibiotics, toxins, antitumors, siderophores…etc. (Schwarzer et al. 2003). Each class will be detailed in the following sections:

### C.1. Antibiotics

Non ribosomal peptides with antimicrobial activity represent half of the peptides found in Norine database. Out of 1179 NRPs found in Norine, 653, grouped in 121 families, are classified as antibiotics in different categories including peptaibols as trichotoxin, glycopeptides as vancomycin, and lipopeptides as iturin and fengycin. Beside their antibiotic activity, some peptides share other activities as antitumor, toxin, surfactant or immuno-modulator.

### C.2. Toxins

Some peptides or lipopeptides have the ability to kill cells. In Norine database, around 266 peptides presented in 48 families are toxins, and most of them share other activities as antibiotics, antitumors or surfactants beside their toxic activity.

### C.3. Antitumors and protease inhibitors

Out of 1179 NRPs found in Norine database, 101 NRPs are antitumors grouped in 34 families. Beside antitumor activity, 75 NRPs found in Norine share other activities including antibiotics, toxin, immunomodulating and surfactants. 39 NRPs are protease inhibitors presented in 13 groups. They are pure and never cross with other activities.

### C.4. Siderophores

Siderophores are polypeptide derivatives that are non ribosomally synthesized by assembly lines constituted of non ribosomal peptides synthetases (NRPSs), while others that are not polypeptides are assembled by different enzymes designated as NIS (NRPS Independent Siderophores) (Challis 2005). They are produced by microorganisms under iron limited conditions. They have the ability to bind and chelate the iron molecules. The non ribosomally produced siderophores are mainly chromopeptides, but can also be lipopeptides or pure peptides. Based on Norine database, 82 NRPs are siderophores and 18 of them are also known as surfactants.

### C.5. Others

Other activities are also linked to other NRPs. These include surfactants (as surfactin), calmodulin (as putative konbamide) antagonist, immunomodulating (as cyclosporin, currently used to prevent rejection following solid organ and bone marrow transplantation) and antiatherogenic.

## D) Modular Non Ribosomal Peptides Synthetases

Non ribosomal peptides are synthesized via a multiple- carrier thiotemplate mechanism on large enzymes called nonribosomal peptide synthetases (NRPSs). They are organized into sets of iterative catalytic units of remarkable size called modules in which each module allows the incorporation of one monomer into the peptidic chain. The order and number of the modules of an NRPS protein are, in many cases, colinear to the amino acid sequence of the corresponding peptide moiety of the final NRPS molecule (“colinearity rule”) (Grünewald and Marahiel 2006). Each module is responsible for specific incorporation of an amino acid or any other building block into the final product (Marahiel et al. 1997; Schwarzer and Marahiel 2001). Each module consists of domains that play enzymatic role for the synthesis of peptide backbone. Modules can

be subdivided into initiation, elongation and termination modules (Fig.8). The initiation module is responsible for the selection and activation of the first amino acid. Initiation modules contain at minimum an A-domain for substrate recognition and a PCP or T domain that holds the activated substrate and serves as the amino acyl donor in the first peptide bond-forming step of the NRPS assembly line. In elongation modules, each module loads its activated substrate onto its PCP-domain. Elongation modules contain at the minimum an A-domain (Dieckmann et al. 1995; May et al. 2001; Mootz and Marahiel 1997; Stachelhaus and Marahiel 1995), a PCP domain (Ehmann et al. 2000; Stachelhaus et al. 1996), and a C-domain (Bergendahl et al. 2002; Stachelhaus et al. 1998). In the termination, the Te-domain (Thio-esterase domain) hydrolyzes the completed polypeptide chain from the PCP-domain of the downstream module.

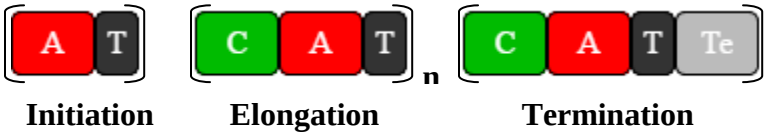



Figure 8: Organization of the domains in each module in NRPS

**D.1. Domain Organization of Non Ribosomal Peptides Synthetases**

NRPS domains are classified into principal and optional domains. Each domain plays essential enzymatic role for the synthesis of peptide backbone. The main catalytic functions are adenylation (A domain), thiolation (T-domain), condensation (C-domain), and thioesterase (Te domain) (Fig.9). The Adenylation domain (~550 aa) is responsible for the selection and activation of the amino acid that makes the product. It determines and controls the entry of the substrates into non ribosomal peptide synthesis. A-domains activate the amino acid substrates as aminoacyl adenylates at the expense of ATP (Dieckmann et al. 1995). The thiolation domain (termed peptide carrier protein (PCP) is a small domain (80-100 aa) which represents the transport unit that accepts the activated amino acid that is covalently tethered to its 4-phosphopantetheine (4-PP) cofactor as thioester (Ehmann et al. 2000; Stachelhaus et al. 1996). This cofactor is post-translationally transferred from Coenzyme A to a conserved serine residue of the carrier protein (PCP) and acts as a flexible arm to allow the bound amino acyl and peptidyl substrate to travel between different catalytic centers (Lambalot et al. 1996; Quadri et al. 1998).

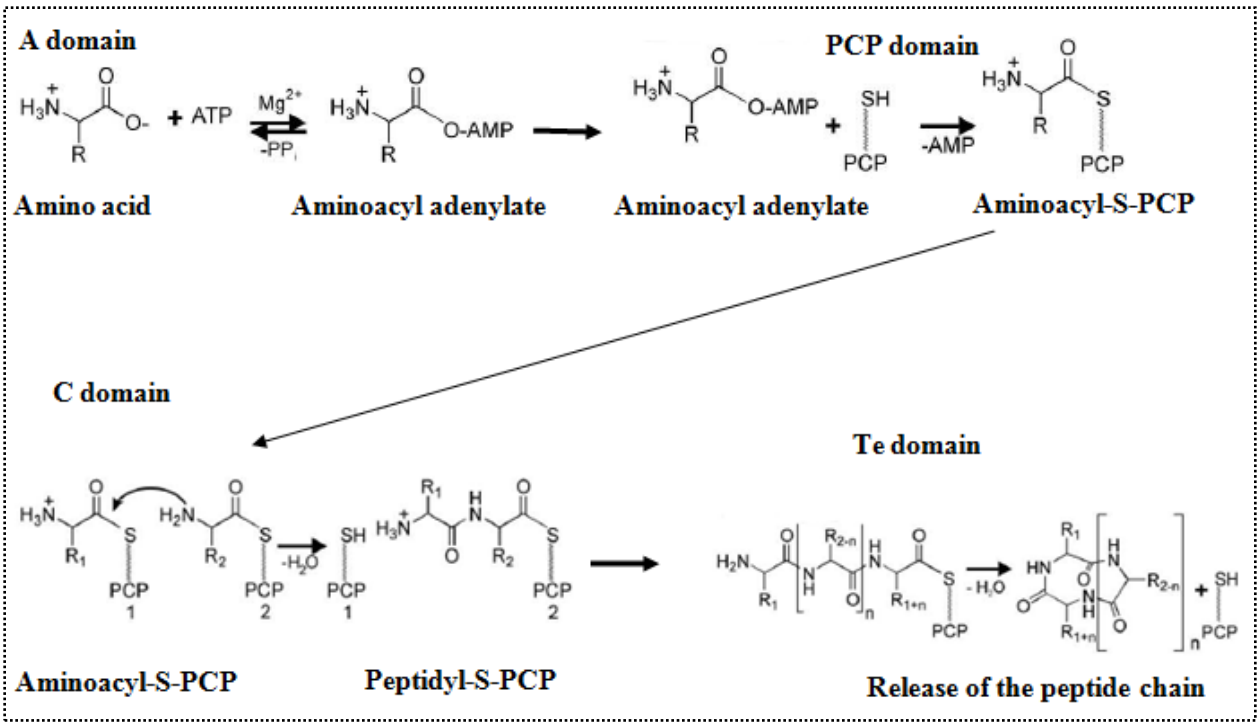



Figure 9: Organization of essential domains involved in the synthesis of nonribosomal peptides (Schwarzer et al. 2003)

The condensation domain (consists of ~ 450 aa) catalyses the formation of the peptide bond between amino acyl substrate tethered to thiolation domains of adjacent modules. (Bergendahl et al. 2002; Stachelhaus et al. 1998). The C domain has two binding sites; one for the nucleophilic acceptor which is aminoacyl-S-4'PP-PCPs and the other one for the electrophilic donor which is peptidyl-S-4'PP-PCP of the N-terminally adjacent module that serve as substrates (Fig.10) (Finking and Marahiel 2004). Different C-domains have been identified and showed strong stereoselectivity (L- or D-amino acid) and the conserved residues driving the stereospecificty of C-domains are observed in binding sites. The C-starter domain acylates the first monomer with β-hydroxyl fatty acid. The $^{L}C_{L}$ that is specific for L configured donor and acceptor catalyze a peptidic bond between two L-monomers. The C domains that are specific for D configured upstream donor and L configured downstream acceptor are classified as $^{D}C_{L}$ domains. The dual C/E domains are involved in both epimerization and condensation. Hetrocyclization (Cy) domains catalyze the both peptide bond formation and cyclization of cysteine and serine residues. The difference between $^{D}C_{L}$ and $^{L}C_{L}$ is based on sequences of 300 amino acids especially a

moderately conserved motif LPxDxxRP is usually observed in $^{L}C_{L}$ domain at the N-terminus (Balibar et al. 2005; Rausch et al. 2007). The dual C/E domains are $^{D}C_{L}$ with epimerase activity. In the assembly line, a Dual C/E domain is located directly after a C-A-T module which activates and incorporates an L amino acid. The module which contains the Dual C/E domain also activates an L-amino acid. Then the Dual domain catalyzes the epimerization of the L-residue into D configuration and subsequently promotes the condensation of those two residues. Beside the active site His-motif which is found in all C domains, Dual E/C domains exhibit a second His-motif, HH[I/L] xxxxGD, which is located close to the N-terminus of the domain (Balibar et al. 2005).

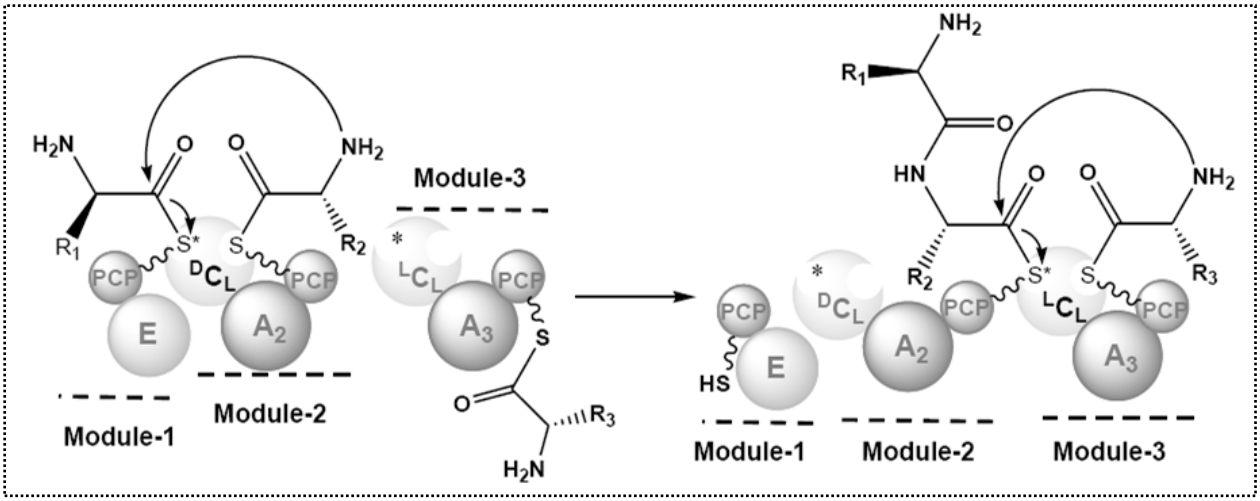



Figure 10: The mechanism of C-domains. The white pockets are the acceptor and donor sites. The acceptor site is marked with an asterisk. $^{L}C_{L}$, condensation between 2 L-monomers; $^{D}C_{L}$, condensation between D-monomer and L-monomer; A, adenylation; PCP, peptidyl carrier protein (Finking and Marahiel 2004).

During the elongation step, the release of the full length peptide from the enzyme is catalyzed by the Te domain (Fig.9) which contains ~250 aa (Fischbach and Walsh 2006; Schwarzer and Marahiel 2001). Product release is achieved by a two-step process that involves an acyl-*O*-Te-enzyme intermediate that is subsequently attacked by either a peptide-internal nucleophile (Kohli et al. 2001) or water (Miller et al. 2001), which results either in a macrocyclic product, as observed in the case of surfactin (Tseng et al. 2002).

Some NRPS contain additional domains that are involved in the biosynthesis of the peptide to modify the structure of the monomer involved in the primary structure or adding some external

compounds to the peptide (Fig. 11). The secondary domains include epimerization domain (E), cyclization (Cy), reduction domain (R), oxidation (Ox), methylation (Me), and formylation domain (F) and addition of fatty acid chain (Mootz et al. 2002a; Schwarzer and Marahiel 2001).

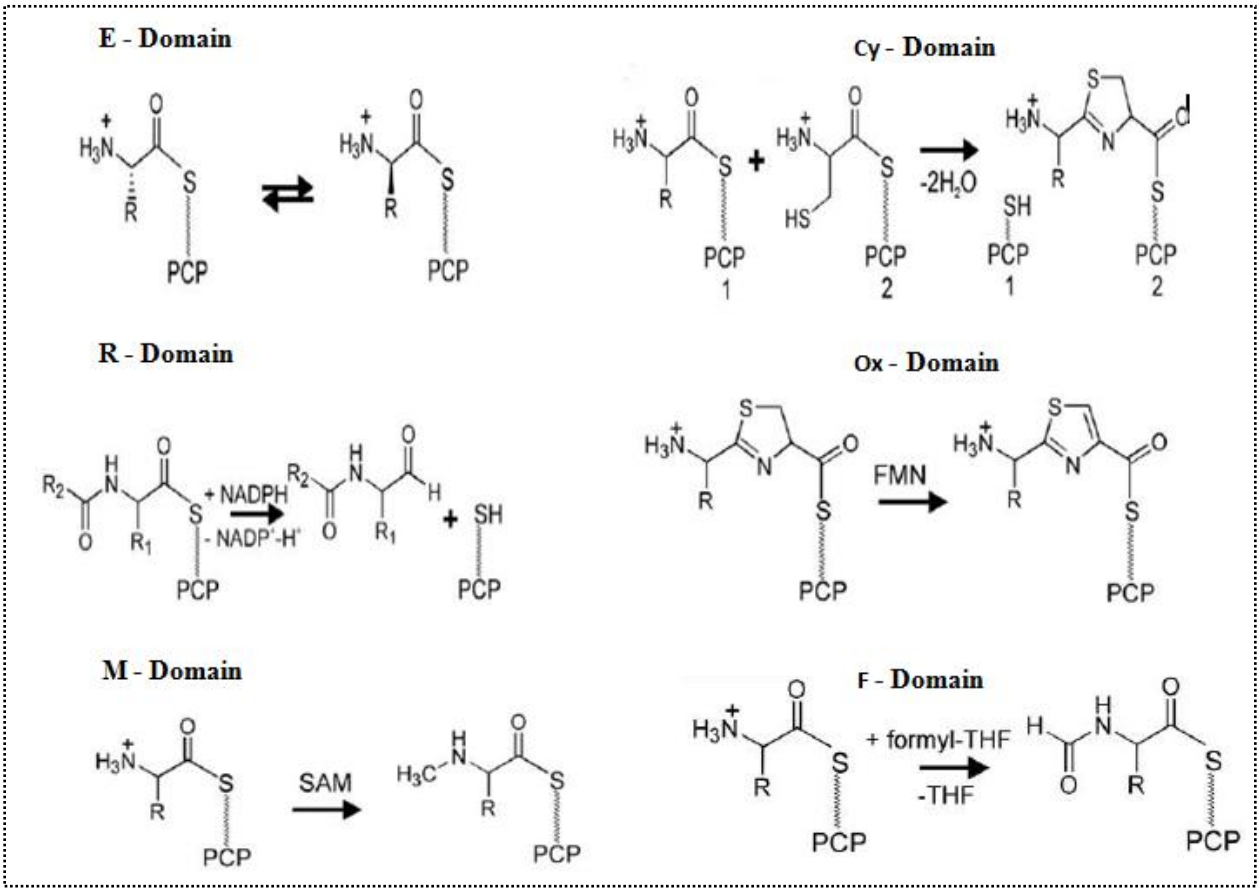



Figure 11: function of the secondary domains involve in the synthesis of NRPs (Schwarzer et al. 2003).

The E-domains are ~450 in size and C–terminal neighbors of the respective module's PCP (Fig.12). They catalyses the racemization of the PCP-bound L-amino acid of the growing polypeptide chain (Stachelhaus and Walsh 2000) (Fig.12). The D-amino acids are very well-known structure elements of non ribosomal peptides. Incorporation the D-amino acids to the growing peptides can be obtained by the presence of E-domain and the specificity of A-domains. For example, the A-domains of cyclosporine and microcystin synthetase are specific for the D-alanine and D-glutamine respectively (Tillett et al. 2000; Weber et al. 1994). Aminoacyl epimerases and peptidyl epimerases are two different types of epimerization domains that

catalyze the epimerization reaction in the initiation and elongation modules respectively (Cosmina et al. 1993).

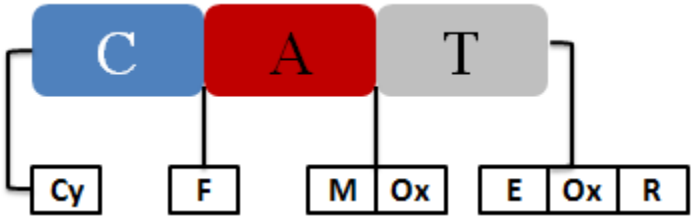



Figure 12: Localization of secondary NRPS domains adapted from (Labby et al. 2015).

The cyclization domains are ~450 aa in size that replace C-domains in some cases and synthesize enzyme bond in NRPS. It is responsible for the ring-formation through the hetrocyclization of functional side chains of the amino acids cysteine, serine and threonine with the peptide backbone. The product of cyclization is thiazoline if the amino acid residue is cysteine or oxazolines if the residue is serine or threonine. The reduction domain located at the extreme C-terminal end and catalyzes the reduction of the PCP-bond peptide to release the mature peptide product reductively with the aid of a NADPH cofactor (Fig. 12). It catalyzes the reduction of acyl thioester into its primary alcohol via an aldehyde intermediate through the four-electron reduction reaction (Silakowski et al. 2000). The oxidation domains can be localized at two different positions within an NRPS module (Fig.12). They can be inserted into the accompanying A-domain and PCP domain (Du et al. 2000a) Figure (12). It catalyzes the thiazoline to thiazole and oxazolines to oxazoles. Two different types of methylation are presented. The first one is the *N-methylation* which is ~450 aa in size that are located next to the A-domains in with the present of cofactor S-adenosylmethionine (SAM) whose methyl group is probably transferred at the aminoacyl stage prior to peptide bond formation (Haese et al. 1993; Schauwecker et al. 2000). The second one is the *C- methylation* which is localized between Cy-domain and the PCP as in the case of yersiniabactin synthetase; the reaction is catalyzed by a C-methyltransferase (C-Mt) with the present of cofactor SAM (Miller et al. 2001). The formylation domain is N-terminal to the A-domain that is responsible for the *N*-formylation with the present of the cofactor *N*-formyltetrahydrofolate (*N*-formyl- THF).

NRPS domains share a number of highly conserved sequence motifs. These “core-motifs” allow the identification of individual domains on the protein level (Table 1) (Schwarzer et al. 2003). Ten residues of Phe A of gramicidin synthetase that are essential for the substrate binding and

control the entery of the specific substrate were identified. They are located at positions 235, 236, 239, 278, 299, 301, 322, 330, 331 and 517 in the conserved motifs A3 to A7 and A10 which make direct contact to phenylalanine. The residues lie in a 100-aa stretch between cores A4 and A5 (Table 1) of the A-domains analysis led to the introduction of the so-called non ribosomal code, which allows the prediction of A-domain selectivity on the basis of its primary sequence (Du et al. 2000b; Stachelhaus et al. 1999a).

Table 1: List of Core-motifs of NRPS-domains (Schwarzer et al. 2003)

| NRPS domains | Motifs | Specified sequences |
|---|---|---|
| A-domains | A1 | L(TS)YxEL |
| | A2 | LKAGxAYL(VL)P(LI)D |
| | A3 | LAYxxYTSG(ST)TGxPKG |
| | A4 | FDxS |
| | A5 | NxYGPTE |
| | A6 | GELxJGx(VL)ARGYL |
| | A7 | Y(RK)TGDL |
| | A8 | GRxPxQVKIRGxRIELGEIE |
| | A9 | LPxYM(IV)P |
| | A10 | NGK(VL)DR |
| PCP-domains | T | LGG(DH)SL |
| C-domains | C1 | SxAQxR(LM)(WY)xL |
| | C2 | RHExLRTxF |
| | C3 | MHHxISDG(WV)S |
| | C4 | YxD(FY)AVW |
| | C5 | (IV)GxFVNT(QL)(CA)xR |
| | C6 | HN)QD(YD)PFE |
| | C7 | RDxSRNPL |
| Te-domains | Te | GxSxG |
| E-domains | E1 | PIQxWF |
| | E2 | HHxISDG(WV)S |
| | E3 | DxLLxAxG |
| | E4 | EGHGRE |
| | E5 | RTVGWFTxxYP(YV)PFE |
| | E6 | PxxGxGYG |
| | E7 | FNYLG(QR) |
| Cy-domains | Cy1 | FPL(TS)xxQxAYxxGR |
| | Cy2 | RHx(IM)L(PAL)x(ND)GxQ |
| | Cy3 | LPxxPxLPLxxxP |
| | Cy4 | (TS)(PA)3x(LAF)6x(IVT)LxxW |
| | Cy5 | (GA)DFTxLxLL |
| | Cy6 | PVVFTSxL |
| | Cy7 | (ST)(QR)TPQVx(LI)D13xWD |
| Ox-domains | Ox1 | KYxYxSxGxxY(PG)VQ |
| | Ox2 | GxxxG(LV)xxGxYYY(HD)P |
| | Ox3 | IxxxYG |
| N-Mt-domains | M1 | VL(DE)xGxGxG |
| | M2 | NELSxYRYxAV |
| | M3 | VExSxARQxGxLD |
| R-domains | R1 | V(L)(L)TG(A)TG(F)(L)GxxLL |
| | R2 | Vx(L)(L)VR(A) |
| | R3 | GPL(G)x(P)x(L)GL |
| | R4 | V(Y)PYxYLxx(P)NVxxT |
| | R5 | GYxxSKW(A)(A)E |
| | R6 | R(P)G |
| | R7 | YxxxxG(LF)LxxP |

## E) NRPS strategies

The mechanism of NRPSs is classified into three classes which are linear NRPSs (type A), iterative NRPSs (type B), and nonlinear NRPSs (type C) (Mootz et al. 2002b). Each class will be detailed in the following sections:

### E.1. Linear biosynthesis

In this type, the number of modules indicates the sequence and number of amino acids incorporated into a non ribosomal peptide. The initiation module lacks the C domain (except for some LP synthetases where the initiation module contains a C-starter domain) and the domains are organized in the order C-A-PCP in the elongation module. The terminal module in most cases contains a Te domain to release the full-length peptide from the enzyme (Keating et al. 2001). The common organization of module and domains is [A-PCP]-[C-A-PCP]n-[C-A-PCP-Te]. Modules can contain optional domains for the modification of the corresponding amino acid. The synthesis can be organized in one or more interacting enzymes. As in case of ACV synthetases ,the penicillin and cephalosporin precursor (Fig. 13A), all modules are organized in one polypeptide chain, whereas in surfactin synthetases, 24 domains organized in seven modules which are distributed over three NRPSs interacting subunits that act in the directional synthesis of surfactin, starting at module 1 and ending at the Te domain of module 7 (Fig. 13 B) (Mootz et al. 2002b).

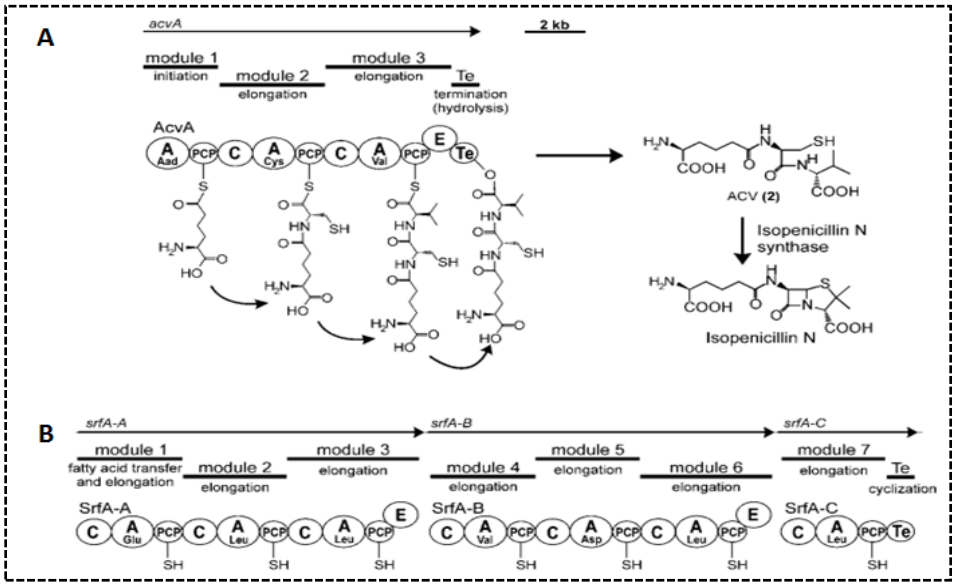



Figure 13: Organization and mode of biosynthesis of linear NRPSs (A) The ACV (B) surfactin (Mootz et al. 2002b).

## E.2. Iterative biosynthesis

In iterative NRPSs, modules or domains can be used more than once in the assembly of one single product. It is applied to build up peptide chains that consist of repeated smaller sequences. For example, gramicidin S is synthesized by a pentapeptide that is stalled on the active-site serine of the Te-domain. The regenerated NRPSs engage in a second round of synthesis and the Te-domain finally catalyzes the head-to-tail cyclization to give the mature product (Kohli et al. 2001). The siderophore enterobactin produced by *Escherichia coli* is composed of three dihydroxybenzoyl (Dhb) –serine (Ser) units. Enterobactin synthesis is organized in two modules. Enterobactin is oligomerized and cyclized on the Te domain (Fig.14) (Gehring et al. 1998).

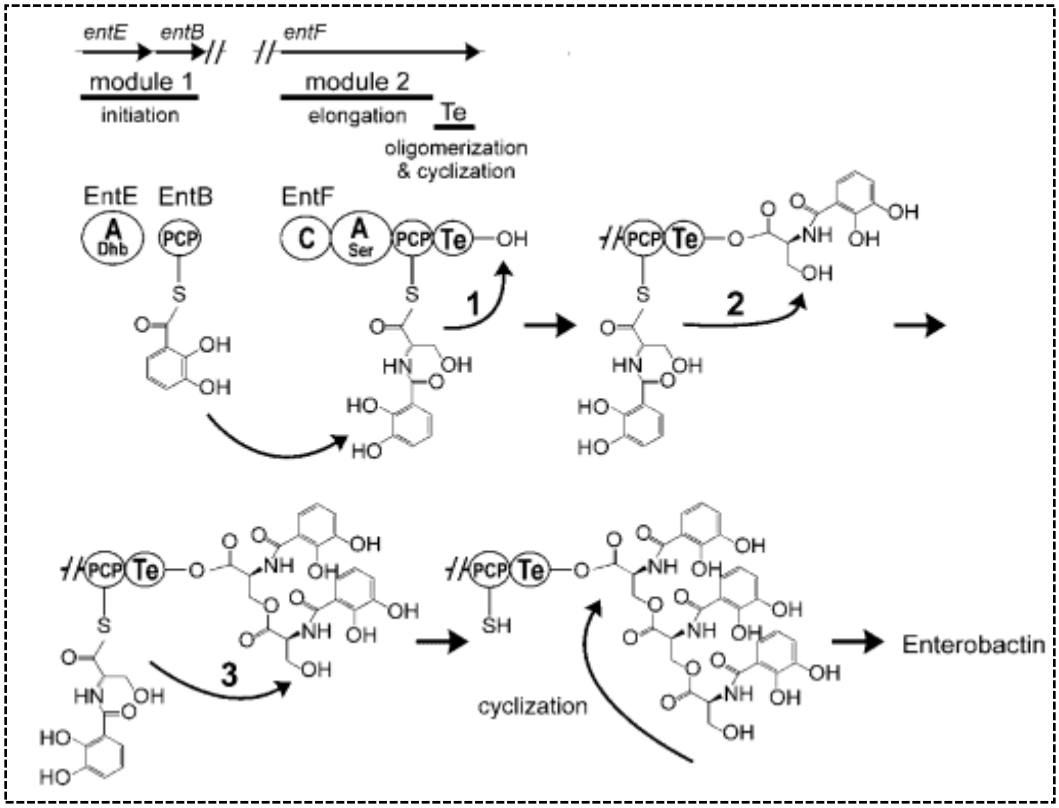



Figure 14: Enterobactin NRPS as example for iterative NRPS (Mootz et al. 2002b).

### E.3. Non linear biosynthesis

This strategy is more complicated and characterized by an arrangement of modules that deviates from the classical (C-A-PCP)n arrangement or by incorporation of small molecules that are not covalently bound to the NRPS template during synthesis (Mootz et al. 2002b). Vibriobactin is synthesized using this complicated strategy by the decoration of norspermidine (NS) with dihydroxybenzoate (DHB) and with two molecules of DHB-mOx (DHB-methyloxyzoline) by the second C-domain of *VibF*. Synthesis of DHB-mOx is realized by transfer of DHB to *VibF*, which condenses it with the heterocyclized threonine (Fig.15) (Finking and Marahiel 2004).

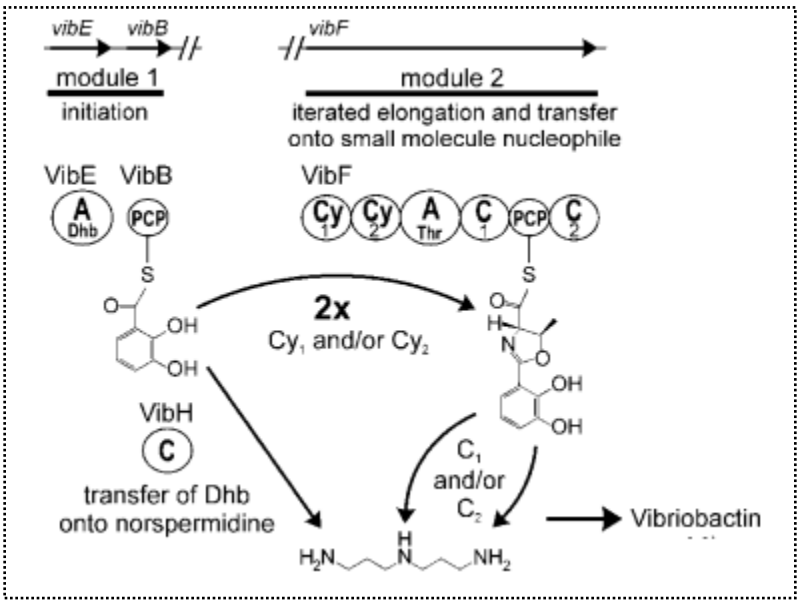



Figure 15: Biosynthesis of Vibriobactin as an example of non linear NRPS strategy (Mootz et al. 2002b).

## F) Modular Polyketides (PKS)

Polyketides are a class of natural products found in bacteria, fungi and plants, and have a great interest in medicine as antibiotics. They are synthesized by polyketide synthases (PKSs) which are modular enzymes working in the same way as NRPS, but the domains and substrates are different (Walsh 2008). The synthesis of a polyketide requires at least four domains that can be organized in modules much like NRPSs. The acyltransferase (AT) selects a carboxylic acid such as acetyl-CoA and methylmalonyl-CoA and holds it on to the acyl carrier protein (ACP) domain. The ketosynthase (KS) domain catalyses the condensation reactions between the upstream acyl-*S*-ACP and the downstream ACP bound carboxy acid. The Te domain release the final product from enzyme (Fig.16) (Hertweck 2009; Staunton and Weissman 2001).

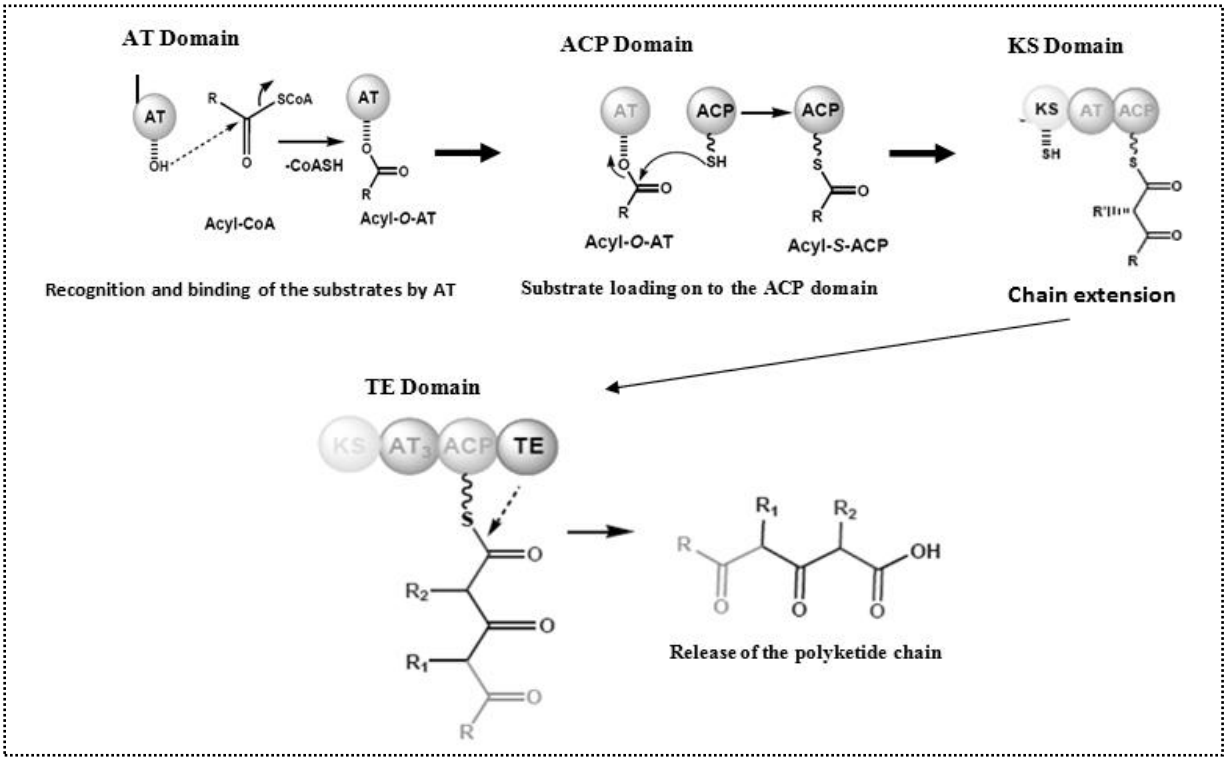



Figure 16: Biosynthesis of Polyketides. AT-acyl transferase; ACP, acyl carrier protein; KS, ketosynthase; Te, thioesterase; CoA, coenzyme (Fischbach and Walsh 2006).

Other domains such as ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) can be inserted to modify the polyketide structure. The KR domain reduces β-ketoacyl-*S*-ACP to β-hydroxyacyl-*S*-ACP which can be dehydrated to α,β-enoyl-*S*-ACP by DH domain finally reduced by the action of an ER domain (Fig.17) (Hertweck 2009).

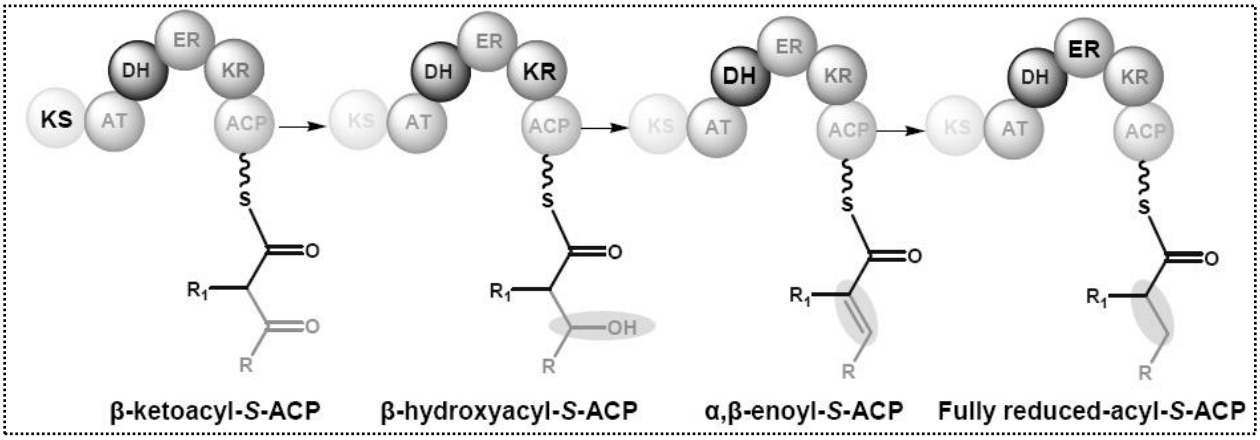



Figure 17: Additional domains involved in the modifications of PKS. AT, acyl transferase; ACP, acyl carrier protein; KS, ketosynthase ; DH, dehydratase; KR, ketoreductase; ER, enoyl reductase (Fischbach and Walsh 2006).

There are three types of PKSs (PKS type I, PKS type II and PKS type III) and each class will be detailed in the following sections:

### F.1. PKS type I

The type I of PKSs are organized into modules, each of which contains the three main domains (Acyltransferase (AT), acyl carrier protein (ACP) and ketosynthase (KS) and the variable set of domains (ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)). Based on their mode of biosynthesis, PKSs type 1 are divided into modules in which active site is used only once and iterative in which active sites are reused repeatedly (Shen 2003). In modular PKSs, the final number of ketide units in the synthesized polyketides is equal to the number of modules.

### F.2. PKS type II

The type II PKSs are multienzyme complexes that carry a single set of iteratively acting activities. The required catalytic domains are located on individual proteins that interact to form a functional PKS enzyme complex (Meier and Burkart 2009).

### F.3. PKS type III

The type III PKS differ from the two other types by not relying on acyl carrier protein domains, they act directly on the acyl CoA substrates (Meier and Burkart 2009). They are essentially iteratively acting condensing enzymes.

### F.4. Hybrids system (NRPS/PKS)

The hybrid system of NRPS/PKS is synthesized via hybrid gene clusters containing genes coding for both NRPS and PKS module-containing proteins. This system is very common and sometimes even more predominant than discrete PKS and NRPS gene clusters (Du et al. 2001). The same catalytic sites are conserved between the hybrid NRPS-PKS and normal NRPS or PKS systems (Fig. 18). The ketoacyl synthase domain in NRPS/PKS hybrids is unique, and that specific interpolypeptide linkers exist at both the C- and N-termini of the NRPS and PKS proteins, which presumably play a critical role in facilitating the transfer of the growing peptide or polyketide intermediate between NRPS and PKS modules in hybrid NRPS-PKS systems.

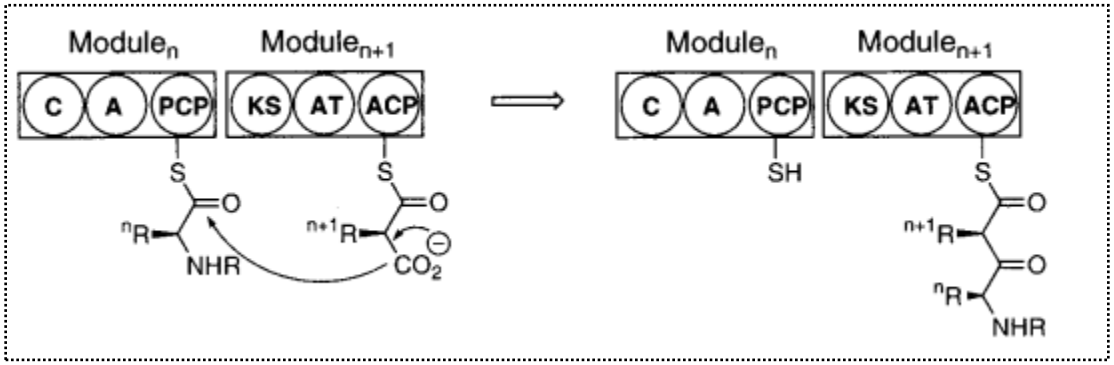



Figure 18: Modular organization of hybrid NRPS-PKS (Du et al. 2001) A, adenylation; PCP, peptidyl carrier protein; KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein.

## G) Strategy to discover new natural products

Discovery of new compounds from natural sources has largely been an adventitious process based on bioactivity screening of microbial extracts combined with bioassay guided identification and natural product structure elucidation. Once the chemical structure of the novel compound is clarified, its biosynthesis can be described on the biochemical and genetic levels. This exploration gives insight into the mechanisms of synthesis like those for non ribosomal peptides. The availability of microbial genome sequences led to the discovery of several biosynthetic gene clusters for which the products are still unknown. As in the case of non ribosomal peptides and polyketides, the bioinformatic analysis of their biosynthesis enzymes might indicate the presence of Aromatic amino or an olefinic partial structure which are known to absorb UV light at specific wavelengths, which led them to be predicted. Therefore, information about these clusters, pathways and metabolites is needed to facilitate the process of natural product discovery. By combining technological development in genomics, bioinformatic and structure biology, translated gene sequence data can be used to rapidly derive structural elements encoded by secondary metabolic gene clusters from microorganisms.

Microbial genome mining has recently guided the discovery of new secondary metabolites such as non ribosomal peptides (NRPs) and Polyketides (PKs). It is a very powerful approach to search for the biosynthetic gene clusters and allow the prediction of the product and identification of the biosynthesis pathway of non ribosomal natural products (Corre and Challis 2009). Different strategies (Fig.19) have been established to predict and isolate the natural compounds including structure and genetic approaches.

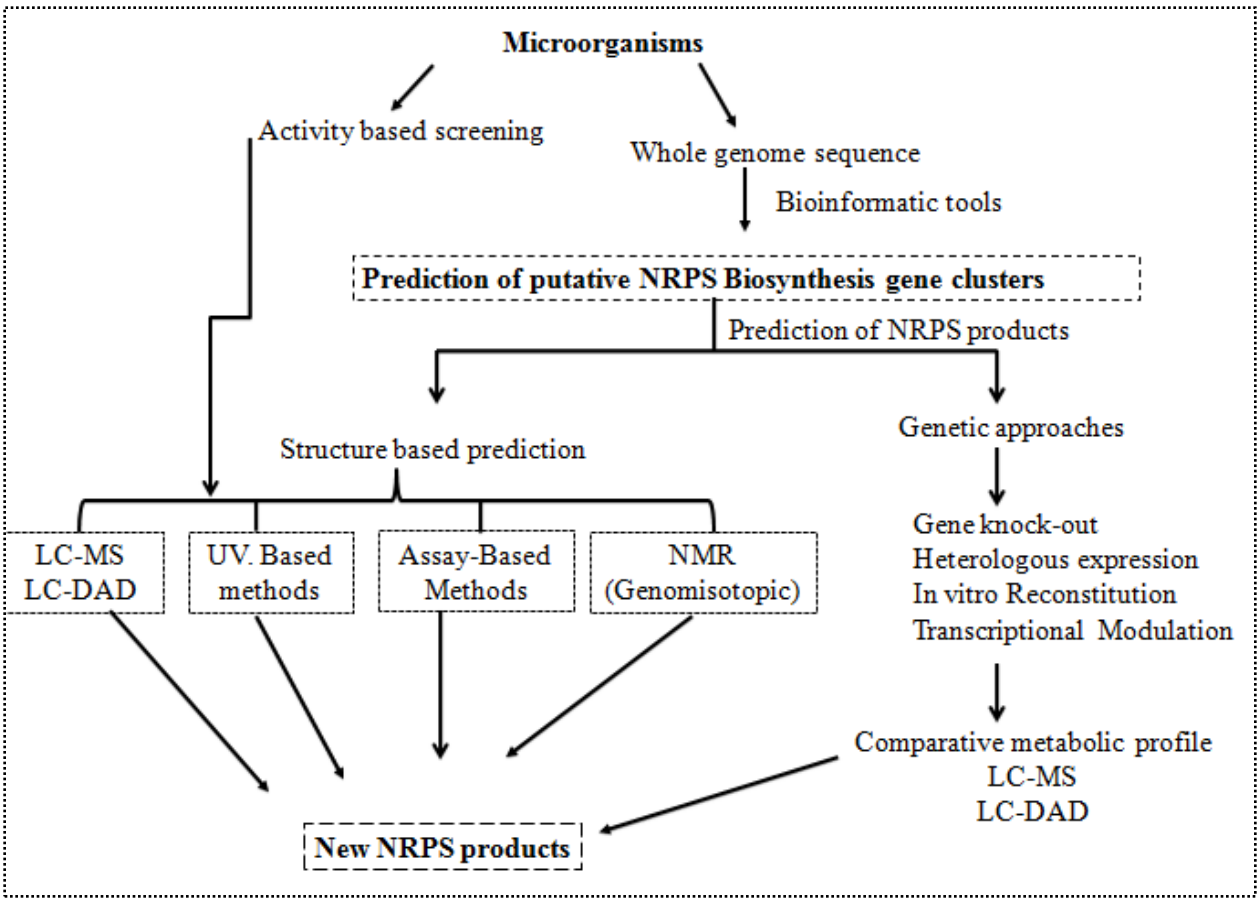



Figure19: Different strategies used to identify the NRPs products. This figure is adapted from (Zerikly and Challis 2009).

### G.1.Structure and Bioactivity based screening

Screening of biological activities and structure features of the metabolic products can lead to the prediction of putative physicochemical properties of targeted metabolites which can be achieved by bioinformatic predictions (Fig.19). This approach is based on the relationship between the organization of gene clusters and the structure of the assembly line products which guide the identification of the putative physicochemical properties of the target product. This includes the variation of culture conditions (media, temperature, PH, etc) to stimulate the expression of the wanted biosynthetic pathway called “One Strain Many Compounds Approach, OSMAC” (Christian et al. 2005). Based on the biological activities, microbial culture extracts can be screened for the occurrence of specific compounds. The most well known example associated with this method is the application of the chrome azurol S (CAS) to detect the presence of siderophores in the medium (Schwyn and Neilands 1987).

By mining the sequenced genome combining with bioinformatic tools, it is also possible to screen the structure feature of the metabolic products. For example, the metabolites with a chromophore can be detected with UV profile that can absorb UV light at specific wavelengths (Nguyen et al. 2008). Genomisotopic approach is also a good strategy to isolate the secondary metabolites based on their structure feature by tracing through selective NMR experiments (Gross et al. 2005).

**G.2. Genetic Approaches**

The availability of genetic information derived from genome sequences facilitated the identification of biosynthesis pathways through genetic engineering to stimulate, augment or to block their production.

Genetic studies include the knockout of particular gene cluster and subsequent comparison of the metabolic profile between the wild type and the mutant strains and in-vitro reconstitution of the genes and their expression in a heterologous host (Meier and Burkart 2009). They are very useful to detect the secondary metabolites. A further possibility is the transcription activating of the biosynthetic gene cluster to get access to the regulatory genes (Bok et al. 2006). Recombinant biosynthetic proteins can also be incubated with their appropriate substrate and the resultant compound is then analyzed further (McClerren et al. 2006).

The knockout of particular gene cluster and subsequent comparison of the metabolic profile between the wild type and the mutant strains is a widely used approach to identify genes that direct the production of known products (Fig.20). This method was successfully applied to discover the coelichelin, a new tripeptide siderophore produced by *Streptomyces coelicolor*, from genome sequence data (Challis and Ravel 2000).

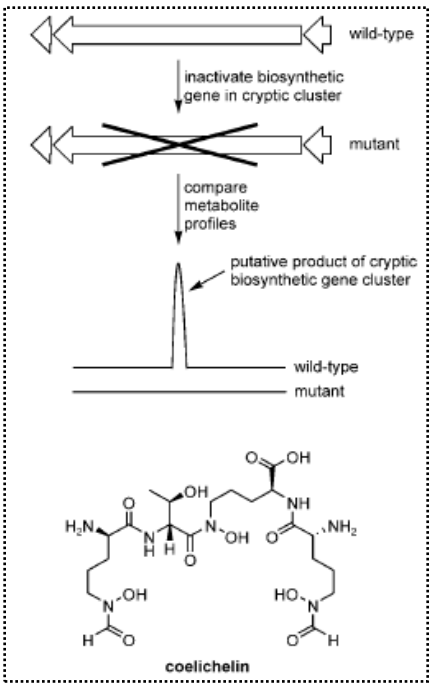



Figure 20: Steps involved in the identification of metabolic products of silent cryptic biosynthetic gene clusters by mutagenesis and comparative metabolic profiling approach (Challis and Ravel 2000).

## H) Specific bioinformatic tools for NRPS analysis

Genome mining is based on two basic steps including the identification of target protein from translated genomic sequences and the prediction of its function. Basic Local Alignment Search Tool (BLAST) is the most useful tool to recover any sequence having at least a short region in common with the query sequence. Several bioinformatics tools (Table 2) are available for the identification of modular genes from gene sequences and also to predict the structures of the associated secondary metabolites (Leclère et al. 2016). Different tools have been developed to analyze and mine PKS and NRPS biosynthetic gene clusters such as ClustScan (Starcevic et al. 2008) and GNP Genome Search (Johnston et al. 2015). Moreover, the workflow Florine, dedicated to identify the NRPS gene clusters in microorganisms including bacteria and fungi has been proposed to facilitate the research strategy (Caradec et al. 2014). This workflow includes the use of bioinformatics tools dedicated to secondary metabolite discovery as antiSMASH (Weber et al. 2015), PKS/NRPS analysis (Bachmann and Ravel 2009), and NRPS-PKS database (Yadav et al. 2003); and the Norine database (Flissi et al. 2016).

Comparative analyses of NRPS gene clusters and enzymes of already known compounds simplified the identification of consensus motifs indicative for the enzymatic activity of the encoded protein. For example, based on Stachelhaus code, it is possible to predict the substrate that will be selected by an A domain of an NRPS module (Stachelhaus et al. 1999b). Nowadays, SEARCHGTr is available to predict glycosyltransferase, enzyme involves in biosynthesis of natural products including NRPs (Kamra et al. 2005).

Currently, different set of tools are used specifically to provide automated specificity prediction for NRPS A-domains such as NRPS predictor 2 (Röttig et al. 2011), PKS/NRPS analysis (Bachmann and Ravel 2009), NRPS/PKS substrate predictor (Baranašić et al. 2014), and NRPS-PKS database (Yadav et al. 2003). The type of C domains, whether it is $^{L}C_{L}$, $^{D}C_{L}$ or dual C/E can also be detected and extracted by NaPDoS (Ziemert et al. 2012). Norine is a unique bioinformatic resource devoted to non ribosomal peptides. Until now, it include more than 1100 NRPs uploaded from scientific literature or directly submitted by researchers (Pupin et al. 2015). The novelty of predicted peptides can be checked by Norine based on the structure-based search which helps to find the similar peptides that already exist. Using these tools, several biosynthetic gene clusters can be identified from a sequenced genome.

Table 2: Specific bioinformatic tools dedicated to NRPS PKS product research

| Software or database | Main function | URL | Reference |
|---|---|---|---|
| AntiSMASH | Analyze BGCs of all SM Domain organization of NRPSPKS | http://antismash.secondarymetabolites.org/ | (Weber et al. 2015) |
| NRPSPKS analysis | Determine domain organization and A domain specificities | http://nrps.igs.umaryland.edu/nrps/ | (Bachmann and Ravel 2009) |
| NRPSpedictor2 | Predict Adomain specificities | http://nrps.informatik.unituebingen.de | (Röttig et al. 2011) |
| NaPDoS | Phylogenetic analysis of PKSKS and NRPSC domains | http://napdos.ucsd.edu | (Ziemert et al. 2012) |
| CLUSEAN | Analyze and mine BGCs of KS/NRPS | https://bitbucket.org/tilmweber/clusean | (Weber et al. 2009) |
| NRPSPKS/SBSPKS | Mine BGCs of PKS and NRPS | http://www.nii.ac.in/~pksdb/sbspks/master.html | (Anand et al. 2010) |
| GNP/Genome Search | Mine and analyze BGCs, mainly PKS/NRPS | http://magarveylab.ca/gnp/#!/genome | (Johnston et al. 2015) |
| ClustScan Professional | Analyze BGCs of PKS/NRPS | http://bioserv.pbf.hr/cms/index.php?page=clustscan | (Starcevic et al. 2008) |
| NP.searcher | Analyze BGCs of PKS/NRPS | http://dna.sherman.lsi.umich.edu/ | (Li et al. 2009) |
| GNP/PRISM | Mine and analyze BGCs, mainly PKS/NRPS, including glycosylations and structure prediction | http://magarveylab.ca/prism | (Skinnider et al. 2015) |
| LSI based Adomain function predictor | Web application to predict Adomain specificities | http://bioserv7.bioinfo.pbf.hr/LSIpredictor/AdomainPrediction.jsp | (Baranašić et al. 2014) |
| NRPS/PKS substrate predictor | Web application to predict Adomain/ATdomain specificities | http://www.cmbi.ru.nl/NRPSPKSsubstratepredictor/ | (Khayatt et al. 2013) |
| NRPSsp | Predict Adomain specificities | http://www.nrpssp.com/ | (Prieto et al. 2012) |
| SEQLNRPS | Predict Adomain specificities | http://services.birc.au.dk/seqlnrps/ | (Knudsen et al. 2015) |
| SEARCHGTr | Predict glycosyltransferase specificities | http://linux1.nii.res.in/~pankaj/gt/gt_DB/html_files/searchgtr.html | (Kamra et al. 2005) |
| DoBISCUIT | Web accessible database of PKS/NRPS BGCs | http://www.bio.nite.go.jp/pks/ | (Ichikawa et al. 2012) |
| ClustScan Database | Web accessible database of PKS/NRPS BGCs | http://csdb.bioserv.pbf.hr/csdb/ClustScanWeb.html | (Diminic et al. 2013) |
| NORINE | Web accessible database on NRPs | http://bioinfo.lifl.fr/norine | (Pupin et al. 2015) |

# I) The genus *Aeromonas*

## I.1. Organism description

*Aeromonas* is a Gram negative, motile, facultative anaerobic, non-spore-forming, rod-shaped, freshwater bacterium. Members of *Aeromonas* genus belong to the class *Gammaproteobacteria*, order *Aeromonadales* and family *Aeromonadaceae* (Martin-Carnahan and Joseph 2005). The classification of *Aeromonas* can be demonstrated based on phenotypic methods including morphological, physiological and biochemical methods and genetic level, which improves with more and more available sequence data. At present, the genus *Aeromonas* contains of more than 25 accepted and validated species, some of them contain subspecies as three under *A. hydrophila* (*hydrophila*, *dhakensis* and *ranae*) and five for *A. salmonicida* (*achromogenes*, subsp. *masoucida*, subsp. *pectinolytica*, subsp. *salmonicida*, and subsp. *smithia*) (Martin-Carnahan and Joseph 2005) (Fig. 21).

## I.2. Diversity and significance of *Aeromonas* strains

Due to their pathogenicity for fishes, different strains of *Aeromonas* received very high economical attention in the last two decades. *Aeromonas* sp. are ubiquitous bacteria found in a variety of aquatic environments worldwide, including well, bottled, flood and heavily polluted water. Most of them are known as mesophilic, except *A. salmonicida* (Optimum growth at 22–28 °C), non-motile and pigmented and are the principal fish pathogens (Fig.22). They can well survive in water, even at low temperatures, allowing growth in food refrigeration (Daskalov 2006). *Aeromonas* species have a broad host spectrum and can cause infections in both cold-and warm-blooded animals, including humans. *A. hydrophila* strains are known to cause disease in both fishes and human beings, whereas *A. salmonicida* is a specific pathogen of salmonid fishes, and is capable of causing disease in healthy fishes at very low levels of infection. *Aeromonas sp*. (especially *A. hydrophila*) has the status of a foodborn pathogen of emerging importance, and has been shown to be a significant cause of infections associated with natural disasters such as hurricanes as Katrina, tsunamis and earthquakes (Hiransuthikul et al. 2005). *A. hydrophila* can also cause a wide range of diseases in poikilothermic and homeothermic animals such as red leg and furunculosis in frogs and fishes. *Aeromonas* infections caused by some strains as *A. salmonicida* produce septicaemia, furunculosis, fin rot, soft tissue rot, ulcerative and haemorrhagic diseases in fish, causing significant mortality in both wild and farmed freshwater

and marine fish species that bring significant economic losses of the aquaculture sector (Gudmundsdóttir and Björnsdóttir 2007; Reith et al. 2008). In human, *Aeromonas* causes wound infection, septicemia, diarrhea, gastroenteritis, meningitis and pneumonia (Joseph et al. 1991). The opportunistic *A. hydrophila* caused notable disagreements in immunocompromised people. It has been revealed that an increasing number of intestinal and extraintestinal diseases documented have been caused by this pathogen, regardless of immunological state. Infections of skin and soft tissues have been well documented and more alarming diseases, such as haemolytic uremic syndrome and necrotising fasciitis (Abuhammour et al. 2006).

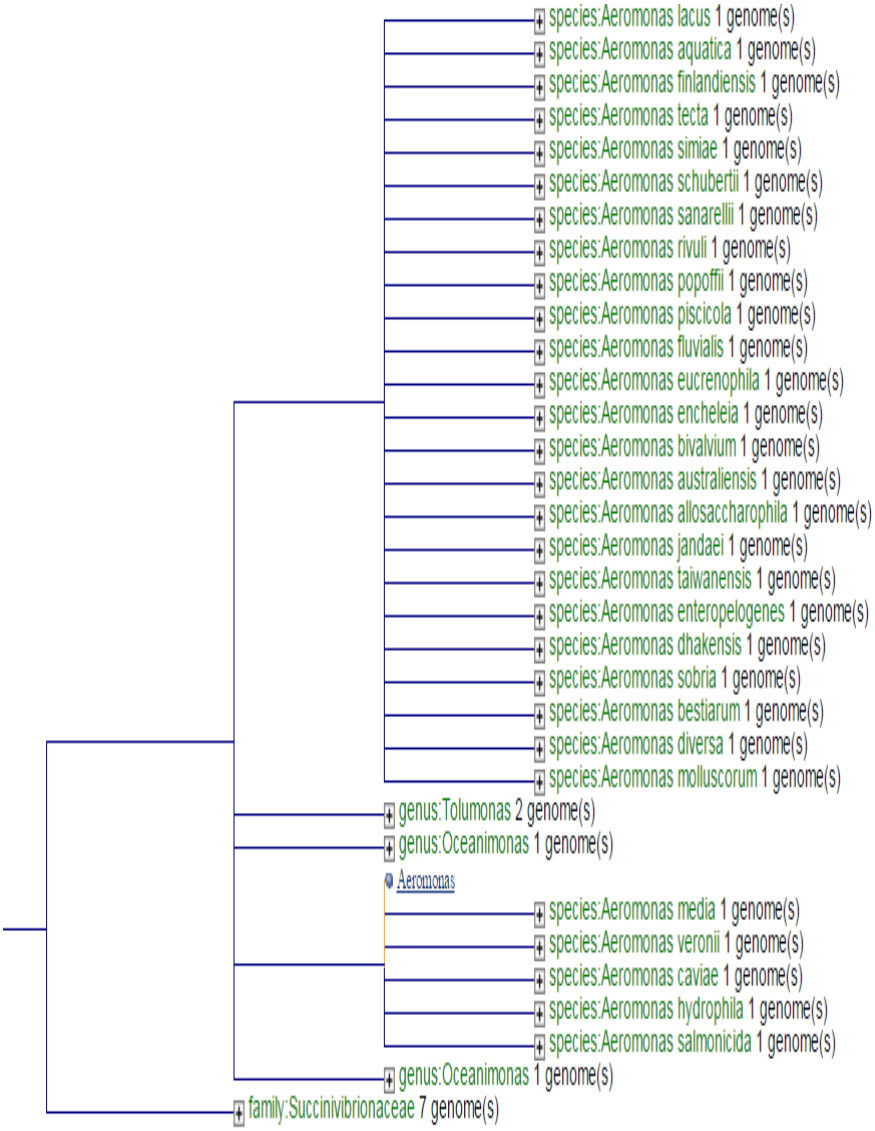



Figure 21: Phylogenetic tree of all sequenced genomes of *Aeromonas* strains available in NCBI

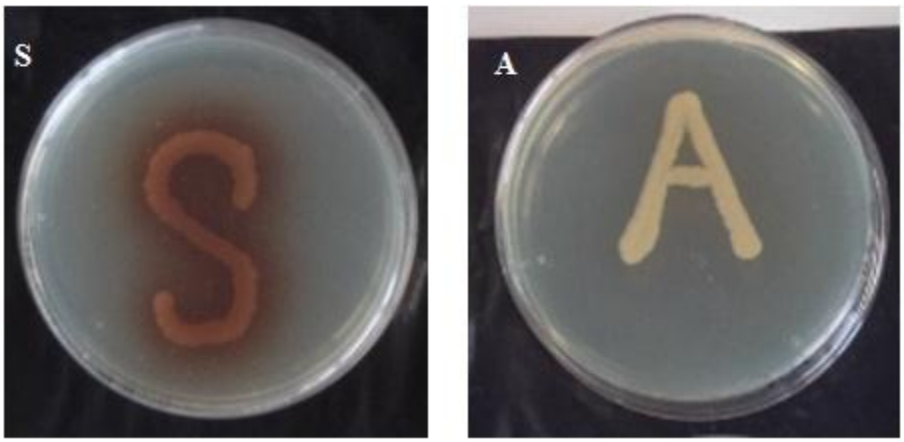



Figure 22: Growth of *Aeromonas* strains on king B medium A: *Aeromonas hydrophila* with salmon colour S: *Aeromonas salmonicida* with brown pigment.

### I.3. Virulence factors of *Aeromonas*

Pathogenicity and virulence of *Aeromonas* depend on different factors that help to adhere to host tissue, enter the host and evade host defence mechanisms. These include the excretion of extracellular virulence factors such as toxins, invasions or adhesion compounds (LPS), S-layers, flagella, communication systems (i.e., quorum sensing), Ferric Uptake Regulator (Fur) and siderophores (Beaz-Hidalgo and Figueras 2013). The pathogenicity is affected by the growth conditions and can vary from laboratory to field conditions. For example, it has been reported that growth of *A. salmonicida* under laboratory conditions leads to a loss of virulence genes (Daher et al. 2011; Dallaire-Dufresne et al. 2014).

Iron is one of the most important micronutrient used by microorganisms, it is essential for their metabolism and required as a cofactor for a large number of enzymes and iron containing proteins. Under conditions of low- iron concentrations, most isolates of *Aeromonas* species produce a siderophore, either amonabactin or the enteric siderophore enterobactin (Barghouthi et al. 1989a; Barghouthi et al. 1989b), both used the catecholate group to chelate iron as they contain the 2,3-dihydroxybenzoic acid (Dhb) monomer as a functional group. Amonabactins are a family of 4 variants, peptide-based catecholate siderophores, produced by *A. hydrophila* in response to low-iron conditions. The biologically active forms are composed of Dhb, Lys, an aromatic residue, either tryptophan (amonabactin T) or phenylalanine (amonabactin P) and a facultative Gly.

Amonabactin is an aeromonad virulence factor as amonabactin-producing strains may acquire iron from the vertebrate serum component Fe-transferrin, whereas enterobactin-producing strains cannot (Massad et al. 1991). Siderophores strongly and specifically bind and solubilize ferric iron and deliver the now bound iron to a specific receptor whereby the iron is engulfed and utilized as a nutritional source (Krewulak and Vogel 2008). The iron acquisition by siderophores is generally regulated at the transcriptional level by the iron-dependent Fur protein which recognizes a consensus 19 bp Fur-box or iron-box on the DNA (Baichoo and Helmann 2002). The iron uptake is mediated by specific receptors. In Gram negative bacteria, it requires an outer membrane receptor, a periplasmic protein (PBP) and an inner membrane ATP-binding cassette (ABC) transporter. The transport of a ferric-siderophore into the periplasmic space also requires energy that is transferred between cytoplamic and outer membranes via three proteins: TonB (26 kDa), ExbB (26 kDa) and ExbD (17 kDa) (Fig.23). Both last ones act together to couple the activity of TonB to the proton gradient of the cytoplasmic membrane (Krewulak and Vogel 2008).

*Aeromonas* spp are important fish pathogens and a persistent threat to the aquaculture sector. Therefore, genome analysis and the production of virulence factor including siderophores are urgently needed in the research for alternative and more effective control method of these fish pathogens.

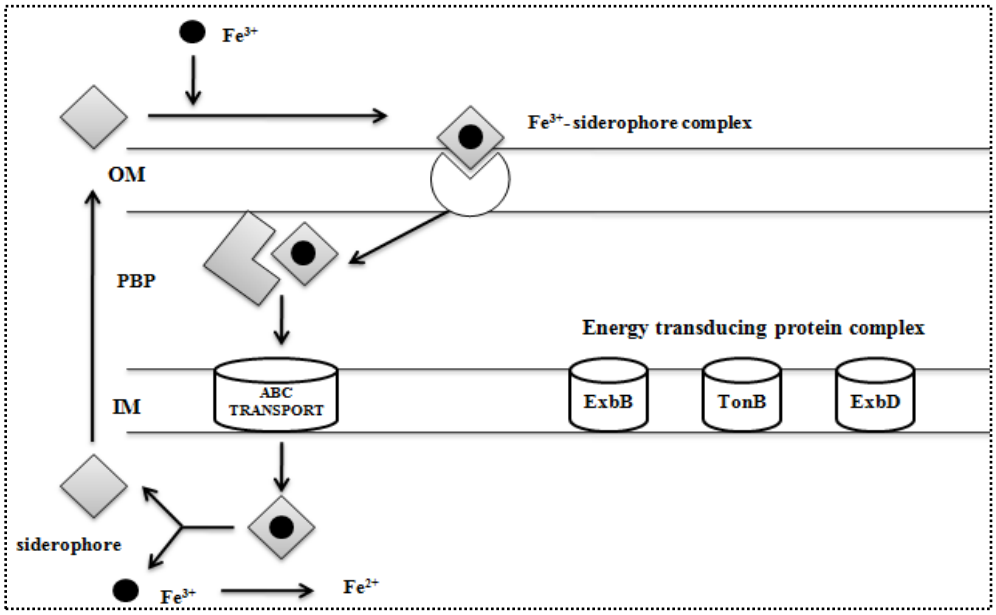



Figure 23: Siderophore uptake in Gram-negative bacteria (Andrews et al. 2003; Cornelis 2010). OM: Outer membrane, IM: Inner membrane, PBP: Periplasmic binding protein

## J) The genus *Burkholderia*

### J.1. Organism description

*Burkholderia* is a Gram negative β-proteobacterium, aerobic, non spore-forming, rod shaped straight or slightly curved (1 to 5 μm in length and 0.5 to 1.0 μm in width) that can be found in different environments including soil, water, plants, animals, and human (Mahenthiralingam et al. 2008b). They are mesophilic and grow in the optimal growth temperature ranges between 20ºC to 37ºC, They are able to utilize a variety of simple and complex carbohydrates including alcohols, amino acids hydrocarbon fuels as carbon sources (Mahenthiralingam 2005). *Burkholderia* firstly was described as a phytopathogenic bacterium, named *Pseudomonas cepacia* that causes rot in onions. Molecular taxonomic analysis including rRNA-DNA, DNA-DNA hybridization techniques and fatty acid analysis led to the transfer of these bacteria to a new genus distinct from *Pseudomonas*, called *Burkholderia* (Yabuuchi et al. 1992). Currently, the genus *Burkholderia* contains more than sixty species; all belonging to proteobacteria (Fig. 24). The presence of high heterogeneity among the different strains of *Burkholderia* leads to identify several new members in the past two decades. The opportunistic pathogen, *B. cepacia*, was repeatedly isolated from patients with cystic fibrosis (CF) (Isles et al. 1984) and then classified into *Burkholderia cepacia* complex (BCC) that consists of 17 species which share 98-100% similarity in their 16S rRNA gene and 94-95% in their *recA* gene sequences (Vanlaere et al. 2009; Vanlaere et al. 2008).

### J.2. Diversity and significance of *Burkholderia* species

*Burkholderia* members are well known for their nutritional versatility, which certainly contributes to their capacity to live in extreme and diverse habitats. The ability to adapt and colonize a wide variety of environments is likely due to their large genomes (4.6 to 9Mb) which are often made of several replicons (generally 3 chromosomes). The presence of multiple insertion sequences which confer genome plasticity could also explain the versatility of the genus *Burkholderia* (Miché et al. 2001). *Burkholderia* species play significant role in the environments as they provide a number of ecological benefits to the environment that has also led to their use for bioremediation processes (Chain et al. 2006; Mahenthiralingam 2005). Some members represent ecological benefits to the environment and actively degrade pollutant chemical as polychlorobiphenyl (PCB), methyl parathion (MP) or polyaromatic hydrocarbons

(PAH) (Liu et al. 2014; Martínez et al. 2007; Schamfuß et al. 2013). It has been shown that *Burkholderia* play an important role in soil organic matter mineralization as they are able to dissolve organic matter from decomposing maple leaves (McNamara and Leff 2004; Štursová et al. 2012).

Some *Burkholderia* species are known as being strong plant-growth promoters (Caballero-Mellado J 2007). It has been shown that *B. brasilensis* and *B. kururiensis* are able to fix nitrogen and nodulate legumes (Baldani et al. 1997; Estrada-De Los Santos et al. 2001; Moulin et al. 2001). *Burkholderia* strains are common members of rhizosphere communities that have been found as environmental and plant beneficial species. For example, *B.tropica*, *B. graminis*, *B. unamae* and *B. silvatlantica* are common rhizobacteria of sugarcane, coffee, maize, pasture, and wheat (Caballero-Mellado et al. 2004; Castro-González et al. 2011; Estrada-De Los Santos et al. 2001; Perin et al. 2006). Beside their presence in the rhizosphere, *B. phytofirmans*, *B. kururiensis* M130 and *B. acidipaudis* were found to exhibit endophytic lifestyles within different plants (Aizawa et al. 2010; Mattos et al. 2008; Sessitsch et al. 2005). The first obligate symbiosis between *Burkholderia* and plants was described between *B. kirki* and *Psychotria kirki* (Carlier and Eberl 2012).

*Burkholderia* represent a promising future source of natural anti fungal compounds due to the incredible diversity of biologically active secondary metabolites they produce. Since the mid of-1990s, members of *Burkholderia* genus have been investigated as a source of novel biological active agents. Among these, pyrrolnitrin (Hammer et al. 1999), CF661 (Quan et al. 2009), phenazines (Cartwright et al. 1995) and a wide range of non ribosomally active peptides including burkholdins (Lin et al. 2012; Tawfik 2010), occidiofungin (Ellis et al. 2012), cepacidines (Lim 1994), AFC-BC11 (Kang et al. 1998) and rhizoxin (Partida-Martinez and Hertweck 2007), which represent the major source of natural products with diverse chemical structure and biological activities produced by *Burkholderia*. Different volatile compounds produced by *Burkholderia* have been reported to possess antifungal activity against a wide range of fungi.

Beside their ecological roles, members of Bcc are opportunistic pathogens. Other members including the species *B. mallei* and *B. pseudomallei* are obligate pathogen as they are causing glanders in animals and melioidosis in humans, respectively (Gilad 2007).

The geographical distribution and versatile ecological properties make these bacteria an interesting model to study, and for putative applications, a good understanding of their occurrence, ecological potential in the environment and their secondary metabolites is necessary.

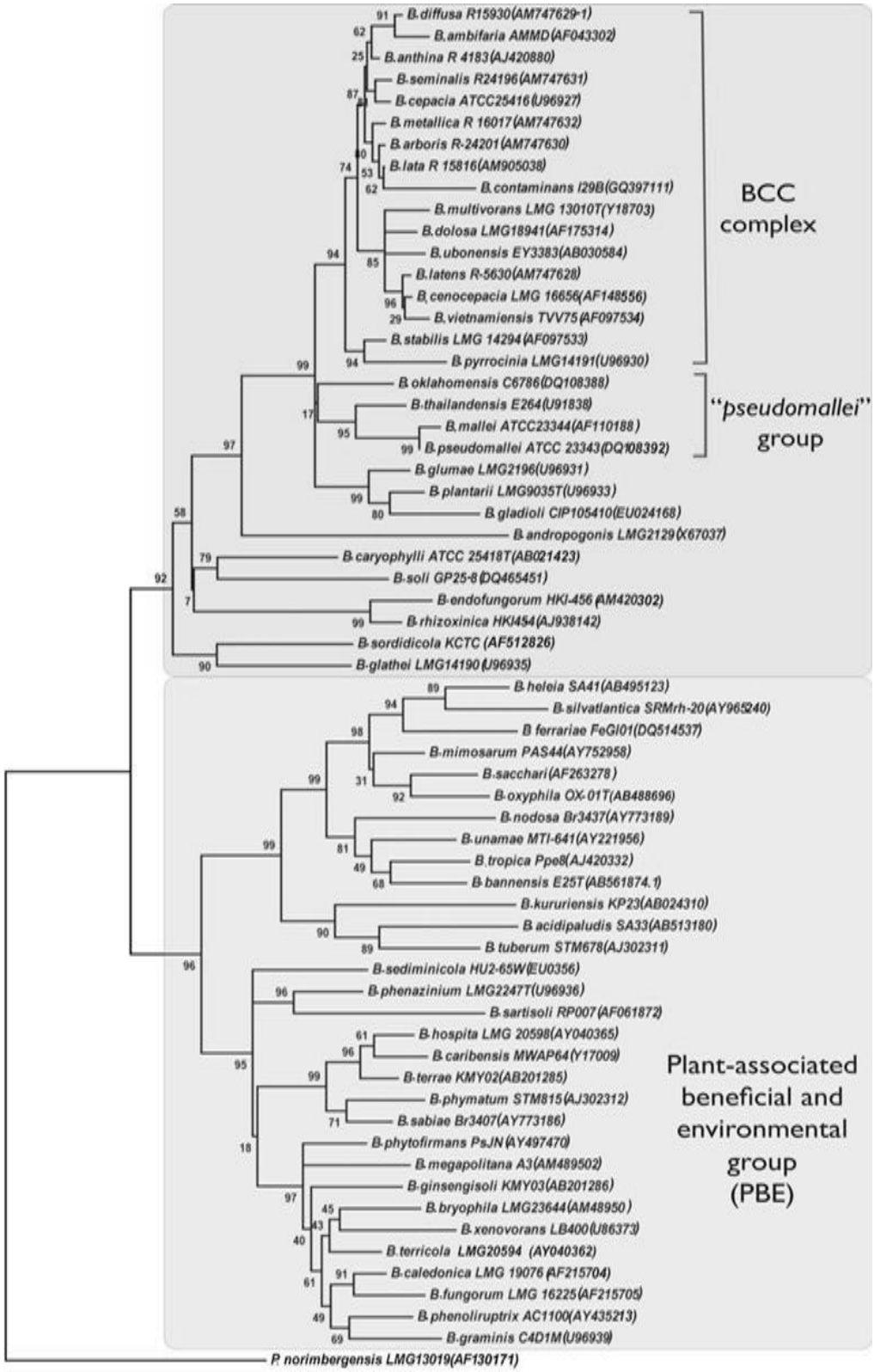



Figure 24: Phylogenetic tree based on 16S rRNA gene sequences of the recognised species of the *Burkholderia* genus (Suárez-Moreno et al. 2011).

# CHAPTER 2- RESULTS
# Identification of a new mechanism of Non Ribosomal Peptide Synthetases

## 2.1. Introduction

The biosynthesis of non ribosomal peptides has extensively been studied over the last years and different ways have been proposed. For that reason, it was necessary to study extensively the synthesis of some NRPs. This will increase knowledge about the mechanisms of NRPs synthesis and guide to discover more complex molecules produced by bacteria and fungi. It has been shown that under conditions of low- iron concentrations, most isolates of *Aeromonas* species produce iron transporting siderophores called amonabactins (Barghouthi et al. 1989a; Barghouthi et al. 1989b). The production of amonabactins by *A. hydrophila* has been reported more than twenty five years ago. Although the gene cluster of amonabactins is known, the mechanism of synthesis using NRPSs is not totally understood. Therefore, the objective of this current study was to investigate in details the biosynthetic gene cluster of amonabactins in *A.hydrophila* and to study the relationship between the domains organization of the NRPSs and the structure of the amonabactins and to validate our hypothesis by using mutants. Moreover, the use of bioinformatic tools led to an overview on the distribution of NRPS genes including siderophores. The biosynthetic way of these metabolites that are probably associated with the pathogenicity of these bacteria, were deciphered in all 30 sequenced genomes of *Aeromonas* strains available in NCBI. Such study is important to improve the knowledge about the mechanisms of non ribosomal peptides synthesis and help to understand the production of virulence factor including siderophores in *Aeromonas* strains.

*Aeromonas* strains including *A. hydrophila* and *A. salmonicida* were grown in iron limited media supplemented or not with amino acids including tryptophan, phenylalanine and glycine to induce the production of siderophores. Culture supernatants were then analyzed for the siderophore production with the CAS reagent to determine the concentrations, and submitted to mass spectrometry analyses (MALDI-TOF and MS-MS) for siderophore identification. To correlate the production of amonabactins to the synthesis operon, the NRPS gene *AHA_2474* (*amoG*) was disrupted by homologous recombination inserting a kanamycin resistance gene between signatures A1 and A9 of the adenylation domain. Kanamycin resistant clones were analyzed by PCR, using different sets of primers to check for correct insertion of the antibiotic resistance marker. In order to understand the synthesis mechanism of amonabactins, we proposed a unique mode of synthesis with iterative, alternative, and optional ways involved in the synthesis. NRPS gene *AHA_2473 (amoH)*, involved in amonabactin biosynthesis by incorporating a glycine

residue, was disrupted. The mutant was then further analyzed. The results are presented in an article published in Applied Microbiology and Biotechnology. The draft of the paper is included herein.

**Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism catalyzes amonabactin synthesis in *Aeromonas***

Qassim Esmaeel[1], Mickael Chevalier[1], Gabrielle Chataigné[1], Rathinasamy Subashkumar[1], Philippe Jacques[1,2], Valérie Leclère[1,3]

[1]: Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, EA 7394-ICV- Institut Charles Viollette, F-59000 Lille, France.

[2]: TERRA Research Centre, Microbial Processes and Interactions (MiPI), Gembloux Agro-Bio Tech University of Liege, B-5030 Gembloux, Belgium

[3] : Inria-Lille Nord Europe, Bonsai team, F-59655 Villeneuve d'Ascq Cedex, France.

Keywords: Aeromonas, amonabactin, nonribosomal peptidesynthetase, siderophore, NRPS

Corresponding author:

Qassim Esmaeel
Phone +33320 43 46 68, Fax +33 328 76 73 56, email: al_hamadi_82@yahoo.com

Appl Microbiol Biotechnol
DOI 10.1007/s00253-016-7773-4

GENOMICS, TRANSCRIPTOMICS, PROTEOMICS

# Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism catalyzes amonabactin synthesis in *Aeromonas*

**Qassim Esmaeel[1] · Mickael Chevalier[1] · Gabrielle Chataigné[1] · Rathinasamy Subashkumar[1] · Philippe Jacques[1,2] · Valérie Leclère[1,3]**

Received: 12 May 2016 / Revised: 25 July 2016 / Accepted: 1 August 2016


**Abstract** Based on the exploration of data generated by genome sequencing, a bioinformatics approach has been chosen to identify the biosynthetic pathway of the siderophores produced by *Aeromonas* species. The amonabactins, considered as a virulence factor, represent a family of four variants of catechol peptidic siderophores containing Dhb, Lys, Gly, and an aromatic residue either Trp or Phe in a D-configuration. The synthesis operon is constituted of seven genes named *amoCEBFAGH* and is iron-regulated. The cluster includes genes encoding proteins involved in the synthesis and incorporation of the Dhb monomer, and genes encoding specific nonribosomal peptide synthetases, which are responsible for the building of the peptidic moiety. The amonabactin assembly line displays a still so far not described atypical mode of synthesis that is iterative, alternative, and optional. A disruption mutant in the adenylation domain of AmoG was unable to synthesize any amonabactin and to grow in iron stress conditions while a deletion of *amoH* resulted in the production of only two over the four forms. The *amo* cluster is widespread among most of the *Aeromonas* species, only few species produces the enterobactin siderophore.

**Keywords** *Aeromonas* · Amonabactin · Nonribosomal peptide synthetase · Siderophore · NRPS

**Electronic supplementary material** The online version of this article (doi:10.1007/s00253-016-7773-4) contains supplementary material, which is available to authorized users.

✉ Qassim Esmaeel
al_hamadi_82@yahoo.com

[1] EA 7394-ICV-Institut Charles Viollette, Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, F-59000 Lille, France

[2] TERRA Research Centre, Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liege, B-5030 Gembloux, Belgium

[3] Inria-Lille Nord Europe, Univ Lille, CNRS, Centrale Lille, UMR9189 – CRIStAL- Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France

## Introduction

*Aeromonas* are Gram-negative, motile, facultative anaerobic, nonspore-forming bacteria. They are among the most common bacteria in aquatic environments through the world (Janda and Abbott 2010). *Aeromonas hydrophila* strains are known to cause disease in both fish and human beings (Janda and Abbott 2010), whereas *Aeromonas salmonicida* is a specific pathogen of salmonid fish (Austin and Austin 2007; Toranzo et al. 2005) and is capable of causing disease in a wide variety of nonsalmonid fish (Beaz-Hidalgo and Figueras 2013; Coscelli et al. 2014; Farto et al. 2011). Pathogenicity and virulence of *A. hydrophila* depend on the ability to produce factors such as toxins, invasion or adhesion compounds, S-layers, flagella, high-affinity iron chelators named siderophores, and indirectly ferric uptake regulator (Fur) as it regulates the production of some virulence factors (Daskalov 2006; Merino et al. 1995; Seshadri et al. 2006; Yu et al. 2005). Many bacterial siderophores are polypeptide derivatives that are nonribosomally synthesized by assembly lines constituted of nonribosomal peptides synthetases (NRPSs), while others that are not polypeptides are assembled by different enzymes designated as NRPS independent siderophores (NIS) (Challis 2005). NRPS and NIS genes are generally found in operons or clusters within the genome. NRPSs are themselves organized in sets of domains such as condensation, adenylation, thiolation, and thioesterase domains, sometimes accompanied by modifying domains as

Published online: 17 August 2016

the epimerization one (Strieker et al. 2010). Few years ago, different subtypes of C-domains have been identified depending on the building blocks involved in the bond formation catalyzed (Rausch et al. 2007). The C-starter domain acylates the first monomer with β-hydroxyl fatty acid, the $^{L}C_{L}$ catalyzes a peptidic bond between two L-monomers, and the $^{D}C_{L}$ joins an L-monomer to a growing peptide ending with a D-monomer. The dual C/E domains are involved in both epimerization and condensation. Hetrocyclization (Cy) domains catalyze both peptide bond formation and cyclization of cysteine or serine residues. The termination of synthesis is catalyzed by the terminal enzyme of the ending module, named thioesterase domain (Te).

In iron-limited conditions, most isolates of *Aeromonas* species produce siderophore, either amonabactin or the enteric siderophore enterobactin (Barghouthi et al. 1989a, b). Amonabactins constitute a family of four coproduced forms, each composed of 2,3-dihydroxy benzoic acid (Dhb or diOH-Bz), lysine, an aromatic residue either phenylalanine (amonabactins P) or tryptophan (amonabactins T), and a glycine only present in two forms over the four (Barghouthi et al. 1989b; Telford et al. 1994; Telford and Raymond 1997). Although the amonabactin structures have been elucidated for several years, poor information remained available about their biosynthesis. However, based on mutants generated by chemical mutagenesis, it was proposed that the amonabactin biosynthetic pathway could be composed of two segments, one producing the Dhb, and the second assembling amonabactin from Dhb and other amino acids (Barghouthi et al. 1989a). A gene called *amoA* by the authors, involved in a first step of the biosynthetic pathway in *A. hydrophila* 495A2, has been identified as being an isochorismate synthetase, allowing the conversion of chorismate into isochorismate which is then transformed into Dhb (Barghouthi et al. 1991). The expression of this gene is controlled by a promoter including a putative iron-regulatory sequence resembling the Fur repressor protein-binding site (Escolar et al. 1999). Later on, in 2008, six biosynthesis genes, organized in a single cluster, were shown to be iron regulated through the Fur regulator. So, they were suspected to probably be implicated in nonribosomal siderophore production by *A. salmonicida* (Najimi et al. 2008). The sequencing of genomes of *A. hydrophila* subsp. *hydrophila* ATCC7966 (Seshadri et al. 2006) and *A. salmonicida* subsp. *salmonicida* A449 (Reith et al. 2008) opened the door for novel track to follow in the discovery of biosynthesis pathways, involving the association of several enzymes, as those leading to build up complex compounds like siderophores.

Therefore, the aim of this study was to decipher the biosynthesis pathway for amonabactins, using a bioinformatics approach to identify the potential synthetases, and to confirm the hypothetical nonribosomal biosynthesis using mutants. Bacterial genome mining via in silico analysis offers an attractive opportunity of discovering new secondary metabolites such as nonribosomal peptides (NRPs) (Van Lanen and Shen 2006). This kind of approach is facilitated, especially as specific bioinformatics tools are now available, including antiSMASH (Weber et al. 2015) or Norine (Flissi et al. 2015). Thus, the availability of whole genome sequences of 30 *Aeromonas* species allowed the identification of NRPS genes including siderophores that are probably associated with the pathogenicity of these bacteria. Relationship between the domain organization of the NRPSs and the structures of the amonabactins revealed a unique mode of synthesis with iterative, alternative, and optional use of the domains.

## Material and methods

### Culture conditions and media

Strains, plasmids, and primers used in this study are listed in Table 1. *A. hydrophila* ATCC 7966 and its mutants, and *Escherichia coli* were routinely grown in Luria-Bertani (LB) medium (per liter: tryptone 10 g; yeast extract 5 g; NaCl 5 g; pH 7.2) at 30 and 37 °C, respectively. *A. salmonicida* subsp. *salmonicida* was grown in nutrient medium (per liter: peptone 5.00 g; meat extract 3.00 g; pH 7.0) at 28 °C. *E. coli* WM3064 was grown in LB containing diaminopimelic acid (DAP) 100 mg $L^{-1}$. *Saccharomyces cerevisiae* InvSc1 was grown on yeast-extract-peptone-dextrose at 30 °C (Shanks et al. 2006). When required, ampicillin (Amp, 50 μg $mL^{-1}$), kanamycin (Km, 100 μg $mL^{-1}$), and gentamicin sulfate (25 μg $mL^{-1}$ for *E. coli* and 100 μg $mL^{-1}$for *A. hydrophila*) were added.

### Siderophore production

For culture in iron-limiting conditions, all glassware was treated with 10 % (*v*/*v*) nitric acid or HCL 6 M. *A. hydrophila* and *A. salmonicida* were grown in casamino acid (CAA)-containing medium and King's B medium (Leclère et al. 2009). Iron-deficient minimum medium (MM9) (Payne 1994) was also used according to the protocol previously described (Esmaeel et al. 2016). When required, 0.3 mM tryptophan, phenylalanine, and glycine were added. Iron-depleted medium was obtained by adding the chelating agent 2,2′-dipyridyl (DIP, 200 μM). For siderophore production, the strains were grown in 50 mL of iron-limited media (CAA, M9) broth for 48 h in 250-mL treated flasks incubated in an orbital shaker (150 rpm). Samples were centrifuged for 10 min at 10,000*g*, and supernatant was then filtered through 0.22-μm membrane (Millipore, Billerica, USA). The solution was stored at −20 °C. Siderophore concentration was determined by universal chemical assay using the chrome azurol-S reagent (CAS) (Schwyn and Neilands 1987) and was expressed as μM EDTA

**Table 1** Strains, plasmids, and primers used in this study

| Strains, plasmids, and primers | Description | Reference or source |
|---|---|---|
| Strains | | |
| *A. hydrophila* subsp. *hydrophila* ATCC 7966 | Wild type | CIP 76.14T |
| *A. salminicida* subsp. *salmonicida* | Wild type | DSM 19634 |
| *A. hydrophila–amoG*$^{-}$ | Mutant with disruption of NRPS gene *AHA_2474* | This study |
| *A. hydrophila-ΔamoH* | Mutant with deletion of NRPS gene *AHA_2473* | This study |
| *Saccharomyces cerevisiae* InvSc1 | Yeast strain for in vivo recombination *MAT*a/*MAT*α *leu*2/*leu*2 *trp*1–289/*trp*1–289 *ura*3–52/*ura*3–52 *his*3-Δ1/*his*3-Δ1 | Invitrogen (Carlsbad, USA) |
| *E. coli* WM3064 | Strain for conjugation; λpir, DAP auxotroph | Saltikov and Newman (2003) |
| *E. coli* DH5α | Host of cloning | ProBioGEM lab stock |
| Plasmids | | |
| pGEM-T easy | PCR cloning vector, Amp$^{r}$ | Promega (Madison, USA) |
| pTnMod-RKm | Km$^{r}$, *mob RP4*, plasposon delivery plasmid, *R6K ori* | Dennis and Zylstra (1998) |
| pAhNRPSKm | 5.6-kb suicide vector containing A fragment of the gene *AHA_2474* and kanamycin resistant marker excised from the plasmid/plasposon pTnMod-RKm | This study |
| pMQ30 | 7.6-kb mobilizable suicide vector used for gene replacement in *Pseudomonas*: *SacB*, *URA3*, Gm$^{r}$ | Shanks et al. (2006) |
| pMQ30Δ*amoH* | pMQ30 containing two 0.8-kb fragments of the NRPS biosynthesis gene *AHA_2473* and *AHA_2472* | This study |
| Primers (5′ → 3′) | | |
| AhA1 | TTGTTGACCTATGGCGAGCTGGAG | This study |
| AhA5 | GCACTCGGTCGGGCCGTACATG | This study |
| AhA9 | CGGCACCATGTATTCGGGCAG | This study |
| AhFo1 | GGGTGGACTGGCTGTTTGCTAC | This study |
| AhRev1 | CAACCGTGGCTCCCTCACTTTC | This study |
| AhFo3 | CGCCTGAGCGAGACGAAATAC | This study |
| AhRev3 | CAAAGCCCTGCTCGAAGAACC | This study |
| KmSalFo | GGGTTCGAAGTCGACGAG | This study |
| KmSalRev | AAAGGTCGACTTATTCAACAAAGC | This study |
| UPAHA2473-F | GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTGTCAGTGAGGGCGGATCAGAT | This study |
| UPAHA2473-R | AGTCCACTCCCCAGACCACCCGTAGCAGTTCCACAGC | This study |
| DownAHA2472-F | GCTGTGGAACTGCTACGGGTGGTCTGGGGAGTGGACT | This study |
| DownAHA2472-R | CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATCAATCCGTTTTCAGCCCAGT | This study |

equivalent/unit of biomass estimated by $OD_{600}$. The color of CAS assay solution changed from blue to yellow when iron-chelating agents are present in sample solutions. CAS agar assay was also performed in accordance with the original protocol (Schwyn and Neilands 1987). Positive results were indicated by the formation of a clear orange zone around the colonies showing a visual change from blue to yellow or orange.

### Structure determination by mass spectroscopy

For the detection of amonabactins in culture supernatant by MALDI-TOF, supernatants were thawed and mixed with a matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid in a 1:2 *v*/*v* solution of $CH_3CN/H_2O$ with 0.1 % trifluoroacetic acid (TFA)) in the ratio 1:9 *v*/*v*. Of this sample solution, 0.5 μL was deposited on the MALDI target and let dry at room temperature. Analysis was made on a Bruker Daltonics Maldi-tof Ultraflex equipped with a nitrogen laser operating at 337 nm, laser power 116.8 μJ, accelerating voltage 20 kV, positive mode, reflector mode, matrix suppression up to 500 *m*/*z*. The instrument was calibrated with a peptide mix ranging from 1046.54 to 3147.47 (Bruker part.206195). The spectrum is obtained with 5 × 30 shots on the sample. MS-MS of amonabactin P750 was obtained on a MS Bruker Esquire HTC quadrupole ion trap mass spectrometer equipped with an electrospray ionization source at atmospheric pressure. The samples were directly infused at 500 μL $h^{-1}$. Mass spectrometer was in positive mode, range 100–800 *m*/*z*. Final spectrum is the average of 50 spectra. Electrospray conditions are as follows: dry temperature 365 °C, dry gas 7 L $min^{-1}$, and nebulizer 20 psi.

### Construction of *A. hydrophila amoG*⁻ mutant

A fragment of the gene *amoG* was amplified by PCR using primers A1 and A9 (Table 1). The PCR fragment was purified from 1 % agarose gel using a QIAquick kit (Qiagen, Hilden, Germany) and cloned into the pGEM-T easy vector (Promega, Madison, USA). The kanamycin resistance marker was excised from the plasposon "pTnMod-RKm" by cleavage with *Ecl*136II followed by purification on agarose gel as previously described (Leclère et al. 2004). The corresponding 1244-bp fragment was subcloned into the blunt end *Stu*I site in the NRPS fragment. The orientation of the kanamycin gene was analyzed by restriction profiles, using *Eco*RI, *Hin*dIII, *Not*I, and *Xho*I enzymes. The final plasmid was named pAhNRPSKm. For origin of replication deletion, the plasmid pAhNRPSKm was digested with *Eco*RI enzyme. Large amounts of the 2572-bp fragment (containing disrupted NRPS) were purified, ligated with DNA ligation kit (Takara, Shiga, Japan) and transformed into *A. hydrophila* competent cells previously prepared as follows: bacteria were first streaked on LB agar plate and a single colony picked and grown in 10 mL of LB medium at 37 °C for overnight with shaking. Two milliliters of this culture was transferred to 100 mL of LB medium and incubated at 37 °C with vigorous shaking at 160 rpm till the cell density reached the absorbance of 0.5–1 at 600 nm. Then, the cells were collected by centrifugation and washed twice with 20 mL of 10 % ice-cold glycerol. Washed cells were resuspended in 100 μL of 10 % ice-cold glycerol and transferred into a sterile Eppendorf tube.

The bacteria were then plated on LB containing Amp and Km and incubated for at least 2 days. Clones resistant to both antibiotics, harboring a characteristic salmon color were selected and further analyzed. For this purpose, PCR was applied on the extracted DNA, using different sets of primers including some primers designed inside and outside of the marker and inside and outside of the NRPS fragment cloned within the plasmid (Fig. S1, Supplementary Material).

### Construction of the *A. hydrophila* Δ*amoH* mutant

The *A. hydrophila* Δ*amoH* mutant was constructed using in vivo homologous recombination in the yeast, *S. cerevisiae* InvSc1 (Shanks et al. 2006). The allelic replacement vector, pMQ30 (Shanks et al. 2006), was used for mutant construction; two regions of the NRPS genes *AHA_2473* (*amoH*) and *AHA_2472* were amplified using primers UPAHA2473_F and UPAHA2473_R (800 bp amplicon) and primers DownAHA2472_F and DownAHA2472_R (827 bp amplicon), respectively (Table 1). The PCR products of *AHA_2473* and *AHA_2472* were cloned flanking each other through in vivo homologous recombination in the yeast, *S. cerevisiae* InvSc1. Plasmid pMQ30-Δ*amoH* was extracted from *S. cerevisiae* InvSc1 and then introduced into *E. coli* WM3064 by electroporation. The plasmid was mobilized into *A. hydrophila* by conjugation, and deletion of *amoH* gene was confirmed by PCR (protocol S1, Supplementary Material).

### Bioinformatics analysis

The genome sequences of 30 *Aeromonas* species available in the NCBI were mined for the presence of NRPS gene clusters by following the strategy described within the Florine workflow (Caradec et al. 2014). Sequences presenting identity with known NRPSs were detected by using the NCBI-Blastp tool with the *A. hydrophila* ATCC 7966 AmoG (YP_856988.1) and AmoH (YP_856987.1) used as a query. The automatically annotated proteins were also fished using a list of keywords including adenylation, synthetase, thiotemplate, phosphopantetheine, nonribosomal, NRPS, siderophores, isochorismatase, dihydroxybenzoate, amonabactin, and enterobactin. Identification and annotation of NRPS gene clusters were detected by antiSMASH (http://antismash.secondarymetabolites.org) (Weber et al.

2015). The modular organization and domain architecture were deciphered using two complementary tools available on the Web including "PKS-NRPS analysis" (http://nrps.igs.umaryland.edu) (Bachmann and Ravel 2009) and "SBSPKS" (http://www.nii.ac.in/~pksdb/sbspks/search_main_pks_nrps.html) (Anand et al. 2010) combined with antiSMASH results. The prediction for a selected monomer by adenylation domain was performed using the same tools combined to the Web-based software NRPS predictor 2 (Röttig et al. 2011). The C-domain types were determined through the identification of specific signatures of the DownSeq (Caradec et al. 2014) combined with the result of natural product domain seeker (NaPDoS) (http://napdos.ucsd.edu) (Ziemert et al. 2012). The D-configuration of monomers was predicted regarding the presence of E domain followed by $^{D}C_{L}$ domains. The structure of predicted peptide was compared to other nonribosomal peptides in Norine database (http://bioinfo.lifl.fr/norine) (Pupin et al. 2015).

## Results

### Production of amonabactins by *A. hydrophila* subsp *hydrophila* ATCC7966

The strain was grown in CAA iron-limited medium supplemented or not with amino acids including tryptophan, phenylalanine, and glycine. Culture-free supernatants were then analyzed with the CAS reagent to determine the concentrations of total siderophores and submitted to mass spectrometry analyses (MALDI-TOF and MS-MS) for siderophore structure identification (Fig. 1 and Fig. S2, Supplementary Material). After 48 h of growth, the strain produced about 33 ± 2 μM eq. EDTA/$OD_{600}$. Final purification of the fractions allowed us to quantify the relative production of the four forms of amonabactins: T732 and T789 representing amonabactins with Trp as aromatic residue and P693 and P750 for amonabactins with Phe monomer. When the strain was grown in CAA medium, the ratio of (P693 + P750)/(T732 + T789) was 85:15 % (Table 2). The same ratio was observed when the medium was supplemented with 0.3 mM phenylalanine but was completely inverted when the strain was grown in the presence of 0.3 mM tryptophan.

### From the *A. hydrophila* genome to a siderophore synthesis cluster

The genome of *A. hydrophila* ATCC 7966, completely sequenced and automatically annotated in 2006 (Seshadri et al. 2006), is available at the NCBI site under accession number NC_008570. 1. Among the 4119 predicted coding sequences (CDS), our attention was attracted on some protein annotations, because they can reflect both siderophore biosynthesis and regulation, or are related to nonribosomal synthetases leading to the identification of a cluster panning from genes *AHA_2479* to *AHA_2473* (Fig. 2).

The cluster is entirely located on the same complement strand, flanked by genes (*AHA_2480* and *AHA_2472*) present on the coding strand. The genes constituting the gene cluster are annotated as "isochorismate synthases," "enterobactin synthase subunit E," "isochorismatase," "nonribosomal peptide synthetase," "2,3-dihydroxybenzoate-2,3-dehydrogenase," "nonribosomal peptide synthetase," and "dimodular nonribosomal peptide synthetase," for genes *AHA_2479* to *AHA_2473*, respectively. Considering the consensus GAGG, a ribosome binding site (RBS) can be found upstream of each ORF. A GTG initiation codon occurs in two over the seven genes (*AHA_2479* and *AHA_2474*) instead of the universal ATG.

The cluster likely represents an iron-regulated operon because a single promoter has been predicted, which includes an iron-box overlapping with the -10 box. Likewise, only one Rho-independent terminator has been found, downstream to the AHA_2473 open reading frame (ORF), making likely possible the transcription termination. Some genes are overlapping on few bp, as *AHA_2477* with *AHA_2476* and *AHA_2474* with *AHA_2473*. The annotations allow the separation in two groups of genes/proteins. The first group includes enzymes for biosynthesis of a Dhb monomer (isochorismate synthetase and Dhb dehydrogenase), and the second group contains proteins that should be involved in a nonribosomal synthesis of a peptidic moiety. All these elements are consistent with an operon involved in a Dhb-containing siderophore synthesis. Moreover, a Fur protein annotated as "ferric uptake regulation protein," encoded by the gene *AHA_1530*, was identified into the protein table related to the genome. It is a protein of 142 amino acids (aa), with a calculated molecular weight of 16 kDa, presenting a similarity score of 80 % with the Fur protein of *E. coli* str. K12 substr. DH10B.

### The amonabactin synthesis operon

According to the annotations, the operon includes at least three potential NRPS genes. The signatures for adenylation, condensation, thiolation, thioesterase, epimerization, methylation, and cyclization domains were researched in the protein sequences of the putative synthetases to determine the modular organization of the NRPS by following Florine workflow previously described (see experimental procedures section). The prediction of the monomer for each adenylation domain obtained from Stachelhaus code and Rausch predictions were combined (Fig. 3).

The protein annotated "isochorismatase" was shown to display high similarity with *E. coli* EntB (58 % identity). In spite of its small size (307 aa), it belongs to the NRPS family but only contains a T domain, as its ortholog EntB

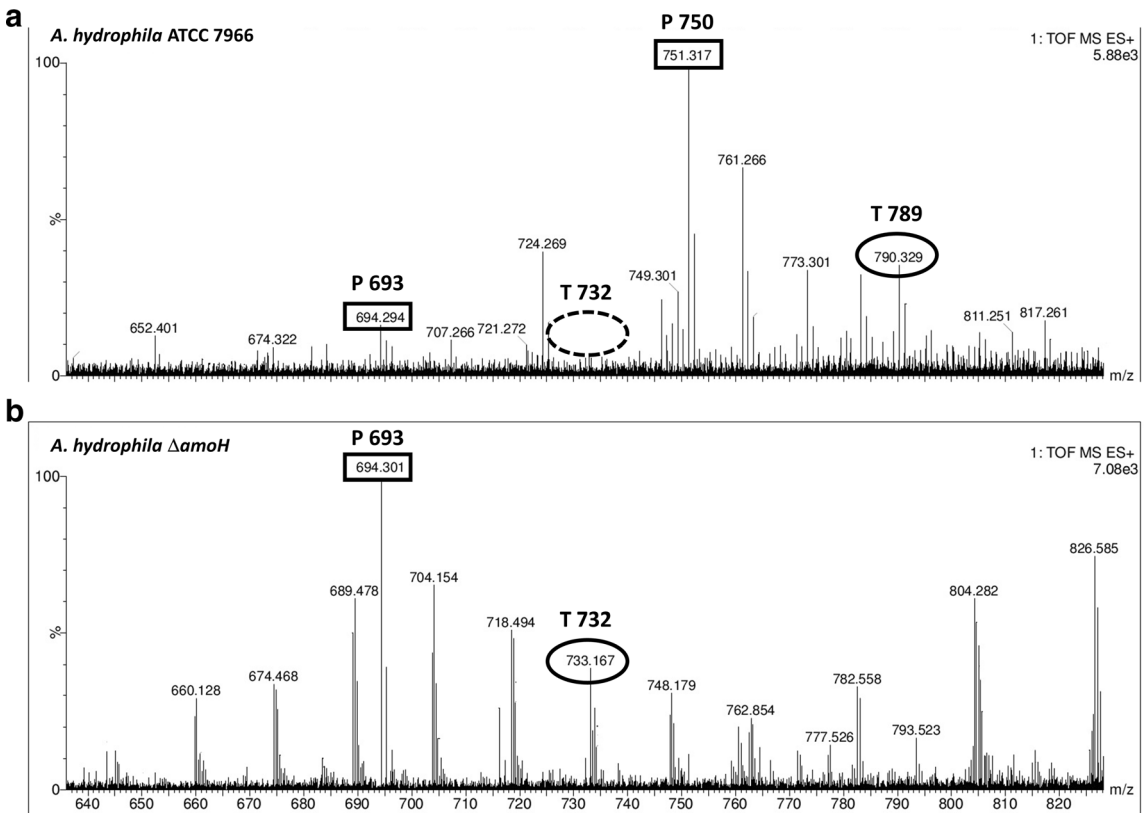



**Fig. 1** MS analysis of amonabactins. **a** *A. hydrophila* ATCC 7966; **b** *A. hydrophila* Δ*amoH* mutant. *Circles* highlight tryptophan-containing amonabactins; *squares* highlight phenylalanine-containing amonabactins. *Dotted line* when the signal is in the noise

in *E. coli*, allowing the tethering of a Dhb monomer on the synthetase involved in enterobactin synthesis. Two other proteins (encoded by genes *AHA_2478* and *AHA_2473* with protein accession numbers YP_856992 and YP_856987 only contain an adenylation domain with prediction of Dhb and Gly, respectively. The protein encoded by the gene *AHA_2476* (YP_856990) is constituted of complete NRPS module including $C_{starter}$-, A-, and T-domains. The A-domain should be able to select Lys or Glu. In the "nonribosomal peptide synthetase" (YP_856988), five domains can be retrieved representing a complete module displaying the following organization: $^{L}C_{L}$-A-T-E-C. The ending C-domain do not clearly harbor specific signatures identified for the different subtypes. The presence of an epimerization domain may lead to the occurrence of a monomer in D-form in the produced peptide but the selected amino acid could not have been predicted. Nested between the epimerization and condensation domain, a signature for the thioesterase can be found. Even if this signature harbor a quite low score when compared to the consensus sequence (0.6), it should be considered because of the lack of any other one in the whole cluster. As shown in Fig. 3, the nonribosomal pathway includes an assembly line composed of four modules, with a global structure [A-T][$C_{start}$-A-T][$^{L}C_{L}$-A-T-E-C][A]. The first module is constituted of the association of two proteins.

**Table 2** Amonabactin ratio

| | CAA | | CAA + Phe 0.3 mM | | CAA + Trp 0.3 mM | |
|---|---|---|---|---|---|---|
| | Intensity | % | Intensity | % | Intensity | % |
| T789 (Trp, Gly) | 813 | 12 | 711 | 14 | 5444 | 61 |
| T732 (Trp) | 165 | 2 | 118 | 2 | 983 | 11 |
| T forms | | 15 | | 16 | | 72 |
| P750 (Phe, Gly) | 4760 | 72 | 3955 | 76 | 2020 | 23 |
| P693 (Phe) | 866 | 13 | 398 | 8 | 501 | 6 |
| P forms | | 85 | | 84 | | 28 |

*CAA* casamino acid medium, *Phe* phenylalanine, *Trp* tryptophan, *Gly* glycin

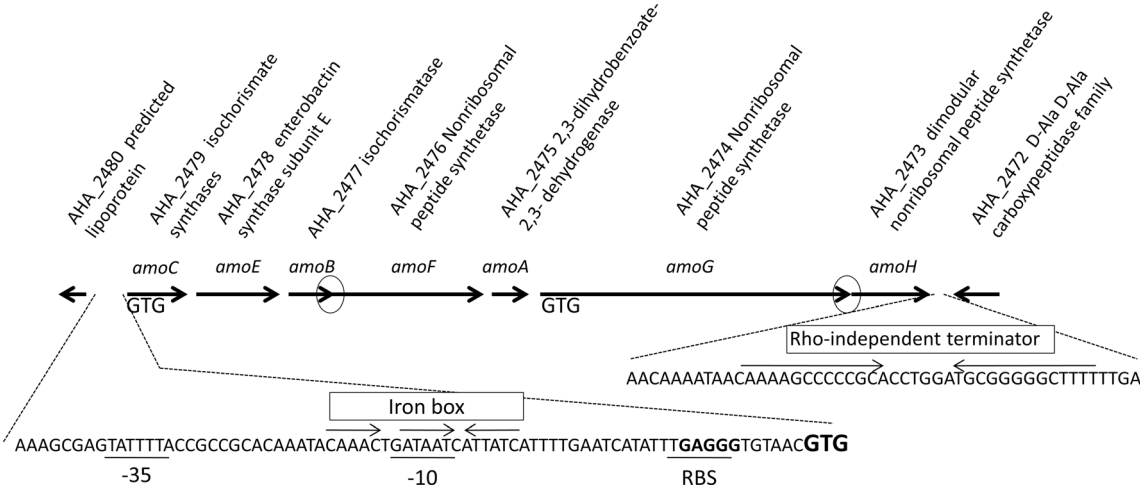



**Fig. 2** Genomic organization of the amonabactin synthesis operon in *A. hydrophila* ATCC7966. The *arrows* represent the genes with their orientation, and gene numbers and annotations on the upper line are those found in the NCBI database. Promoter sequences -35, -10, iron box, the RBS of the first gene, and the Rho-independent terminator at the end of the transcript unit are indicated. The use of CTG instead of ATG is mentioned. The *circles* highlight overlapping ORFs

The modules 2 and 3 are separated by a non-NRPS Dhb dehydrogenase. Considering the predicted selected monomers, the peptide produced should contain Dhb, Lys or Glu, Gly, and an unknown amino acid present in a D-form. This indicates that the operon may be involved in the biosynthesis of the amonabactins, the unpredicted monomer being either Phe, or Trp. The genes of the operon were named *amoC*, *amoE*, *amoB*, *amoF*, *amoA*, *amoG*, and

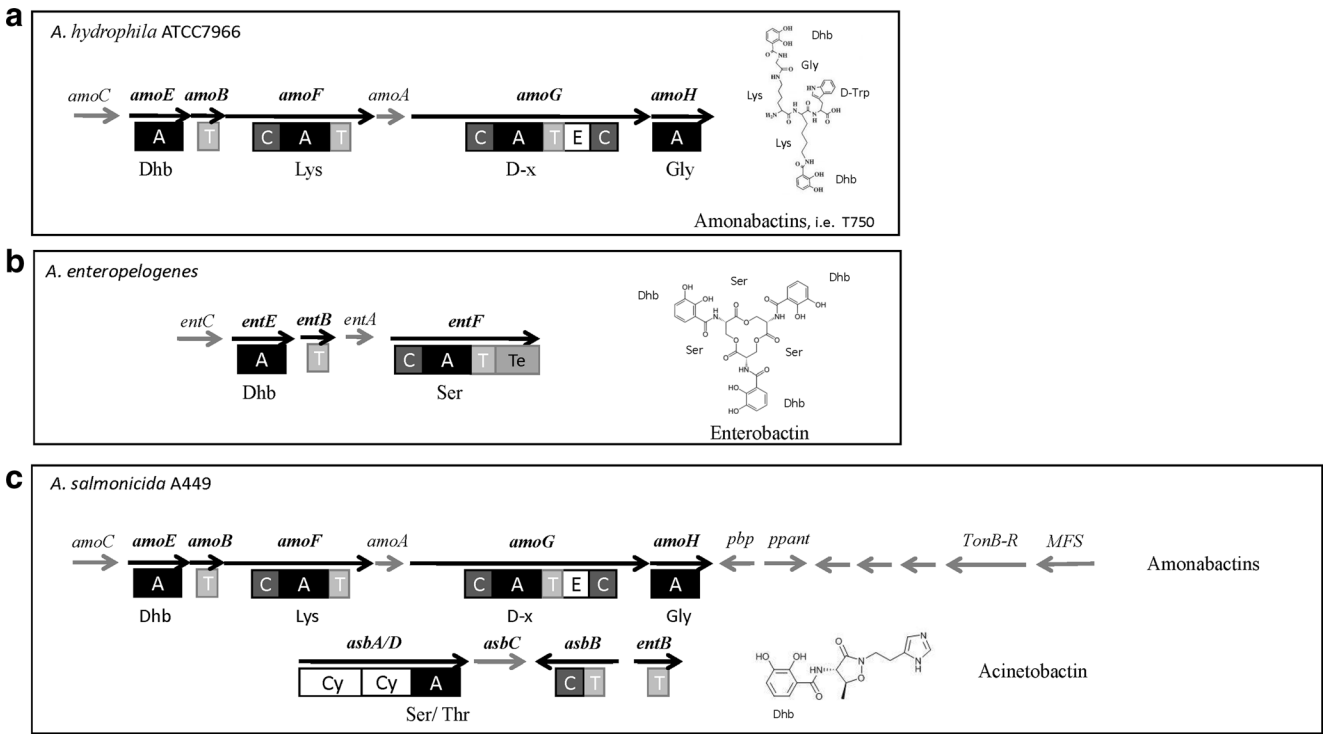



**Fig. 3** Siderophore synthesis pathways and siderophore structures. The NRPS domains (*A*, adenylation; *C*, condensation; *T*, thiolation; *Te*, thioesterase; *E*, epimerization; *Cy*, cyclization) are indicated by *boxes*. *amo* amonabactin synthetase, *ent* enterobactin synthetase, *asb* acinetobactin synthetase. In *bold* are the genes coding for NRPS domain containing proteins. Predicted monomers specificity by A-domains are indicated under the corresponding boxes. *pbp* periplasmic binding protein, *ppant* phosphopantetheinyl transferase, *TonB-R* TonB dependent receptor, *MFS* major facilitator superfamily transporter

*amoH*, according to the order of their location on the genome. The products of *amoE*, *amoB*, *amoF*, *amoG*, and *amoH* belong to NRPS family.

### *AmoG* is necessary for the production of the four amonabactins

To correlate the production of amonabactins to the synthesis operon, the gene *amoG* was disrupted by homologous recombination inserting a kanamycin resistance gene between signatures A1 and A9 of the adenylation domain. Kanamycin-resistant clones were further analyzed by PCR, using different sets of primers (Fig. S1, Supplementary Material) to check for correct insertion of the antibiotic marker. A disrupted-mutant M was then tested for its ability to grow under iron-limiting conditions. *A. hydrophila* wild type (A) and the mutant (M) were streaked on plates of King's B medium with or without the iron-chelating agent DIP. Contrary to the wild type, the mutant strain was unable to grow in the presence of 200 μM DIP (Fig. 4). On KB and KB with 600 μM $FeCl_3$ plates only, both wild type (A) and mutant (M) appear with a characteristic salmon color. A growth of the mutant M was also recovered when the DIP-containing medium was complemented by a supernatant obtained from *A. hydrophila* wild type cultured in iron-limiting broth (Fig. 4).

### *AmoH* is not required for siderophore production

In order to understand the involvement of the standing alone A-domain of AmoH, the gene was deleted by in vivo homologous recombination in the yeast, *S. cerevisiae* InvSc1. Both the *A. hydrophila* wild type and *amoH* mutant were tested for their siderophore production induced by iron chelation.

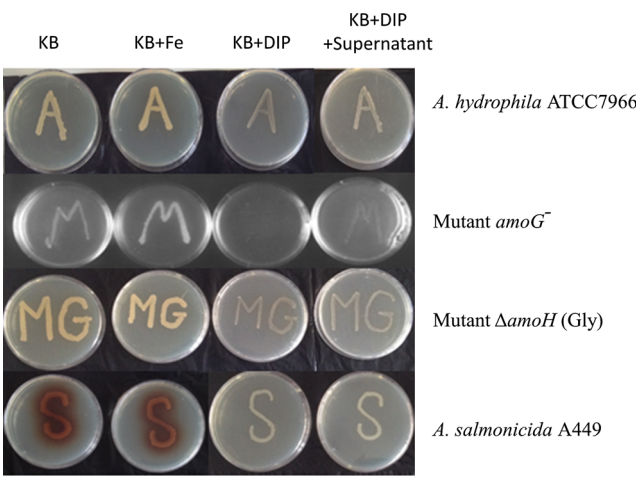



**Fig. 4** Growth of *A. hydrophila* wt and mutants and *A. salmonicida* plated on different solid media. *KB* King's B medium, *KB+Fe* KB added with 60 μM $FeCl_3$, *KB+DIP* KB added with 200 μM 2,2′-dipyridyl (DIP), *KB+DIP+supernatant* KB with 200 μM DIP and growth supernatant from *A. hydrophila* ATCC7966 grown in iron-limiting broth

Contrary to the *amoG* mutant, the *amoH* mutant remains able to grow on a medium containing the iron-chelating agent DIP. The phenotype is very similar to the wild-type strain one (Fig. 4). The supernatant of the *amo*H mutant grown in iron-limited broth was then analyzed by mass spectrometry. The results showed that the deletion of *amoH* gene led to the production of only two forms of amonabactins corresponding to the forms without Gly (P693 and T732) (Fig. 1).

### Siderophore production by *A. salmonicida* A449

The tabular format of the annotations of the genome of *A. salmonicida* A449 was queried using the keyword "peptide synthetase." Four proteins were pointed out, two of them are annotated "nonibosomal peptide synthetase" and the remaining are simply annotated "peptide synthetase." The genes belong to two distinct clusters. The first cluster is ortholog to the amonabactin biosynthesis one identified in *A. hydrophila* 7966, including *amoC* to *amoH* (Fig. 3 and Table S1, Supplementary Material). The modular organization of the NRPSs is the same, as the prediction of the amino acid residue activated by the different adenylation domains, and the C-domain subtypes. In *A. salmonicida*, the operon is boarded by a series of genes predicted to be involved in the siderophore transport, the proteins being annotated ABC transporter, permease, TonB-dependent receptor, and MFS transporter. A gene encoding a 4′-phosphopantetheinyl protein (ppant) is also present. As expected, *A. salmonicida* was able to grow in King's B medium supplemented with iron-chelating agent DIP, due to amonabactin production (Fig. 4). On KB medium, supplemented or not with iron 660 μM, the bacteria are brown pigmented.

The second cluster includes several genes for siderophore transport and genes encoding for NRPS. The domain organization indicated that the larger protein contains only one A-domain with a specificity for Ser/Thr and two heterocyclization domains (Fig. 3). Another protein harbors a C- and a T-domains, and the third protein has a single T-domain.

### Distribution of siderophore biosynthetic gene cluster in *Aeromonas* strains

The genome sequences of 30 *Aeromonas* species are currently available on NCBI database (for some species as *A. hydrophila*, several strain genomes are published). Twelve of them are finished, but the remaining can be explored even if the data are not definitive. The genomes are represented by one chromosome with a size of about 5 Mb. No plasmid is present except for *A. salmonicida* subsp. *salmonicida* A449 containing five plasmids. The genomes were screened for the presence of NRPS genes, especially those responsible for the biosynthesis of amonabactins,

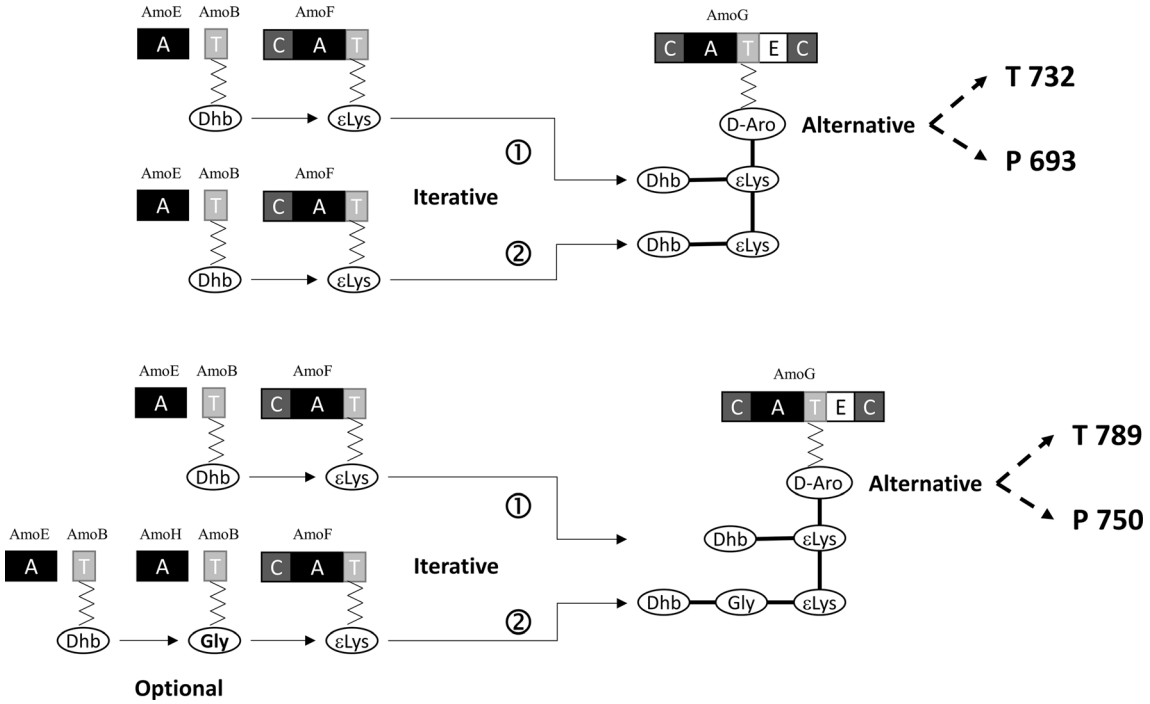



**Fig. 5** Schematic representation of the iterative, alternative, and optional mode of synthesis of amonabactins. εLys mentions that the Dhb is linked onto the εNH2 function of the Lys residue. Aro designs aromatic residue either Trp or Phe. Iterative, alternative, and optional are placed where they occur

following the Florine workflow previously described. For all the putative NRPSs detected, the domain organization was designed; the C-domain subtypes including $C_{start}$, $^{L}C_{L}$, $^{D}C_{L}$, and Cy were determined; and the prediction of the most probable selected monomer was predicted for each A-domain. A gene cluster responsible for synthesis of amonabactins was identified in the genome of 17 *Aeromonas* sequenced genomes (Table S1, Supplementary Material). For *Aeromonas popoffii* and *A. hydrophila*, it only contains the biosynthetic part, while for the remaining strains, it also includes genes involved in transport and the gene encoding the ppant, as found in the *A. salmonicida* A449 genome. For *Aeromonas australiensis*, *Aeromonas Enteropelogenes*, and *Aeromonas lacus*, no biosynthetic gene cluster for amonabactin was found but a cluster for enterobactin synthesis, including *entD*, *entC*, *entE*, *entB*, *entA*, and *entF* genes (Fig. 3). For the eight remaining strains, no genes encoding NRPS involved in siderophore synthesis or any other product have been identified (Table S1, Supplementary Material).

## Discussion

Today, the genus *Aeromonas* is regarded, not only as important pathogen for fish and frogs but also as etiologic agent responsible for a variety of infectious complications in both immunocompetent and immunocompromised persons (Janda and Abbott 2010). The increasing availability of whole *Aeromonas* genome sequences can be explored with bioinformatic tools, and the analysis in silico enables the deciphering of biosynthetic pathways of virulence factors especially siderophores. Indeed, following the Florine workflow (Caradec et al. 2014) allowed us to identify the assembly line involved in the production of amonabactins. Amonabactins are a family of four related siderophores, coproduced by different *Aeromonas* species. The advantage of producing four forms of one siderophore remains unclear, and this issue has to be further explored to understand the ecological role for these compounds. The biochemical structures have been elucidated for a long time (Barghouthi et al. 1989b), but the biosynthesis pathway is not. Before the complete sequencing of the genome of *A. hydrophila* ATCC 7966 (Seshadri et al. 2006), only one gene implicated in amonabactin biosynthesis has been identified and separately sequenced (Barghouthi et al. 1991). The exploration of the genomic data allowed us to point out a single operon enabling the amonabactin synthesis. This operon is composed of genes we have named *amoCEBFAGH*, where the *amoC*, *E*, *B*, and *amoA* were named related to their orthologs *entC*, *E*, *B*, and *entA* involved in enterobactin synthesis. Consequently, *amoC*, involved in the synthesis of Dhb, a key monomer of a series of peptidic siderophores, is the one previously described and named *amoA* by Barghouthi et al. (1991). The same cluster also includes genes involved in the transport of the siderophore, except for *A. popoffii* and *A. hydrophila*. For the latter, the TonB-dependent receptor has already been identified elsewhere on the genome (Funahashi et al. 2013). Two TonB-dependent systems are annotated corresponding to TonB, ExbD, and

ExbB proteins, respectively. For the first system, the proteins' accession numbers are YP_857923, 857925, and 857926, and for the second system, they are YP_858665, 858666, and 858667. The presence of a Dhb monomer in a peptide nonribosomally synthesized gives an indication for a highly probable siderophore function (Caboche et al. 2010), as the presence of a Fur-box in the promoter driving the operon expression. The mode of biosynthesis for Dhb-containing siderophore generally involves iterative mechanism because several hydroxyl functions harbored by the Dhb monomer are needed to chelate iron. This is for example the case for bacillibactin produced by *Bacillus subtilis* (May et al. 2001), cepaciachelin produced by *Burkholderia ambifaria* AMMD (Esmaeel et al. 2016), and enterobactin produced by *E. coli* (Gehring et al. 1998). The construction of both *amoG* and *amoH* mutants allowed us to identify an atypical mode of biosynthesis for the production of amonabactins. It was qualified of (i) alternative as Trp or Phe can be recruited by the same A-domain of *amoG*, (ii) iterative because *amoE* and *amoF* react twice to link onto Trp/Phe two fragments composed of a Dhb linked by a peptidic bond on the $\varepsilon NH_2$ of a Lys residue, and (iii) optional because the Gly residue is introduced in the Lys side chain in only two over the four amonabactins (Fig. 5). The nonribosomal synthesis of Dhb-containing siderophores is rarely linear, but the amonabactin synthesis is the more complex use of a single assembly line leading to the coproduction of four variants of a siderophore currently described.

For *A. salmonicida* subsp. *salmonicida* A449 and *A. hydrophila* YL17, a second NRPS gene cluster involved in the production of another siderophores, was also found. This cluster was recently described for the synthesis of the Dhb-containing siderophore acinetobactin by *A. salmonicida* (Balado et al. 2015). The *asb* cluster has been likely acquired through horizontal transfer and *amoC*, *E*, *B*, and *amoA* are shared for biosynthesis of both acinetobactin and amonabactin (Balado et al. 2015). Indeed, we only found it in genomes where the complete *amo* cluster is also present.

Most isolates of *Aeromonas* species are known to produce a siderophore, either amonabactin or the enteric siderophore enterobactin, both using the catecholate group to chelate iron as they contain the Dhb monomer as a functional group (Massad et al. 1994). Among the 30 species explored, the amonabactin synthesis gene cluster is widespread and only three genomes harbor the enterobactin gene cluster. A role in virulence has been proposed for amonabactin and not for enterobactin (Massad et al. 1991).

**Acknowledgments** This work was supported by the University of Lille 1, the INTERREG IV program France-Wallonie-Vlaanderen (Phytobio project), the bioinformatics platform bilille and Inria. QE received financial support from Sana'a University (Yemen).

**Compliance with ethical standards**

**Funding** The work was funded by the University of Lille 1, the INTERREG IV program France-Wallonie-Vlaanderen (Phytobio project), the bioinformatics platform bilille, and Inria. QE received financial support from Sana'a University (Yemen).

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

Anand S, Prasad MVR, Yadav G, Kumar N, Shehara J, Ansari MZ, Mohanty D (2010) SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res 38:W487–W496. doi:10.1093/nar/gkq340

Austin B, Austin DA (2007) Bacterial fish pathogens: diseases of farmed and wild fish, 4th edn. Praxis Publishing Ltd, Chichester

Bachmann BO, Ravel J (2009) Chapter 8. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Meth Enzymol 458:181–217

Balado M, Souto A, Vences A, Careaga VP, Valderrama K, Segade Y, Rodríguez J, Osorio CR, Jiménez C, Lemos ML (2015) Two catechol siderophores, acinetobactin and amonabactin, are simultaneously produced by *Aeromonas salmonicida* subsp. *salmonicida* sharing part of the biosynthetic pathway. ACS Chem Biol 10(12): 2850–2860. doi:10.1021/acschembio.5b00624

Barghouthi S, Young R, Arceneaux JEL, Byers BR (1989a) Physiological control of amonabactin biosynthesis in *Aeromonas hydrophila*. Biol Met 2(3):155–160. doi:10.1007/bf01142554

Barghouthi S, Young R, Olson MO, Arceneaux JE, Clem LW, Byers BR (1989b) Amonabactin, a novel tryptophan- or phenylalanine-containing phenolate siderophore in *Aeromonas hydrophila*. J Bacteriol 171(4):1811–1816

Barghouthi S, Payne SM, Arceneaux JE, Byers BR (1991) Cloning, mutagenesis, and nucleotide sequence of a siderophore biosynthetic gene (*amoA*) from *Aeromonas hydrophila*. J Bacteriol 173(16): 5121–5128

Beaz-Hidalgo R, Figueras MJ (2013) *Aeromonas* spp. whole genomes and virulence factors implicated in fish disease. J Fish Dis 36(4): 371–388. doi:10.1111/jfd.12025

Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. J Bacteriol 192(19):5143–5150. doi:10.1128/jb.00315-10

Caradec T, Pupin M, Vanvlassenbroeck A, Devignes M-D, Smaïl-Tabbone M, Jacques P, Leclère V (2014) Prediction of monomer isomery in Florine: a workflow dedicated to nonribosomal peptide discovery. PLoS One 9(1):e85667. doi:10.1371/journal.pone.0085667

Challis GL (2005) A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. Chembiochem 6(4):601–611. doi:10.1002/cbic.200400283

Coscelli GA, Bermúdez R, Losada AP, Faílde LD, Santos Y, Quiroga MI (2014) Acute *Aeromonas salmonicida* infection in turbot (*Scophthalmus maximus* L.). Histopathological and immunohistochemical studies. aquac 430(0):79–85. doi:10.1016/j.aquaculture.2014.04.002

Daskalov H (2006) The importance of *Aeromonas hydrophila* in food safety. Food Control 17(6):474–483. doi:10.1016/j.foodcont.2005.02.009

Dennis JJ, Zylstra GJ (1998) Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes. Appl Environ Microbiol 64(7):2710–2715

Escolar L, Pérez-Martín J, de Lorenzo V (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. J Bacteriol 181(20):6223–6229

Esmaeel Q, Pupin M, Kieu NP, Chataigné G, Béchet M, Deravel J, Krier F, Höfte M, Jacques P, Leclère V (2016) *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. MicrobiolOpen 5(3): 512–526. doi:10.1002/mbo3.347

Farto R, Milton D, Bermúdez M, Nieto T (2011) Colonization of turbot tissues by virulent and avirulent *Aeromonas salmonicida* subsp. *salmonicida* strains during infection. Dis Aquat Org 95(2):167–173. doi:10.3354/dao02342

Flissi A, Dufresne Y, Michalik J, Tonon L, Janot S, Noé L, Jacques P, Leclère V, Pupin M (2015) Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. Nucleic Acids Res 44(D1):D1113–D1118. doi:10.1093/nar/gkv1143

Funahashi T, Tanabe T, Miyamoto K, Tsujibo H, Maki J, Yamamoto S (2013) Characterization of a gene encoding the outer membrane receptor for ferric enterobactin in *Aeromonas hydrophila* ATCC 7966 T. Biosci Biotechnol Biochem 77(2):353–360. doi:10.1271/bbb.120774

Gehring AM, Mori I, Walsh CT (1998) Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. Biochemistry 37(8):2648–2659. doi:10.1021/bi9726584

Janda JM, Abbott SL (2010) The genus *Aeromonas*: taxonomy, pathogenicity, and infection. Clin Microbiol Rev 23(1):35–73. doi:10.1128/cmr.00039-09

Leclère V, Béchet M, Blondeau R (2004) Functional significance of a periplasmic Mn-superoxide dismutase from *Aeromonas hydrophila*. J Appl Microbiol 96(4):828–833. doi:10.1111/j.1365-2672.2004.02231.x

Leclère V, Beaufort S, Dessoy S, Dehottay P, Jacques P (2009) Development of a biological test to evaluate the bioavailability of iron in culture media. J Appl Microbiol 107(5):1598–1605. doi:10.1111/j.1365-2672.2009.04345.x

Massad G, Arceneaux JEL, Byers BR (1991) Acquisition of iron from host sources by mesophilic *Aeromonas* species. J Gen Microbiol 137(2):237–241. doi:10.1099/00221287-137-2-237

Massad G, Arceneaux JEL, Byers BR (1994) Diversity of siderophore genes encoding biosynthesis of 2,3-dihydroxybenzoic acid in *Aeromonas* spp. Biometals 7(3):227–236. doi:10.1007/bf00149553

May JJ, Wendrich TM, Marahiel MA (2001) The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. J Biol Chem 276(10):7209–7217. doi:10.1074/jbc.M009140200

Merino S, Rubires X, Knøchel S, Tomás JM (1995) Emerging pathogens: *Aeromonas* spp. Int J Food Microbiol 28(2):157–168. doi:10.1016/0168-1605(95)00054-2

Najimi M, Lemos M, Osorio C (2008) Identification of siderophore biosynthesis genes essential for growth of *Aeromonas salmonicida* under iron limitation conditions. Appl Environ Microbiol 74:2341–2348. doi:10.1128/AEM.02728-07

Payne SM (1994) Detection, isolation, and characterization of siderophores. Meth Enzymol 235:329–344

Pupin M, Esmaeel Q, Flissi A, Dufresne Y, Jacques P, Leclère V (2015) Norine: a powerful resource for novel nonribosomal peptide discovery. Synth Syst Biotechnol 1(2):89–94. doi:10.1016/j.synbio.2015.11.001

Rausch C, Hoof I, Weber T, Wohlleben W, Huson D (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evolutionary Biol 7(1):78. doi:10.1186/1471-2148-7-78

Reith M, Singh R, Curtis B, Boyd J, Bouevitch A, Kimball J, Munholland J, Murphy C, Sarty D, Williams J, Nash J, Johnson S, Brown L (2008) The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. BMC Genom 9(1):427. doi:10.1186/1471-2164-9-427

Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O (2011) NRPSpredictor 2- a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. doi:10.1093/nar/gkr323

Saltikov CW, Newman DK (2003) Genetic identification of a respiratory arsenate reductase. Proc Natl Acad Sci 100(19):10983–10988. doi:10.1073/pnas.1834303100

Schwyn B, Neilands JB (1987) Universal chemical assay for the detection and determination of siderophores. Anal Biochem 160(1):47–56. doi:10.1016/0003-2697(87)90612-9

Seshadri R, Joseph SW, Chopra AK, Sha J, Shaw J, Graf J, Haft D, Wu M, Ren Q, Rosovitz MJ, Madupu R, Tallon L, Kim M, Jin S, Vuong H, Stine OC, Ali A, Horneman AJ, Heidelberg JF (2006) Genome sequence of *Aeromonas hydrophila* ATCC 7966 T. J Bacteriol 188(23):8272–8282. doi:10.1128/jb.00621-06

Shanks RMQ, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA (2006) *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from Gram-negative bacteria. Appl Environ Microbiol 72(7):5027–5036. doi:10.1128/aem.00682-06

Strieker M, Tanović A, Marahiel MA (2010) Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol 20(2):234–240. doi:10.1016/j.sbi.2010.01.009

Telford JR, Raymond KN (1997) Amonabactin: a family of novel siderophores from a pathogenic bacterium. JBIC 2(6):750–761. doi:10.1007/s007750050191

Telford JR, Leary JA, Tunstad LMG, Byers BR, Raymond KN (1994) Amonabactin: characterization of a series of siderophores from *Aeromonas hydrophila*. J Am Chem Soc 116(10):4499–4500. doi:10.1021/ja00089a058

Toranzo AE, Magariños B, Romalde JL (2005) A review of the main bacterial fish diseases in mariculture systems. Aquac 246(1–4):37–61. doi:10.1016/j.aquaculture.2005.01.002

Van Lanen SG, Shen B (2006) Microbial genomics for the improvement of natural product discovery. Curr Opin Microbiol 9(3):252–260. doi:10.1016/j.mib.2006.04.002

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller 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. doi:10.1093/nar/gkv437

Yu HB, Zhang YL, Lau YL, Yao F, Vilches S, Merino S, Tomas JM, Howard SP, Leung KY (2005) Identification and characterization of putative virulence genes and gene clusters in *Aeromonas hydrophila* PPD134/91. Appl Environ Microbiol 71(8):4469–4477. doi:10.1128/aem.71.8.4469-4477.2005

Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS One 7(3):e34064. doi:10.1371/journal.pone.0034064

**Applied Microbiology and Biotechnology**

# Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism catalyzes amonabactin synthesis in *Aeromonas*

Qassim Esmaeel[1], Mickael Chevalier[1], Gabrielle Chataigné[1], Rathinasamy Subashkumar[1], Philippe Jacques[1, 2], Valérie Leclère[1, 3]

[1]: Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, EA 7394-ICV- Institut Charles Viollette, F-59000 Lille, France.

[2]: TERRA Research Centre, Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liege, B-5030 Gembloux, Belgium

[3]: Inria-Lille Nord Europe, Bonsai team, F-59655 Villeneuve d'Ascq Cedex, France.

Corresponding author:
Qassim Esmaeel
Phone +33320 43 46 68, Fax +33 328 76 73 56, email: al_hamadi_82@yahoo.com

**Supplementary Material**

**Table S1:** Presence of homologous genes involved in sideophore biosynthesis among 30 *Aeromonas* species

| **Amonabctin gene cluster** | *AmoC* | *AmoE* | *AmoB* | *AmoF* | *AmoA* | *AmoG* | *AmoH* | *ABC Tpter* | *ppant* | *ABC tpter* | *permease* | *permease* | *TonB R* | *MFS* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *A. hydrophila* ATCC7966 | + | + | + | + | + | + | + | | | | | | | |
| *A. popoffii* | + | + | + | + | + | + | + | + | + | | | | | |
| *A. hydrophila* YL17 | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. salmonicida* A449 | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. aquatica* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. bestiarum* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. bivalvium* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. caviae* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. dhakensis* AAK1 | + | + | + | + | + | + | + | + | + | + | + | + | + | |
| *A. encheleia* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. eucrenophila* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. media* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. piscicola* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. rivuli* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. sanarellii* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. schubertii* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. taiwanensis* | + | + | + | + | + | + | + | + | + | + | + | + | + | |
| *A. tecta* | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| *A. molluscorum* | + | + | | | | + | | | | | | | | |

| **Enterobactin gene cluster** | Ent D | EntC | EntE | EntB | EntA | EntF |
|---|---|---|---|---|---|---|
| *A. australiensis* | + | + | + | + | + | + |
| *A. enteropelogenes* | + | + | + | + | + | + |
| *A. lacus* | + | + | + | + | + | + |

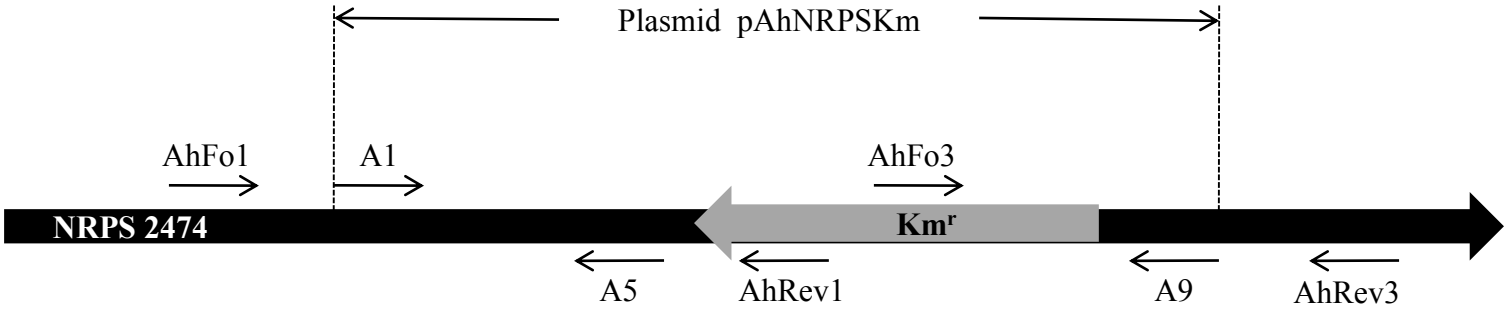
| **Acinetobactin gene cluster** | *asbB* | *asbD* |
|---|---|---|
| *A. hydrophila* YL17 | + | + |
| *A. salmonicida* subsp. *salmonicida* A449 | + | + |
| *A. dhakensis* | + | + |
| *A. salmonicida* subsp. *masoucida* NBRC 13784 | | + |
| *A. salmonicida subsp. pectinolytica* | + | + |
| *A. hydrophila* SSU | + | + |

**No NRPS gene cluster**

*A. allosaccharophila*
*A. diversa*
*A. finlandiensis*
*A. fluvialis*
*A. jandaei*
*A. simiae*
*A. sobria*
*A. veronii*

"+" indicates the presence of homologous

**A**




**B**

| | Ah 7966 (A) | Mutant (M) | Plasmid (P) |
|---|---|---|---|
| **A1-A5** | 786 | 786 | 786 |
| A1: TTGTTGACCTATGGCGAGCTGGAG | | | |
| A5: GCACTCGGTCGGGCCGTACATG | | | |
| **A1-A9** | 1308 | 2552 | 2552 |
| A1: TTGTTGACCTATGGCGAGCTGGAG | | | |
| A9: CGGCACCATGTATTCGGGCAG | | | |
| **AhFo1-AhRev1** | NA | 1360 | NA |
| AhFo1: GGGTGGACTGGCTGTTTGCTAC | | | |
| AhRev1: CAACCGTGGCTCCCTCACTTTC | | | |
| **AhFo3-AhRev3** | NA | 1317 | NA |
| AhFo3: CGCCTGAGCGAGACGAAATAC | | | |
| AhRev3: CAAAGCCCTGCTCGAAGAACC | | | |

**C**

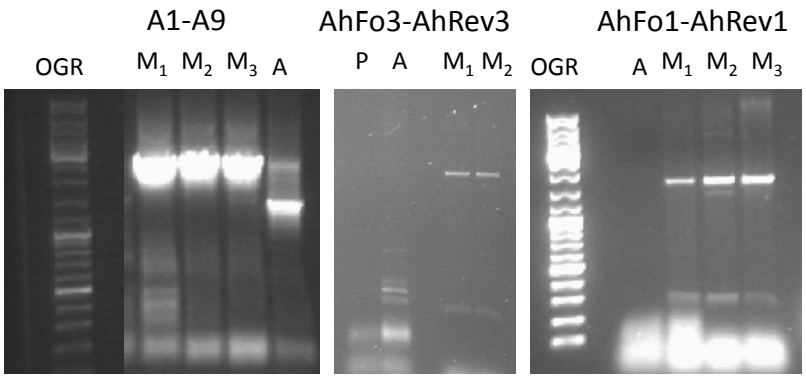
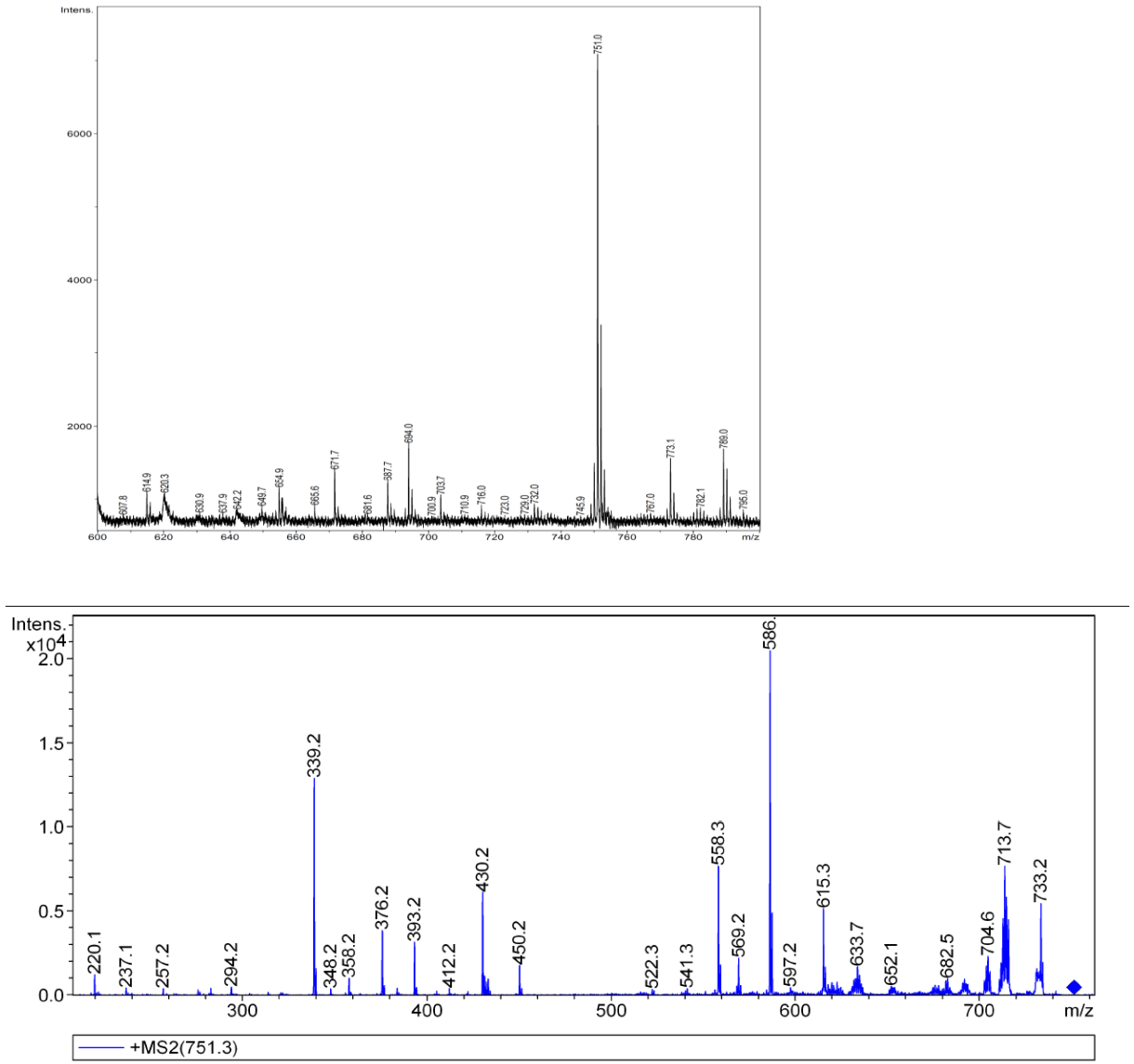


**Fig. S1.** Analysis of disruption of *amoG* gene

**A**: Schematic representation of the NRPS fragment disrupted by Km resistance cassette ($Km^r$). The primers are represented by arrows, where they hybridize. The DNA fragment included into the plasmid used for homologous recombination is presented between dot lines.

**B**: Table representing the primer sequences, together with sizes of expected fragments after PCR applied on DNA of *A. hydrophila* ATCC7966 (Ah 7966), the mutant obtained by homologous recombination (M) and the plasmid pAhNRPSKm. NA: no amplification.

**C**: PCR results using different sets of primers. OGR: O gene'ruler (molecular weight marker); A: *A. hydrophila* ATCC7966; M1, M2, and M3: three different clones of disrupted mutant.



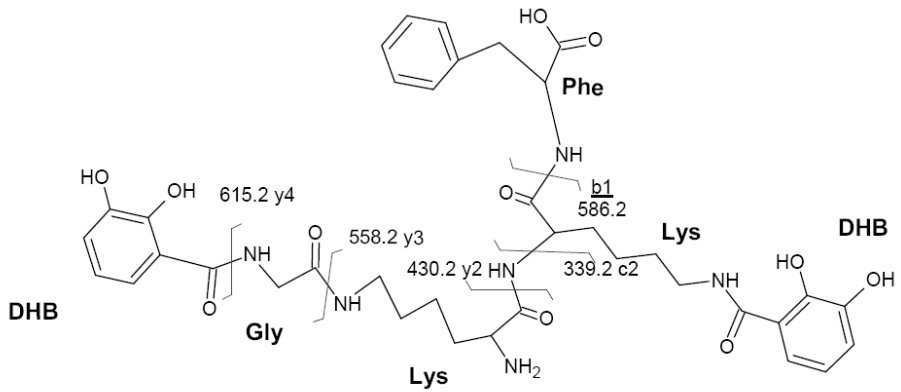




**Fig.S2:** MS-MS analysis of amonabactin T750

## Protocol S1- Construction of *A. hydrophila* -Δ*amoH* mutant

### Construction of pMQ30 *ΔamoH*

The allelic replacement vector, pMQ30 (Shanks et al., 2006) was used for mutant construction. Two regions of the NRPS genes *AHA_2473* (*amoH*) and *AHA_2472* were amplified using primers UPAHA2473_F and UPAHA2473_R (800 bp amplicon) and primers DownAHA2472_F and DownAHA2472_R (827 bp amplicon), respectively. These primers contained recombination sites for the plasmid. The PCR mix consisted of 25 µL of PCR Master Mix (Thermo Scientific Fermentas, Waltham, USA), 10 µL Q-solution (Qiagen, Hilden, Germany), 1.25 µL of each primer (each at 20 µM), 7.5 µL water and 5 µL of genomic DNA. The reaction mixture was subjected to the following thermal cycles: one cycle at 94°C for 3min; 30 cycles (94°C, 30 s; 60°C, 45 s; 72°C, 2 min) and a final extension at 72°C for 10 min. Plasmid pMQ30 was purified using GeneJet plasmid miniprep kit (Thermo Scientific Fermentas, Waltham, USA). A 5 µL sample was checked on a 1% agarose gel. Subsequently, the purified plasmid was digested by a mix of 9.5 µL milliQ water, 4 µL Tango yellow buffer (2X), 5 µL plasmid pMQ30, 1 µL *Bam*HI and 0.5 µL *Eco*RI. The mix was incubated for 2h at 37°C. 5 µL of the plasmid digest was checked on 1% agarose and the concentration of the digest was also measured.

### In vivo cloning in *S. cerevisiae* InvSc1

The PCR products of *AHA_2473* and *AHA_2472* were cloned flanking each other through in vivo homologous recombination in the yeast, *S. cerevisiae* InvSc1.,grown overnight in yeast peptone dextrose (YPD) at 30°C. Then, 0.5 mL culture was centrifuged for 1 min at 3,000 rpm. The cells pellet was washed with 0.5 mL lazy bones solution and the following was added to the mix: 20 µL of carrier DNA (2 mg/mL), 45 µL of each PCR products UP and DOWN, 5 µL digested pMQ30 plasmid. The mix was homogenized for 1 min on a Vortex and incubated overnight at room temperature. The mixture was subjected to heat shock for 12 min at 42°C. Then, cells were centrifuged for 1 min at 3,000 rpm, washed with 0.6 mL of TE buffer and the resulting pellet was further redissolved in 0.6 mL of TE buffer, and the cells were plated on SD-uracil medium. Further, plasmid was isolated using miniprep kit (Fermentas, Waltham, USA). 10 µL of the eluted plasmid was checked on 1% agarose gel.

**Construction of *A. hydrophila -ΔamoH* deficient mutant**

Plasmid pMQ30Δ*amoH* was introduced into *E. coli* WM3064 by electroporation. 2 µL of pMQ30Δ*amoH* was added to 50 µL of thawed *E. coli* WM3064 competent cells. Mix was transferred to an electroporation cuvette and electroporated at 2.5 kV/200 ohms/25 µF. Electroporated cells were transferred into 1 mL of LB containing 2% of glucose and incubated in a rotary shaker for 1 h at 37°C. The cells were plated on LB containing 25 µg $mL^{-1}$ of gentamicin sulphate and 100 mg $L^{-1}$ of DAP (diaminopimelic acid). Colony PCR was conducted using primers UpAHA2473_F and DownAHA2472_R to confirm the success of electroporation.

The plasmid was mobilized into *A. hydrophila* by conjugation. The donor strain *E. coli* WM3064+ pMQ30Δ*amoH* was grown in LB broth containing Gm at 25 µg $mL^{-1}$ and DAP at 100 mg $L^{-1}$, respectively. The acceptor strain *A. hydrophila* was grown without antibiotics. The cells were pelleted for 2 min at 5,000 rpm, the supernatant was discarded and the cells were washed with 1.5 mL of LB. To do conjugation, 200 µL of acceptor strain was mixed with 200 µL of donor strain. This mix was spotted on LB plates containing 100 mg $L^{-1}$ DAP, left to dry, after which plates were incubated overnight at 37 °C. After incubation, the cell mass were collected and dissolved in 1 mL of LB. For each conjugate, dilutions up to $10^{-3}$ were made, plated on LB plates containing 100 µg $mL^{-1}$ of Gm and incubated overnight at 37°C. Colonies chosen were streaked on new LB plates with Gm at 100µg $mL^{-1}$. Colony PCR was done to confirm the presence of the plasmid in the merodiploid strain (cell contains plasmid) using primers UpAHA2473_F and DownAHA2472_R. To obtain the 2nd recombination, one colony was cultured on LB and incubated overnight at 37 °C in rotary shaker. Culture were diluted up to $10^{-5}$ and 100 µL aliquot was plated on LB plates containing 10% sucrose. Selected colonies were streak out on new LB plates without gentamicin. Colony PCR of these colonies was carried out using primers UpAHA2473_F and DownAHA2472_R to confirm and detect the deleted mutants.

# CHAPTER 3- RESULTS

## Screening for new non ribosomal peptides

## 3.1. Introduction

Microorganisms are a prolific source of natural products extremely valuable in drug discovery and agriculture researches. They represent a promising basis for the development of substances with novel activities. Non ribosomal peptide is one class of natural products that attract a great interest and applications in medicine and agriculture. The ability to produce nonribosomal bioactive peptides is widely spread among microorganisms including bacteria and fungi. They display a biodiversity with a broad range of biological activities and complex chemical structures Microbial genome mining has recently guided the discovery of new secondary metabolites such as nonribosomal peptides (NRPs) and polyketides (PKs). This was efficient for the discovery of cyclic lipopeptides as those produced by *Bacillus* and *Pseudomonas,* which may play an important role in sustainable agriculture, used as biopesticides suppressing plant pathogens, or promoting plant growth. Therefore, the objectives of this study were to develop a global strategy to screen all the potential NRPSs gene clusters in 48 gapless genomes of *Burkholderia* species and design the organization domains including the C-domain sub-types and predict the most probable produced peptides for all detected NRPSs gene clusters. Clusters potentially involved in the synthesis of new biocontrol agents were further analyzed by structural and genetic approaches to support the prediction obtained *in silico*.

### A. Identification of new non-ribosomal peptides from *Burkholderia*

In this study, 48 gapless complete genomes of *Burkholderia* available in the NCBI were extensively screened *in silico* to identify all NRPSs gene clusters, especially those producing siderophores and lipopeptides, by following Florine, a workflow we have developed especially with this aim. For all the putative NRPSs, the domains organization were designed, the C-domains sub-type was determined, and the prediction of the most probable produced peptides was performed *in silico* using specific bioinformatics tools. The studied strains include species playing ecological roles related to xenobiosis, biocontrol, promotion of plant-growth, biological nitrogen fixation, as well as pathogenic strains for animals and humans as some of those belonging to the Bcc group. They were grouped into three phylogenetic clades based on the alignment of their 16S rRNA sequences.

Analyses of genomes have revealed a numerous examples of NRPS biosynthesis gene clusters, with the potential to direct the production of novel NRPs. Although most strains produce the main siderophore ornibactin (Bcc group) or malleobactin ("mallei" group), a cluster corresponding to a new siderophore, called phymabactin, was identified in the genome of *B. phymatum* STM815. In the course of the mining of *B.ambifaria* AMMD, a cluster of 8 genes encoding for cepaciachelin pathway was identified. The putative *nrps* gene-containing cluster responsible for the antifungal activity was also identified in the genomes of both *B. ambifaria* AMMD and *Burkholderia* sp. KJ006. Results revealed large NRPS- encoding genes, as well as regulator, decorating, and transporter-encoding gene clusters. To identify the cluster dedicated to antifungal activity against soilborne pathogens, we constructed a burkholdin-deficient mutant, using *in vivo* homologous recombination in the yeast, *S. cerevisiae* InvSc1.

Cyclic lipopeptides (CLPs) are versatile molecules composed of a fatty acid tail linked to a cyclic oligopeptide head. The gene cluster of CLPs are characterized by the presence of two specific signatures including the C-starter domain, the condensation (C) domain of the first NRPS module that catalyses N-acylation of the first amino acid in the CLP module, and a tandem of Te domain involved in the cyclization of the CLP molecule. Analysis of the genomes revealed that all strains belonging to the *B. pseudomallei* species have the NRPS gene cluster that exhibits several features of NRPS genes responsible for CLP biosynthesis.

Analysis the genome of *B. rhizoxinica* HKI 454 revealed that the genome is composed of one chromosome and two plasmids that harbor 13 NRPS gene clusters responsible for the synthesis of different peptides potentially implicated in biocontrol.

This study revealed the significant potential of the genus *Burkholderia* to produce new NRPs, speeding the screening for new sources of antimicrobial agents, potentially useful for their biocontrol properties. Such a study will help to improve strategies of natural product discovery and gave insights on the non-ribosomal synthesis of *Burkholderia* exemplified by the identification of dual C/E domains in lipopeptide NRPSs, as frequently found in *Pseudomonas* strains.

The results are presented as the following article published in MirobiologyOpen. The table 1 that is not very clear in the published paper is presented in the end of supplementary material (page 98).

MicrobiologyOpen

Open Access

ORIGINAL RESEARCH

# *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis

Qassim Esmaeel[1], Maude Pupin[2,3], Nam Phuong Kieu[4], Gabrielle Chataigné[1], Max Béchet[1], Jovana Deravel[1], François Krier[1], Monica Höfte[4], Philippe Jacques[1] & Valérie Leclère[1,2,3]

[1]University Lille, INRA, ISA, University Artois, University Littoral Côte d'Opale, EA 7394-ICV - Institut Charles Viollette, F-59000 Lille, France
[2]University Lille, CNRS, Centrale Lille, UMR 9189-CRIStAL, Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
[3]Bonsai Team, Inria-Lille Nord Europe, F-59655 Villeneuve d'Ascq Cedex, France
[4]Laboratory of Phytopathology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

**Keywords**
*Burkholderia*, genome mining, lipopeptide, NRPS, siderophore.

**Correspondence**
Valérie Leclère, Institut Charles Viollette, ProBioGEM team, Bât. Polytech'lille, Avenue Langevin, Université de Lille1, F-59655 Villeneuve d'Ascq, Cedex, France.
Tel: +33320 43 46 68;
E-mail: valerie.leclere@univ-lille1.fr

**Funding Information**
The work was supported by the University of Lille 1, the INTERREG IV program France-Wallonie-Vlaanderen (Phytobio project), the bioinformatics platform Bilille, and Inria. QE received financial support from Sana'a University (Yemen).

Received: 17 December 2015; Revised: 26 January 2016; Accepted: 3 February 2016

doi: 10.1002/mbo3.347

## Abstract

*Burkholderia* is an important genus encompassing a variety of species, including pathogenic strains as well as strains that promote plant growth. We have carried out a global strategy, which combined two complementary approaches. The first one is genome guided with deep analysis of genome sequences and the second one is assay guided with experiments to support the predictions obtained in silico. This efficient screening for new secondary metabolites, performed on 48 gapless genomes of *Burkholderia* species, revealed a total of 161 clusters containing nonribosomal peptide synthetases (NRPSs), with the potential to synthesize at least 11 novel products. Most of them are siderophores or lipopeptides, two classes of products with potential application in biocontrol. The strategy led to the identification, for the first time, of the cluster for cepaciachelin biosynthesis in the genome of *Burkholderia ambifaria* AMMD and a cluster corresponding to a new malleobactin-like siderophore, called phymabactin, was identified in *Burkholderia phymatum* STM815 genome. In both cases, the siderophore was produced when the strain was grown in iron-limited conditions. Elsewhere, the cluster for the antifungal burkholdin was detected in the genome *of B. ambifaria* AMMD and also *Burkholderia* sp. KJ006. *Burkholderia pseudomallei* strains harbor the genetic potential to produce a novel lipopeptide called burkhomycin, containing a peptidyl moiety of 12 monomers. A mixture of lipopeptides produced by *Burkholderia rhizoxinica* lowered the surface tension of the supernatant from 70 to 27 $mN \cdot m^{-1}$. The production of nonribosomal secondary metabolites seems related to the three phylogenetic groups obtained from 16S rRNA sequences. Moreover, the genome-mining approach gave new insights into the nonribosomal synthesis exemplified by the identification of dual C/E domains in lipopeptide NRPSs, up to now essentially found in *Pseudomonas* strains.

## Introduction

The *Burkholderia* genus includes more than 60 species that colonize a wide range of environments including soil, water, plants, animal, and human (Mahenthiralingam et al. 2008). They are well known for their nutritional versatility, which certainly contributes to their capacity to live in extreme and diverse habitats and which has led to their use in biocontrol, bioremediation processes, and biodegradation of pollutants (Mahenthiralingam and Goldberg 2005). Some *Burkholderia* species are involved in promoting plant or nitrogen-fixing bacteria (Caballero-Mellado et al. 2007). On the other hand, several species of *Burkholderia*, including those from *B. cepacia* complex group (Bcc) as well as *B. gladioli* and *B. fungorum*, represent a significant threat to the life of immunocompromised individuals,

especially for those who are suffering from cystic fibrosis or chronic granulomatous diseases (Coenye et al. 2001). *Burkholderia pseudomallei* and *Burkholderia mallei* are the only known members of the genus *Burkholderia* that are primary pathogens in humans and animals, causing melioidosis in humans (Cheng and Currie 2005) and glanders in horses (Nierman et al. 2004). The ability to adapt and colonize a wide variety of environments is likely due to an unusually large, complex, and variable genome (4.6–9 Mb), split into up to three chromosomes and large plasmids. Furthermore, some species have been found to secrete a variety of extracellular enzymes with proteolytic, lipolytic, and hemolytic activities (Vial et al. 2007), together with secondary metabolites. These include siderophores called ornibactins and malleobactins (Franke et al. 2014), hemolytic peptides called cepalysins (Abe and Nakazawa 1994), antifungal agents as pyrrolnitrin (Hammer et al. 1999), cepapafungins, and related compounds cepacidines A1 and A2 (Lim et al. 1994). More recently, antifungal cyclic lipopeptides (CLPs) called occidiofungins/burkholdines (bks) have been isolated from strains belonging to the Bcc group (Gu et al. 2009; Tawfik et al. 2010). Some secondary metabolites including siderophores and lipopeptides can be produced nonribosomally by NonRibosomal Peptide Synthetases called NRPS (Marahiel 2009). These large synthetases are organized into sets of domains that constitute modules containing the information needed to complete an elongation step in an original peptide biosynthesis. The main catalytic functions are responsible for the activation of an amino acid residue (adenylation, -A domain), the transfer of the corresponding adenylate to the enzyme-bound 4-phosphopantetheinyl cofactor (thiolation, -T domain), the peptide bond formation (condensation, -C domain), and the release of the peptide from the NRPS (thioesterase, -Te domain). Optional domains of NRPS modules can modify the amino acids by different reactions such as *N*-methylation, epimerization (-E domain), or heterocyclization.

A genome mining was conducted with the aim of identifying biosynthetic clusters for secondary metabolites, but in this study, we have chosen to only present the clusters including at least one gene encoding NRPS. The domain organization of each NRPS protein was designed, the C-domain sub-types were determined, and the prediction of the most probable produced peptides was performed in silico using specific bioinformatics tools included in the Florine workflow previously described (Caradec et al. 2014). We specially describe the identification of siderophores and lipopeptides because of their possible use in biocontrol (Sharma and Johri 2003; Roongsawang et al. 2011). Many of the structure and activity predictions obtained in silico are corroborated by experimental supports as biological activities from wild-type strains and deletion mutants, iron chelating measurements, surface tension, or MS detections of the compounds in growth supernatants.

## Materials and Methods

### In silico detection of NRPS genes within sequenced genomes

Forty-eight gapless complete genomes of *Burkholderia* strains available in the NCBI were mined by following the strategy described within the Florine workflow (Caradec et al. 2014). Sequences presenting identity with known NRPSs were detected by using the NCBI-Blastp tool with the *B. cepacia* GG4 OrbJ used as a query (YP_006615894). The automatically annotated proteins were also fished using a list of keywords including adenylation, synthetase, thiotemplate, phosphopantetheine, nonribosomal, NRPS, siderophores, ornibactin, pyochelin, syringomycin, malleobactin, polyketide, and acyl carrier protein (ACP). Identification and annotation of NRPS gene clusters in all selected strains were also detected by antiSMASH (http://antismash.secondarymetabolites.org) (Weber et al. 2015).

Protein sequences of all genes encoding NRPSs were obtained from NCBI databases. The modular organization and domain architecture were deciphered combining the results from NRPS-PKS analysis website (http://nrps.igs.umaryland.edu/nrps) (Bachmann and Ravel 2009) and Structure-Based Sequence Analysis of Polyketide Synthases (SBSPKS) (http://www.nii.ac.in/~pksdb/sbspks/search_main_pks_nrps.html) (Anand et al. 2010) with antiSMASH results. The specificity of the A domain was conducted by using the web-based software NRPSpredictor2 (http://nrps.informatik.uni-tuebingen.de) (Röttig et al. 2011). The C-domain types were determined by identification of specific signatures of the DownSeq (Caradec et al. 2014). The D-configuration of monomers of nonribosomal peptide was predicted regarding the presence of E followed by $^{D}C_{L}$ domains or dual C/E- domains.

The structures of the predicted peptides were compared to other nonribosomal peptides through the structural search tool of the Norine database (bioinfo.lifl.fr/norine) (Caboche et al. 2008, 2009).

### Culture conditions and media

Strains, plasmids, and primers used in this study are listed in Table S1. Molds (*Fusarium oxysporum*, *Galactomyces geotrichum*, *Botrytis cinerea*, and *Rhizoctonia solani*) were gown on PDA (Sigma-Aldrich, St Louis, MO). *Burkholderia* strains, *Micrococcus luteus*, *Listeria innocua*, and *Escherichia coli* were routinely grown at 37°C on Luria Bertani (LB) medium. *Candida albicans* was grown on Sabouraud agar at 37°C and *Saccharomyces cerevisiae* was grown on

yeast-extract–peptone–dextrose at 30°C. For lipopeptide production, strains were grown in Landy medium buffered with MOPS 100 mmol/L (Landy et al. 1948). When required, gentamycin sulfate (Sigma Aldricht) was added at a concentration of 25 $\mu$g mL$^{-1}$ for *E. coli* and at 300 $\mu$g mL$^{-1}$ for *B. ambifaria* AMMD *wt* and mutant strains.

## Construction of AMMD mutant strains

*Burkholderia ambifaria* AMMD-Δbamb_6472 was constructed using in vivo homologous recombination in the yeast, *S. cerevisiae* InvSc1 (Shanks et al. 2006). To construct *B. ambifaria* AMMD-Δbamb_6472, two regions of the NRPS gene Bamb_6472 were amplified using primers Up6472-F and Up6472-R (1015 bp amplicon) and primers Down6472-F and Down6472-R (887 bp amplicon), respectively. Deletion of NRPS gene Bamb_6472 was confirmed by PCR (See Protocole S2 for details).

## Biological activity assays

### Antifungal activity

Antifungal activity was demonstrated by contact antagonism assay. PDA plates were inoculated with 5 mm of agar plug of *F. oxysporum*, *G. geotrichum*, *B. cinerea*, and *R. solani* grown for 2 days on PDA. The plates were inoculated with the bacterium and incubated at 25°C for 5–7 days.

### Antiyeast activity

Antiyeast activity was performed by the agar drop plate method. Erlenmeyer flask containing 100 mL of PDB were inoculated with 1.5 mL of bacterial suspension containing about $10^8$ CFU mL$^{-1}$ and incubated under rotary shaking (160 rpm) at 37°C for 5 days. Then, the culture was centrifuged at 10,000*g* for 10 min at 4°C. The supernatant was concentrated 10 times via speed vacuum, and then 50 $\mu$L of this supernatant were deposited on 14 mL solid Sabouraud medium inoculated with target strains including *C. albicans* and *S. cerevisiae*. After 24 h of incubation at 30°C, the activity was estimated by measuring the diameter of the zone of inhibition of target strains.

### Hemolytic activity

The hemolytic activity was assessed at 30°C on LB supplemented with 5% (vol/vol) horse blood. Supernatant (50 $\mu$L) were spotted on the plates that were incubated for 48 h at 30°C. Positive results were detected by the formation of a clear zone around the wells.

## MALDI-ToF mass spectrometry

Bacterial sample were cultured in liquid medium and incubated at 37°C for 5 days. Supernatant was mixed with a matrix solution (10 mg/mL cyano-4-hydroxycinnamic acid in 70% water, 30% acetonitrile, and 0.1% TFA). The samples were homogeneized on a Vortex and centrifuged at 4500 g. For classical analysis 1 $\mu$L of sample solution was spotted onto a MALDI-ToF MTP 384 target plate (Bruker Daltonik GmbH, Leipzig, Germany) according to the procedure of the dried-droplet preparation. Mass profiles experiments were analyzed with an Ultraflex MALDI-ToF/ToF mass spectrometer (Bruker, Bremen, Germany) equipped with a smartbeam laser. Samples were analyzed using an accelerating voltage of 25 kV and matrix suppression in deflexion mode at m/z 750. The laser power was set to just above the threshold of ionization (around 35%). Spectra were acquired in reflector positive mode in the range of 400 at 3000 Da. Each spectrum was the result of 2000 laser shots per m/z segment per sample delivered in 10 sets of 50 shots distributed in three different locations on the surface of the matrix spot. The instrument was externally calibrated in positive reflector mode using bradykinin [M+H]+ 757.3991, angiotensin II [M+H]+ 1046.5418, angiotensin I [M+H]+ 1296.6848, substance P [M+H]+ 1347.7354, bombesin [M+H]+ 1619.8223, and ACTH (1-17) [M+H]+ 2093.0862.

## Production of siderophores

Strains were propagated three times in iron-deficient minimum medium (MM9) (Payne 1994) supplemented with 10% of casamino acids previously treated, first with 3% 8-hydroxyquinoline in chloroform to remove contaminating iron and then with chloroform to remove remaining 8-hydroxyquinoline. Then, $MgCl_2$ 1 mmol/L and $CaCl_2$ 0.1 mmol/L were added. All glassware was rinsed with HCl 6 mol/L. CAS (Chrome Azurol S) liquid and agar assays were performed in accordance to the original protocol (Schwyn and Neilands 1987).

## Surface tension measurements

The surface tension was measured according to the De Nouy methodology using a tensiometer TD1 (Lauda, Königshofen, Germany) as previously described (Leclère et al. 2006).

## Phylogeny tree

The 16S rRNA sequences were extracted from RDP database (Cole et al. 2009) and submitted to MEGA6 program (Tamura et al. 2013). The tree was built using the neighbor-joining method (Saitou and Nei 1987). The evolutionary

distances were computed using the maximum composite likelihood method (Tamura et al. 2004). The rate variation among sites was modeled with a gamma distribution with five rate categories. Bootstrap analysis with 1000 replicates was performed to assess the support of the clusters (Felsenstein 1985). All positions containing gaps and missing data were eliminated. *Cupriavidus taiwanensis* was used as an outgroup reference.

## Results

### Overview on *Burkholderia* sequenced genomes

Forty-eight gapless complete genomes of *Burkholderia* strains available in 2014 within the NCBI genome bank were mined for NRPSs. To evaluate the relevancy of sequenced strains among the *Burkholderia* genus, the 16S rRNA sequences were extracted and aligned to build a phylogenetic tree of all sequenced strains (Fig. 1). The considered strains for NRPS seeking are representative of the genus as they include species playing ecological role related to xenobiosis, biocontrol, promotion of plant growth, biological nitrogen fixation, as well as pathogenic strains for animals as some of those belonging to the Bcc group. Elsewhere, the tree clearly shows that the strains are distributed into three clades. The first clade (clade I) includes *B. mallei*, *B. pseudomallei*, and the rarely pathogenic *B. thailandensis* strains, the second clade (clade II) gathers the Bcc group strains. The remaining species are contained in the clade III.

All the strains have their genome split into up to three chromosomes added with one or two plasmids, exceptionally three (*B.* sp.YI23), four (*B. gladioli* BSR3 and *B. glumae* BGR1), or five (*B. vietnamiensis* G4). The whole genome sizes span from 3.75 to 9.73 Mb with most frequent values between 7 and 9 Mb (33 strains).

Except for *B. phenoliruptrix* BR3459a and *B.* sp. CCGE1002, genes for at least one NRPS could have been detected in the genomic sequence of the 48 studied strains, whatever their lifestyle or virulence ability (Table 1).

### Identification of a novel malleobactin-like siderophore

A cluster of genes including those for NRPS responsible for ornibactin or malleobactin synthesis has been found in the genome of 41 strains (Table 1) (Fig. 2A). The cluster is present on the larger chromosome (chromosome 1), except for *B. phytofirmans* PsJN and *B. xenovorans* LB400 where it is located on chromosome 2, and for *B.* sp. YI23 and *B.* sp. RPE64 harboring the cluster on a plasmid. A complete cluster includes genes responsible for siderophore biosynthesis (coding for 2 NRPSs and accessory enzymes modifying the monomers of the peptidic chain), genes implicated in the uptake (siderophore export and ferri-siderophore TonB dependent receptor) and the *orbS* or *mbaF* genes coding for an extracytoplasmic sigma70 factor implicated in the regulation (Table S2). The NRPSs responsible for the siderophore backbone are encoded by *orbI-orbJ* and *mbaA-mbaB* in the case of ornibactin and malleobactin synthesis, respectively. The proteins harbor the same domain organization including 4 adenylation domains (A1-A4). The specificity code for the A1 domain is variable: Leu for ornibactin and Bht (beta-hydroxytyrosin) for malleobactin (Fig. 2A). According to the presence of an acyl transferase gene (*orbL*) together with the Leu selectivity of A1 domain, ornibactin is predicted to be produced by all the strains belonging to the Bcc group (*B. ambifaria*, *B. cenocepia*, *B. cepacia*, *B. lata*, *B. multivorans*, *B. vietnamiensis* and *B.* sp. KJ006) and two strains belonging to the clade III (YI23 and RPE64). The malleobactin synthetic cluster detected with the presence of an accessory enzyme gene (*mbaM*) and the Bht selectivity for A1 is found in the genomes of two strains belonging to the clade III (*B. phytofirmans* PsJN and *B. xenovorans* LB400), and the strains from clade I, except *B. thailandensis* 2002721723.

*Burkholderia phymatum* STM815 belonging to clade III, possesses on chromosome 2, a cluster with some characteristics for malleobactin synthesis and uptake (absence of the acyl transferase gene *orbL*), together with features related to ornibactin biosynthesis (absence of *mbaM* and presence of *orbK*). Moreover, the predicted specificities of A1 and A4 domains of the NRPS are Asp and Cys, respectively. These predictions are different from those obtained both with ornibactin and malleobactin NRPS A1 and A4 domains (Fig. 2A). The strain *B. phymatum* STM815 was grown under iron-limited conditions and was analyzed for siderophore production. The CAS agar assay was positive showing an orange halo surrounding the strain (Fig. S1) and compounds with a mass comprised between 445 Da and 773 Da were detected and fragmented from supernatant. No fragmentation spectrum could have been assigned to a known siderophore belonging to ornibactin or malleobactin family. We suggest to name this probably new malleobactin-like siderophore phymabactin and the corresponding *nrps* genes *phmA* and *phmB* (Fig. 2) (Table S2).

### Identification of a cluster involved in cepaciachelin synthesis

Pyochelin synthetase is constituted of two NRPSs of about 1360 and 1890 AA, harboring a specific domain architecture that includes two cyclization domains (Cy), associated

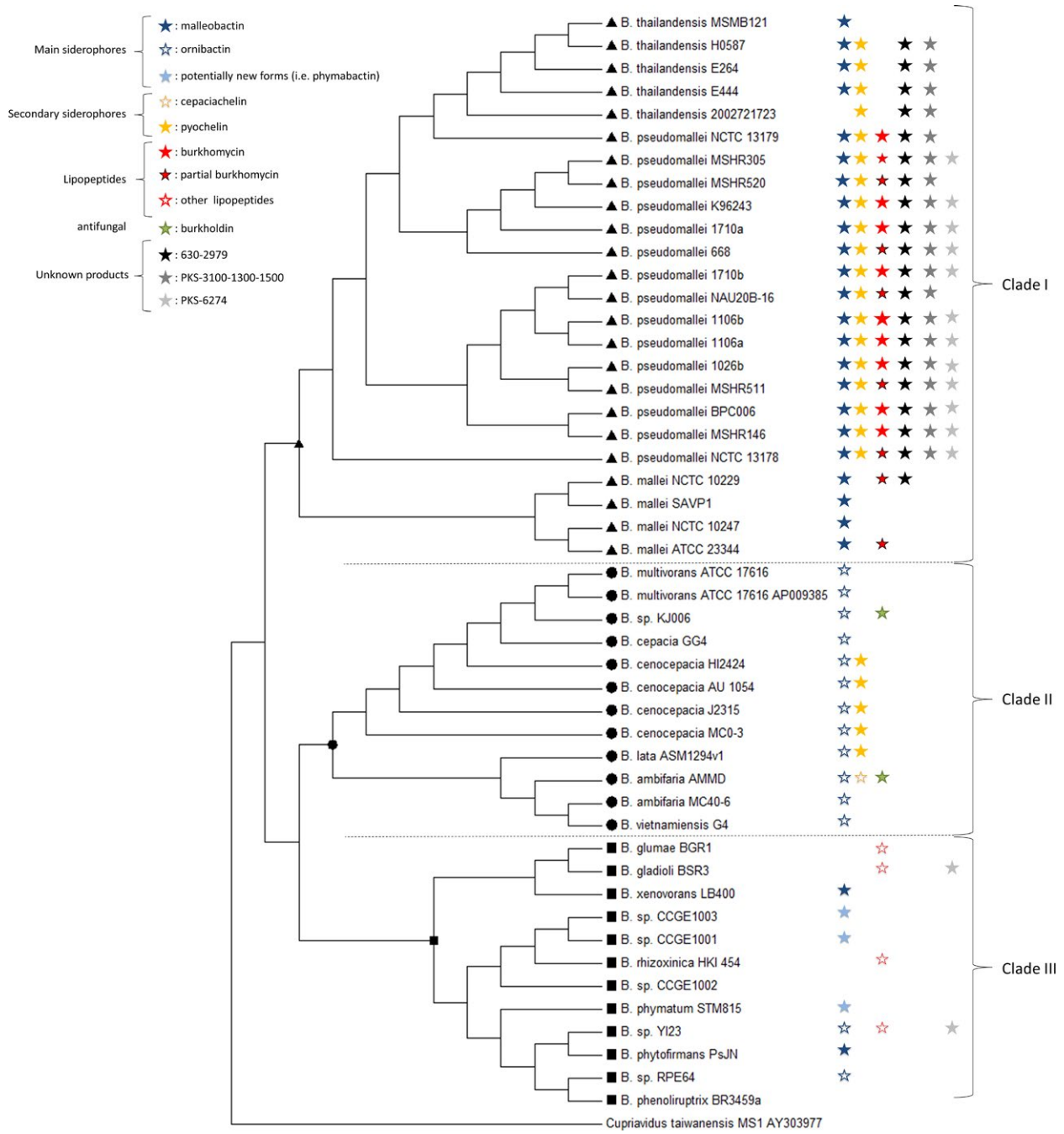



**Figure 1.** Phylogenetic tree of the *Bukholderia* strains with complete genome sequenced. Performed using MEGA6 (Tamura et al. 2013), the tree was built using the neighbor-joining method (Saitou and Nei 1987). The bootstrap analysis (Felsenstein 1985) with 1000 replicates was performed to assess the support of the clusters. The analysis involved 48 16S rRNA sequences and all positions containing gaps and missing data were eliminated. ▲Clade I, ●Clade II, ■Clade III, colored symbols representing the different clusters are presented in front of each strain.

with two stand-alone proteins containing a Te and an A domain, respectively (Fig. 2B) (Michel et al. 2007; Gasser et al. 2015). Corresponding clustered genes have been found in the chromosome 2 of 19 and 5 strains belonging to clades I and II, respectively (*B. lata* ASM1294v1, all *B. cenocepia*, *B. pseudomallei* and *B. thailandensis* strains except *B. thailandensis* MSMB121) (Table 1). Except for *B. thailandensis* 2002721723 for which no malleobactin synthesis genes have been found, pyochelin synthesis clusters are always present together with ornibactin or

**Table 1.** Overview on NRPS identified by genome mining on *Burkholderia* strains.

| | | | | | | Main siderophore | |
|---|---|---|---|---|---|---|---|
| Phylogenic group | *Burkholderia* species | Genome size (Mb) | Number of chromosomes | Number of plasmids | Lifestyle | Ornibactin | Malleobactin |
| Clade I | *B. thailandensis* MSMB121 | 6.73 | 2 | 0 | Saprophytic, human pathogen (rare) | | Chr1 |
| | *B. thailandensis* H0587 | 6.76 | 2 | 0 | Saprophytic, human pathogen (rare) | | Chr1 |
| | *B. thailandensis* E264 | 6.72 | 2 | 0 | Saprophytic, human pathogen (rare) | | Chr1 |
| | *B. thailandensis* E444 | 6.65 | 2 | 0 | Saprophytic, human pathogen (rare) | | Chr1 |
| | *B. thailandensis* 2002721723 | 6.57 | 2 | 0 | Saprophytic, human pathogen (rare) | | |
| | *B. pseudomallei* NCTC 13179 | 7.33 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* MSHR305 | 7.42 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* MSHR520 | 7.45 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* K96243 | 7.25 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* 1710a | 7.33 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* 668 | 7.04 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* 1710b | 7.30 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* NAU20B-16 | 7.31 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* 1106b | 7.21 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* 1106a | 7.08 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* 1026b | 7.23 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* MSHR511 | 7.31 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* BPC006 | 7.15 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* MSHR146 | 7.31 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. pseudomallei* NCTC 13178 | 7.39 | 2 | 0 | Human pathogen | | Chr1 |
| | *B. mallei* NCTC 10229 | 5.74 | 2 | 0 | Human and animal pathogen | | Chr1 |
| | *B. mallei* SAVP1 | 5.23 | 2 | 0 | Human and animal pathogen | | Chr1 |
| | *B. mallei* NCTC 10247 | 5.84 | 2 | 0 | Human and animal pathogen | | Chr1 |
| | *B. mallei* ATCC 23344 | 5.83 | 2 | 0 | Human and animal pathogen | | Chr1 |
| Clade II | *B. multivorans* ATCC 17616 | 7.01 | 3 | 1 | Human pathogen | Chr1 | |
| | *B. multivorans* ATCC 17617 | 7.01 | 3 | 1 | Human pathogen | Chr1 | |
| | *B.* sp. KJ006 | 6.63 | 3 | 1 | Biocontrol, mutualist | Chr1 | |
| | *B. cepacia* GG4 | 6.47 | 2 | 0 | Bioremediation, human pathogen | Chr1 | |
| | *B. cenocepacia* HI2424 | 7.70 | 3 | 1 | Human pathogen | Chr1 | |
| | *B. cenocepacia* AU 1054 | 7.28 | 3 | 0 | Human pathogen | Chr1 | |
| | *B. cenocepacia* J2315 | 8.05 | 3 | 1 | Human pathogen | Chr1 | |
| | *B. cenocepacia* MC0-3 | 7.97 | 3 | 0 | Plant pathogen, human pathogen | Chr1 | |
| | *B. lata* ASM1294v1 | 8.68 | 3 | 0 | Human pathogen | Chr1 | |
| | *B. ambifaria* AMMD | 7.53 | 3 | 1 | Biocontrol | Chr1 | |
| | *B. ambifaria* MC40-6 | 7.64 | 3 | 1 | Human pathogen | Chr1 | |
| | *B. vietnamiensis* G4 | 8.39 | 3 | 5 | Mutualist, human pathogen | Chr1 | |

| Secondary siderophore | | Lipopeptides | | Antifungal | Unknown products | | | |
|---|---|---|---|---|---|---|---|---|
| Pyochelin | Cepaciachelin | Burkhomycin (CLP12) | Cstart_Te (LP) | Burkholdin | "630-2979" AT-CATECAT | PKS-3100_1300_ 1500 | PKS-NRPS "6274" | NRPS |
| Chr2 | | | | | Chr1 | Chr2 | | |
| Chr2 | | | | | Chr1 | Chr2 | | |
| Chr2 | | | | | Chr1 | Chr2 | | |
| Chr2 | | | | | Chr1 | Chr2 | | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | | |
| Chr2 | | Chr2(11) | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2(9) | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2(5) | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2(7) | | | Chr1 | Chr2 | | |
| Chr2 | | Chr2 (S) | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2(10) | | | Chr1 | Chr2 | Chr1(S) | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1 | |
| Chr2 | | Chr2 | | | Chr1 | Chr2 | Chr1(S) | |
| Chr2 | | Chr2(10) | | | Chr1 | Chr2 | Chr1 | |
| | | Chr2(5) | | | Chr1 | | | |
| | | | | | Chr1 | | | |
| | | Chr2(5) | | | | | | |
| | | | | Chr3 | | | | |
| Chr2 | | | | | | | | |
| Chr2 | | | | | | | | |
| Chr2 (P) | | | | | | | | |
| Chr2 | | | | | | | | |
| Chr2 | | | | | | | | |
| | Chr1 | | | Chr3 | | | | |

**Table 1.** (*Continued*)

| Phylogenic group | *Burkholderia* species | Genome size (Mb) | Number of chromosomes | Number of plasmids | Lifestyle | Main siderophore: Ornibactin | Main siderophore: Malleobactin |
|---|---|---|---|---|---|---|---|
| Clade III | *B. glumae* BGR1 | 7.28 | 2 | 4 | Plant pathogen | | |
| | *B. gladioli* BSR3 | 9.05 | 2 | 4 | Plant pathogen | | |
| | *B. xenovorans* LB400 | 9.73 | 3 | 0 | Mutualist | | Chr2 |
| | *B.* sp. CCGE1003 | 7.04 | 2 | 0 | Mutualist | *Chr1* | |
| | *B.* sp. CCGE1001 | 6.83 | 2 | 0 | Mutualist | *Chr1* | |
| | *B. rhizoxinica* HKI 454 | 3.75 | 1 | 2 | Mutualist | | |
| | *B.* sp. CCGE1002 | 7.88 | 3 | 1 | Mutualist | | |
| | *B. phymatum* STM815 | 8.68 | 2 | 2 | Mutualist, nitrogen fixation | *Chr2* | |
| | *B.* sp. YI23 | 8.89 | 3 | 3 | Bioremediator | Plasmid | |
| | *B. phytofirmans* PsJN | 8.21 | 2 | 1 | Biocontrol | | Chr2 |
| | *B.* sp. RPE64 | 6.96 | 3 | 2 | Mutualist | Plasmid | |
| | *B. phenoliruptrix* BR3459a | 7.65 | 2 | 1 | Mutualist | | |

Potentially new forms are in italics. Chr, chromosome on which the cluster is; *numbers in brackets indicate the number of monomers expected in the product when different from the model; (P), partial cluster; (S), the NRPS is split into more proteins than in the model; LP, lipopeptide.
*white box is for Chr1; light grey box is for Chr2; dark grey box is for Chr3

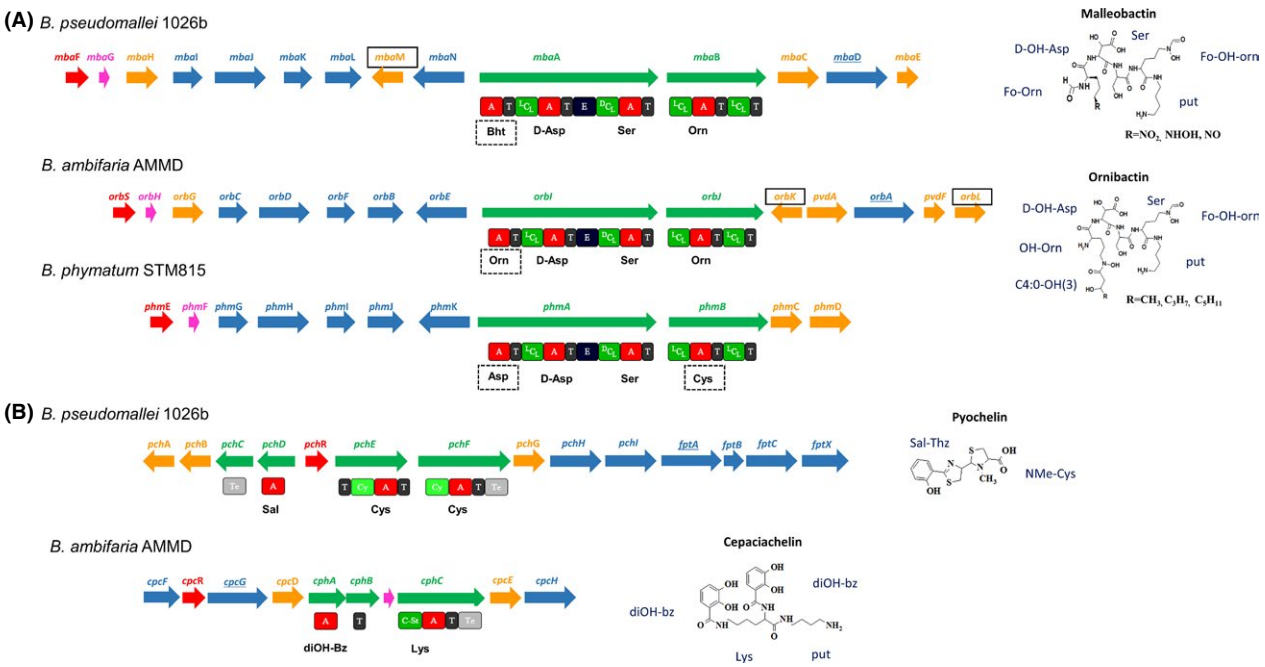



**Figure 2.** Siderophore biosynthesis clusters. Clusters are presented on the left side and the products on the right side. Genes are represented by arrows colored as follows: green for NRPS, orange for accessory enzymes, red for regulatory proteins, blue for uptake participation, and pink for unknown function. Genes encoding TonB-dependent receptor are underlined. The domain organization of the NRPSs is shown below arrows. A, adenylation domain; Cst, C-starter; $^{L}C_{L}$, condensation between two L-monomers; $^{D}C_{L}$, condensation between D-monomer and L-monomer; Cy, heterocyclisation domain; E, epimerization domain; T, thiolation domain; Te, thioesterase domain; predicted amino acid specificity is shown under each A domain; Bht, beta–hydroxy-tyrosin; diOH-bz, 2, 3-dihydroxybenzoic acid; sal, salicylic acid. (A) Main siderophores. Genes and monomers specific for malleobactin or ornibactin synthesis are framed and highlighted by dotted squares, respectively. (B) Secondary siderophores.

| Secondary siderophore | | Lipopeptides | | Antifungal | Unknown products | | | |
|---|---|---|---|---|---|---|---|---|
| Pyochelin | Cepaciachelin | Burkhomycin (CLP12) | Cstart_Te (LP) | Burkholdin | "630-2979" AT-CATECAT | PKS-3100_1300_ 1500 | PKS-NRPS "6274" | NRPS |
| | | | Chr2 (LP7) | | | | | Chr2 (4a + 4b) |
| | | | Chr2 (LP7 + LP5) | | | | Chr1 | Chr2 (4a + 4b) |
| | | | Chr1 + plasmid (13 LPs) | | | | | |
| | | | Chr3(6) | | | | Chr3 | |

malleobactin synthesis genes. No relationship between occurrence of the pyochelin cluster and the belonging to the clade I or II can be established.

In the chromosome 1 of *B. ambifaria* AMMD, a second cluster characteristic of a siderophore synthesis pathway has been identified beside the cluster for ornibactin synthesis. This cluster includes a gene for a TonB-dependent receptor, genes to build up a diOH-bz (2,3-dihydroxy benzoic acid) monomer and two *nrps* genes coding for a complex constituted of one stand-alone A domain and a protein with a [Cstarter-A-Te] domain architecture (Fig. 2B). The specificity code for the isolated A domain is diOH-bz monomer with a score of 90%. The specificity code for the second A domain is Lys. The organization of the genes and the predictions associated to both A domains are highly compatible with the cepaciachelin siderophore synthesis. The strain was grown under iron-limited conditions allowing the production of siderophores (Fig. S1A). Mass spectrometry analysis of the supernatant revealed the coproduction of two compounds with masses corresponding to C4-ornibactin ([M + H+] + Fe = 759; [M + Na+] + Fe = 787) and cepaciachelin ([M + H+] = 789; [M + Na+] = 511) (Fig. S1B). The fragmentation of the latter compounds revealed the presence of diOH-bz and putrescine residues. The production of both siderophores (ornibactin and cepaciachelin) was totally abolished by addition of iron into the culture broth (not shown).

## Identification of a novel lipopeptide synthetase containing dual C/E domains

A complete cluster corresponding to a lipopeptide synthesis pathway has been found in chromosome 2 in the genomes of 9/15 strains belonging to the *pseudomallei* species (Table 1) (Fig. 3). The NRPS is generally constituted of three proteins containing three, four, and five modules, respectively, except for strains 1106b, 1710a, and NCTC13179 for which the mis-annotated synthetase spans on six, four, and five proteins, respectively. The NRPSs are generally annotated as "hypothetical protein," "(non ribosomal) peptide synthetase," or "syringomycin synthetase." The synthetase composed of 12 [C-A-T] modules is predicted to produce a lipopeptide containing 12 monomers in the peptidic moiety. The first C domain has been identified as a Cstarter and the C domains of modules 2, 4, 6, 8, and 11 harbor the dual C/E signature found in CLP domains of *Pseudomonas*. This putative CLP12 is predicted to be FA_D-Ser_Glu_D-Ser_Arg_D-Thr_Leu_D-Dab_Thr_Thr_D-Glu_Gly_Val (where FA is for Fatty Acid and Dab for diaminobutyric acid). Structure search query performed in the Norine database revealed no peptide displaying similar pattern, even when lowering the threshold to six monomers on the 12 of the peptidic moiety. We suggest to call this potential lipopeptide burkhomycin and to name the corresponding NRPSs BkmA, BkmB, and BkmC (Fig. 3).

The cluster was only partially found in the genome of the six remaining *B. pseudomallei* strains (MSRH 305, 511, 520, NAU20B-16, 668 and NCTC 13178) and in the genomes of two *B. mallei* strains (ATCC 23344 and NCTC 10229) (Fig. 3).

All these NRPSs are characterized by the presence of a Cstarter, a Te tandem ending the assembly line, the lack of any E domain, and the presence of dual C/E domains. A gene encoding a protein of 927 AA, assumed to be an aminotransferase, is present within all the CLP biosynthetic clusters.

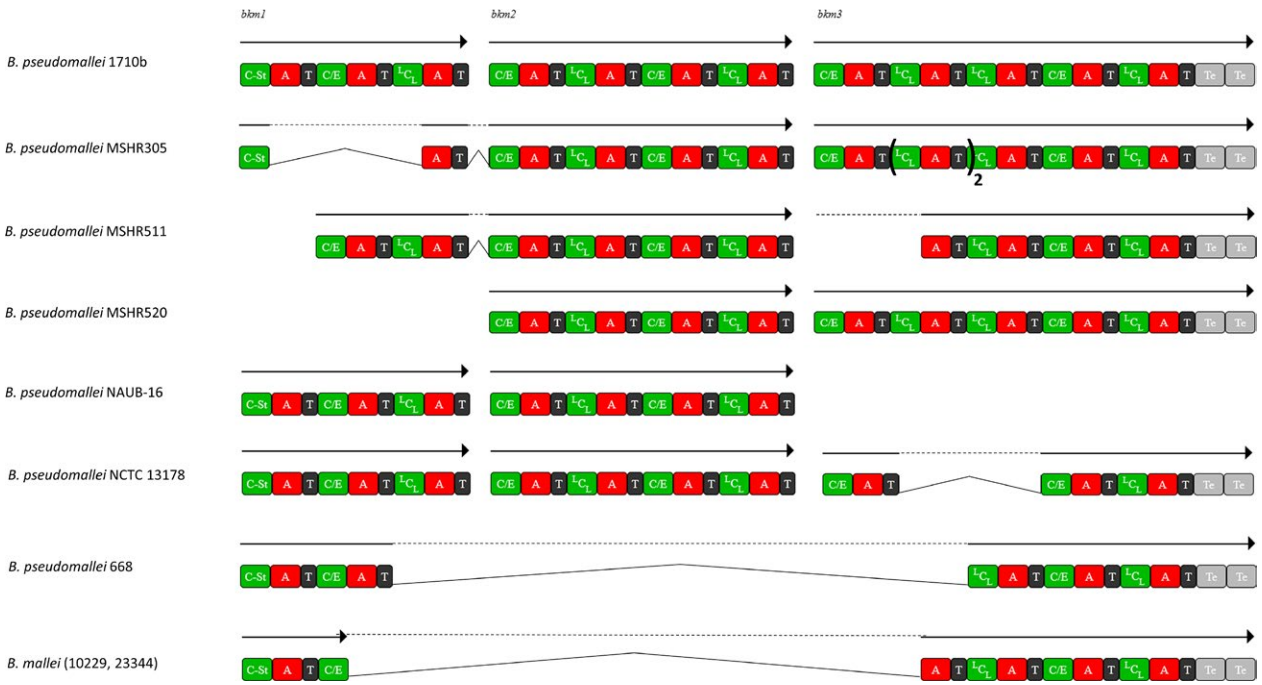



**Figure 3.** Burkhomycin biosynthesis gene clusters. *Burkholderia pseudomallei* 1710b is represented as a model for the complete cluster found in most strains. All the partial clusters are drawn below. Genes are represented by arrows, dotted lines are for lacking genomic sequences, lines represent the fusions in the proteins. Similar domains (adenylation with the same prediction and condensation with the same type) have been aligned. A, adenylation domain; C/E, dual condensation domain catalyzing both epimerization and condensation; $^{L}C_{L}$, condensation between two L-monomers; Cst, Cstarter for condensation between fatty acid and L-monomer; T, thiolation domain; Te, thioesterase domain.

## Lipopeptide production by *B. glumae*, *B. gladioli*, and *B. rhizoxinica*

*Burkholderia glumae* BGR1 and *B. gladioli* BSR3, both have a cluster of genes encoding a NRPS potentially producing a lipoheptapeptide (Table 1). Indeed, this protein can be divided into seven modules, the module 1 includes a Cstarter and modules 2, 3, 4, and 6 contain a dual C/E domain. For *B. gladioli* BSR3, the predicted peptidic moiety is D-Thr_D-Pro_D-Gln_Gly_D-X_Phe_Pro, where X represents a position for which no reliable prediction was obtained. For *B. glumae* BGR1, the same structure was predicted except that a D-Ser is predicted for position 3.

The higher number of NRPS gene clusters was found in *B. rhizoxinica* HKI454. The genome of this strain is constituted of one chromosome and two large plasmids pBRH01 and pBRH02, containing 2949, 784, and 203 genes, respectively. The genome mining using the keyword "nonribosomal" revealed at least 31 potentially interesting genes, 12 located on the chromosome and 19 on the plasmid pBRH01. The NRPS clusters represent 4% of the total genome length. We have further analyzed the 19 proteins with a size larger than 1000 AA and we also considered smaller proteins when they were supposed to be included in cluster also containing larger NRPSs. Finally, 13 clusters were pointed out, 10 of them being harbored by the plasmid pBRH01 (Table S3). All the NRPSs contain at least one dual C/E domain and nine of them start with a Cstarter domain, without any ambiguity. The synthetases are predicted to direct the synthesis of 13 different lipopeptides spanning from two to ten monomers within the peptidic moiety. For clusters 10, 11, and 13, the synthetases contain additional methylation domains. The strain was grown in different media (LB, TSB, TSB-glycerol, nutrient broth, Landy). The production of compounds with surfactant properties was assessed by surface tension measurements. The compounds produced by the strain grown in Landy medium displayed the higher activity lowering the surface tension from 70 to 27 $mN{\cdot}m^{-1}$. Compared to results obtained with purified surfactin, the curve corresponding to surface tension versus supernatant dilution clearly shows that a mixture of several active compounds is produced by *B. rhizoxinica* HK1454 (Fig. S2). The supernatant was then analyzed by MALDI-ToF. A group of peaks, characteristic of lipopeptide detection with different fatty acid chain lengths, with mass differences of 14 Da (1522 and 1536), was observed. (Fig. S2). The same supernatant was tested for antimicrobial activities. An antibacterial activity was observed against *Micrococcus luteus* and

*Listeria innocua*. A moderate inhibition was observed against the yeast *C. albicans* ATCC10231, and no activity at all against fungi *F. oxysporum*, *B. cinerea*, and *R solani* (not shown).

### Identification of the antifungal Bk cluster in the genomes of *B. ambifaria* AMMD and *B.* sp.KJ006

The genome mining revealed an *nrps*-containing cluster located in chromosome 3 of 2 strains: *B. ambifaria* AMMD and *B. sp*. KJ006. This large cluster spanning over 49 000 nt also includes genes coding for regulatory, decorating, and transporter proteins. The proteins exhibit several typical features of NRPS/PKS hybrids involved in Bk biosynthesis (Fig. 4). Eight modules are present, containing NRPS domains as C, A, E, T, and PKS domains as acetytransferase (AT), ACP, ketosynthetase (KS), and ketoreductase (KR) domains. The predicted compound can have the following structure Ile_PKM_Ser_D-Trp_D-Ser_Gly_Asn_D-Ser where the PKM represent a monomer incorporated by the PKS domains. This predicted peptide looks like cyclic glycolipopeptides belonging to the bk/occidiofungin family. Different types of condensation domains have been detected: $^{L}C_{L}$, $^{D}C_{L}$ following an E domain, and an hybrid condensation domain (noted "C") allowing the condensation of a PKM to an NRP monomer (Fig. 4). The presence of a gene encoding a glycosyl transferase suggests that the product can be glysosylated.

A large 9530-bp fragment within gene Bamb_6472 of *B. ambifaria* AMMD was deleted using in vivo homologous recombination in the yeast, *S. cerevisiae* InvSc1 leading to AMMD-Δbamb_6472 mutant (Protocole S2).

Both the AMMD wild type and mutant were tested for biological activities potentially related to the Bk production. Compared to the AMMD wild-type strain, the AMMD-Δbamb_6472-deficient mutant failed to inhibit growth of *C. albicans* and *S. cerevisiae*, exhibited a reduced hemolytic activity, and antagonistic effect on *F. oxysporum*, *G. geotrichum*, *B. cinerea*, and *R solani* (Fig. S3). Culture supernatants from wild-type and deletion mutant both grown in LB broth at 37°C were analyzed by MALDI-ToF. Two peaks observed for m/z 1200.5 and 1222.65 corresponding to the [M+H] and [M+Na] forms of a compound with a mass of 1999.5 Da were only detected in the wild-type supernatant (Protocole S2). Moreover, MALDI-ToF mass spectrometry analysis of culture supernatants confirmed that the production of an antifungal compound with a mass of 1999.5 Da was completely abolished in AMMD-Δbamb_6472 mutant. Considering their belonging to the biosynthesis pathway for bk, the genes were named *bks* for Bk synthesis (Fig. 4).

## Discussion

The comprehensive mining carried out on genome sequences has revealed numerous clusters of genes involved in the synthesis of secondary metabolites including genes coding for PKS, NRPS, and hybrids. Among them, more

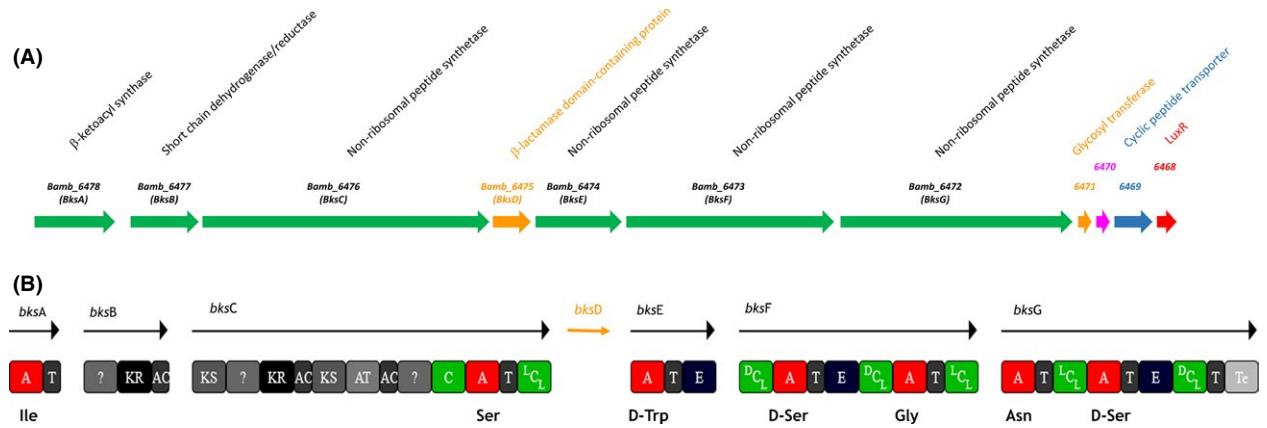



**Figure 4.** Gene cluster responsible for burkholdin (Bk) biosynthesis in *Burkholderia ambifaria* AMMD. (A) all genes involved in Bk synthesis; gene numbers are above the arrows, annotations from NCBI are reported above the gene numbers. Different colors are used to distinguish between the functions: green is for modular enzymes, orange for accessory enzymes, blue for transport, red for regulation, and pink for hypothetical protein. (B) Modular organization of Bk synthesis proteins (Bks). ?, nonidentified domains; A, adenylation domain; C, hybrid condensation domain allowing condensation between a PKS monomer and an NRPS monomer; $^{L}C_{L}$, condensation between two L-monomers; $^{D}C_{L}$, condensation between D-monomer and L-monomer; T, thiolation domain; E, epimerization domain; Te, thioesterase domain; AT, acetytransferase domain; AC, acyl carrier protein domain; KS, ketosynthetase domain; KR, ketoreductase domain. Predicted amino acid specificity is shown under each A domain.

than 160 clusters contain at least one *nrps* gene. With only two exceptions, such *nrps* containing clusters were found in all the explored genomes, whatever phylogenetic distance (Fig. 1) and the bacterial lifestyle (Table 1). Indeed, *Burkholderia* is an intriguing and important genus encompassing a variety of species and strains, ranging from highly pathogenic organisms to strains that promote plant growth (Mahenthiralingam et al. 2008). Some of these properties can be related to the production of secondary metabolites, especially NRPs as those displaying antifungal activities as Bks (Lin et al. 2012). In this article, we have focused on nonribosomal siderophores and lipopeptides because of their potential role in biocontrol, but the annotations we have assigned to some genes will also be very helpful to infer the "*Burkholderia* genome database," a specific tool dedicated to cystis fibrosis research community and aimed to facilitate comparative analysis (Winsor et al. 2008). Elsewhere, the present work completes previous analysis as the annotation of the genome of *B. rhizoxinica* HKI 454 (Lackner et al. 2011). Especially, with this example, we have shown the relevancy of combining approaches both in silico and in wetlab.

The genomic mining for new NRPs highlighted the presence of clusters specific for strains belonging to clade I. This includes a cluster annotated "630-2979" and two hybrid PKS-NRPS "3100-1300-1500" and "6274," named according to the protein sizes (Table 1, Fig. 1). None of the three predicted peptides share any pattern with the known curated peptides annotated in the Norine database. We have no other clues leading to the prediction for siderophore activity or lipopeptide structure and the concerned strains are pathogenic. We can guess that the products may be involved in the virulence and will not be useful in biocontrol, but this has to be further supported experimentally. Moreover, as they seemed to be only produced by bacteria belonging to clade I, they can be considered as maker of the *pseudomallei* species.

As expected, a cluster for biosynthesis of malleobactin and ornibactin was pointed out from the genomes of bacteria belonging to clade I and Bcc group, respectively (Franke et al. 2014, 2015). However, potentially new forms for a main siderophore have been identified in other strains. These includes a malleobactin-like produced by *B. phymatum* we have called phymabactin, and compounds putatively produced by *B. sp*. CCG1001 and *B. sp*. CCG1003. The latter NRPSs contain 6 and 7 modules allowing the production of compounds displaying no resemblance with already known peptides but the clusters are highly suspected to produce a siderophore synthetase because of the presence of genes coding for a TonB-dependent siderophore receptor and an MbtH domain-containing protein (Lautru et al. 2007).

As found in some *Pseudomonas* strains, a pyochelin synthesis cluster has been identified in the genome of strains also producing ornibactin or malleobactin. The cosynthesis of more than one siderophore by bacteria is not well understood, but it has been suggested that beside their ability to provide access to iron, pyochelin, whose affinity for iron is much lower (Cox et al. 1981), may have alternative roles including antibacterial activity (Cornelis and Matthijs 2002; Adler et al. 2012). The pyochelin production may also be associated with the virulence of bacteria (Sokol and Woods 1988). *Burkholderia ambifaria* AMMD does not produce pyochelin, but we have found a cluster of eight genes, located on chromosome 1, directing a siderophore synthesis. The identified cluster includes a gene for a TonB-dependent receptor, genes encoding enzymes necessary to build up a diOH-bz monomer and two *nrps* genes, the predicted monomer for the first one being diOh-bz. All these elements evidenced for the relationship between the biosynthesis cluster and the cepaciachelin known to be produced *B. ambifaria* strain PHP7 (LMG 11351) (Barelmann et al. 1996). Indeed, although the structure of cepaciachelin has been known for a long time (Barelmann et al. 1996), the mechanism of its synthesis was not yet reported. This relationship between genes and the product was supported by MS detection, performed on culture supernatants of the strain grown under iron-limited conditions allowing the production of siderophores (Fig. S1). As the cepciachelin is a nonribosomal peptide, the siderophore has been introduced into the Norine database, the id NOR01254 was attributed.

CLPs are versatile molecules composed of a fatty acid tail linked to a cyclic oligopeptide head, considered as efficient weapons for plant disease biocontrol (Ongena and Jacques 2008). Nearly all CLPs characterized to date have been identified from *Pseudomonas* and *Bacillus* species (Roongsawang et al. 2011). This is comforted with the Norine querying indicating that 139 peptides belonging to 15 families are lipopeptides with surfactant activity. Among them, 76 are produced by *Bacillus* strains and 46 by *Pseudomonas* strains and only 17 are produced by *Serratia*, *Marinobacter*, and *Halomonas*. With the genome mining of *Burkholderia* strains, we have found clusters of genes coding for NRPSs starting with Cstarter and ending with a tandem of Te domains, sharing these features with NRPS of *Pseudomonas* CLPs (De Bruijn et al. 2007). Furthermore, those NRPSs were lacking E domains, and some of the C domains were identified as dual C/E domains. These dual C/E domains were described for the first time in the NRPS responsible for arthrofactin production in *Pseudomonas* (Balibar et al. 2005). Since then, they have been detected in many NRPSs catalyzing the synthesis of CLPs in *Pseudomonas* (Roongsawang et al. 2011) and more recently in *Xanthomonas* (Royer et al. 2013). In all these cases, two types of NRPS clusters coexist in the same strain. The first one contains E domains

followed by $^{D}C_{L}$, generally directing the synthesis of a siderophore (i.e., pyoverdin), while the second type of NRPS containing dual C/E domains is involved in CLP synthesis. Strains can even synthesize several CLPs, with dual C/E domains for all the NRPSs (Pauwelyn et al. 2013; D'aes et al. 2014). Thus, *B. pseudomallei* strains can produce malleobactin where an Asp residue is in D-configuration due to an E domain, and burkhomycin CLP, where five monomers can be epimerized by dual C/E domains. It is quite different with *B. glumae*, *B. gladioli*, and *B. rhizoxinica* because no NRP harboring siderophore activity has been identified. The co-occurrence of two distinct types of NRPs (with E domains vs. with dual C/E domains) in strains sharing the same soil habitat has to be further studied in terms of evolution, also considering the favored horizontal exchanges during the colonization of same ecological niches.

A total of 13 NRPS gene clusters were found in the plasmids (*Burkholderia* sp. I23, *B.* sp. RPE64, *B. glumae* BGR1, *B.gladioli* BSR3, and *B. rhizoxinica* HKI 454). Up to now, this singular feature had only been described in *B. rhizoxinica* (Lackner et al. 2011) and may be specific to the genus *Burkholderia*.

## Acknowledgments

The work was supported by the University of Lille 1, the INTERREG IV program France-Wallonie-Vlaanderen (Phytobio project), the bioinformatics platform Bilille, and Inria. Q. E. received financial support from Sana'a University (Yemen).

## Conflict of Interest

None declared.

### References

Abe, M., and T. Nakazawa. 1994. Characterization of haemolytic and antifungal substance, cepalycin, from *Pseudomonas cepacia*. Microbiol. Immunol. 38:1–9.

Adler, C., N. S. Corbalán, M. R. Seyedsayamdost, M. F. Pomares, R. E. de Cristóbal, J. Clardy, et al. 2012. Catecholate siderophores protect bacteria from pyochelin toxicity. PLoS ONE 7:e46754.

Anand, S., M. V. R. Prasad, G. Yadav, N. Kumar, J. Shehara, M. Z. Ansari, et al. 2010. SBSPKS: structure based sequence analysis of polyketide synthases. Nucl. Acids Res. 38:W487–W496.

Bachmann, B. O., and J. Ravel. 2009. In silico prediction of microbial secondary metabolic pathways from DNA sequence data. Meth. Enzymol. 458:181–217.

Balibar, C., F. Vaillancourt, and C. Walsh. 2005. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. Chem. Biol. 12:1189–1200.

Barelmann, I., J. M. Meyer, K. Taraz, and H. Budzikiewicz. 1996. Cepaciachelin, a new catecholate siderophore from *Burkholderia* (*Pseudomonas*) *cepacia*. Z. Naturforsch. 51C:627–630.

Caballero-Mellado, J., J. Onofre-Lemus, P. Estrada-de Los Santos, and L. Martínez-Aguilar. 2007. The tomato rhizosphere, an environment rich in nitrogen-fixing *Burkholderia* species with capabilities of interest for agriculture and bioremediation. Appl. Environ. Microbiol. 73:5308–5319.

Caboche, S., M. Pupin, V. Leclère, A. Fontaine, P. Jacques, and G. Kucherov. 2008. NORINE: a database of nonribosomal peptides. Nucl. Acids Res. 36:D326–D331.

Caboche, S., M. Pupin, V. Leclère, P. Jacques, and G. Kucherov. 2009. Structural pattern matching of nonribosomal peptides. BMC Struct. Biol. 9:15.

Caradec, T., M. Pupin, A. Vanvlassenbroeck, M.-D. Devignes, M. Smaïl-Tabbone, P. Jacques, et al. 2014. Prediction of monomer isomery in Florine: a workflow dedicated to nonribosomal peptide discovery. PLoS ONE 9:e85667.

Cheng, A. C., and B. J. Currie. 2005. Melioidosis: epidemiology, pathophysiology, and management. Clin. Microbiol. Rev. 18:383–416.

Coenye, T., P. Vandamme, J. R. W. Govan, and J. J. LiPuma. 2001. Taxonomy and identification of the *Burkholderia cepacia* complex. J. Clin. Microbiol. 39:3427–3436.

Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, et al. 2009. The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucl. Acids Res. 37:D141–D145.

Cornelis, P., and S. Matthijs. 2002. Diversity of siderophore-mediated iron uptake systems in fluorescent pseudomonads: not only pyoverdines. Environ. Microbiol. 4:787–798.

Cox, C. D., K. L. Jr Rinehart, M. L. Moore, and J. C. Jr Cook. 1981. Pyochelin: novel structure of an iron-chelating growth promoter for *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA 78:4256–4260.

D'aes, J., N. P. Kieu, V. Leclère, C. Tokarski, F. E. Olorunleke, K. De Maeyer, et al. 2014. To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain *Pseudomonas* CMR12a. Environ. Microbiol. 16: 2282–2300.

De Bruijn, I., M. De Kock, M. Yang, P. De Waard, T. Van Beek, and J. Raaijmakers. 2007. Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species. Mol. Microbiol. 63:417–428.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Franke, J., K. Ishida, and C. Hertweck. 2014. Evolution of siderophore pathways in human pathogenic bacteria. J. Am. Chem. Soc. 136:5599–5602.

Franke, J., K. Ishida, and C. Hertweck. 2015. Plasticity of the malleobactin pathway and its impact on siderophore action in human pathogenic bacteria. Chem. Eur. J. 21:8010–8014.

Gasser, V., L. Guillon, O. Cunrath, and I. J. Schalk. 2015. Cellular organization of siderophore biosynthesis in *Pseudomonas aeruginosa*: evidence for siderosomes. J. Inorg. Chem. 148:27–34.

Gu, G., L. Smith, N. Wang, H. Wang, and S. E. Lu. 2009. Biosynthesis of an antifungal oligopeptide in *Burkholderia contaminans* strain MS14. Biochem. Biophys. Res. Commun. 380:328–332.

Hammer, P. E., W. Burd, D. S. Hill, J. M. Ligon, and K. H. van Pée. 1999. Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains. FEMS Microbiol. Lett. 180:39–44.

Lackner, G., N. Moebius, L. P. Partida-Martinez, S. Boland, and C. Hertweck. 2011. Evolution of an endofungal lifestyle: deductions from the *Burkholderia rhizoxinica* genome. BMC Genom. 12:210.

Landy, M., G. H. Warren, S. B. Rosenman, and L. G. Colio. 1948. Bacillomycin – an antibiotic from *Bacillus subtilis* active against pathogenic fungi. Proc. Soc. Biol. Med. 67:539–541.

Lautru, S., D. Oves-Costales, J. L. Pernodet, and G. L. Challis. 2007. MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. Microbiology 153:1405–1412.

Leclère, V., R. Marti, M. Béchet, P. Fickers, and P. Jacques. 2006. The lipopeptides mycosubtilin and surfactin enhance spreading of *Bacillus subtilis* strains by their surface-active properties. Arch. Microbiol. 186:475–483.

Lim, Y., S. J. Suh, S. Kim, B. Hyun, C. Kim, and C. H. Lee. 1994. Cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. II. Physico-chemical properties and structure lucidation. J. Antibiot. 47:1406–1416.

Lin, Z., J. O. Falkinham, K. A. Tawfik, P. Jeffs, B. Bray, G. Dubay, et al. 2012. Burkholdines from *Burkholderia ambifaria*: antifungal agents and possible virulence factors. J. Nat. Prod. 75:1518–1523.

Mahenthiralingam, E., and J. B. Goldberg. 2005. The multifarious, multireplicon *Burkholderia cepacia* complex. Nat. Rev. Microbiol. 3:144–156.

Mahenthiralingam, E., A. Baldwin, and C. G. Dowson. 2008. *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. J. Appl. Microbiol. 104:1539–1551.

Marahiel, M. A. 2009. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. J. Pept. Sci. 15:799–807.

Michel, L., A. Bachelard, and C. Reimmann. 2007. Ferripyochelin uptake genes are involved in pyochelin-mediated signalling in *Pseudomonas aeruginosa*. Microbiology 153:1508–1518.

Nierman, W. C., D. DeShazer, H. S. Kim, H. Tettelin, K. E. Nelson, T. Feldblyum, et al. 2004. Structural flexibility in the *Burkholderia mallei* genome. Proc. Natl. Acad. Sci. USA 101:14246–14251.

Ongena, M., and P. Jacques. 2008. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol. 16:115–125.

Pauwelyn, E., C. J. Huang, M. Ongena, V. Leclère, P. Jacques, P. Bleyaert, et al. 2013. New linear lipopeptides produced by *Pseudomonas cichorii* SF1-54 are involved in virulence, swarming motility, and biofilm formation. Mol. Plant Microbe Interac. 26:585–598.

Payne, S. M. 1994. Detection, isolation, and characterization of siderophores. Methods Enzymol. 235:329–344.

Roongsawang, N., K. Washio, and M. Morikawa. 2011. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int. J. Mol. Sci. 12:141–172.

Röttig, M., M. H. Medema, K. Blin, T. Weber, C. Rausch, and O. Kohlbacher. 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucl. Acids Res. 39:W362–W367.

Royer, M., R. Koebnik, M. Marguerettaz, V. Barbe, G. Robin, C. Brin, et al. 2013. Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. BMC Genom. 14:658.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Schwyn, B., and J. B. Neilands. 1987. Universal chemical assay for the detection and determination of siderophores. Anal. Biochem. 160:47–56.

Shanks, R. M. Q., N. C. Caiazza, S. M. Hinsa, C. M. Toutain, and G. A. O'Toole. 2006. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from Gram-negative bacteria. Appl. Environ. Microbiol. 72:5027–5036.

Sharma, A., and B. N. Johri. 2003. Growth promoting influence of siderophore-producing *Pseudomonas* strains GRP3A and PRS9 in maize (Zea mays L.) under iron limiting conditions. Microbiol. Res. 158:243–248.

Sokol, P. A., and D. E. Woods. 1988. Effect of pyochelin on *Pseudomonas cepacia* respiratory infections. Microb. Pathog. 5:197–205.

Tamura, K., M. Nei, and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. USA 101:11030–11035.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729.

Tawfik, K. A., P. Jeffs, B. Bray, G. Dubay, J. O. Falkinham, M. Mesbah, et al. 2010. Burkholdines 1097 and 1229, potent antifungal peptides from *Burkholderia ambifaria* 2.2N. Org. Lett. 12:664–666.

Vial, L., M. Groleau, V. Dekimpe, and E. Deziel. 2007. *Burkholderia* diversity and versatility: an inventory of the extracellular products. J. Microbiol. Biotechnol. 17:1407–1429.

Weber, T., K. Blin, S. Duddela, D. Krug, H.U. Kim, R. Bruccoleri, et al. 2015. antiSMASH 3.0 — a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucl Acids Res. 43:W237–W243.

Winsor, G. L., B. Khaira, T. Van Rossum, R. Lo, M. D. Whiteside, and F. S. L. Brinkman. 2008. The *Burkholderia* genome database: facilitating flexible queries and comparative analyses. Bioinformatics 24:2803–2804.

## Supporting Information

Additional supporting information may be found in the online version of this article:

**Figure S1.** Detection of siderophores by CAS assays and MALDI-ToF. Left: *Burkholderia phymatum* STM85, right: *Burkholderia ambifaria* AMMD. (A) CAS assays. The color change from blue to orange indicates the presence of iron-chelating compounds. (B) MALDI-ToF spectra. Peaks corresponding to cepaciachelin and ornibactin are pointed out by arrows.

**Figure S2.** Detection of lipopeptides produced by *Burkholderia rhizoxinica* HKI 454. The strain was grown in Landy medium. (A) Surface tension, (B) MALDI-ToF. A group of peaks were observed, characteristic of lipopeptide detection with different fatty acid chain lengths, with mass differences of 14 Da (1522 and 1536).

**Figure S3.** Biological activities of *Burkholderia ambifaria* AMMD. (A) antifungal activity. *Burkholderia ambifaria* AMMD was inoculated as a line in front of the fungal target on PDA medium. The fungi are mentioned under the plates. Plates were photographed 5 days after inoculation. (B) Antiyeast activity. The yeasts are mentioned under the plates. 1: AMMD wild type, 2: ΔBamb_6472 mutant, 3: sterile medium, 4: ethanol 70%. (C) Hemolytic activity on horse bood (5%) LB medium, 50 $\mu$L of supernatant obtained after 48 h of growth were used to fill the wells. 1: AMMD wild type, 2: ΔBamb_6472 mutant.

**Table S1.** Strains, plasmids, and primers used in this study.

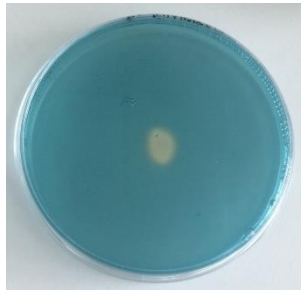
**Table S2.** Ornibactin, malleobactin, and phymabactin gene clusters. "+" and "−" mention the relative orientation of the genes. Numbers represent the size of the proteins (in AA). Colors represent the role of the protein in the siderophore production: green for NRPS biosynthesis, orange for accessory enzymes, blue for uptake, red for regulation, and yellow for unknown function.

**Table S3.** NRPSs of *Burkholderia rhizoxinica* HKI 454. A, adenylation domain; $^{L}C_{L}$, condensation between two L-monomers; $^{D}C_{L}$, condensation between D-monomer and L-monomer; C/E, dual condensation domain catalyzing both epimerization and condensation; T, thiolation domain; E, epimerization domain; Te, Thioesterase domain; "?", nonidentified domain; LP, lipopeptide.

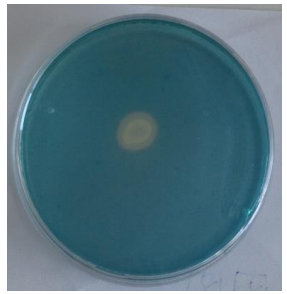
**Protocol S1.** Construction of a *B. ambifaria* AMMD mutant deleted in burkholdin (Bk) synthesis. The supplementary file describes the construction of a mutant strain deleted in the gene *bks*G (Bamb_6472) to support the relationship between the cluster *bks* and the production of an antifungal compound belonging to Bk family.

# Figure S1

A

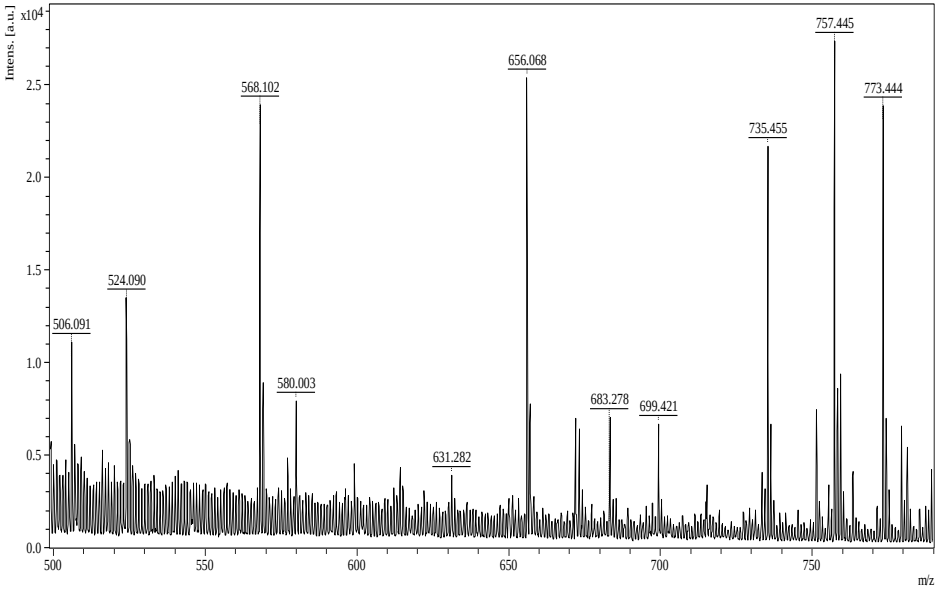


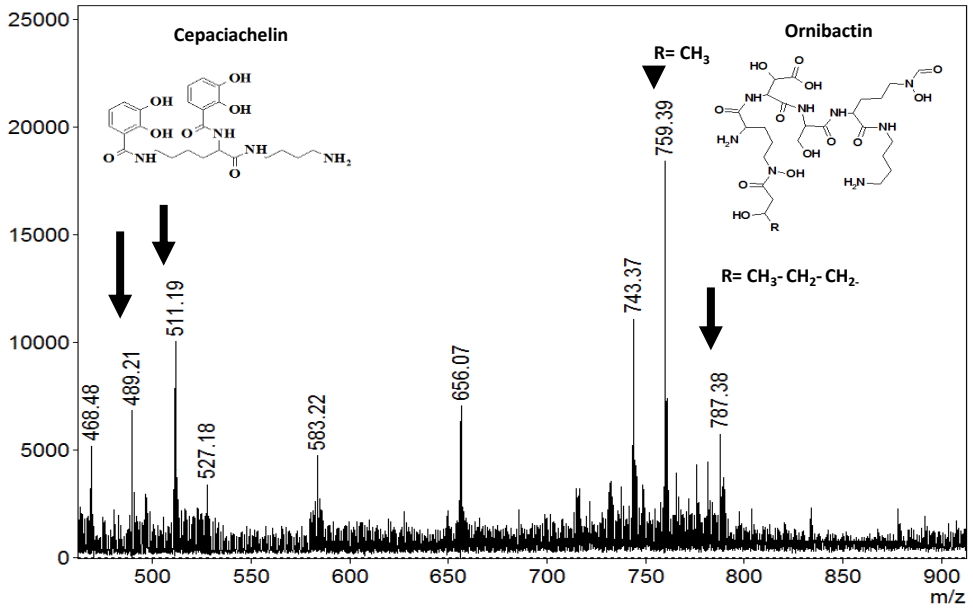
*B. phymatum* STM85



*B. ambifaria* AMMD

B

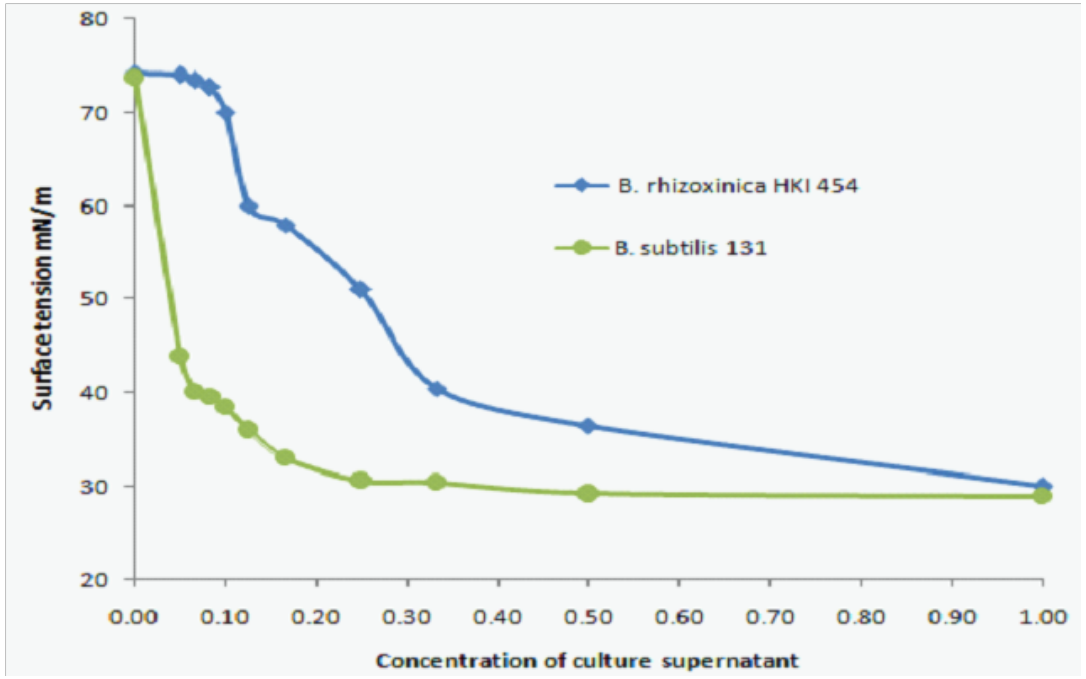


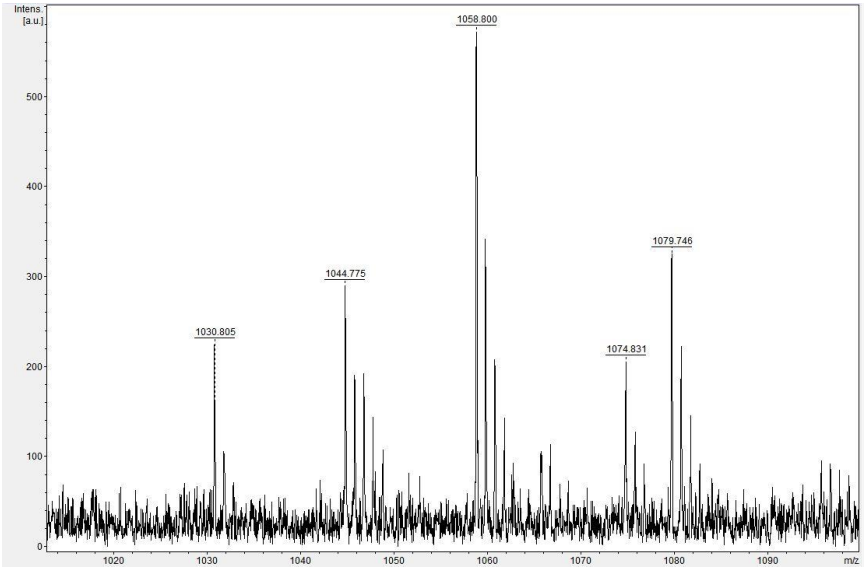
# Figure S2

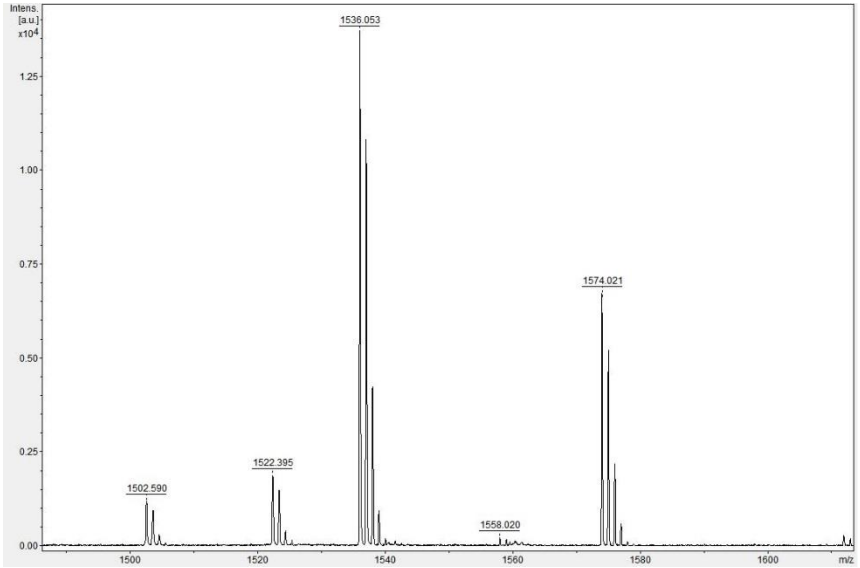


A


Surface tension mN/m
80
70
60
50
40
30
20
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Concentration of culture supernatant
B. rhizoxinica HKI 454
B. subtilis 131

B

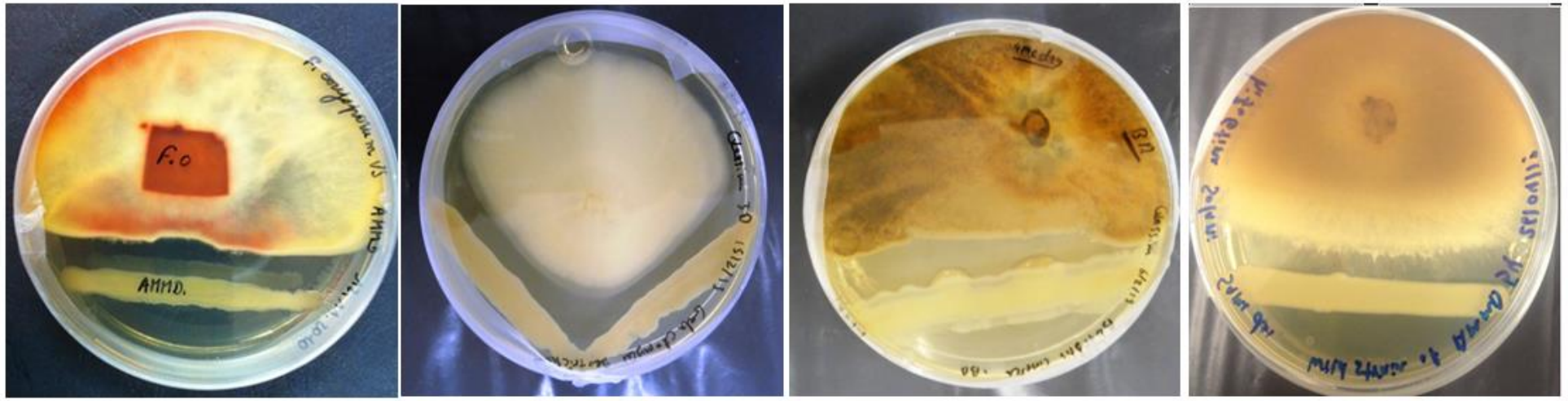

Intens. [a.u.]
1058.800
1079.746
1044.775
1030.805
1074.831
m/z


Intens. [a.u.] x10^4
1536.053
1574.021
1522.395
1502.590
1558.020
m/z

# Figure S3

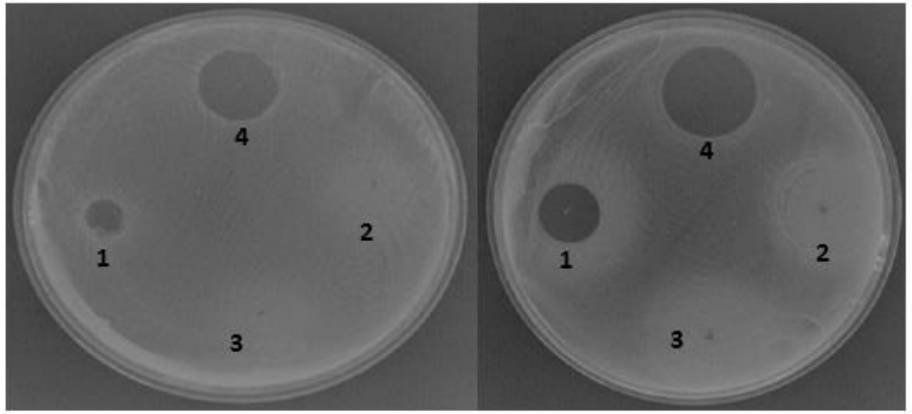


A




*Fusarium oxysporum* *Galactomyces geotrichum* *Botrytis cinerea* *Rhizoctonia solani*

B




*Saccharomyces cerevisiae* *Candida albicans*

C

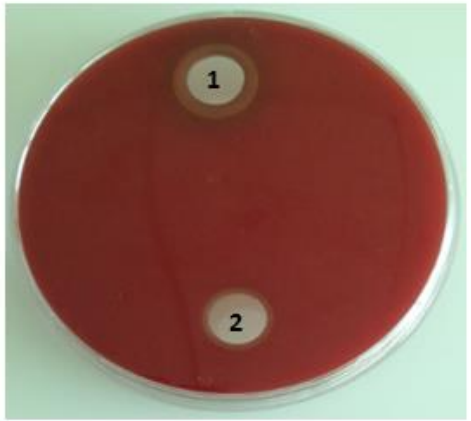


Table S1

| **Strains, plasmids, and primers** | **Description** | **Reference or source** |
|---|---|---|
| **Strains** | | |
| *B. ambifaria* AMMD LMG 19182 | Wild type; $Gm^S$ | LMG |
| *B.ambifaria* AMMD-Δbamb_6472 | Mutant with deletion of NRPS gene bamb_6472 | This study |
| *B. phymatum* STM815 | | DSMZ 17167 |
| *B. rhizoxinica* HKI 454 | | DSMZ 19002 |
| *Escherichia coli WM3064* | Strain for conjugation; λ pir, DAP auxotroph | Saltikov and Newman, 2003 |
| *Escherichia coli* DH5α | Host of cloning | ProBioGEM lab stock |
| *Listeria innocua 51742* | Target for antibacterial activity | ATCC |
| *Micrococcus luteus* | Target for antibacterial activity | ProBioGEM lab stock |
| *Candida albicans* ATCC10231 | Target yeast strain for fungicide testing | ProBioGEM lab stock |
| *Saccharomyces cerevisiae* | Wild type | ProBioGEM lab stock |
| *Saccharomyces cerevisiae* InvSc1 | Yeast strain for *in vivo* recombination *MAT* a/*MAT* α *leu* 2/*leu* 2 *trp* 1-289/*trp* 1-289 *ura* 3-52/*ura* 3-52 *his* 3-Δ1/*his* 3-Δ1 | Invitrogen |
| *Botrytis cinerea* R16 | Phytopathogen | ProBioGEM lab stock |
| *Fusarium oxysporum* | Phytopathogen | ProBioGEM lab stock |
| *Galactomyces geotrichum* MUCL2859 | Phytopathogen | ProBioGEM lab stock |
| *Rhizoctonia solani* S010-1 | Phytopathogen | ProBioGEM lab stock |
| **Plasmids** | | |
| pMQ30 | 7.6 kb mobilizable suicide vector used for gene replacement in *pseudomonas* : SacB, URA3, $Gm^r$ | Shanks *et al* ., 2006 |
| pMQ30Δ6472 | pMQ30 containing two fragments of 1kb of the NRPS biosynthesis gene bamb_6472 | This study |
| **Primers (5'-----> 3')** | | |
| Up6472-F | *GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTG* CGATCCAGTACCGCGACTAC | This study |
| Up6472-R | *TCGGAAGGGAATAGGTCAGC* TGATCGGTGACCAGTACGTT | This study |
| Down6472-F | *AACGTACTGGTCACCGATCA* GCTGACCTATTCCCTTCCGA | This study |
| Down6472-R | *CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGAT* TCCTTTTGCACAGGTTGACG | This study |

Table S2

| siderophore | strain | Extracytoplasmic σ70 (regulation) | mbtH | SyrP like | ABC transporter | transporter permease | ferric iron reductase | periplasmic binding protein | Hypothetical Protein | cyclic peptide transporter | NRPS | NRPS | Hypothetical Protein | Ornithin monooxygenase | TonB dpdt receptor | Folate formyltransferase | ornibactin biosynthesis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ***mba* F** | ***mba* G** | ***mba* H** | ***mba* I** | ***mba* J** | ***mba* K** | ***mba* L** | ***mba* M** | ***mba* N** | ***mba* A** | ***mba* B** | | ***mba* C** | ***mba* D** | ***mba* E** | |
| malleobactin | mallei ATCC 23344 | +207 | +80 | +338 | +296 | +706 | +268 | +342 | -255 | -562 | 0 | +1732 | 0 | +468 | +741 | ? | 0 |
| malleobactin | mallei NCTC 10229 | +237 | +80 | +333 | +296 | +706 | +268 | +398 | -255 | -562 | +3294 | +1732 | 0 | +517 | +752 | ? | 0 |
| malleobactin | mallei NCTC 10247 | +237 | +80 | +333 | +296 | +706 | +268 | +398 | -255 | -562 | +3294 | +1732 | 0 | +468 | +736 | ? | 0 |
| malleobactin | mallei SAVP1 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191 | -562 | +3297 | +1732 | 0 | +468 | +736 | ? | 0 |
| malleobactin | phytofirmans PsJN | +223 | +79 | +344 | +288 | +702 | +272 | +342 | -95 | -588 | +3230 | +1674 | 0 | +456 | +745 | +280 | 0 |
| malleobactin | pseudomallei 1026b | +257 | +80 | +338 | +296 | +706 | +264 | +342 | -191 | -581 | +3287 | +1732 | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei 1106a | +237 | +80 | +333 | +296 | +706 | +268 | +398 | +250 | -562 | +3290 | +1741 | 0 | +468 | +753 | +279 | 0 |
| malleobactin | pseudomallei 1106B | +237 | +80 | +338 | +296 | +706 | +264 | +385 | -191(+123) | -562 | +3290 | +1741 | 0 | +468 | +753 | ? | 0 |
| malleobactin | pseudomallei 1710a | +237 | +82 | +338 | +296 | +706 | +264 | +342 | -220(+59) | -562 | +1480+1908 | +1739 | 0 | +458 | +753 | ? | 0 |
| malleobactin | pseudomallei 1710b | +237 | +82 | + 338 | + 296 | + 706 | + 268 | +429 | - 267 | -562 | +3287(+3293) | +1739(+1772) | 0 | +495 | +737 | +279 | 0 |
| malleobactin | pseudomallei 668 | +237 | +82 | +338 | +296 | +706 | +264 | +398 | -253 | -562 | +3290 | +1745 | 0 | +468 (MbaA) | +753 | +279 | 0 |
| malleobactin | pseudomallei BPC006 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191(+129) | -581 | +3291 | +1732 | 0 | +517 | +737 | ? | 0 |
| malleobactin | pseudomallei K96243 | +207 | +80 | +338 | +296 | +706 | +268 | +382 | -191 | -581 | +3290 | +1748 | 0 | +468 | +753 | +279 | 0 |
| malleobactin | pseudomallei MSHR146 | +237 | +80 | +336 | +296 | +706 | +268 | +398 | -191 | -562 | +3288 | 0 | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei MSHR305 | +237 | +80 | +340 | +296 | +706 | +268 | +398 | -191 | -576 | +3292 | +1736 | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei MSHR346 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191 | -562(+59) | +3300 | +1739 | 0 | +517 | +737 | +279 | 0 |
| malleobactin | pseudomallei MSHR511 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191 | -562 | +3288 | +1739 | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei MSHR520 | +237 | +80 | +340 | +296 | +706 | +268 | +398 | -191 | -562 | +3293,6 (pseudo) | +1737,3 (pseudo) | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei NAU20B-16 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191 | -562 | +1828 | +1739 | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei NCTC 13178 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191 | -562 | +3295 | +1739 | 0 | +468 | +737 | +279 | 0 |
| malleobactin | pseudomallei NCTC 13179 | +237 | +80 | +338 | +296 | +706 | +268 | +398 | -191 | -562 | +3300 | +1738 | 0 | +468 | +736 | +279 | 0 |
| malleobactin | thailandensis E264-1 | +237 | +82 | +363 | +302 | +659 | +264 | +419 | -191 | -562 | +3296 | +1772 | 0 | +504 | +738 | +248 | 0 |
| malleobactin | thailandensis E444 | +237 | +82 | +338 | +302 | +706 | +264 | +397 | -191 | -562 | +3296 | +1763 | 0 | +478 | +738 | +279 | 0 |
| malleobactin | thailandensis H0587 | +237 | +82 | +338 | +305 | +706 | +264 | +397 | -191 | -562 | +3305 | +1763 | 0 | +478 | +738 | +279 | 0 |
| malleobactin | thailandensis MSMB121 | +230 | +80 | +338 | +306 | +710 | +268 | +354 | 0 | -562 | +3326 | +1702 | 0 | +473 | +739 | +279 | 0 |
| malleobactin | xenovorans LB400 | +223 | +105 | +340 | +288 | +704 | +272 | +276 | 0 | -562 | +3180 | +1675 | 0 | +456 | -727 | +280 | 0 |
| | | | | | | | | | | | ***phm* A** | ***phm* B** | | | | | |
| phymabactin | phymatum STM815 | +244 | +77 | 0 | +281 | +708 | +264 | +341 | 0 | -597 | +3224 | +1659 | +341 | +462 | 0 | 0 | 0 |
| | | ***orb* S** | ***orb* H** | ***orb* G** | ***orb* C** | ***orb* D** | ***orb* F** | **orbB** | | ***orb* E** | ***orb* I** | ***orb* J** | ***orb* K** | ***pvdA*** | ***orb* A** | ***pvd* F** | ***orb* L** |
| ornibactin | ambifaria AMMD | +222 | +80 | +339 | + 283 | + 697 | +266 | +345 | 0 | - 581 | + 3227 | + 1670 | - 341 | + 458 | + 750 | +279 | +338 |
| ornibactin | ambifaria MC40-6 | +222 | +80 | +339 | +283 | + 697 | +266 | +345 | 0 | - 581 | + 3227 | + 1669 | + 341 | + 458 | + 753 | +279 | +338 |
| ornibactin | cenocepia AU 1054 | +222 | +80 | +335 | +282 | +696 | +266 | +341 | 0 | -607 | +3231 | +1662 | +343 | 458 | +754 | +279 | +338 |
| ornibactin | cenocepia HI2424 | +222 | 80 | +335 | +282 | +696 | +266 | +341 | 0 | -607 | +3231 | +1662 | +343 | +458 | +754 | +279 | +338 |
| ornibactin | cenocepia J2315 | +222 | +80 | +335 | +268 | +696 | +266 | +297 | 0 | -581 | +3222 | +1669 | +345 | +458 | +755 | +279 | +340 |
| ornibactin | cenocepia MC0-3 | +222 | +80 | +335 | +282 | +696 | +266 | +341 | 0 | -607 | +3221 | +1665 | +343 | +458 | +754 | +279 | +338 |
| ornibactin | cepacia GC4 | +222 | +80 | +339 | +290 | +697 | +266 | +344 | 0 | -581 | +3232 | +1661 | +341 | +458 | +765 | +279 | +338 |
| ornibactin | lata ASM1294v1 | +218 | +80 | +335 | +282 | +696 | +267 | +345 | 0 | -581 | +3219 | +1663 | +344 | +458 | +750 | +279 | +338 |
| ornibactin | multivorans ATCC 17616-1 | +215 | +80 | +336 | +280 | +696 | +266 | +344 | 0 | -581 | +3219 | +1658 | +338 | +454 | +734 | +279 | +336 |
| ornibactin | multivorans ATCC 17616-2 | +213 | +80 | +336 | +280 | +696 | +266 | +344 | 0 | -581 | +3219 | +1658 | +338 | +454 | +739 | +279 | +336 |
| ornibactin | sp. KJ006 | +221 | +80 | 0 | 283 | +703 | +269 | +349 | 0 | -579 | +3259 | +1661 | 0 | +437 | +757 | +279 | +336 |
| ornibactin | vietnamiensis G4 | +221 | +80 | +339 | +283 | +703 | +269 | +349 | 0 | -579 | +3242 | +1661 | 0 | +459 | +773 | +279 | +336 |
| ornibactin | sp. RPE64 | 221 | 99 | +334 | +278 | +699 | +259 | +283 | +43? | -583 | +3181 | +1655 | 0 | +454 | +750 | +281 | +319 |
| ornibactin | sp. YI23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -582 | +3167 | 1651 | 0 | +453 | +728 | +277 | +303 |

NRPS biosynthesis
Accessory enzymes
Uptake
Regulation
Unknown function

Table S3

| Cluster | putative product | gene | genetic support | NCBI annotation | strand | protein length (AA) | Protein-ID | domain architecture |
|---|---|---|---|---|---|---|---|---|
| 1 | LP7 | RBRH_01504 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 7658 | YP_004021982.1 | $C_{start}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-C/E-A-T-C/E-A-T-$^{L}C_{L}$-A-T-C/E-A-T-Te |
| 2 | LP6 | RBRH_00429 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 6184 | YP_004022028.1 | A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-C/E-A-T-Te |
| 3 | LP3 | RBRH_00451 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 181 | YP_004022050 | $C_{starter}$ |
| | | RBRH_00452 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 2773 | YP_004022052.1 | A-T-$^{L}C_{L}$-A-T-C/E-A-T-Te |
| 4 | LP4 | RBRH_00484 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 4512 | YP_004022085.1 | $C_{starter}$-A-T-C/E-A-T-$^{L}C_{L}$-A-T-C/E-A-T-Te |
| 5 | LP3 | RBRH_00572 | plasmid pBRH01 | non-ribosomal peptide synthetase module | minus | 227 | YP_004022168.1 | T |
| | | RBRH_00574 | plasmid pBRH01 | peptide synthetase | minus | 799 | YP_004022169.1 | ?-A |
| | | RBRH_00575 | plasmid pBRH01 | Carveol dehydrogenase | minus | 274 | YP_004022170.1 | KR |
| | | RBRH_00576 | plasmid pBRH01 | thioesterase | minus | 237 | YP_004022171.1 | Te |
| | | RBRH_00578 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 2515 | YP_004022173.1 | $C_{starter}$-A-T-$^{L}C_{L}$-A-T-Te |
| 6 | LP3 | RBRH_00622 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 2505 | YP_004022223.1 | $C_{starter}$-A-T-C/E-A-T-$^{L}C_{L}$ |
| | | RBRH_04248 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 394 | YP_004022224.1 | A |
| | | RBRH_00623 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 402 | YP_004022225.1 | T |
| 7 | LP2 | RBRH_00274 | plasmid pBRH01 | non-ribosomal peptide synthetase module | plus | 2537 | YP_004022375.1 | ?-A-T-C/E-A-T-Te |
| 8 | LP3 | RBRH_00260 | plasmid pBRH01 | non-ribosomal peptide synthetase module | minus | 3431 | YP_004022385.1 | $C_{starter}$-A-T-$^{L}C_{L}$-A-T-C/E-A-T-Te |
| 9 | LP6 | RBRH_01792 | plasmid pBRH01 | non-ribosomal peptide synthetase module | minus | 6591 | YP_004022432.1 | $C_{starter}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-Te |
| 10 | LP8 | RBRH_04314 | plasmid pBRH01 | non-ribosomal peptide synthetase module | minus | 1407 | YP_004022559.1 | T-$^{L}C_{L}$-A-T-Te |
| | | RBRH_02787 | plasmid pBRH01 | non-ribosomal peptide synthetase module | minus | 4500 | YP_004022560.1 | A-T-$^{L}C_{L}$-A-M-T-$^{L}C_{L}$-A-T-C/E-A-M |
| | | RBRH_04279 | plasmid pBRH01 | non-ribosomal peptide synthetase module | minus | 4100 | YP_004022561.1 | $C_{starter}$-A-M-T-C/E-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$ |
| 11 | LP | RBRH_02642 | chromosome 1 | Non-ribosomal peptide synthetase modules | plus | 2133 | YP_004029261.1 | A-M-T-C/E |
| | | RBRH_04173 | chromosome 1 | Non-ribosomal peptide synthetase modules | plus | 872 | YP_004029262.1 | A-T-$^{L}C_{L}$ |
| 12 | LP8 | RBRH_03984 | chromosome 1 | Non-ribosomal peptide synthetase modules | minus | 3538 | YP_004029595.1 | T-$^{L}C_{L}$-A-T-C/E-A-T-$^{L}C_{L}$-A-T-Te |
| | | RBRH_04313 | chromosome 1 | Non-ribosomal peptide synthetase modules | minus | 2340 | YP_004029596.1 | T-$^{L}C_{L}$-A-T-C/E-A |
| | | RBRH_03867 | chromosome 1 | Non-ribosomal peptide synthetase modules | minus | 2865 | YP_004029597.1 | $C_{starter}$-A-T-$^{L}C_{L}$-A-T-C |
| | | RBRH_03865 | chromosome 1 | hypothetical protein | plus | 118 | YP_004029598.1 | T |
| | | RBRH_03864 | chromosome 1 | D-alanine-activating enzyme | plus | 534 | YP_004029599.1 | A |
| 13 | LP10 | RBRH_01506 | chromosome 1 | Non-ribosomal peptide synthetase modules | plus | 5791 | YP_004029720.1 | A-T-C/E-A-T-$^{L}C_{L}$-A-M-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T-Te |
| | | RBRH_04125 | chromosome 1 | Non-ribosomal peptide synthetase modules | minus | 1086 | YP_004029722.1 | A-T-Te |
| | | RBRH_00209 | chromosome 1 | Non-ribosomal peptide synthetase modules | minus | 4454 | YP_004029723.1 | $C_{starter}$-A-T-C/E-A-T-$^{L}C_{L}$-A-T-$^{L}C_{L}$-A-T |

**Protocol S1- Construction of a burkholdin deleted mutant of *B. ambifaria* AMMD**

To identify the cluster dedicated to antimicrobial and haemolytic activity, we constructed a burkholdins-deficient mutant, *B.ambifaria* AMMD-Δbamb_6472, by deleting a 9,530-bp fragment within the burkholdins–encoding gene *bamb_6472*.

**1- Construction of pMQ30-Δ6472**

The allelic replacement vector PMQ30 was used for mutant construction. Two regions of the NRPS gene *bamb_6472* were amplified using primers Up6472-F and Up6472-R (1,015 bp amplicon) and primers Down6472-F and Down6472-R (887 bp amplicon), respectively. These primers contained recombination sites for the plasmid. The PCR mix consisted of 25 μL of PCR Master Mix (Thermo Scientific Fermentas), 10 μL Q-solution (Qiagen), 1.25 μL of each primer (each at 20 μM), 7.5 μL water and 5 μL of genomic DNA. The reaction mixture was subject to the following thermal cycles: one cycle at 94°C for 3 min; 30 cycles (94°C, 30 s; 60°C, 45 s; 72°C, 2 min) and a final extension step at 72°C for 10 min. Plasmid pMQ30 was purified using GeneJet plasmid miniprep kit (Thermo Scientific Fermentas). A 5 μL sample was checked on a 1% agarose gel. Subsequently, the purified plasmid was digested by a mix of 9.5 μL milliQ water, 4 μL Tango yellow buffer (2 x), 5 μL plasmid pMQ30, 1 μL *Bam*HI and 0.5 μL *EcoR*I. The mix was incubated for 2 h at 37°C. 5 μL of the plasmid digest was checked on 1% agarose and the concentration of the digest was also measured.

**2- *In vivo* cloning in *S. cerevisiae* InvSc1**

The PCR products of *bamb_6472* were cloned flanking each other through *in vivo* homologous recombination in the yeast, *S. cerevisiae* InvSc1.,grown overnight in yeast peptone dextrose (YPD) at 30°C. Then, 0.5 mL culture was centrifuged for 1 min at 3,000 rpm. The cells pellet was washed with 0.5 mL lazy bones solution and the following was added to the mix: 20 μL of carrier DNA (2 mg/mL), 45 μL of each PCR products UP and DOWN, 5 μL digested pMQ30 plasmid. The mix was homogeneized for 1 min on a Vortex and incubated overnight at room temperature. The mixture was subjected to heat shock for 12 min at 42°C. Then, cells were centrifuged for 1 min at 3,000 rpm, washed with 0.6 mL of TE buffer and the resulting pellet was further redissolved in 0.6 mL of TE buffer, and the cells were plated on SD-uracil medium.

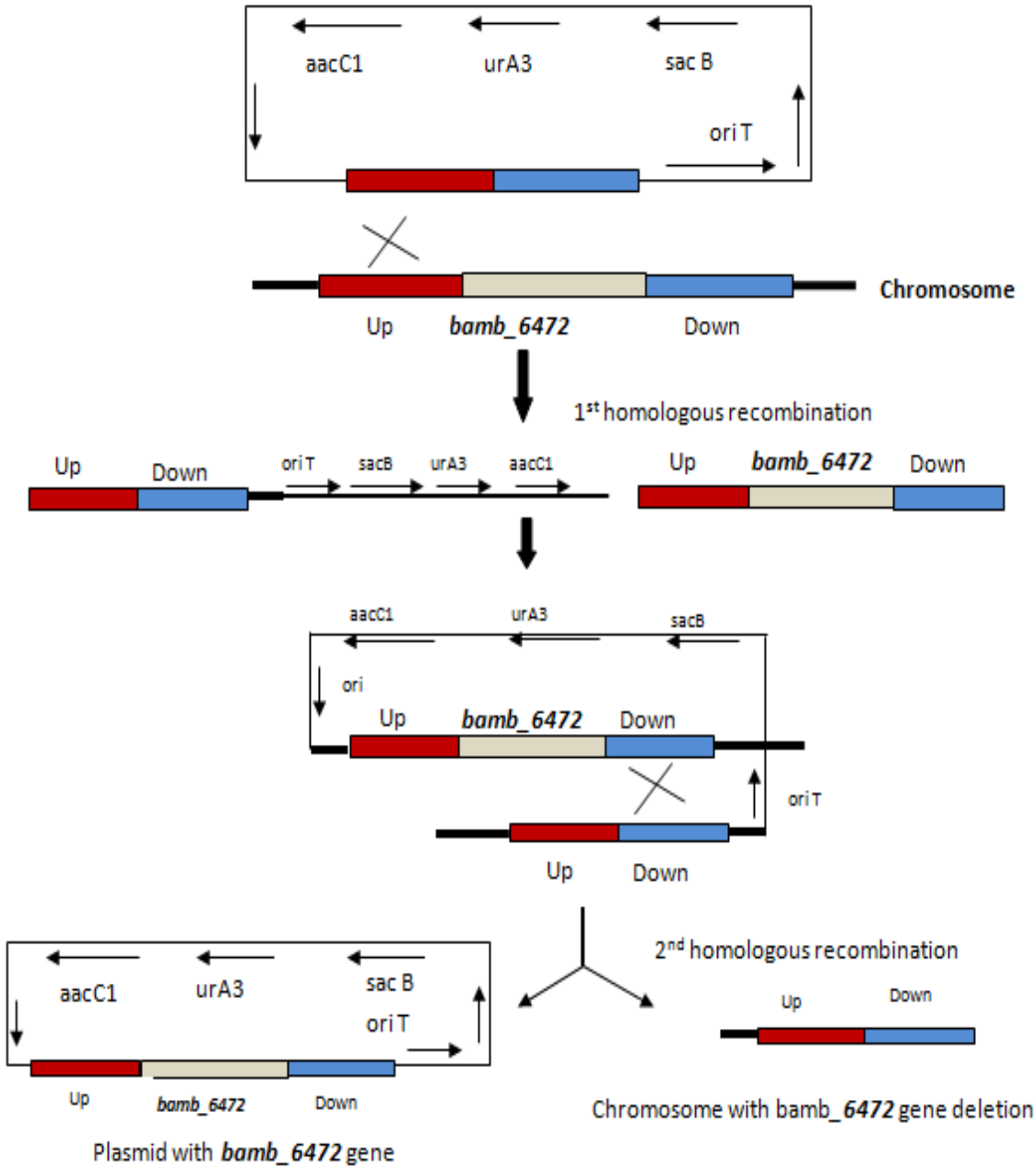
Further, plasmid was isolated using miniprep kit (Fermentas). 10 µL of the eluted plasmid was checked on 1% agarose gel.

**3- Construction of AMMD- deficient mutant**

Plasmid pMQ30-Δ6472 was introduced into *E. coli* WM3064 by electroporation. 2 µl of pMQ30 Δ6472 was added to 50 µl of thawed *E. coli* WM3064 competent cells. Mix was transferred to an electroporation cuvette and electroporated at 2.5 kV/200 ohms/25 µF. Electroporated cells were transferred into 1 ml of LB containing 2% of glucose and incubated in a rotary shaker for 1 h at 37°C. The cells were plated on LB containing 20 µg/ ml of gentamicin sulfate and 100mg/L of DAP (Diaminopimelic acid). Colony PCR was conducted using primers Up6472-F and Down6472-R to confirm the success of electroporation.

The plasmid was mobilized into *B. ambifaria* AMMD by conjugation. The donor strain *E. coli* WM3064+pMQ30-Δ6472 were grown in LB broth containing Gm 20 µg/ml and DAP 100mg/L. The acceptor strain, AMMD, was grown without antibiotics. The cells were centrifuged for 2 min at 5000 rpm, the supernatant was discarded and the cells were washed with 1. 5 ml LB. For conjugation, 200 µl of acceptor strain was mixed with 200 µL of donor strain. This mix was spotted on LB plates containing 100 mg/L DAP, dried, and then plates were incubated overnight at 37 °C. Cell mass were then collected and suspended in 1 ml LB. For each conjugate, dilutions up to 10-3 were performed, plated on LB containing gentamicin 300 µg/ml and incubated overnight at 37 °C. Colonies were streaked on new LB plates with gentamicin 300 µg/ml. Colony PCR was performed to confirm the presence of the plasmid in merodiploid strain (cell containing plasmid) using primers Up6472-F and Down6472-R. One colony was cultured on LB and incubated overnight at 37 °C in a rotary shaker. Culture was diluted up to 10-5 and 100 µl were plated on LB containing 10% sucrose. Selected colonies were streak out on new LB plates without gentamicin.

Figure S2-1: Knock-out mutant for burkholdins/ occidiofungin production in *B. ambifaria* AMMD




Colony PCR of these colonies was carried out using primers Up6472-F and Down6472-R to confirm and detect deleted mutants Fig. S5-2.

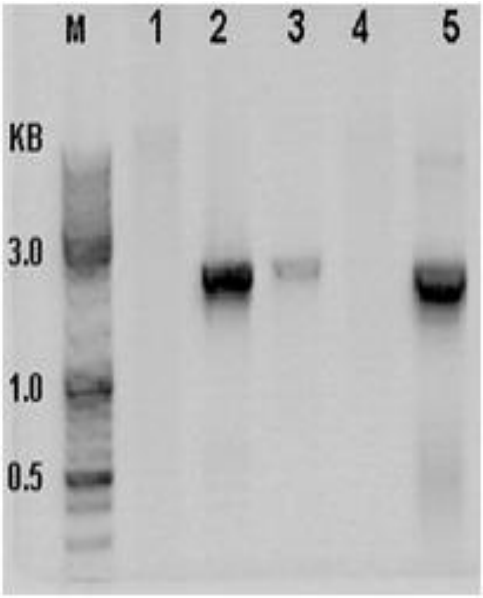
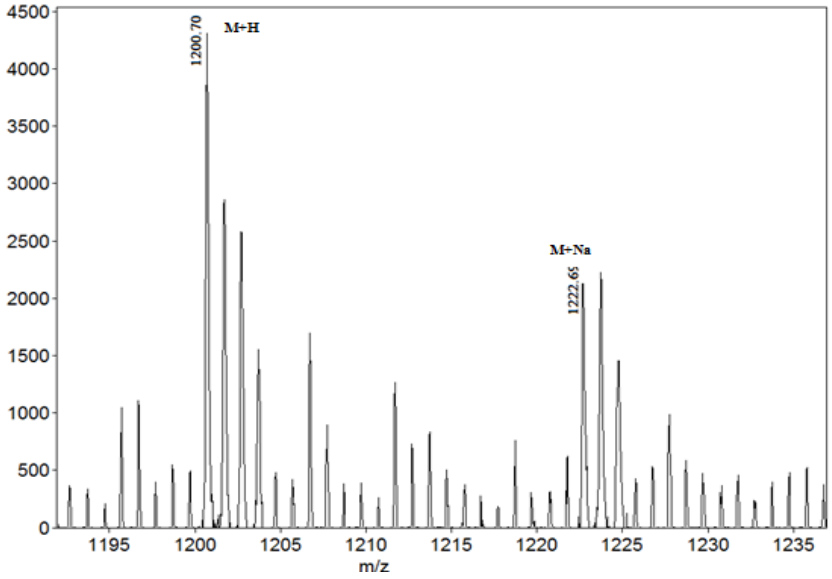



Fig. S2-2: Colony PCR analysis to confirm the deletion of bamb_6472 gene in the second recombination.

Lane M: 10 kb O'geneRuler (Thermo Fisher Fermentas)
Lane 1: wild-type showing no amplification (Fragment too long)
Lanes 2-3 : colonies of AMMDΔbamb_6472 allowing the corresponding amplification of 2000 bp
Lane 4 : Negative PCR
Lane 5 : pMQ30-Δbamb_6472 in recombination with bamb_6472 up and down.

Fig. S2-3 : MALDI-ToF detection of burkholdin produced by *B. ambifaria* AMMD




Shanks, R.M.Q., Caiazza, N.C., Hinsa, S.M., Toutain, C.M. and O'Toole, G.A. (2006) *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Appl. Environm. Microbiol*. **72**, 5027-5036.

Saltikov, C.W. and Newman, D.K. (2003) Genetic identification of a respiratory arsenate reductase. *Proc. Natl. Acad. Sci*. **16**, 10983-10988.

# Table 1: Overview

| phylogenic group | *Burkholderia* species | genome size (Mb) | Nb of chromosomes | Nb of plasmids | Lifestyle | Ornibactin | Malleobactin | pyochelin | Cepaciachelin | burkholdin | Burkhomycin (CLP12) | Cstart_Te (LP) | "630 - 2979" AT - CATECA1 | PKS-3100_1300_1500 | PKS-NRPS "6274" | NRPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | **Main siderophore** | | **Secondary siderophore** | | antifungal | **Lipopeptides** | | **Unknown products** | | | |
| mallei | *B. thailandensis* MSMB121 | 6.73 | 2 | 0 | Saprophytic, human pathogen (rare) | | **Xmo1** | | | | | | | ? | | |
| mallei | *B. thailandensis* H0587 | 6.77 | 2 | 0 | Saprophytic, human pathogen (rare) | | **Xmo1** | **Xmo2** | | | | | **Xmo1** | **Xmo2** | | |
| mallei | *B. thailandensis* E264 | 6.72 | 2 | 0 | Saprophytic, human pathogen (rare) | | **Xmo1** | **Xmo2** | | | | | **Xmo1** | **Xmo2** | | |
| mallei | *B. thailandensis* E444 | 6.65 | 2 | 0 | Saprophytic, human pathogen (rare) | | **Xmo1** | **Xmo2** | | | | | **Xmo1** | **Xmo2** | | |
| mallei | *B. thailandensis* 2002721723 | 6.58 | 2 | 0 | Saprophytic, human pathogen (rare) | | | **Xmo2** | | | | | **Xmo1** | **Xmo2** | | |
| mallei | *B. pseudomallei* NCTC 13179 | 7.34 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | | |
| mallei | *B. pseudomallei* MSHR305 | 7.43 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo1(11)** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* MSHR520 | 7.45 | 2 | 0 | Human pathogen | | | **Xmo2** | | | **Xmo2(9)** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* K96243 | 7.25 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* 1710a | 7.33 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* 668 | 7.04 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2(5)** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* 1710b | 7.31 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* NAU20B-16 | 7.31 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2(7)** | | **Xmo1** | **Xmo2** | | |
| mallei | *B. pseudomallei* 1106b | 7.21 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2 (S)** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* 1106a | 7.09 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* 1026b | 7.23 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* MSHR511 | 7.32 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2(10)** | | **Xmo1** | **Xmo2** | **Xmo1(S)** | |
| mallei | *B. pseudomallei* BPC006 | 7.16 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. pseudomallei* MSHR146 | 7.31 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2** | | **Xmo1** | **Xmo2** | **Xmo1(S)** | |
| mallei | *B. pseudomallei* NCTC 13178 | 7.39 | 2 | 0 | Human pathogen | | **Xmo1** | **Xmo2** | | | **Xmo2(10)** | | **Xmo1** | **Xmo2** | **Xmo1** | |
| mallei | *B. malle* i NCTC 10229 | 5.74 | 2 | 0 | Human and Animal pathogen | | **Xmo1** | | | | **Xmo2(4)** | | **Xmo1** | | | |
| mallei | *B. mallei* SAVP1 | 5.23 | 2 | 0 | Human and Animal pathogen | | **Xmo1** | | | | | | | | | |
| mallei | *B. mallei* NCTC 10247 | 5.85 | 2 | 0 | Human and Animal pathogen | | **Xmo1** | | | | | | | | | |
| mallei | *B. malle* i ATCC 23344 | 5.84 | 2 | 0 | Human and Animal pathogen | | **Xmo1** | | | | **Xmo2(4)** | | | | | |
| Bcc | *B. multivorans* ATCC 17616 | 7.01 | 3 | 1 | Human pathogen | **Xmo1** | | | | | | | | | | |
| Bcc | *B. multivorans* ATCC 17617 | 7.01 | 3 | 1 | Human pathogen | **Xmo2** | | | | | | | | | | |
| Bcc | *B. sp* . KJ006 | 6.63 | 3 | 1 | Biocontrol, mutualist | **Xmo1** | | | | **Xmo3** | | | | | | |
| Bcc | *B. cepacia* GG4 | 6.47 | 2 | 0 | Bioremediation, Human pathogen | **Xmo1** | | | | | | | | | | |
| Bcc | *B. cenocepacia* HI2424 | 7.70 | 3 | 1 | Human pathogen | **Xmo1** | | **Xmo2** | | | | | | | | |
| Bcc | *B. cenocepacia* AU 1054 | 7.28 | 3 | 0 | Human pathogen | **Xmo1** | | **Xmo2** | | | | | | | | |
| Bcc | *B. cenocepacia* J2315 | 8.06 | 3 | 1 | Human pathogen | **Xmo1** | | **Xmo2 (P)** | | | | | | | | |
| Bcc | *B. cenocepacia* MC0-3 | 7.97 | 3 | 0 | plant pathogen, human pathogen | **Xmo1** | | **Xmo2** | | | | | | | | |
| Bcc | *B. lata* ASM1294v1 | 8.68 | 3 | 0 | Human pathogen | **Xmo1** | | **Xmo2** | | | | | | | | |
| Bcc | *B. ambifaria* AMMD | 7.53 | 3 | 1 | Biocontrol | **Xmo1** | | | **Xmo1** | **Xmo3** | | | | | | |
| Bcc | *B. ambifaria* MC40-6 | 7.64 | 3 | 1 | Human pathogen | **Xmo1** | | | | | | | | | | |
| Bcc | *B. vietnamiensis* G4 | 8.39 | 3 | 5 | mutualist, human pathogen | **Xmo1** | | | | | | | | | | |
| others | *B. glumae* BGR1 | 7.28 | 2 | 4 | plant pathogen | | | | | | | **Xmo2 (LP7)** | | | | **Xmo2 (4a + 4b)** |
| others | *B. gladioli* BSR3 | 9.05 | 2 | 4 | plant pathogen | | | | | | | **Xmo2 (LP7 + LP5)** | | | **Xmo1** | **Xmo2 (4a + 4b)** |
| others | *B. xenovorans* LB400 | 9.73 | 3 | 0 | Mutualist | | **Xmo2** | | | | | | | | | |
| others | *B. sp.* CCGE1003 | 7.04 | 2 | 0 | Mutualist | ***Xmo1*** | | | | | | | | | | |
| others | *B. sp.* CCGE1001 | 6.83 | 2 | 0 | Mutualist | ***Xmo1*** | | | | | | | | | | |
| others | *B. rhizoxinica* HKI 454 | 3.75 | 1 | 2 | Mutualist | | | | | | | **(13 LPs)** | | | | |
| others | *B. sp.* CCGE1002 | 7.88 | 3 | 1 | Mutualist | | | | | | | | | | | |
| others | *B. phymatum* STM815 | 8.68 | 2 | 2 | Mutualist, Nitrogen fixation | ***Xmo2*** | | | | | | | | | | |
| others | *B. sp.* YI23 | 8.90 | 3 | 3 | Bioremediator | **Plasmid** | | | | | | **Xmo3(6)** | | | **Xmo3** | |
| others | *B. phytofirmans* PsJN | 8.21 | 2 | 1 | Biocontrol | | **Xmo2** | | | | | | | | | |
| others | *B. sp.* RPE64 | 6.96 | 3 | 2 | Mutualist | **Plasmid** | | | | | | | | | | |
| others | *B. phenoliruptrix* BR3459a | 7.65 | 2 | 1 | Mutualist | | | | | | | | | | | |

## B. Norine: a powerful resource for novel nonribosomal peptide discovery

The search for novel microbial secondary metabolites has been motivated by the need for new effective agents that can influence the growth of other microorganisms. NRPs represent a wide diversity of natural compounds with different biological activities potentially exploited by industries in diverse areas such as phytosanitary sector, cosmetics or health. Many of these compounds can be discovered using bioinformatic tools. With this aim, availability of a tool gathering all known NRPs as the Norine database, was very useful in the identification of natural compounds as NRPs.

Norine is a bioinformatic resources that include the unique database dedicated to NRPs, associated with computational tools for their analysis. It currently contains more than 1100 peptides annotated based on their monomeric structure. It allows structure comparison of NRPs and help the prediction of the biological activities of a given NRP. With the aim to facilitate the process of entering NRPs into the database, My Norine, an interface for peptide submission and modification by contributors/curators, has been developed recently to facilitate the entry of new peptides and enrich the database.

Although the structure of cepaciachelin, a siderophore produced by *B.ambifaria* AMMD, has been known for more than twenty years (Barelm I 1996), surprisingly the mechanism of synthesis was not yet reported. In the course of a search of *B.ambifaria* AMMD genome for gene cluster encoding novel NRPS cluster, a cluster of 8 genes encode for cepaciachelin was identified. It is a small compound composed of only 4 monomers: one lysine is bond to one putrescin and two residues of di-hydoxy-benzoic acid (noted Dhb or diOH-bz). This cluster include a gene for a TonB-dependent receptor, genes to build up a diOH-bz monomer and two nrps genes allowing the incorporation of a diOH-bz and a lys.

In the current work, cepaciachelin was used as a study case for peptide submission in Norine database. In the first step, I introduced cepaciachelin such as in Norine database with a putative status following the identification of the synthetic pathway including nrps gene (as described in the preceding paragraph), I changed the status into curated. The attributed Norine ID was NOR01254. Cepaciachelin represents the fifth curated diOH-bz containing peptide with a siderophore activity annotated in the Norine database.

This user case was described in the following article published in the first issue of a journal named Synthetic and Systems Biotechnology, in response to the Editor Tilmann Weber's invitation.

## Norine: a powerful resource for novel nonribosomal peptide discovery

Pupin[1,2], M., Esmaeel[3], Q., Flissi[1,2], A., Dufresne[1,2], Y., Jacques[3,4], P. and Leclère[1,2,3,❖], V.

[1] : Univ Lille 1, CRIStAL, UMR CNRS 9189, F-59655 Villeneuve d'Ascq Cedex, France

[2] : Inria-Lille Nord Europe, F-59655 Villeneuve d'Ascq Cedex, France

[3] : Univ Lille 1, ICV Charles Viollette Institute EA 7394, ProBioGEM, Polytech'Lille, Avenue Langevin, F-59655 Villeneuve d'Ascq Cedex, France

[4] : Microbial Processes and Interactions Laboratory, TERRA Research, Gembloux Agro-Bio Tech-University of Liege, Passage des Déportés, 2, 5030 Gembloux, Belgium

❖: corresponding author
Tel: +33320 43 46 68

Mail: valerie.leclere@univ-lille1.fr

Synthetic and Systems Biotechnology 1 (2016) 89–94

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: keaipublishing.com/synbio

# Norine: A powerful resource for novel nonribosomal peptide discovery

M. Pupin [a,b], Q. Esmaeel [c], A. Flissi [a,b], Y. Dufresne [a,b], P. Jacques [c,d], V. Leclère [a,b,c,*]

[a] Univ Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
[b] Inria-Lille Nord Europe, Villeneuve d'Ascq Cedex F-59655, France
[c] Univ Lille, INRA, ISA, Univ Artois, Univ Littoral Côte d'Opale, EA 7394 - ICV - Institut Charles Viollette, F-59000 Lille, France
[d] Bioindustry Unit, Gembloux Agro-Bio Tech-University of Liege, Passage des Déportés, 2, Gembloux 5030, Belgium

ARTICLE INFO

*Article history:*
Received 15 October 2015
Received in revised form 26 November 2015
Accepted 26 November 2015
Available online

*Keywords:*
Norine
Nonribosomal peptides
Secondary metabolites
Database

ABSTRACT

Since its first release in 2008, Norine remains the unique resource completely devoted to nonribosomal peptides (NRPs). They are very attractive microbial secondary metabolites, displaying a remarkable diversity of structure and functions. Norine (http://bioinfo.lifl.fr/NRP) includes a database now containing more than 1160 annotated peptides and user-friendly interfaces enabling the querying of the database, through the annotations or the structure of the peptides. Dedicated tools are associated for structural comparison of the compounds and prediction of their biological activities. In this paper, we start by describing the knowledgebase and the dedicated tools. We then present some user cases to show how useful Norine is for the discovery of novel nonribosomal peptides.



## 1. Introduction

Nonribosomal peptides (NRPs) are attractive natural compounds because of their numerous biological activities potentially exploited by industries in diverse areas such as phytosanitary sector, cosmetics or health. They are produced by microorganisms (including bacteria and fungi) through specialized biosynthetic pathways. NRPs are biosynthesized by enzymatic modular complexes called NonRibosomal Peptide Synthetases (NRPSs) working as multidomain assembly lines.[1] The mode of synthesis leads to the production of compounds displaying a broad range of structures. Indeed, if some of them look like classical peptides because they are linear, most of them are more complex, including one or more cycles and branches. Moreover, those peptides are composed of monomers that are not limited to the 20 proteinogenic amino acids. Up to now, we have identified more than 530 building blocks composing the different NRPs. The structural biodiversity is also due to the monomer modifications occurring during the synthesis made by the NRPSs themselves or performed post synthesis by accessory enzymes (also named tailoring or decorating enzymes). Famous examples for NRPs are the antibiotics penicillin,[2] bacitracin and vancomycin,[3] or the immunosuppressor cyclosporine.[4] In addition, some NRPs show antitumor activity such as Dactinomycin.[5] A current worrying public health issue is to find and develop new drugs to overcome multi-resistant pathogens. Therefore, it is important to develop bioinformatics tools for secondary metabolite discovery, such as antiSMASH[6] and tools especially dedicated to NRPSs and NRPs, such as Florine,[7] NaPDos,[8] and Norine.[9,10] The development of Norine was first motivated by the availability of computational tools allowing structure comparison of all NRPs,[11] in spite of their complexity. For this purpose, we needed a database gathering all known NRPs, annotated according to their monomeric structure (i.e. monomer composition and 2D topology). Until now, the Norine team screened the literature to enter new peptides and annotated them manually. To get a more complete database, we have recently opened it to crowdsourcing through an easy-to-use web-based application.[10] Moreover, a semi-automatic process to extract data from external sources is currently under development.

## 2. The Norine database

### 2.1. Description and querying

Norine is a platform that includes the unique database dedicated to NRPs, associated with computational tools for their analysis. It has gained an international recognition thanks to high quality and manually curated annotations. Containing about 700 annotated NRPs for its first release in 2008, Norine database now contains more than 1160 NRPs that are clustered into 214 families, and composed of

* Corresponding author. Univ Lille, ICV, Charles Viollette Institute, PrBioGEM team, Polytech'Lille, Avenue Langevin, Villeneuve d'Ascq Cedex F-59655, France. Tel.: +33 320 43 46 68.
*E-mail address:* valerie.leclere@univ-lille1.fr (V. Leclère).
Peer review under responsibility of KeAi Communications Co., Ltd.

http://dx.doi.org/10.1016/j.synbio.2015.11.001

at least 530 distinct monomers. Among the peptides, 73.5% are tagged with the "curated" status, which means that their nonribosomal origin is supported experimentally (for example due to identified NRPS) while 26.5% are annotated with "putative" status due to only presumed nonribosomal origin (often based on structural features). Two thirds of the peptides are cyclic or partially cyclic or contain at least one cycle. The sizes of the peptides range from 2 to 26 monomers, if polytheonamide is excluded, which was described as being the biggest NRP with 49 monomers for a long time but recently was identified to be an RiPP (Ribosomally synthesized and post-translationally modified peptide).[12] Thus, in the near future, a third category will be created to tag all deprecated peptides when the hypothetical NRPS origin is finally excluded.

Each peptide page includes a comprehensive description of the peptide with the name, activities and structural atomic and monomeric details. The monomeric structure can be automatically obtained through the integrated smiles2monomers tool (s2m) when SMILES are available.[13] When identified, links to UniProt (for synthetases), PDB and PubChem (for structural data on the peptides) are provided. Moreover, a direct link to the NRPS gene clusters annotated in MIBiG[14] will be added soon.

Norine is queried from all over the world by biologists and biochemists to further analyze the nonribosomal peptides they study. For example, Desriac et al.[15] queried Norine to predict the antibacterial activity of a putative NRP produced by *Pseudoalteromonas*, while Bills et al.[16] used Norine to investigate the structural differences between bacterial and fungal NRPs. Indeed, for this purpose, the Norine platform provides visualization and editing applets for monomeric structure as well as tools to compare monomeric structures. Currently, Norine can be queried either by annotations (through "general search" tab) or by structural information (through "structure search" tab) of the peptides.

### 2.1.1. General search

Norine provides a basic interface that enables to query the database and search for peptides by combining multiple criteria, such as the name of the peptide, the Norine ID, the biological activities, the structure type, the producing organism, or the title or authors of references associated to the NRP. The main advantage of this interface is that it allows users to extract data and get statistics according to different criteria. For example, one can query for all siderophores produced by "any bacteria" (check "siderophore" in the "activity" field and enter "bacteria" in the "organism search" field), or all peptides with a linear structure, or simply search for all NRPs produced by the genus "*Pseudomonas*" (enter "pseudomonas" in the "organism search" field) (see the results in Fig. 1). The first output is a list of all the peptides corresponding to the criteria selected, classified by families. A click on a peptide name directs to the peptide page containing all details on the compound. Moreover, a click on the pie chart icon located above the list of results provides graphical output (Fig. 1). Pie charts and diagrams enable to filter the obtained results in order to refine them, by clicking on a slice, for example by structure type or monomers size.

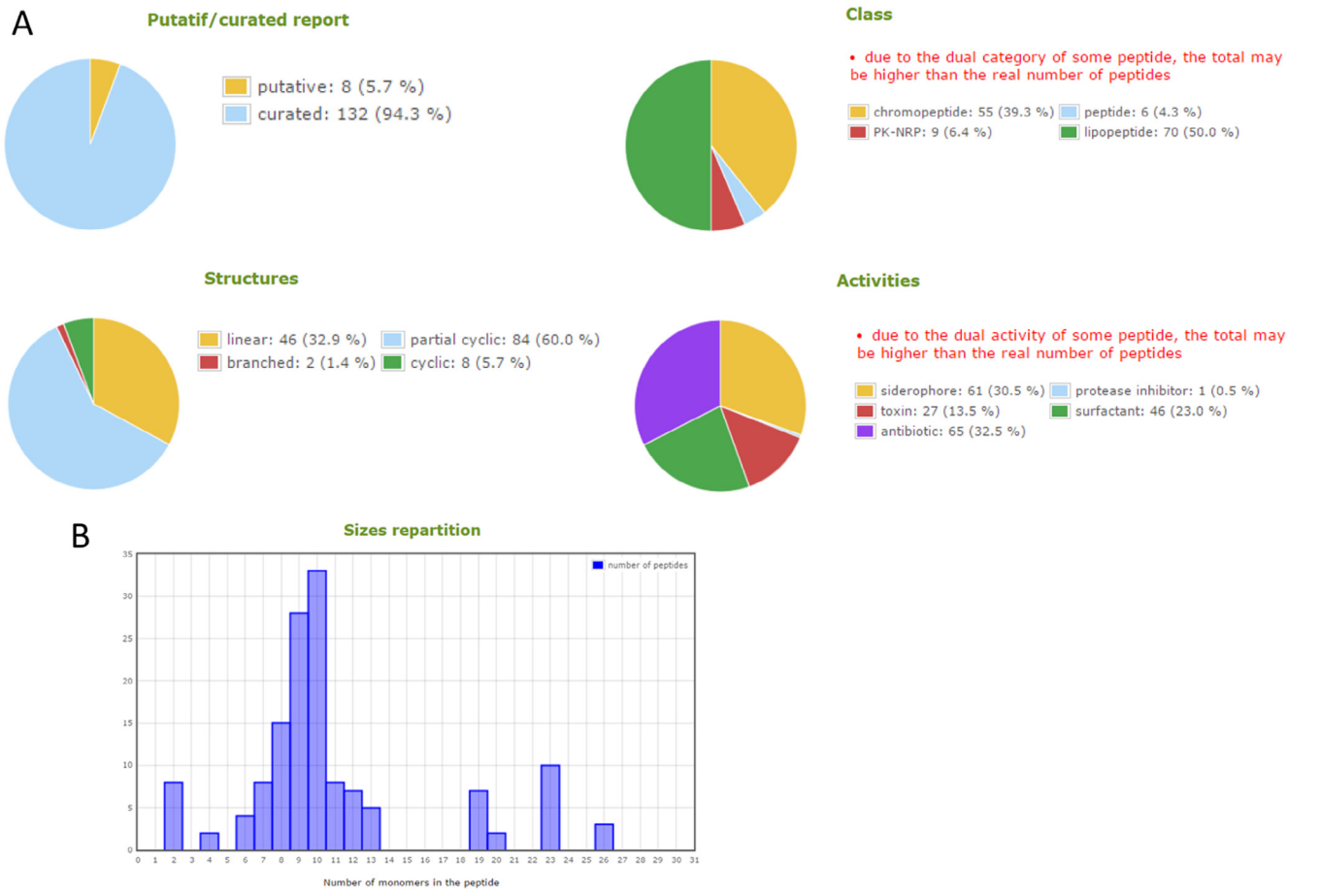



**Fig. 1.** Graphical output provided with *Pseudomonas* query in "organism search" form. (A) Pie charts representing the percentages of the nonribosomal peptides produced by *Pseudomonas*, according to their status, their class, their structure types and their activities. (B) Histogram representing size distribution of the peptides produced by *Pseudomonas*. For lipopeptides, the fatty acid is considered as one monomer.

*2.1.2. Structure search*

In addition to search through annotations, Norine proposes efficient structure search tools based on different algorithms: monomeric composition fingerprint (MCFP),[17] structure-based search for pattern comparison and similarity-based search.[11,18] These enable to find peptides containing a given list of monomers or a given 2D-pattern for structural comparisons. In Norine, a specific syntax is used (the NOR format) to describe the two-dimension graph of an NRP, taking into account the topology of the molecule (linear, cyclic, branched, etc.). In the string representation used by the computational tools (i.e. "Val,Orn,Leu,D-Phe,Pro@1@0,2@1,3@2,4@3"), the monomers are listed, separated by commas; the @ character symbolizes the links between numbered monomers (explained in more details in the "structure search" part of the help tab). However, in most use-cases the string can be automatically generated with the graphical editor applet provided in the structure search form. Users only have to draw a peptide (or fragments of a peptide) by picking the monomers in the list proposed on the left side, and connecting the monomers, once they are placed on the drawing area. Fig. 2 illustrates an example of structure search results obtained for the linear peptide "Val_Orn_Leu_D-Phe_Pro" using the representation in NOR format generated by the graphical editor.

Advantages of structure-based search are manifold. First, using the graph representation of an NRP in Norine structure-based search tools enables to find similar NRPs that can be variants, or identical compounds from a structure point of view independent of their names/annotations. Second, we are convinced that the diversity of the biological activities of NRPs comes from their monomeric composition and the diversity of their structures.[18] That is the reason why the structure-based search tools can help predict biological activities of an NRP. Indeed, similar NRPs probably share common properties such as their known activities. Finally, the structure comparison tools may be helpful to annotate NRPS coding genes/clusters within a microorganism genome sequence. Indeed, peptides with similar monomeric structures may be produced by NRPSs with close modular organization.

Examples presented below further illustrate the different use cases.

*2.2. Norine database is now open to crowdsourcing*

With the development of high throughput technologies dedicated to the screening of secondary metabolites, the number of published descriptions of new NRPs is increasing exponentially. Considering that researchers are the best experts to annotate the NRPs they are working on, we have decided to open Norine to crowdsourcing.[10] We have facilitated the process of entering NRPs into the database by developing an interface for peptide submission and modification by contributors/curators. In order to use this MyNorine interface, users firstly register by creating an account. Standardized forms are provided to submit new peptides or update records for existing peptide entries. The entered data are thereafter reviewed by validators of the Norine team. That process is crucial to ensure that peptides stored in Norine are expert

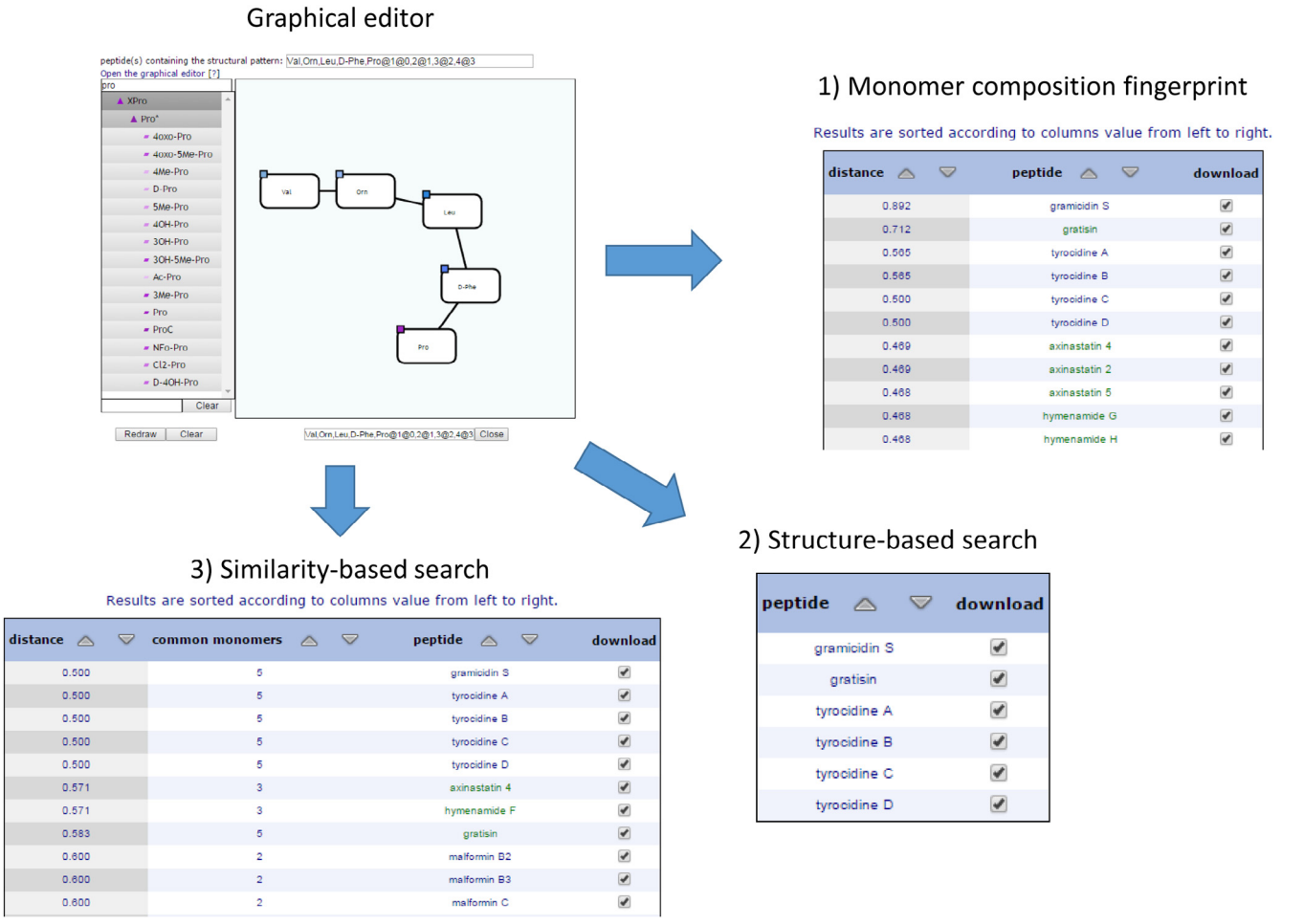



**Fig. 2.** Structure search results. Screenshots of the results obtained using the linear pentapeptide "Val_Orn_Leu_D-Phe_Pro" as a pattern (drawn as graph and transformed in NOR format by the graphical editor) for structural comparison.

validated as this guarantees the quality of the data. The contributions will help enrich the Norine database. The contributors will be mentioned as the authors of the entry.

## 3. Use cases

### 3.1. *Identification of novel CLPs produced by* Pseudomonas *CMR12a*

With the aim of discovering new cyclic lipopeptides (CLPs) with potential biocontrol activity, a combination of chemical structure analysis and *in silico* analysis of the genes encoding NRPSs was carried out on *Pseudomonas* CMR12a.[19] The strain was shown to produce two components originally named CLP1 and CLP2 with 18 and 10 amino acid monomers within the peptide backbone, respectively. The structures of both compounds were elucidated and compared to the structure of all the peptides stored in Norine, using the "structure search" interface. A peptide named orfamide B was identified in the database that matched exactly to the peptide sequence of CLP2 (Fig. 3). CLP1 was identified as being a new member of the tolaasin group,[20] displaying only one substitution on the monomer at position 6 (Fig. 3). Thus, the tolaasin group, which comprises at least 11 CLPs produced by different *Pseudomonas* strains (7 tolaasins, 2 corpeptins and 2 fuscopeptins), was extended with this new member, named sessilin according to its involvement in biofilm formation.[19] This example demonstrates the relevancy of structure search tool of the Norine resource to evaluate the novelty of peptides detected during a screening for active secondary metabolites.

### 3.2. *Gratisin shares a pattern with the well-known gramicidin S*

Gratisin is an undecapeptide with antibiotic activity produced by *Brevibacillus brevis* formerly named *Bacillus brevis* Y-33. Considering its primary cyclic structure,[21] including a D-enantiomer of phenylalanine, and the presence of the nonproteogenic amino-acid ornithine, it was assumed to be nonribosomally synthesized. An entry was created in the Norine database with a "putative" status because no NRPS associated with gratisin biosynthesis was known. The structure-based search returned 5 peptides sharing a pattern constituted of the pentapeptide motif "Val,Orn, Leu, D-Phe, Pro". All the 5 peptides display antibiotic activity and are produced by *Bacillus* strains: four belonging to tyrocidin family (tyrocidins A, B, C and D) and the fifth one is the decapeptide gramicidin S (Fig. 2), in which the same pattern is repeated twice due to an iterative mode of biosynthesis[1,22] (Fig. 4). Even if the synthetase has not yet been identified in any sequenced *Bacillus* genome, we can guess that gratisin is nonribosomally synthesized with an NRPS also using an iterative mode of biosynthesis because it also contains a repeated motif. This example shows that the structure comparison of the final peptide may give insights into the biosynthetic pathway and thus contributes to the prediction of the modular organization of its producing NRPS. This can facilitate the identification of the genes or clusters directly involved in the production of such metabolites using genome-mining approaches.

### 3.3. *Cepaciachelin: from putative to curated status*

Twenty years ago, the structure of a siderophore produced by *Burkholderia ambifaria* strain PHP7 (LMG 11351) was elucidated.[23] It is a small compound composed of only 4 monomers: one lysine is bond to one putrescine and to two residues of di-hydroxy-benzoic acid (abbreviated Dhb or diOH-Bz). For siderophores a nonribosomal origin can not be systematically attributed because some of them, like anguibactin or enterobactin,[24] are synthesized by NRPSs, whereas others, like desferrioxamine, are built up by other enzymes.[25] During a genome-mining analysis, we have identified a gene cluster within the genome of *B. ambifaria* AMMD that is responsible for the cepaciachelin production (personal communication). Norine was directly queried with the NRP sequence predicted by antiSMASH, resulting in a hit against cepaciachelin. As there is a functional confirmation that cepaciachelin is indeed an NRP, the status Norine entry now could be updated from "putative" to "curated" (Norine ID = NOR01254). Cepaciachelin represents the fifth curated diOH-Bz containing peptide with a siderophore activity annotated in the Norine database.

## 4. Conclusion

Norine (http://bioinfo.lifl.fr/NRP) is a freely available and unique resource dedicated to nonribosomal peptides (NRPs).[10] A user-friendly interface allows easy browsing, annotation, structure searching and downloading of the NRPs and their monomers. To discover new natural products, Norine may be the final step of a workflow, which is aimed at detecting the potential for new NRP biosynthesis from genomic data.[7] In case where compounds are identified as being new NRPs or variants of an existing family, researchers can now submit them directly to Norine with the easy-to-use MyNorine interface. The scientific community will contribute to and benefit from the enriched resource, improving the screening for NRPs with biological or medical applications.

## Acknowledgments

The authors are grateful to University of Lille 1, Inria-Lille Nord Europe, and bililleplateform for their financial support.

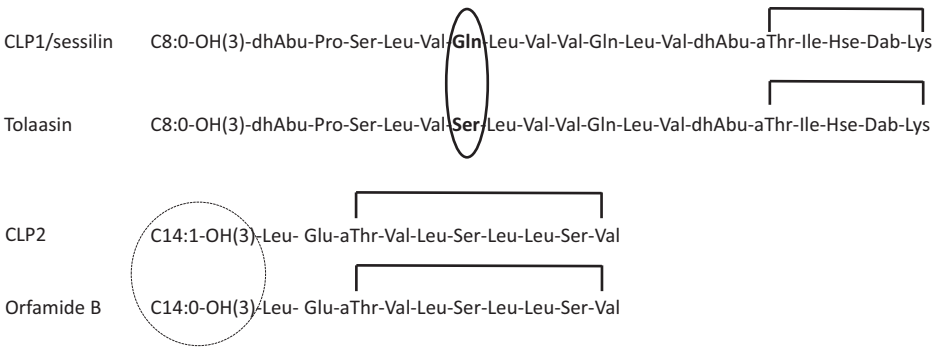



**Fig. 3.** Structural comparison of sessilin and CLP2 produced by *Pseudomonas* CMR12a with tolaasin and orfamide. The monomers have been aligned, the full lines represent the cyclization within the peptidic part. The variable amino acids in the peptide moiety are in bold and surrounded, small variability within the acyl moiety is highlighted by a dotted circle.

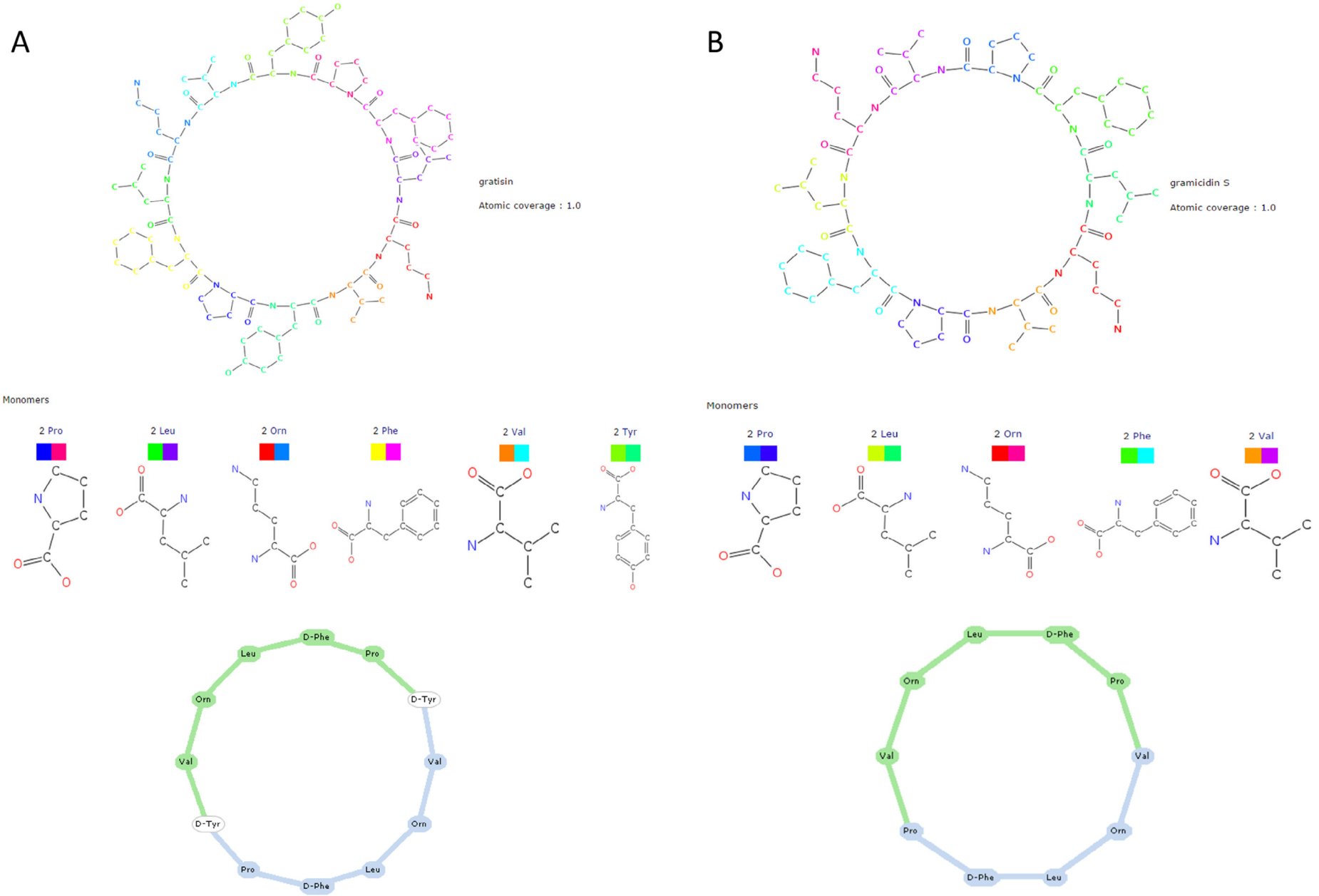



**Fig. 4.** Structure comparison at the monomeric level. (A) Gratisin, (B) Gramicidin S. Top: monomeric representation as returned by the integrated smiles2monomers tool (s2m). Middle: monomers composing the peptides identified by s2m. Bottom: schematic representation highlighting the repetition of a common pentapeptide motif and the differences between both molecules.

## References


1. Hur GH, Vickery CR, Burkart MD. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* 2012;**29**:1074–98.
2. Queener S. Molecular biology of penicillin and cephalosporin biosynthesis. *Antimicrob Agents Chemother* 1990;**34**:943–8.
3. Hubbard BK, Walsh CT. Vancomycin assembly: nature's ways. *Angew Chem Int Ed Engl* 2003;**42**:730–65.
4. Liu H, Wang Y, Li S. Advanced delivery of ciclosporin A: present state and perspective. *Expert Opin Drug Deliv* 2007;**44**:349–58.
5. Li JJ, Huang HH, Shen J, Jiang JY, Pan F, Yu ST, et al. Multimodal therapy for adult Wilms' tumor: an experience from one centre. *Clin Transl Oncol* 2011;**13**:672–6.
6. Weber T, Blin K, Duddela S, Krug D, Uk Kim H, Bruccoleri R, et al. antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 2015;**43**:W237–43.
7. Caradec T, Pupin M, Vanvlassenbroeck A, Devignes MD, Smaïl-Tabbone M, Jacques P, et al. Prediction of monomer isomery in Florine: a workflow dedicated to nonribosomal peptide discovery. *PLoS ONE* 2014;**9**:e85667.
8. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The natural product domain seeker NaPDoS: a phylogeny based bioinformatics tool to classify secondary metabolite gene diversity. *PLoS ONE* 2012;**7**:e34064.
9. Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, Kucherov G. NORINE: a database of nonribosomal peptides. *Nucleic Acids Res* 2008;**36**:D326–31.
10. Flissi A, Dufresne Y, Michalik J, Tonon L, Janot S, Noé L, et al. Norine, the knowledgebase dedicated to nonribosomal peptides, is now open to crowdsourcing. *Nucleic Acids Res* 2015;doi:10.1093/nar/gkv1143.
11. Caboche S, Pupin M, Leclère V, Jacques P, Kucherov G. Structural pattern matching of nonribosomal peptides. *BMC Struct Biol* 2009;**9**:15.
12. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 2013;**30**:108–60.
13. Dufresne Y, Noé L, Leclère V, Pupin M. Smiles2Monomers: a link between chemical and biological structures for polymers. *J Cheminform* 2015;**7**:62.
14. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 2015;**11**:625–31.
15. Desriac F, Jégou C, Balnois E, Brillet B, Le Chevalier P, Fleury Y. Antimicrobial peptides from marine proteobacteria. *Mar Drugs* 2013;**11**:3632–60.
16. Bills G, Li Y, Chen L, Yue Q, Niu XM, An Z. New insights into the echinocandins and other fungal non-ribosomal peptides and peptaibiotics. *Nat Prod Rep* 2014;**31**:1348–75.
17. Abdo A, Caboche S, Leclère V, Jacques P, Pupin M. A new fingerprint to predict nonribosomal peptides activity. *J Comput Aided Mol Des* 2012;**26**:1187–94.
18. Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P. Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *J Bacteriol* 2010;**192**:5143–50.
19. D'aes J, Kieu NP, Leclère V, Tokarski C, Olorunleke FE, De Maeyer K, et al. To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain *Pseudomonas* CMR12a. *Environ Microbiol* 2014;**16**:2282–300.
20. Roongsawang N, Washio K, Morikawa M. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. *Int J Mol Sci* 2011;**12**:141–72.
21. Tamaki M, Takimoto M, Sofuku S, Muramatsu I. Synthetic studies on gratisin. *J Antibiot* 1983;**6**:751–2.
22. Hoyer KM, Mahler C, Marahiel M. The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization. *Chem Biol* 2007;**14**:13–22.
23. Barelmann I, Meyer JM, Taraz K, Budzikiewicz H. Cepaciachelin, a new catecholate siderophore from *Burkholderia* (*Pseudomonas*) *cepacia*. *Z Naturforsch [C]* 1996;**51**:627–30.
24. Crosa JH, Walsh CT. Genetics and assembly line enzymology of siderophores biosynthesis in bacteria. *Microbiol Mol Biol Rev* 2002;**66**:223–49.
25. Challis GL. A widely distributed bacteria pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chembiochem* 2005;**6**:601–11.

## C. Non ribosomal peptides and polyketides of *Burkholderia* as a new source of bioactive compounds potentially implicated in agriculture and pharmaceutical

*Burkholderia* are soil bacteria able to use a wide range of organic compounds and produce a wide range of potent natural products including NRPs and polyketides with a great interest of uses as antibiotics, biocontrol agents, anticancer, and toxins.

Bacterial genome mining via *in silico* analysis offers an attractive opportunity of discovering new secondary metabolites such as nonribosomal peptides (NRPs) and Polyketides (PKs) (Van Lanen and Shen 2006). Thus, the availability of whole genome sequences of *burkholderia* species and computing abilities such as multiple sequences alignment have allowed the identification of an important number of NRPS and PKS genes clusters that are responsible for the synthesis of many natural products.

Due to their wide geographical distribution and versatile ecological properties including their significance role as pathogenic, plant promoting growth, and biocontrol, *Burkholderia* have received a lot of scientific attentions in recent years to be investigated in more details as producers of new non ribosomally peptides and polyketides potentially involve as new sources of many natural products.

During the comprehensive screening of NRPS gene clusters in 48 gapless complete genomes of *Burkholderia*; many hybrid clusters NRPS-PKS, involved in the synthesis of many natural compounds, were detected. For that reason, the current review aimed to deal with all the potential NRPs and polyketides discovered from *Burkholderia* by genome mining approaches. Their antagonistic activities on other organisms including pathogenic fungi and bacteria are also included as well as their positive effect on plant growth and their potential uses in agriculture and pharmaceutical sectors are also discussed.

The review presented in the following section completes the previous paragraph focused on siderophores and lipopeptides.

# Non ribosomal peptides and polyketides of *Burkholderia* as a new source of bioactive compounds potentially implicated in agriculture and pharmaceutical

Qassim Esmaeel[1], Philippe Jacques[2], and Valérie Leclère[1, 3]

[1] :Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, EA 7394-ICV- Institut Charles Viollette, F-59000 Lille, France.

[2]: TERRA Research Centre, Microbial Processes and Interactions (MiPI), Gembloux Agro-Bio Tech University of Liege, B-5030 Gembloux, Belgium

[3] : Inria-Lille Nord Europe, Bonsai team, F-59655 Villeneuve d'Ascq Cedex, France.

❖: Corresponding author
Tel: +33320 43 46 68

Mail: valerie.leclere@univ-lille1.fr

## Abstract

*Burkholderia* are soil bacteria able to use a wide range of organic compounds and produce a several secondary metabolites. Among these, NonRibosomal Peptides (NRPs) and PolyKetides (PKs) are commonly synthesized in a ribosome- independent way via modular megaenzymes called NRPS and PKS. The aim of this review is to highlight all the modular gene clusters found in 48 strains of *Burkholderia* that are directly implicated in the biosynthesis of NRPs and PKs. The screening was performed *in silico* following the strategy including Florine workflow we have developed with this aim. Thus, 228 gene clusters were detected, among them, 31% are likely to produce siderophores, 15% showed the typical organization of different lipopeptides assembly lines, and 54% may be involved in synthesis of other products through NRPS or typical hybrid NRPS-PKS complexes. Their biological activities and potential uses in agriculture and pharmaceutical were also discussed.


Key words: *Burkholderia*, secondary metabolites, PKS, NRPS, Florine, Lipopeptides

## 1. Introduction

The genus *Burkholderia* consists of over sixty species, all belonging to proteobacteria. They live in various ecological niches including soil, water, plant as well as animal and human [1]. The first isolates of *Burkholderia* were described by William Burkholder in 1950 who classified them as *Pseudomonas cepacia*. Two decade ago, molecular taxonomic analysis led to the transfer of these bacteria to a new genus distinct from *Pseudomonas*, called *Burkholderia* [2]. Different species of *Burkholderia* are well known for their roles in the environments including biodegradation of pollutants, bioremediation processes [3]. For example, *B. xenovorans* strain LB400 is one of the most effective polychlorinated biphenyl (PCB) degraders known so far [4]. *Burkholderia* contains species pathogenic for plants [5], as well as species involved in promoting plant growth used as biocontrol agents [6-7] or as nitrogen fixing bacteria [8]. Plant roots can be nodulated by several species of *Burkholderia* such as *B. nodosa*, *B. mimosarum*, and *B. phymatum* [9-10]. Some strains related to *B. cepacia* complex (Bcc) as well as *B. gladioli* and *B. fungorum* are pathogenic especially for those who are suffering from cystic fibrosis or chronic granulomatous diseases [11]. *B. pseudomallei* and *B. mallei* are considered the primary cause of melioidosis in humans [12] and glander in horses [13]. The capacity of *Burkholderia* strains to live and colonize a wide variety of environments is likely due to an unusually large, complex, and variable genome (4.6 to 9 Mb), split into up to three chromosomes and large plasmids.

*Burkholderia* members live in different conditions, and to survive under such conditions, they often produced a wide variety of of extracellular enzymes for reasons linked to the adaptation to different ecological niches and continued existence [14]. These include enzymes with proteolytic, lipolytic, and haemolytic activities [15] and secondary metabolites, especially those produced by modular megaenzymes as non ribosomal peptides [16, 7] and polyketides [17].

Over the last years, the importance of non ribosomal compounds produced by microorganisms for mediating relation between bacteria and other organisms has become more and more evident. Moreover the need for new antimicrobial agents motivates the search of new microbial metabolites to fight the microbial multidrug-resistance. In addition to their influence on plant growth and development, non ribosomal peptides (NRPs) have also been shown to reduce the growth of phytopathogenic fungi [18]. The antifungal activity of these compounds against a wide range of phytopathogenic fungi including *Botrytis cinerea* and *Mycosphaerella figiensis* has attracted intense interest as a new class of antifungal agents. Bacterial genome mining via *in*

*silico* analysis offers an attractive opportunity to discover new secondary metabolites such as NRPs and Polyketides (PKs) [19]. Thus, the availability of whole genome sequences of *Burkholderia* species and development of specific bioinformatics tools have allowed the identification of an important number of modular genes responsible for the synthesis of many natural compounds including lipopeptides, siderophores and antimicrobial metabolites. This review covers all the potential NRPs and PKs discovered from *Burkholderia*. In addition to the screening of all the modular gene clusters found in the complete genome sequences of 48 strains of *Burkholderia*, which is the main focus of this review, other known products and their biosynthetic genes and their biological activities and applications are also included.

## 2. Genome mining of *Burkholderia*

Biosynthetic gene clusters are known to be involved in a wide variety of enzymatic pathways that produce specialized metabolites in bacteria, fungi and plants. Therefore, information about these clusters, pathways and metabolites is needed to facilitate the process of natural product discovery. By combining technological development in genomics, bioinformatics and structure biology, translated gene sequence data can be used to rapidly predict structural elements encoded by secondary metabolic gene clusters from microorganisms. Discovery of new compounds from natural sources has largely been an adventitious process based on bioactivity screening of microbial extracts combined with bioassay guided identification and natural product structure elucidation. Once the chemical structure of the novel compound clarified, its biosynthesis can be described at the biochemical and genetic levels. Theses exploration gives insight into the mechanisms of synthesis like those for non ribosomal peptides and polyketides.

Genome mining on microbial genomes has recently guided the discovery of new secondary metabolites such as NRPs and PKs. The availability of whole bacterial genome sequencing combined with bioinformatics approaches facilitated the understanding of secondary metabolites biosynthesis and gave new insight into bacterial metabolism. This guided the discovery of a large number of new biosynthetic pathways of natural products. Genome analyses have shown that members of *Burkholderia* are an untapped reservoir of non ribosomal peptides and polyketides [20-21, 18, 22]. However, up to now only a few NRPSs and PKs have been characterized for a given members of *Burkholderia* due to the restricted conditions used that are conductive to the production of NRPs and PKs. Indeed, the majority of *Burkholderia* biosynthetic genes cluster seem to be silent under standard laboratory conditions.

Analyzing the complete genome sequenced available in NCBI of 48 strains of *Burkholderia* species revealed that *Burkholderia* has high potential for production of NRPs and hybrid NRPS-PKs products that might have fantastic applications in medicine as drug candidate and agriculture as biocontrol agents. All the clusters identified were either NRPS or hybrid NRPS-PKS and no pure PKS cluster was detected. Around 228 gene clusters have been identified via *in silico* analysis following Florine, a workflow especially developed with the aim to identify *in silico* all NRPS gene clusters in microorganisms including bacteria and fungi [23]. Of these, 31% are likely to encode for assembly lines leading to the production of siderophores, 15% show the typical organization allowing the synthesis of different lipopeptides including LP that exhibit the feature

of [C-start-TE], antifungal LP, and CLP close to those produced by *Pseudomonas*. Among other products are NRPS or typical NRPS-PKS hybrids involved in the synthesis of bactobolins, thailanstatins, burkholdacs, and malleilactone (Fig. 1)

The workflow Florine includes the use of bioinformatics tools dedicated to secondary metabolite discovery as antiSMASH [24], NRPS & PKS analysis, and NapDos [25]. The first steps are related to general identification of secondary metabolites and the following steps are dedicated to NRPs. The final step of the workflow corresponding to the checking for novelty of the peptides, using Norine, a unique resource devoted to NRPs (http://bioinfo.lifl.fr/NRP) [26]. Some predicted gene clusters contain genes that encode typical accessory enzymes, transporter genes, and core enzymes that involve directly in the synthesis of predicted peptides.

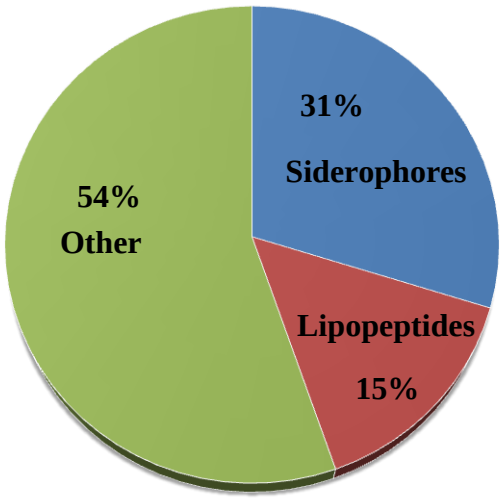



Figure 1: Schematic representation of NRPS and hybrid NRPS-PKS gene clusters found in the genomes of 48 strains of *Burkholderia*

### 3. Biosynthesis of NRPs and PKs in *Burkholderia*

The biosynthetic mechanisms and regulation system of NRPs have been extensively studied over the few years. They are synthesized in a ribosome- independent way via megaenzymes called non ribosomal peptides synthetases (NRPSs) [27]. NRPSs are multidomain enzymes that catalyze the formation of natural products via the reaction that catalyze the assembly of both proteinogenic and non-proteinogenic amino acids in a modular fashion. Genes encoding the

enzymes incorporating monomers or modifying the different residues are, most of the time, organized in clusters. These large synthetases are organized in sets of domains as condensation (C), adenylation (A), thiolation (T) and thioesterase (TE) domains. However, these basic domains only synthesize the raw peptide in their respective biosynthetic pathways. Additional domains as epimerization, oxidation, and methylation are usually involved in the transforming peptides into more structurally peptides. In some cases tailoring enzymes are also needed [28]. Transporter genes might also be present, such as efflux pumps that help the export of NRPs. Production of NRPs and PKs by *Burkholderia* produce is reflected by the high number of NRPS and PKS genes encoded in their genomes [20]. NRPs of *Burkholderia* can be classified into different groups based on their structure and biological activities (Table.1). The main groups contain lipopeptides, siderophores, and hybrid PKS-NRPSs with biosynthetic gene cluster composed of PKS and NRPS genes. The biosynthesis gene cluster of NRPS may contain pathway-specific regulators and co-regulators that control only the gene cluster in which they reside. Sometimes the regulation can be encoded by global regulators gene located elsewhere on the genome [29]. The incorporation of D-amino acids in NRPSs products in bacteria can be achieved through two different mechanisms. The first include the presence of E domains that is responsible for epimerization of L-amino acids to D-forms. The second mechanism includes dual C/E domains that catalyze both condensation and epimerization. These dual C/E domains were first identified in the NRPSs producing cyclic lipopeptides of *Pseudomonas* [30] and were recently also detected in similar NRPSs in *Burkholderia* [31, 20].

Polyketides are another type of secondary metabolites synthesized by modular enzymes named polyketide synthases (PKSs). PKSs work in the same way as NRPSs, but the domains and substrates are different [32]. The synthesis of a polyketide requires at least four domains that can be organized in modules like NRPSs. The acyltransferase (AT) selects a carboxy acid such as acetyl-CoA or methylmalonyl-CoA and holds it on to the acyl carrier protein (ACP) domain. The ketosynthase (KS) domain catalyzes the condensation reaction between the upstream acyl-S-ACP and the downstream ACP bound carboxy acid. The TE domain releases the final product from enzyme [33-34]. Secondary domains can also be involved in the synthesis, including ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) can be inserted to modify the polyketides structure during the synthesis [33].

### 3.1. The role of NRPs and PKs of *Burkholderia*

Secondary metabolites including NRPs and PKs are produced by bacteria for reasons associated with cellular defense, survival, competition and long-term prosperity [14]. When the resources are limited, the prosperity of any given species is typically improved by the antagonism or death of competitors, grazers and parasites. Understanding the nature and the production of these molecules in the ecological niches(s) of *Burkholderia* can guide the determination of their biological activities. NRPs and PKs of *Burkholderia* represent a wide range of natural products involved in different applications including medicine and agriculture as lipopeptides that act as biopesticides to defend an ecological niche from competitors and promote growth of plant (Table1). NRPs represented by lipopeptides are considerably less toxic toward plant, animal, and human and promising sources as it can replace the use of highly toxic pesticides, especially with antifungal activities. For example, it has been reported that *B.ambifaria* secretes extremely bioactive compounds with antagonistic activities against wide range of phytopathogenic fungi [35]. *Burkholderia* NRPs act also as virulence factors as some strains belonging to *B. cepacia* complex and *B. pseudomallei* deploy NRPSs as virulence factors during infection of their respective hosts [28, 31].

Table 1: Diversity of biological activities of NRPs and PKs compounds produced by *Burkholderia*

| NRPs compounds | | Mass | Produced strains | Biological activities | References |
|---|---|---|---|---|---|
| **Lipopeptides** | Xylocandin A1 | 1215 | *B. cepacia* ATCC 39277 | Antifungal<br>Anti yeast | [36] |
| | Xylocandin A2 | 1199 | | | |
| | Xylocandin B1 | 1229 | | | |
| | Xylocandin B2 | 1213 | | | |
| | Cepacidin A1 | 1215 | *B. cepacia* AF 2001 | Antifungal<br>Immunosuppressive | [6] |
| | Cepacidin A2 | 1200 | | | |
| | Burkholdins | 1213, 1215, 1119 | *B.ambifaria* 2.2N | Antifungal | [35] |
| | | 1229 | *B. ambifaria* 2.2N | | [18] |
| | Occidiofungin | 1215.5, 1199.5 | *B. contaminans* MS14 | Antifungal, anti yeast | [37] |
| | Burkholdin/ occidiofungin | ND | *B.ambifaria* AMMD | Antifungal | [28] |
| | AFC-BC-11 | 733 | *B. cepacia* BC11 | Antifungal | [38] |
| **Hybrid PKS-NRPS** | Malleilactone Burkholderic acid | 306 | *B. thailandensis* E264<br>*B. mallei* ATCC23344<br>*B. pseudomallei* K96243 | cytotoxicity<br>antibacterial | [39] |
| | Thailandamides | 691 | *B. thailandensis* E264 | Antiproliferative | [40] |
| | Thailandpsins/ Burkholdacs | 561 | *B. thailandensis* E264 | HDAC inhibitors | [41] |
| | Thailanstatins | 535 | *B. thailandensis* MSMB43 | Potent antiproliferative inhibits pre-mRNA splicing | [42] |
| | Bactobolins | 246 | *B. thailandensis* E264 | Antibacterial | [43] |
| | Glidobactins | 522 | *Burkholderia* spp. K481-B101 | Anticancer<br>antifungal | [44] |
| | Rhizoxin | 625 | *Burkholderia rhizoxina* | Antimitotic | [45] |
| | Rhizonin | 796 | *Burkholderia rhizoxina* | Mycotoxin | [46] |
| **siderophores** | Ornibactin | 680, 708, 736 | Bcc strains | Iron chelating | [47] |
| | Malleobactin | 622 | *B. pseudomallei* | Iron chelating<br>Antibacterial | [48-49] |
| | Pyochelin | 324 | *B. pseudomallei*<br>Bcc strains | Iron chelating | [48] |
| | Cepaciachelin | 488 | *B. ambifaria* AMMD | Iron chelating | [50] |

### 3.2. Lipopeptides of *Burkholderia*

Lipopeptides are biological compounds produced by a wide variety of microorganisms. They are composed of a hydrophilic peptidic chain part linked to hydrophobic fatty acid portion in the molecule [51]. They exhibit different biological properties including antimicrobial, surface activity, anti-enzymatic activity as well as swarming motility and biofilm formation [52-54]. The lipopeptides have received large attention in the last decade due to their wide activities potentially useful for agriculture, pharmaceutical, food, and chemical industries. *Burkholderia* is well known for its capacity to produce two types of lipopeptides including the CLPs that exhibit several feature such as the presence of a starter (Cstart) domain, and two terminal thioesterase domains [20]. The E domain in CLPs is replaced by and dual C /E domains, a unique form that found in CLPs of *Pseudomonas* [30] or *Xanthomonas* [55]. Other lipopeptides like burkholdins are glycolipopeptides with a reduced fatty acid part and probably without any biosurfactant activity are also produced by member of *Burkholderia*. The incorporation of D-monomer in this LPs is directed by the E domain.

Various strains of *Burkholderia* including *B. ambifaria*, *B. contaminans*, and *B. cepacia* have been reported to produce lipopeptides such as burkholdins or occidiofungins [18, 35]. Most lipopeptides from *Burkholderia* can be classified into four groups: burkholdins, occdiofungins, cepacidin, and xylocandin. Some members of *Burkholderia* are able to produce two to three families of lipopeptides [18]. It has been shown that all members related to *B. pseudomallei* produced lipopeptides called malleipeptin [31] or burkhomycin [20] similar to those produced by *Pseudomonas* as their organization domains start by $C_{starter}$ and contain dual C/E domains and ending with a tandem of Te-domains. In addition to these four main families, several other lipopeptides have been also identified in *Burkholderia* species including *B. rhizoxinica*, *Burkholderia glumae* BGR1, *Burkholderia gladioli* BSR3, *Burkholderia sp*. YI23 [20, 56].

#### 3.2.1. Biosynthesis of burkholdins

Burkholdins (Fig.2.1A) are cyclic glycolipopeptides isolated from different strains of *Burkholderia* including *B. ambifaria*, *B. contaminans*, and *B. cepacia* [18]. Burkholdins act as strong antifungal against wide range of phytopathogens by interfering with cell membrane [57]. Biological activity assay showed that burkholdins are regarded as new promising antifungal compounds that can be implicated in different sectors including pharmaceutical and agriculture as biopesticides. They are also considered as virulence factor due to their role in causing

infection by secretion of potent haemolytic toxins which are required for full virulence [28].The structure of burkholdins is composed of eight monomers including glycine, asparagine, 3-hydroxyasparagine (3OH-Asn), 3-hydroxytyrosine (3-OH-Tyr); DAB; two units of serine; and a fatty acyl amino acid (FAA). Different forms of burkholdins (Table. 1) have been reported based on the presence or absence of xylose and the replacement of Asn by DAB [35]. It was shown that burkholdins 1 and 2 share the same amino acids of xylocandins B1 and C1 and the only difference is that modified FAA in xylocandin is its hydroxylation pattern.

The NRPS gene cluster of burkholdins have been identified in *B. vietnamiensis* DBO1 by [28]. Our previous studies revealed the presence of high similar gene cluster biosynthetic gene cluster of burkholdin in the genome of *B. ambifaria* AMMD and *B.* sp.KJ006 [20]. The gene cluster spanning over 49 000 nt from ORFs *Bamb_6466* to *Bamb_6478* contains six NRPS/PKS hybrids genes. The whole cluster encode eight modules, combining NRPS domains as C, A, E, T, and PKS domains as AT, ACP, KS, and KR domains. The cluster also includes genes coding for regulatory, decorating, and transporter proteins (Fig. 3A).

### 3.2.2. Biosynthesis of occidiofungins

Occidiofungins (Fig. 2. 2A) are non ribosomal cyclic lipopeptides consisting of 8 amino acids and fatty acid chain and xylose. High resolution mass spectrometry analysis confirmed the elucidation of complete structure of occidiofungin that composed of Asn1-novel amino acid 2 (NAA2)-Trp3-β-hydroxyTyr4-Lys5-Gly6-Asn7-Ser8. Occidiofungin exists in two forms including the monoisotopic mass 1,199.543 Da (Occidiofungin A) and 1,215.518 Da (Occidiofungin B) [57]. This structure endows the peptides potent antifungal activities against variety of phytopathogenic fungi including *R. solani*, *F. oxysporum*, and *Pythium* spp with MIC values of 15-50 µg/mL. Occdiofungin was also shown to be effective against animal pathogenic fungi especially against those causing invasive pulmonary aspergillosis and dermatophytosis as *Microsporum gypseum* and *Trichophyton mentagrophytes* [57]. Beside its antifungal activity, occdiofungin was shown to have antagonism against different strains of *Candida* including *C. albicans, C. glabrata, C. tropicalis*, and *C. parapsilosis* [58]. This makes occidiofungin potentially useful for management of fungal diseases. The structure and mechanism of action of occidiofungins have been reported to be similar to burkholdins [59]. It has been shown that occidiofungin interacts with cell wall by formation of intracellular inclusion and make significant

change in hyphal morphology and of targeted fungi [57]. The gene cluster of occidiofungin synthetase has been identified in the genome of *B. contaminans* MS14 [37]. They used transposon mutagenesis to disrupt the production of occidiofungin as shown by the elimination of antifungal activities of mutant MS14MT18 compared to the wild type strain.

### 3.2.3. Biosynthesis of xylocandins and cepacidines

Xylocandins (Figure 2.3A) are cyclic lipopeptides isolated from *B. cepacia* (originally *P. cepacia*) [60] composed of 8 monomers including serine (2 residues), glycine, asparagine (1- 3 residues), β-hydroxytyrosine, 2, 4-diaminobutyric acid, and an unusual amino acid with the formula $C18H_{37}NO_5$[36]. Different structures of xylocandins have been reported including A1, A2, B1, B2, C1, C2, D1, and D2, with the following mass (m /z) of 1,215; 1,199; 1,229; 1,213; 1,097; 1,081; 1,083; and 1,067 Da, respectively. Xylocandins carries strong antifungal activity against a panel of dermatophytes and yeast including *Candida albicans* [60]. Cepacidines (Fig.2.4A) A and B share the same structure of xylocandin A1 and A2. They are glycol lipopeptides produced by *B. cepacia* AF 2001 [61]. Cepacidine was found to carry strong antifungal activity against phytopathogenic fungi including *Pythium ultimum* on the cotton and cucumbers, *Plasmopora veticola* on grapes and *Septoria nodorum* and *Fusarium culmorum* on wheat [62]. Beside its antifungal activities, cepacidine possesses a potent immunosuppressive activity and play significant role to suppress the activation of B lymphocytes [63]. As xylocandin and cepacidine are related compounds to burkholdin and occidiofungin, the biosynthetic gene cluster seem to be similar to each other.

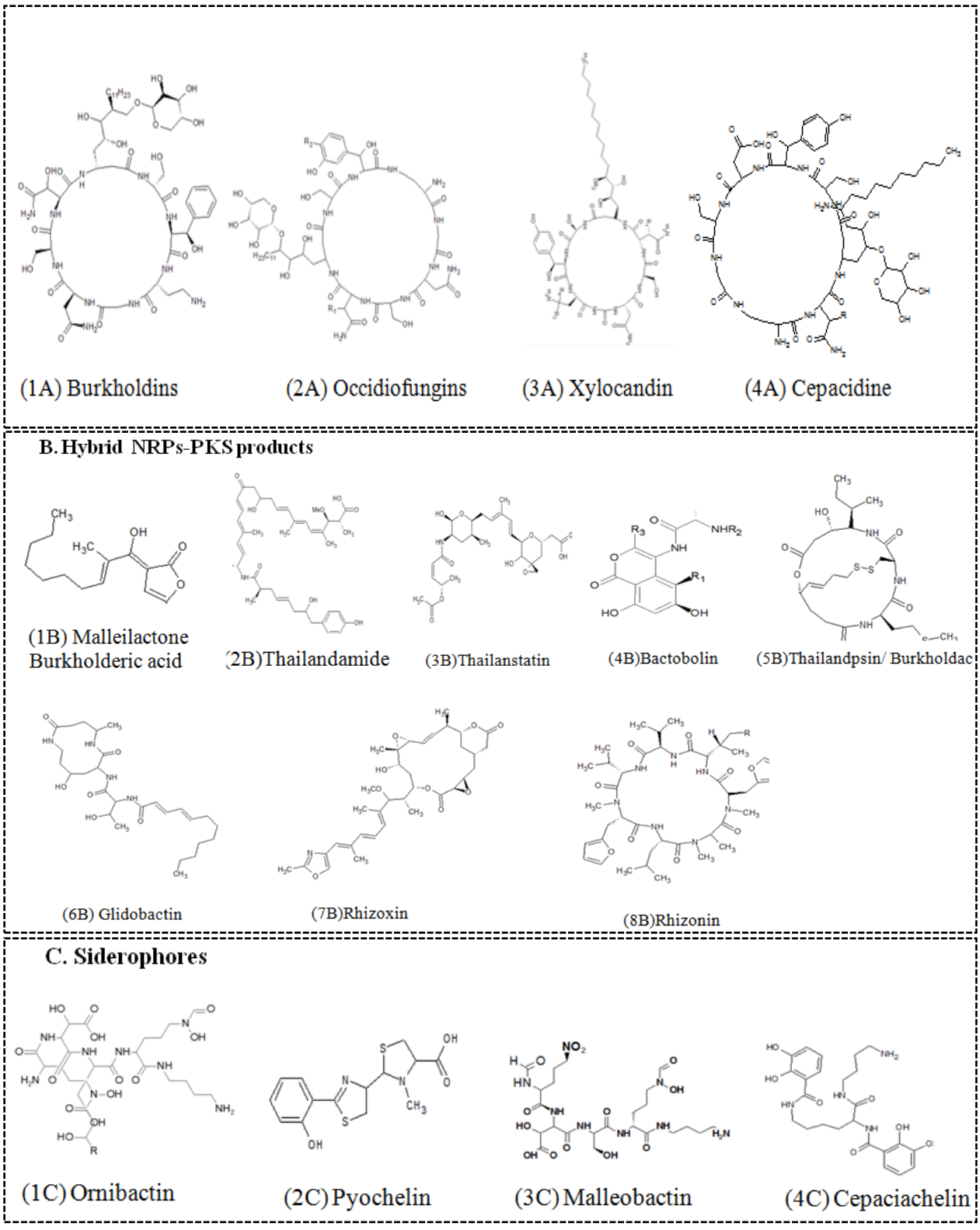



Figure 2: Structure of NRPs and PKs produced by *Burkholderia*

### 3.2.4. Biosynthesis of lipopeptides AFC-BC-11

They are lipopeptides isolated form *B. cepacia* BC11 and possess a potent antifungal activity against a range of soil fungi including *Rhizoctonia solani* that cause damping-off in cotton [38]. Analysis of purified peaks by mass spectroscopy combined with NMR revealed that AFC- BC-11 have mass of 733 Da and consist of peptidic chain of four amino acids (glycine, lysine, phenylalanine, and diaminobutyric acid) and a fatty acyl substituent. Comparing to cepacidin and xylocandin, AFC-BC-11 was found to be smaller (733 versus 1100 Da) and lacks xylose. Moreover, AFC-BC-11 is not active against *Saccharomyces cerevesiae*, *Aspergillus niger*, and dermatophytes [38]. Four genes (*afcA*, *afcB*, *afcC*, and *afcD*) involved in the production of BC-11 were characterized and gave insight into the structure and biosynthesis of AFC-BC1.

## 3.3. Biosynthesis of hybrid PKS-NRPS peptides

The hybrid clusters include in the same complex both NRPS and PKS domains harboured by the same protein or not. This system is very common and sometimes even more predominant than discrete PKS and NRPS gene clusters [64]. The final products are composed of proteinogenic and nonproteinogenic amino acids combined with carbon skeleton. The genomes of *Burkholderia*, especially strains belonging to *thailandensis*, *pseudomallei* and *mallei* species are mainly rich in hybrid biosynthetic gene clusters that are responsible for synthesis of a wide variety of clinically important products that have different activities including anticancer and antimicrobial [39-40, 42]. *B. thailandensis*, a nonvirulent member of the genus that is closely related to *B. pseudomallei* has become a model organism for studying the expression of secondary metabolites as well as virulence factors.

### 3.3.1. Malleilactone/ burkholderic acid

Malleilactone or burkholderic acid (Fig 2. 1B) is a secondary metabolite produced by strains belonging to *mallei*, *pseudomallei*, and *thailandensis* species. The biosynthetic gene cluster of burkholdic acid has been identified in the genome of *B. mallei* ATCC23344, *B. pseudomallei* K96243, and *B. thailandensis* E264 [39, 17]. The synthetases harbour13 open reading frames that exhibit features of NRPS/PKS hybrids enzymes responsible for the synthesis of malleilactone [17] or burkholderic acid [39]. The ORFs are annotated as LuxR-type regulator, NRPS/PKS (Fig. 3B), methyltransferase, hydroxylase, aminotransferase, dehydrogenase,

NRPS/PKS, ketol acid reductoisomerase, FkbH-like protein, decarboxylase acyl CoA ligase, unknown, acyltransferase, acyl CoA ligase, and efflux carrier protein. By following the strategy we have developed to detect the biosynthetic gene clusters in the complete sequenced genomes of *Burkholderia*, the cluster of malleilactone was found in 21 strains belonging to *B. pseudomallei*, *B. mallei*, and *B. thailandensis*. Burkholderic acid was found to have weak to moderate cytotoxicity and no antimicrobial activities has been noted [39], while Brady group found that malleilactone possess weak to moderate antibacterial activities against some strains of Gram-positive bacteria and human cell line [17]. Malleilactone/burkholderic acid might be involved in microbial communication as their structure share similarity with bacterial sensing molecules like A-factor and N-acyl homoserine lactones [39].

### 3.3.2. Thailandamides

In the genome of *B. thailandensis* E264, 17 modules organized in six open reading frames (ORFS) were identified to exhibits several typical features of NRPS/PKS hybrids responsible for the biosynthesis of thailandamide (Fig.2.2B) [40]. The cluster contains also a tandem acyl transferase (AT)-enoyl reductase (ER), and tailoring enzymes. Thailandamide was shown to have moderate antiproliferative properties against human tumor cell lines [40].

### 3.3.3. Thailandpsins/ burkholdacs

Thailandpsins or burkholdac (Fig.2.5B) are bicyclic depsipeptide isolated from *B. thailandensis* E264. They are analogues to FK228, an anticancer drug approved by FDA to the refractory cutaneous T-cell lymphoma [65]. They are different from each other by having different amino acid side chains. Thailandpsin was shown to have a broad spectrum of growth inhibition activities and considered as potent HDAC inhibitors, mainly towards the class I human HDACs (HDAC1-3) [41]. Analysis of 48 complete genome sequences of *Burkholderia* showed the presence of thailandpsin gene cluster in 4 strains related to *B. thailandensis* (Fig. 3C). The cluster of thailandpsin [41] or burkholdac [22] contains 13 genes that include 8 modules hybrid PKS-NRPS, as well as enzymes involve in the modifying and regulation (Fig. 3C).

### 3.3.4. Thailanstatins

Thailanstatins (Fig.2.3B) are natural products produced by *B. thailandensis* MSMB43, belonging FR901464 family inhibiting pre-mRNA splicing [66-67]. The cluster contains 15 genes including

9 module hybrid NRPS/PKS (Fig. 3D), responsible for the synthesis of thailanstatin as well as accessory enzymes and gene involved in the regulation [42]. Thailanstatins are considered as a novel therapeutic candidate due to its potential activities against glucocorticoid receptor splicing process [68] and its potent antiproliferative activities in human cell lines [42].

### 3.3.5. Bactobolins

Bactobolins (Fig.2.4B) are NRPS-PKS hybrids antibiotic produced by *B. thailandensis* E264. They have a potent antibacterial activities against some human pathogen bacteria including *S. aureus* (MRSA) and *V. parahemolyticus* LM5674 with MICs values of < 1 µg/mL [43]. It was shown that bactobolin is regulated by Ac-HSL, the typical quorum sensing signals in proteobacteria. They are tripeptide (Ala-Ala–(OH-Cl2-Val) fused to a C6-polyketide. The gene cluster is spanning about 44 Kb that contain hybrid PKS/NRPS genes including genes synthesis of OH-Cl2-Val, Ala-Ala and the C6 polyketide (Fig. 3E). The cluster contains as well genes involved in regulation, metabolite/product transporters, tailoring reactions and genes of unknown function [43, 69].

### 3.3.6. Glidobactins

Glidobactins (Fig.2.6B) from *Burkholderia* spp. K481-B101 [70] (originally assigned to *Plyangium brachysporum*) also known as cepafungins from *B. cepacia* [44], are cytotoxic acylated tripeptides composed of two unusual amino acids erythro-4- hydroxy-L-lysine and 4(S)-amino-2(E)-pentenoic acid linked to an L-threonine residue [70]. Beside their potential activity as anticancer, glidobactins were shown to have a wide inhibitory action against yeast and fungi [71, 44]. In silico analysis of gene clusters of 48 strains of *Burkholderia*, the cluster involved in the synthesis of glidobactins was found in 18 strains belonging to *pseudomallei* and *mallei* species. This cluster shares the same domain organization to the cluster that previously described by Dudler group [70]. The gene cluster (Fig. 3F), composed of NRPSs and PKSs genes, containing genes named *glbA* to *glbH* involved in the synthesis of glidobactins, has been identified in the genome of *Burkholderia* K481- B101, *B. mallei* and *B. pseudomallei* strains [70].

### 3.3.7. Rhizoxin and rhizonin

Rhizoxin (Fig.2.7B) isolated from *Pseudomonas fluorescens* Pf-5 [72] and *Burkholderia rhizoxina* [73], is a potent macrocyclic polyketide which plays a key role on rice seedling blight. Production of rhizoxin by *Pseudomonas fluorescens* Pf-5

Rhizoxin has a strong antimitotic activity and is considered as a potential antitumour drug candidate [74]. The gene cluster spanning over ~81 kb, contains ten ORFs (*rhiA*– *rhiJ)* including five NRPS-PKS modules that are directly involved in the synthesis of rhizoxin [45].

Rhizonin (Fig.2.8B) is a cyclic nonribosomal heptapeptide isolated from *B. rhizoxina*, a symbiotic bacteria that resides within the fungal mycelium of *Rhizopus microsporus* [46].

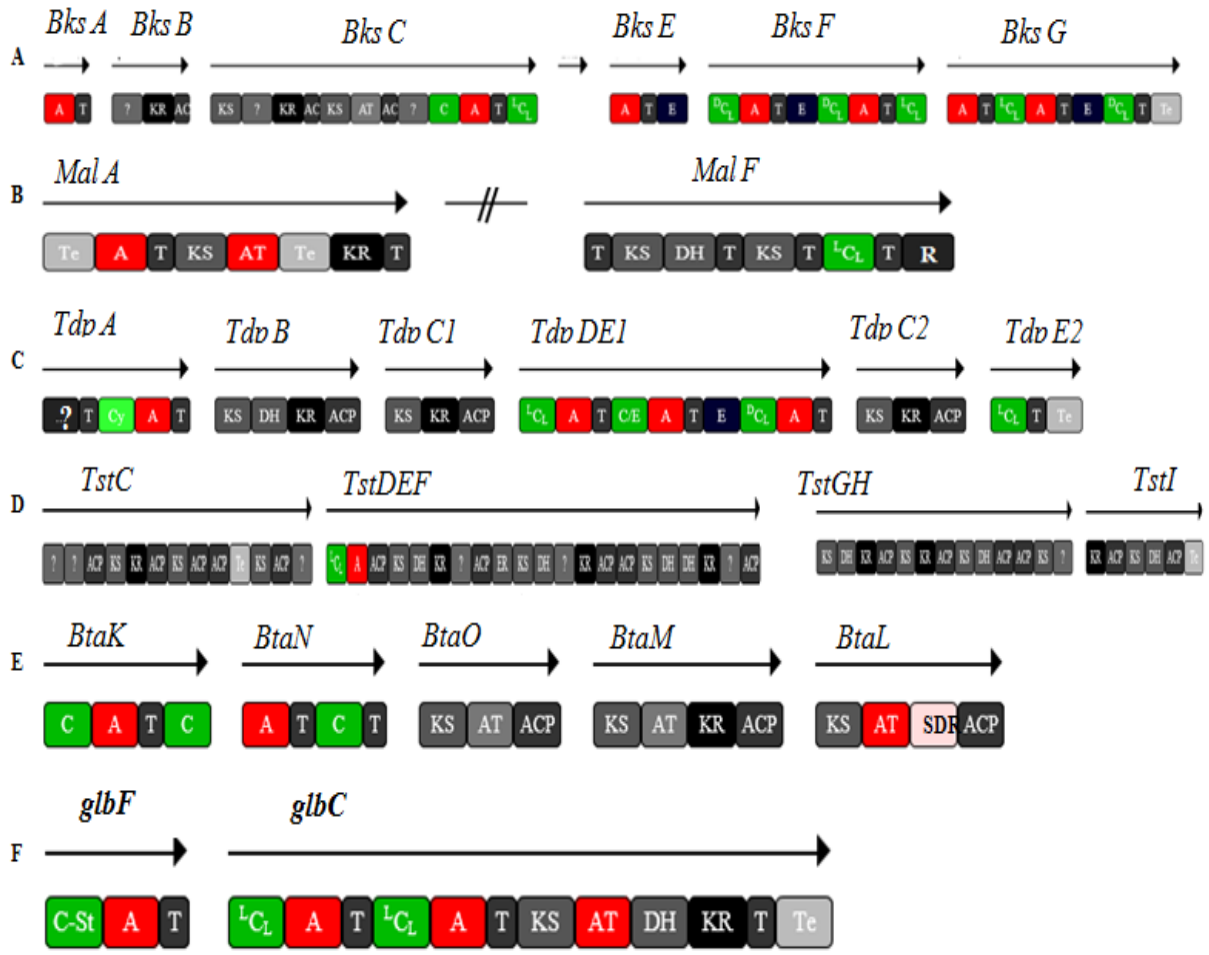



Figure 3: Biosynthesis gene clusters of different NRPS-PKS products of *Burkholderia* A: Lipopeptides presented by burkholdin, B; Malleilactone, C; Thailandpsins, D; Thailanstatins, E; Bactobolin, F; Glidobactin.

A: adenylation domain, Cst: Cstarter for condensation between fatty acid and L-monomer $^{L}C_{L}$: condensation between 2 L-monomers, $^{D}C_{L}$: condensation between D-monomer and L-monomer, KS: ketosynthetase, T: thiolation domain, E: epimerization domain, Te: thioesterase domain, AT: acetytransferase, ACP: acyl carrier protein domain, KR: ketoreductase, DH: dehydratase, R: reductase domain.

### 3.4. Siderophores

Iron is one of the most essential microelement that presents in two states including ferric ($Fe^{3+}$) and ferrous ($Fe^{2+}$). The importance of iron is largely due to its intervention in the synthesis of many vital components and because it is required as a cofactor for many enzymes and iron containing protein [75]. The availability of iron in the environments is limited due to its presence in the insoluble ferric form in oxic environments at neutral pH [76]. Siderophores biosynthesis in bacteria is tightly controlled by a complex network of regulatory enzymes that respond to iron limited conditions. Therefore, microorganisms including bacteria have evolved different mechanisms to transport and control iron concentration inside the cell. Among them siderophores, low molecular weight molecules displaying high affinity to scavenge and solubilize iron from the surrounding environments and transport the ferric iron into the cell [77]. The siderophore with ferric iron form complexes are recognized via specific receptors in the outer membrane and then moved through a periplasmic binding protein to the ABC transporters in the cytoplasmic membrane. Finally the iron is released from complex by reduction of ferric iron and siderophores are recycled by the cells (Fig. 4) [77, 76]. This process is mediated by an energy transduction protein complex including TonB, ExbB, and ExbD [78]. Several siderophores have been described and classified based on their functional groups including catechols, hydroxymates and α-hydoxycarboxylates [79]. Members of *Burkholderia* are well known to overcome iron limitation in the host environment by production of different siderophores that are classified into two different forms including hydroxamates and catecholates based on their chemical structure of their chelating group. In *Burkholderia*, these include ornibactin, malleobactin, pyochelin, and cepaciachelin (Fig.3. C) [15]. The production of siderophores by bacteria can also be beneficial for plants as these compounds can inhibit the growth of pathogenic fungi [80].

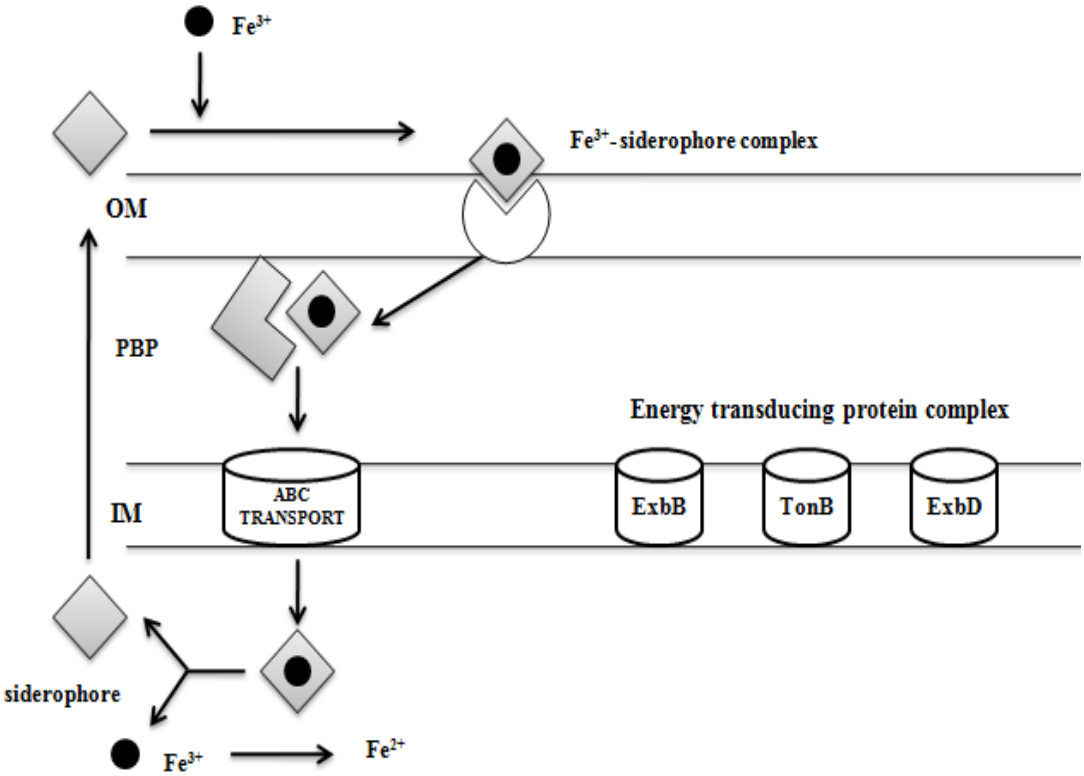



Figure 4: Siderophore uptake in Gram-negative bacteria as *Burkholderia* [76, 78]. OM: Outer membrane, IM: Inner membrane, PBP: Periplasmic binding protein

### 3.4.1. Ornibactin

The siderophores ornibactins (Fig. 2.1C) are linear hydroxamate-hydroxycarboxylates that are composed of four residues. The importance of ornibactin is linked to its essential role of iron transport and growth of member of Bcc strains [47]. As it is important for iron uptake, it is considered as a virulence factor in model organisms that cause chronic lung infections [81]. One NRPS gene cluster involved in the biosynthesis of ornibactin has been identified in *B. cenocepacia* K56-2 [82]. The gene cluster is composed of two NRPS genes *orbI* and *orbJ*. Ornithin N-monooxygenase is present as an accessory enzyme to modify the monomers of the peptidic chain. The cluster contains also genes implicated in the uptake (siderophore export and ferri-siderophore TonB dependent receptor) and the *orbS* genes coding for an extracytoplasmic sigma70 factor implicated in the regulation.

### 3.4.2. Malleobactin

Malleobactin (Fig. 2.3C) is very similar to ornibactin produced by members of *B. pseudomallei*. Both siderophores share the same structure and mechanism of synthesis [48-49]. The biosynthesis gene cluster of malleobactin is composed of two NRPS genes including *mbaI* and *mbaJ*. This cluster harbors the same organization as ornibactin assembly line, as well as the same organization domains for the NRPSs. The difference between the both clusters is first based on the prediction for the first adenylation domain: it is a Leu for ornibactin and Bht (beta-hydroxy-tyrosin) for malleobactin. Secondly, the occurrence of specific genes: two genes annotated "hypothetical protein" (*orbK*) and "ornibactin biosynthesis" (*orbL*) are specific of the ornibactin cluster while another gene annotated "hypothetical protein" (*mbaM*) is specific of the malleobactin one[20].

### 3.4.3. Pyochelin

Pyochelin (Fig. 2.2C) is a secondary siderophore produced by members of *Burkholderia* and *Pseudomonas*. It was first described in *P. aeruginosa* grown under iron limited conditions [83, 48]. It is formed from the condensation of salicylic acid and two molecules of cysteine. The gene cluster responsible for the biosynthesis of pyochelin is composed of two NRPSs of about 1360 and 1890 AA, harbouring a specific domain architecture that includes two cyclization domains (Cy), associated with two stand-alone proteins containing a TE or an A domain (Fig. 2D). It has been shown that pyochelin is always produced as a secondary siderophore together with ornibactin or malleobactin, except for *B. thailandensis* 2002721723 for which no malleobactin synthesis cluster has been found [20]. Pyochelin is present in two interconvertible stereoisomers pyochelin I and II [84] and it has been shown that pyochelin binds iron with a stoichiometry of two pyochelin molecules per ferric iron [85]. Beside its role in iron acquisition, pyochelin was found to be active against some bacterial strains [86].

### 3.4.4. Cepaciachelin

Cepaciachelin (Fig.2.4C) is a catecholate siderophore produced by *B. ambifaria* AMMD [50]. Until now, its role remains an understood. It may play a key role in iron transport and hence the growth of *Burkholderia* strains. In our previous work on *B. ambifaria* AMMD, we have identified the NRPS gene cluster of cepaciachelin that is composed of two NRPS genes, a gene for a TonB-dependent receptor, and genes necessary to build up a diOH-bz (2,3-dihydroxy

benzoic acid) monomer. Additional genes encoding enzymes such as decorating, tailoring or accessory enzymes are also present.

### 3.5. Other NRPS products

BTH-II0204-207 is a potent PDE4 inhibitor isolated from *B. thailandensis* E264 and *B. pseudomallei* K96243 [21]. The NRPS gene cluster was identified in the genome of *B. thailandensis* E264. Four genes spanning from Il0204 to IL0207) were found to include the function required for the synthesis of BTH-II0204-207. These genes are annotated as "nonribosomal peptide synthetase (NRPS)", a "dehydratase", "an isomerase", and "a methyltransferase". The synthesis of BTH-II0204-207 is initiated with two phenylalanine unite predicted by the NRPS genes.

During the genome mining of 48 genomes of *Burkholderia*, we have found that all the strains belonging to *pseudomallei* and *thailandensis* species (except MSMB1212) harbor two specific clusters. The first one found on chromosome 1, includes two NRPSs of 630 AA and 2979 AA with respective architecture [A-T] and [$^{L}C_{L}$-A-T-E-$^{D}C_{L}$-A-T]. The corresponding genes are separated by two genes; one of them is sometimes annotated as a facilitator transporter. There is no well identified TE-domain. The prediction for the produced peptide is Val_ D-Ala_Arg with scores of 60%, 60% and 70%, respectively (Fig.5A). The second one is on chromosome 2 (Fig.5B) and includes three modular megasynthases: one hybrid PKS-NRPS of about 3 000 residues, and two NRPSs of about 1 300 and 1 500 residues. The PKS-NRPS and the following NRPS are separated by a single gene potentially coding for an accessory enzyme of 338 AA, annotated oxygenase or SyrP, similar to the one found in the malleobactin (*mbaH*) and ornibactin (*orbG*) clusters. The architecture was defined as [KS-AT-ACP-?-$^{L}C_{L}$-A-T-E] [$^{L}C_{L}$-A-T-TE] [$^{D}C_{L}$-A-T-$^{L}C_{L}$]. The predicted compound should be PKM_D-Asp_Cys_Gln, regarding the presence of E-domain probably followed by the NRPS starting with $^{D}C_{L}$-domain. This cluster also contains a SyrP-like gene coding for an accessory enzyme of about 350 AA. On chromosome 1, *B. pseudomallei* strains also carries a gene coding for a PKS-NRPS hybrid of about 6274 AA containing one PKS module constituted of CAL and ACP domains, followed by 5 [$^{L}C_{L}$-A-T] modules (Fig. 5C). The cluster may produce an unknown compound with a putative peptidic moiety: Leu_Leu_X_X_Pro, where the "X" indicates that no significant prediction was

obtained. This cluster is present neither in the *B. thailandensis* nor *B. mallei* strains. None of the three peptides share any pattern with the known curated peptides found in Norine.

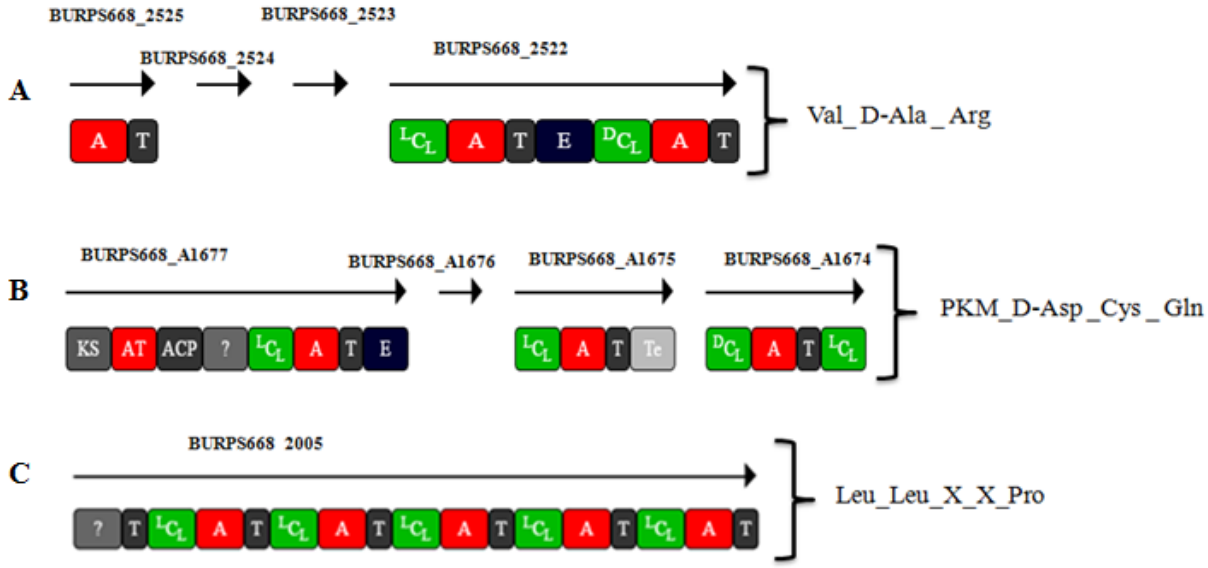



Figure 5: Domains organization of three NRPS gene clusters found in all strains belonging to *pseudomallei* and *thailandensis* species. A: adenylation domain, $^{L}C_{L}$: condensation between 2 L-monomers, $^{D}C_{L}$: condensation between D-monomer and L-monomer, KS: ketosynthetase, T: thiolation domain, E: epimerization domain, Te: thioesterase domain, AT: acetytransferase, PKM represent a monomer incorporated by the PKS domains

### 4. Regulation of *Burkholderia* NRPSs biosynthesis

Biosynthesis of extracellular molecules in bacteria is regulated by different factors including environmental and physiological conditions as well as quorum sensing (QS) which is essential for bacteria species to interact and communicate between each other and regulate gene expression [87]. Biosynthesis of NRPs and PKs in *Burkholderia* is strongly controlled by a complex network of regulatory proteins that respond to specific parameters, including carbon source and nitrogen source. It has been shown that *Burkholderia* uses QS system that relies on LuxI and LuxR represented by *CepI* and *CepR* respectively, to activate several genes that are involved in the production of different extracellular virulence factors [88]. It has been shown that CepI produces two N-acylhomoserine lactones (AHLs) including N-hexanoylhomoserine lactone (C6-HSL) and N-octanoylhomoserine lactone (C8-HSL) which is the most abundant one that

binds to CepR that activates or repress the transcription of target genes[89]. *CepI* was shown to play essential role in the synthesis of NRPs of *Burkholderia*as and the comparison between different strains including wild types and mutants in QS led to the conclusion that QS is associated with production of antifungal peptides [90]and siderophores [88].

**Conclusions**

Microorganisms produce a wide variety of natural secondary metabolites based on their environmental conditions including growth stage and culture media. The search for novel microbial secondary metabolites has been motivated by the need for new activities, as microbial agents can influence the growth of other microorganisms. *Burkholderia* are soil bacteria able to use a wide range of organic compounds and produce a wide range of potent antimicrobial agents. *Burkholderia* strains live in different conditions, and to survive under such conditions, they often produce compounds with unique biological properties that are attractive for researchers. Among these compounds NRPs and PKs that have numerous applications as biocontrol, antibiotics, and anticancer. *Burkholderia* should be investigated in more details as producers of new NRPs and potentially developed as new sources of antimicrobial agents. For that reason, this review aimed to highlight and screen all the biosynthetic gene clusters found in 48 strains of *Burkholderia* that are involved in the synthesis of known and unknown NRPs and PKs as well other products that we did not detected in our analysis are also included.

Several studies have reported the significant effect of NRPs of *Burkholderia* on plant growth promotion. Different molecules have been suggested to account for the plant growth promoting effects of NRPs, such as burkholdins. In addition to their influence on plant growth and development, NRPs have also been shown to inhibit fungal growth. The best well known example of such lipopeptides mediated fungal inhibition is the production of occidiofungin by *B. ambifaria*, whose direct contribution in biocontrol has been demonstrated [57].

Overall, the genus *Burkholderia* constitutes a promising reservoir of new NRPS and PKS gene clusters responsible for the biosynthesis of several secondary metabolites that potentially provide a very useful guidance for future exploration as a new source of natural products.

## References

[1] Mahenthiralingam E, Baldwin A, Dowson CG. *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. J. appl. Microbiol 2008;104:1539-51.

[2] Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, et al. Proposal of *Burkholderia* gen. nov. and Transfer of Seven Species of the Genus *Pseudomonas* Homology Group II to the New Genus, with the Type Species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. Microbiol. Immunol. 1992;36:1251-75.

[3] Mahenthiralingam E, Urban TA, Goldberg JB. The multifarious, multireplicon *Burkholderia cepacia* complex. Nat. Rev. Microbiol. 2005;3:144-56.

[4] Chain PS, Denef VJ, Konstantinidis KT, Vergez LM, Agulló L, Reyes VL, et al. *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. Proc. Natl. Acad. Sci. U. S. A. 2006;103:15280-7.

[5] Jeong Y, Kim J, Kim S, Kang Y, Nagamatsu T, Hwang I. Toxoflavin produced by *Burkholderia glumae* causing rice grain rot is responsible for inducing bacterial wilt in many field crops. Plant Dis 2003;87:890-5.

[6] Lim Y, Suh J-w, Kim S, Hyun B, Kim C, Lee C-h. Cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. II. Physico-chemical properties and structure elucidation. J. Antibiot. 1994;47:1406-16.

[7] Abe M, Nakazawa T. Characterization of hemolytic and antifungal substance, cepalycin, from *Pseudomonas cepacia*. Microbiol. Immunol. 1994;38:1-9.

[8] Caballero-Mellado J, Onofre-Lemus J, Estrada-de los Santos P, Martínez-Aguilar L. The tomato rhizosphere, an environment rich in nitrogen-fixing *Burkholderia species* with capabilities of interest for agriculture and bioremediation. Appl. Environ. Microbiol. 2007;73:5308-19.

[9] Chen W-M, James EK, Coenye T, Chou J-H, Barrios E, De Faria SM, et al. *Burkholderia mimosarum* sp. nov., isolated from root nodules of Mimosa spp. from Taiwan and South America. Int. J. Syst. Evol. Microbiol. 2006;56:1847-51.

[10] Chen W-M, De Faria SM, James EK, Elliott GN, Lin K-Y, Chou J-H, et al. *Burkholderia nodosa* sp. nov., isolated from root nodules of the woody Brazilian legumes *Mimosa bimucronata* and *Mimosa scabrella*. Int. J. Syst. Evol. Microbiol. 2007;57:1055-9.

[11] Coenye T, Vandamme P, Govan JR, LiPuma JJ. Taxonomy and identification of the *Burkholderia cepacia* complex. J. Clin. Microbiol. 2001;39:3427-36.

[12] Cheng AC, Currie BJ. Melioidosis: epidemiology, pathophysiology, and management. Clin. Microbiol. Rev. 2005;18:383-416.

[13] Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T, et al. Structural flexibility in the *Burkholderia mallei* genome. Proc. Natl. Acad. Sci. U. S. A. 2004;101:14246-51.

[14] Czárán TL, Hoekstra RF, Pagie L. Chemical warfare between microbes promotes biodiversity. Proc. Natl. Acad. Sci 2002;99:786-90.

[15] Vial L, Groleau M-C, Dekimpe V, Déziel É. *Burkholderia* diversity and versatility: an inventory of the extracellular products. J. Microbiol. Biotechnol. 2007;17:1407-29.

[16] Franke J, Ishida K, Hertweck C. Evolution of Siderophore Pathways in Human Pathogenic Bacteria. J. Am. Chem. Soc. 2014;136:5599-602.

[17] Biggins JB, Ternei MA, Brady SF. Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of *Burkholderia pseudomallei* group pathogens. J. Am. Chem. Soc. 2012;134:13192-5.

[18] Tawfik KA, Jeffs P, Bray B, Dubay G, Falkinham III JO, Mesbah M, et al. Burkholdines 1097 and 1229, potent antifungal peptides from *Burkholderia ambifaria* 2.2 N. Org. Lett. 2010;12:664-6.

[19] Van Lanen SG, Shen B. Microbial genomics for the improvement of natural product discovery. Curr. Opin. Microbiol. 2006;9:252-60.

[20] Esmaeel Q, Pupin M, Kieu NP, Chataigné G, Béchet M, Deravel J, et al. *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. MicrobiologyOpen 2016.

[21] Biggins JB, Liu X, Feng Z, Brady SF. Metabolites from the induced expression of cryptic single operons found in the genome of *Burkholderia pseudomallei*. J. Am. Chem. Soc. 2011;133:1638-41.

[22] Biggins JB, Gleber CD, Brady SF. Acyldepsipeptide HDAC inhibitor production induced in *Burkholderia thailandensis*. Org. Lett. 2011;13:1536-9.

[23] Caradec T, Pupin M, Vanvlassenbroeck A, Devignes M-D, Smaïl-Tabbone M, Jacques P, et al. Prediction of monomer isomery in Florine: A workflow dedicated to nonribosomal peptide discovery. PLoS ONE 2014;9:e85667.

[24] Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucl. Acids. Res 2015.

[25] Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. PLoS ONE 2012;7:e34064.

[26] Pupin M, Esmaeel Q, Flissi A, Dufresne Y, Jacques P, Leclère V. Norine : a powerful resource for novel nonribosomal peptide discovery. Synth. Syst. Biotechnol 2015.

[27] Marahiel MA. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. J. Pept. Sci. 2009;15:799-807.

[28] Thomson ELS, Dennis JJ. A *Burkholderia cepacia* complex non-ribosomal peptide-synthesized toxin is hemolytic and required for full virulence. Virulence 2012;3:286-98.

[29] Bibb MJ. Regulation of secondary metabolism in *Streptomycetes*. Curr Opin Microbiol 2005;8:208-15.

[30] Balibar CJ, Vaillancourt FH, Walsh CT. Generation of D Amino Acid Residues in Assembly of Arthrofactin by Dual Condensation/Epimerization Domains. Chem Biol 2005;12:1189-200.

[31] Biggins JB, Kang H-S, Ternei MA, DeShazer D, Brady SF. The chemical arsenal of *Burkholderia pseudomallei* is essential for pathogenicity. J. Am. Chem. Soc. 2014;136:9484-90.

[32] Walsh CT. The Chemical Versatility of Natural-Product Assembly Lines. Acc. Chem. Res. 2008;41:4-10.

[33] Hertweck C. The Biosynthetic Logic of Polyketide Diversity. Angew. Chem., Int. Ed. 2009;48:4688-716.

[34] Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 2001;18:380-416.

[35] Lin Z, Falkinham III JO, Tawfik KA, Jeffs P, Bray B, Dubay G, et al. Burkholdines from *Burkholderia ambifaria*: antifungal agents and possible virulence factors. J. Nat. Prod. 2012;75:1518-23.

[36] Bisacchi GS, Hockstein DR, Koster WH, Parker WL, Rathnum ML, Unger SE. Xylocandin: A new complex of antifungal peptides. II. Structural studies and chemical modifications. J. Antibiot. 1987;40:1520-9.

[37] Gu G, Smith L, Wang N, Wang H, Lu S-E. Biosynthesis of an antifungal oligopeptide in *Burkholderia contaminans* strain MS14. Biochem. Biophys. Res. Commun. 2009;380:328-32.

[38] Kang Y, Carlson R, Tharpe W, Schell MA. Characterization of Genes Involved in Biosynthesis of a Novel Antibiotic from *Burkholderia cepacia* BC11 and Their Role in Biological Control of *Rhizoctonia solani*. Appl. Environ. Microbiol 1998;64:3939-47.

[39] Franke J, Ishida K, Hertweck C. Genomics-Driven Discovery of Burkholderic Acid, a Noncanonical, Cryptic Polyketide from Human Pathogenic *Burkholderia* Species. Angew. Chem., Int. Ed. 2012;51:11611-5.

[40] Ishida K, Lincke T, Behnken S, Hertweck C. Induced biosynthesis of cryptic polyketide metabolites in a *Burkholderia thailandensis* quorum sensing mutant. J. Am. Chem. Soc. 2010;132:13966-8.

[41] Wang C, Henkes LM, Doughty LB, He M, Wang D, Meyer-Almes F-J, et al. Thailandepsins: bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. J. Nat. Prod. 2011;74:2031-8.

[42] Liu X, Biswas S, Berg MG, Antapli CM, Xie F, Wang Q, et al. Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative agents from *Burkholderia thailandensis* MSMB43. J. Nat. Prod. 2013;76:685-93.

[43] Seyedsayamdost MR, Chandler JR, Blodgett JAV, Lima PS, Duerkop BA, Oinuma K-I, et al. Quorum-Sensing-Regulated Bactobolin Production by *Burkholderia thailandensis* E264. Org. lett 2010;12:716-9.

[44] SHOJI JI, Hinoo H, Kato T, Hattori T, Hirooka K, Tawara K, et al. Isolation of cepafungins I, II and III from *Pseudomonas* species. J. Antibiot. 1990;43:783-7.

[45] Partida-Martinez LP, Hertweck C. A gene cluster encoding rhizoxin biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus Rhizopus microsporus. ChemBioChem 2007;8:41-5.

[46] Partida-Martinez LP, de Looß CF, Ishida K, Ishida M, Roth M, Buder K, et al. Rhizonin, the first mycotoxin isolated from the zygomycota, is not a fungal metabolite but is produced by bacterial endosymbionts. Appl. Environ. Microbiol. 2007;73:793-7.

[47] Darling P, Chan M, Cox AD, Sokol PA. Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia*. Infect. Immun. 1998;66:874-7.

[48] Thomas MS. Iron acquisition mechanisms of the *Burkholderia cepacia* complex. Biometals 2007;20:431-52.

[49] Franke J, Ishida K, Ishida-Ito M, Hertweck C. Nitro versus hydroxamate in siderophores of pathogenic bacteria: effect of missing hydroxylamine protection in malleobactin biosynthesis. Angew. Chem., Int. Ed. 2013;52:8271-5.

[50] Barelmann I, Meyer J-M, Taraz K, Budzikiewicz H. Cepaciachelin, A New Catecholate Siderophore From *Burkholderia* (*Pseudomonas*) *Cepacia****. Zeitschrift für Naturforschung C 1996;51:627-30.

[51] Ongena M, Jacques P. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 2008;16:115-25.

[52] Rodrigues L, Banat IM, Teixeira J, Oliveira R. Biosurfactants: potential applications in medicine. J Antimicrob Chemother 2006;57:609-18.

[53] Bais HP, Fall R, Vivanco JM. Biocontrol of *Bacillus subtilis* against infection of Arabidopsis roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. Plant Physiol 2004;134:307-19.

[54] Kearns DB, Losick R. Swarming motility in undomesticated *Bacillus subtilis*. Mol Microbiol 2003;49:581-90.

[55] Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin G, Brin C, et al. Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. BMC Genomics 2013;14:658.

[56] Lackner G, Moebius N, Partida-Martinez LP, Boland S, Hertweck C. Evolution of an endofungal Lifestyle: Deductions from the *Burkholderia rhizoxinica* Genome. BMC Genomics 2011;12:1-13.

[57] Lu S-E, Novak J, Austin FW, Gu G, Ellis D, Kirk M, et al. Occidiofungin, a Unique Antifungal Glycopeptide Produced by a strain of *Burkholderia contaminans*. Biochem 2009;48:8312-21.

[58] Ellis D, Gosai J, Emrick C, Heintz R, Romans L, Gordon D, et al. Occidiofungin's Chemical Stability and In Vitro Potency against *Candida* Species. Antimicrob. Agents Chemother. 2012;56:765-9.

[59] Gu G, Smith L, Liu A, Lu S-E. Genetic and Biochemical Map for the Biosynthesis of Occidiofungin, an Antifungal Produced by *Burkholderia contaminans* Strain MS14. Appl. Environ. Microbiol. 2011;77:6189-98.

[60] Meyers E, Bisacchi G, Dean L, Liu W, Minassian B, Slusarchyk D, et al. Xylocandin: A new complex of antifungal peptides. I. Taxonomy, isolation and biological activity. J. Antibiot. 1987;40:1515-9.

[61] Lee C-h, Kim S, Hyun B, Suh J-w, Yon C, Kim C, et al. Cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. I. Taxonomy, production, isolation and biological activity. J. Antibiot. 1994;47:1402-5.

[62] Lee C-H, Kempf H-J, Lim Y, Cho Y-H. Biocontrol activity of *Pseudomonas cepacia* AF2001 and anthelmintic activity of its novel metabolite, cepacidine A. J. Microbiol. Biotechnol. 2000;10:568-71.

[63] LEE C-H, SUH J-W, CHO Y-H. Immunosuppressive activity of cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. J. Microbiol. Biotechnol 1999;9:672-4.

[64] Du L, Sánchez C, Shen B. Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. Metab. Eng. 2001;3:78-95.

[65] VanderMolen KM, McCulloch W, Pearce CJ, Oberlies NH. Romidepsin (Istodax, NSC 630176, FR901228, FK228, depsipeptide): a natural product recently approved for cutaneous T-cell lymphoma. J. Antibiot 2011;64:525-31.

[66] Nakajima H, Hori Y, Terano H, OKUHARA M, MANDA T, MATSUMOTO S, et al. New antitumor substances, FR901463, FR901464 and FR901465. II. Activities against experimental tumors in mice and mechanism of action. J. Antibiot. 1996;49:1204-11.

[67] Nakajima H, Takase S, Terano H, TANAKA H. New antitumor substances, FR901463, FR901464 and FR901465. III. Structures of FR901463, FR901464 and FR901465. J. Antibiot. 1997;50:96-9.

[68] Jain A, Liu X, Wordinger RJ, Yorio T, Cheng Y-Q, Clark AF. Effects of thailanstatins on glucocorticoid response in trabecular meshwork and steroid-induced glaucoma. Invest. Ophthalmol. Visual Sci. 2013;54:3137.

[69] Carr G, Seyedsayamdost MR, Chandler JR, Greenberg EP, Clardy J. Sources of Diversity in Bactobolin Biosynthesis by *Burkholderia thailandensis* E264. Org. Lett. 2011;13:3048-51.

[70] Schellenberg B, Bigler L, Dudler R. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. Environ. Microbiol. 2007;9:1640-50.

[71] Oka M, Nishiyama Y, Ohta S, Kamei H, Konishi M, Miyaki T, et al. Glidobactins A, B and C, new antitumor antibiotics. I. Production, isolation, chemical properties and biological activity. J. Antibiot. 1988;41:1331-7.

[72] Loper JE, Henkels MD, Shaffer BT, Valeriote FA, Gross H. Isolation and identification of rhizoxin analogs from *Pseudomonas fluorescens* Pf-5 by using a genomic mining strategy. Appl Environ Microbiol 2008;74.

[73] Partida-Martinez LP, Hertweck C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. Nat 2005;437:884-8.

[74] Lafontaine JA, Provencal DP, Gardelli C, Leahy JW. Enantioselective total synthesis of the antitumor macrolide rhizoxin D. J. Organic. chem 2003;68:4215-34.

[75] Van Vliet AH, Wooldridge KG, Ketley JM. Iron-responsive gene regulation in a *Campylobacter jejuni* fur mutant. J. Bacteriol 1998;180:5291-8.

[76] Andrews SC, Robinson AK, Rodríguez-Quiñones F. Bacterial iron homeostasis. FEMS Microbiol. Rev. 2003;27:215-37.

[77] Crosa JH. Genetics and molecular biology of siderophore-mediated iron transport in bacteria. Microbiol. Rev. 1989;53:517-30.

[78] Cornelis P. Iron uptake and metabolism in *Pseudomonads*. Appl. Microbiol. Biotechnol 2010;86:1637-45.

[79] Hider RC, Kong X. Chemistry and biology of siderophores. Nat. Prod. Rep. 2010;27:637-57.

[80] Sharma A, Johri B. Growth promoting influence of siderophore-producing *Pseudomonas* strains GRP3A and PRS 9 in maize (*Zea mays* L.) under iron limiting conditions. Microbiol. Res. 2003;158:243-8.

[81] Visser M, Majumdar S, Hani E, Sokol P. Importance of the ornibactin and pyochelin siderophore transport systems in *Burkholderia cenocepacia* lung infections. Infect. Immun. 2004;72:2850-7.

[82] Sokol PA, Darling P, Woods DE, Mahenthiralingam E, Kooi C. Role of Ornibactin Biosynthesis in the Virulence of *Burkholderia cepacia*: Characterization of pvdA, the Gene Encoding l-Ornithine N(5)-Oxygenase. Infect. Immun. 1999;67:4443-55.

[83] Cox CD, Graham R. Isolation of an iron-binding compound from *Pseudomonas aeruginosa*. J. Bacteriol. 1979;137:357-64.

[84] Schlegel K, Taraz K, Budzikiewicz H. The stereoisomers of pyochelin, a siderophore of *Pseudomonas aeruginosa*. Bio*Metals* 2004;17:409-14.

[85] Tseng C-F, Burger A, Mislin GL, Schalk IJ, Yu SS-F, Chan SI, et al. Bacterial siderophores: the solution stoichiometry and coordination of the Fe (III) complexes of pyochelin and related compounds. JBIC, J. Biol. Inorg. Chem. 2006;11:419-32.

[86] Adler C, Corbalán NS, Seyedsayamdost MR, Pomares MF, de Cristóbal RE, Clardy J, et al. Catecholate Siderophores Protect Bacteria from Pyochelin Toxicity. PLoS ONE 2012;7:e46754.

[87] Reading NC, Sperandio V. Quorum sensing: the many languages of bacteria. FEMS Microbiol. Lett. 2006;254:1-11.

[88] Lewenza S, Conway B, Greenberg E, Sokol PA. Quorum sensing in *Burkholderia cepacia*: identification of the LuxRI homologs CepRI. J. Bacteriol 1999;181:748-56.

[89] Huber B, Riedel K, Hentzer M, Heydorn A, Gotschlich A, Givskov M, et al. The cep quorum-sensing system of *Burkholderia cepacia* H111 controls biofilm formation and swarming motility. Microbiol 2001;147:2517-28.

[90] Zhou H, Yao F, Roberts DP, Lessie TG. AHL-deficient mutants of *Burkholderia ambifaria* BC-F have decreased antifungal activity. Curr. Microbiol. 2003;47:0174-9.

# Chapter 4- General Discussions and Conclusions

Over the last years, the importance of non ribosomal compounds produced by microorganisms received very high scientific attentions due to their biological activities including antibiotics, anticancer, and biocontrol agents as lipopeptides produced by *Pseudomonas* (D'Aes et al. 2014) or *Bacillus* (Leclère et al. 2006) to be used as biopesticides to suppress or kill phytopathogenic fungi. Moreover the need for new antimicrobial agents motivates the search of new NRPs to fight the microbial multidrug-resistance. In addition to the discovery of NRPs with useful activities, and to produce large amount of such interesting compounds, it is necessary to study their biosynthetic genes and pathways. This will also be relevant for the industry to improve the production process.

The biosynthetic mechanisms of NRPs have been extensively studied over the few years. Different pathways have been proposed ranging from simple (as a linear) to the most complicated one (as a non linear) (Mootz et al. 2002b). However, the incorporation of different monomers into growing peptidic chain still not yet been clarified. In this context, the deciphering of NRPS gene clusters involved in the synthesis of amonabactins, siderophores produced by *Aeromonas* (Barghouthi et al. 1989a; Barghouthi et al. 1989b), gave deep insight in the comprehension of the mechanism of NRPSs to build natural compounds including siderophores. The amonabactins represents a family of four variants of catechol peptidic siderophores containing Dhb, Lys, Gly, and an aromatic residue (either Trp or Phe) in a D-configuration (Barghouthi et al. 1989a; Barghouthi et al. 1989b). According to the amonabactins structure, several hydroxyl functions harboured by the Dhb monomer are needed in the synthesis to chelate iron (Massad et al. 1994). Until now, the mode of biosynthesis for Dhb-containing siderophore generally involves iterative mechanisms where the modules or domains can be used more than once in the assembly of one single product. This is for example the case for enterobactin produced by *E. coli* (Gehring et al. 1998) and bacillibactin produced by *B. subtilis* (May et al. 2001). The nonribosomal synthesis of Dhb-containing siderophores is rarely linear but the amonabactin synthesis is the more complex use of a single assembly line leading to the co-production of four variants of a siderophore currently described. Thus, the availability of whole genome sequences of 30 *Aeromonas* species combined with bioinformatic tools allowed the identification of NRPS genes including siderophores that are probably associated with the pathogenicity of these bacteria. Based on the metabolic profile of the constructed mutants of *A. hydrophila*, the current study gave new insight into the mechanisms of synthesis of NRPs

through the identification of a unique mode of synthesis with iterative, alternative, and optional use of the domains. It is considered to the first evidence of an optional pathway involved in the introduction of Gly residue to Lys side chain in only 2 over the 4 amonabactins synthesis.

Analyses of domains organization of putative NRPSs of *Burkholderia* gave an interesting new insight into the mechanism of synthetases exemplified by the detection of dual C/E domains in NRPS involved in the production of CLPs by *Burkho*lderia. In CLPs, the incorporation of D-monomers within their active peptides is directed by dual C/E domains that replace the E domains responsible for epimerization in other NRPs (Pauwelyn et al. 2013). In the assembly line, a Dual C/E domain is located directly after a C-A-T module which activates and incorporates an L amino acid. The module which contains the Dual C/E domain also activates an L-amino acid. Then the Dual domain catalyzes the epimerization of the L-residue into D configuration and subsequently promotes the condensation of those two residues. These C-domain sub-type was first found in CLPs of *Pseudomonas* spp. (Balibar et al. 2005) and more recently *Xanthomonas* (Royer et al. 2013). To date, most of all CLPs characterized have been identified from *Pseudomonas* and *Bacillus* species (Roongsawang et al. 2011) and we have shown that *Burkholderia* can now be considered as a source of such biosurfactants putatively useful in biocontrol as weapons against plant disease (Ongena and Jacques 2008).

The understanding of biosynthetic pathways afford new clarification in the comprehension of the potential of the NRPS to build natural compounds and which in turn leads to increase knowledge about the mechanisms of synthetases and facilitate the screening and development of new NRPSs with novel activities.

Microbial genome-mining is a very powerful approach for the discovery of new secondary metabolites as in the case of *Burkholderia*, where we have detected biosynthetic gene clusters for known products such as cepaciachelin, discovered new siderophores (phymabactin) and lipopeptides (burkhomycin). We have also identified strains probably of interest, regarding their potential of secondary metabolite production. However, this kind of screening is still limited due to technological barriers. For example, for some strains we have detected burkhomycin NRPSs split into several proteins, or incomplete clusters, probably due to DNA sequencing errors or problems encountered during genome assembly, emphasized by the modular organization of genes. More accurate predictions could be obtained after re-sequencing of the most relevant clusters for which mis-annotation or mis-prediction is supposed. Anyway,

the annotations we have assigned to some genes (as those for phymabactin, cepaciachelin or burkhomycin productions) will be very helpful to infer the "*Burkholderia* genome database", a specific tool dedicated to cystic fibrosis research community and aimed to facilitate comparative analysis (Winsor et al. 2008). However, the *in silico* prediction have to be comforted by wetlab experiments, and biological activities of the secondary metabolites tested *in vivo* and *Burkholderia* should be investigated in more details as producers of new nonribosomal peptides and potentially developed as new sources of antimicrobial agents, useful for their biocontrol properties. For example, further investigations as biofilm formation, *in vivo* test, and swarming activity should be applied as tests of *B. rhizoxinica* on plant pathosystem as they were previously made with other strains including *Bacillus* (Leclère et al. 2006) or *Pseudomonas* (D'Aes et al. 2014).

Overall, the present study highlights the known products and help to discover a range of new non ribosomal peptides. It showed the significant potential of the genus *Burkholderia* as a promising source of bioactive compounds. The genus was first classified in *Pseudomonas,* another important source of secondary metabolites with useful applications as phenazines (Mavrodi et al. 2006) and viscosin (de Bruijn et al. 2007). Many compounds produced by *Burkholderia* were already known including burkholdins (Lin et al. 2012; Tawfik 2010), occdiofungins (Ellis et al. 2012), cepacidines (Lim 1994), and AFC-BC11 (Kang et al. 1998), but in *silico* approach complete the screening as it allows identification of secondary metabolites gene clusters even if they are not expressed (Corre and Challis 2009). By mining the sequenced genome combining with bioinformatic tools, it is also possible to detect the structure feature of the metabolic products. The knockout of particular gene cluster and subsequent comparison of the metabolic profile between the wild type and the mutant strains is a widely used approach to identify genes that direct the production of known products (Challis and Ravel 2000).

It is interesting to note that a screening can also give new insights on the mechanism of nonribosomal synthetases. Through my work, I pointed out two examples. The first one is the detection of dual C/E domains in NRPSs involved in the production of cyclic lipopeptides by *Burkholderia.* The second one is the identification of a unique use of domains and modules in the pathway responsible for synthesis of amonabactins in *A. hydrophila*. Applying the same kind of screening on other genera may probably lead to the identification of new insights on the NRPS mechanisms, increasing the knowledge on the synthesis of secondary metabolites by modular

enzymes. This can be of great importance in the future development of biotechnologies for the production of compounds useful in different sectors as health or phytosanitary area.

# Chapter 5: Materials and Methods

## 5. A. Microbiological methods

### 5. A.1. Microbial growth media and solutions

All media used in the current study are listed in Table 1. Free iron limited media are listed with siderophores activity section.

Table 1: List and composition of growth media and solution used in this study

| Media | Composition for 1L of medium |
|---|---|
| LB | 10 g tryptone, 5 g yeast extract and 10 g NaCL. pH is adjusted to 7.0 with NaOH. LB agar is prepared by the addition of 17 g of agar. |
| LB sucrose | It is prepared by adding 10% of sucrose (w/v) to LB medium. |
| Blood agar | It is prepared by adding 5 % of horse or sheep blood to LB medium |
| SOB medium | 20 g bactotryptone, 5 g yeast extract, 10 mM NaCl and 2.5 mM KCl |
| SOC medium | 20 mM glucose 20 mM Mg+2 (10 mM MgSO4 + 10 mM MgCl2) to SOB medium |
| MM9 medium | 0.3 g l_1 $KH_2PO_4$, 0.5 $gL^{-1}$, NaCl, 1.0 $gL^{-1}$ , $NH_4Cl$, 6.0 g $L^{-1}$,NaOH and 30.24 g $L^{-1}$Pipes [piperazine-1,4-bis(2-ethanesulphonic acid). |
| Potato dextrose agar (PDA, Difco) | Prepared by suspending PDA 39 g. It is autoclaved at 110°C for 15 min. |
| Yeast Extract-Peptone-Dextrose (YPD) | Bacto-yeast extract10 g, bacto-peptone 20 g, water 900 mL, 20% dextrose 100 mL<br>20% dextrose was autoclaved separately. |
| Sabouraud dextrose agar | 65 g of the medium in 1L of purified water and autoclave at 121°C for 15 minutes |
| Mueller Hinton | 38 g of the medium in 1Lof purified water and autoclave at 121°C for 15 minutes. Muller Hinton agar is prepared by the addition of 17 g of agar |
| Landy medium | 20 g glucose, 5 g glutamic acid, 0.5 g $MgSO_4$. $7H_2O$, 0.5 g KCl, 1 g K2HPO4, 1 g yeast extract , 0.4 mg $Mn(SO_4)3$, 1.2 mg $Fe(SO_4)_3$ and 1.6 mg $CuSO_4$, 100 mM 3-[N-morpholino]-propane sulfonic acid (MOPS) as a buffer. pH adjusted to 7.0 with 5M KOH and medium autoclaved at 110°C for 30 min. Sometimes glucose was replaced by glycerol |
| MGY medium | M9 minimal medium supplemented with 1.25 g $L^{-1}$ yeast extract and 10 g $L^{-1}$ glycerol) |
| TSB M9 | 10 g $L^{-1}$ glycerol, 3 g $L^{-1}$ yeast extract, 15 g $L^{-1}$ tryptone soy broth, M9 salts |
| Bacto tryptic soy broth | 40 g in 1 Lof purified, tryptic soy agar is prepared by the addition of 17 g of agar. For *Burkholderia rhizoxinica* TSB was supplemented with glycerol 10 g $L^{-1}$ and 5% defibrinated sheep blood. |
| Nutrient broth | 5.0 g, meat extract 3.0 g. agar is prepared by the addition of 17 g of agar. pH was adjusted to 7.0. |
| 830 medium | 0.50g Yeast extract, 0.50 g Protease Peptone , 0.50 g Casamino acids, 0.50 g glucose, 0.50g Soluble starch, 0.30 g Na-pyruvate, 0.30 g $K_2HPO_4$, 0.05 g $MgSO_4$ x 7 H2O. PH was adjusted to 7.2 with crystalline $K_2HPO_4$ or $KH_2PO_4$. |

All buffers and solutions used in this study are listed in table 2

Table 2: List of buffers and solutions used

| Buffers &Solutions | **Compositions** |
|---|---|
| MEB buffer | 1 mM $MgCl_2$ and 1mM HEPES, adjust the pH to 7.0; autoclaved at 120 ºC and store at 4ºC |
| TBE buffer | To prepare 10X TBE buffer in 1 liter, dissolve 108 g Tris, 27.5g Boric acid in 800 ml of distilled water. Add 40 mL 0.5M $Na_2EDTA$ (pH 8.0) then adjust volume to 1 liter and store at room temperature. |
| 1M LiAc | 10.2 g LiAc.2 H2O + 80 mL distilled water; adjust pH to 7.5 with acetic acid; adjust volume to 100 mL with distilled water. |
| 1M Tris-HCl | 12.1 g Trizma base + 80 mL distilled water; adjust pH to 8.0 with HCl; adjust volume to 100 mL with distilled water. |
| 0.1M EDTA | 3.72 g ethylenediamine tetraacetic acid, disodium salt + 80 mL distilled water; adjust pH to 7.0 with sodium hydroxide; adjust volume to 100 mL. |
| 50% PEG | 100 g polyethylene glycol 4000; to adjust to 200 mL with distilled water |
| Lazy Bones (40% PLT) | For 200 mL, mix 160 mL 50% PEG stock solution + 20 mL 1M LiAc stock solution + 2 mL Tris-HCl stock solution + 2 mL EDTA stock solution + 16 mL distilled water |
| TE solution | For 100 mL, mix 1 mL 1M Tris-HCl stock solution + 1 mL 0.1M EDTA stock solution + 98 mL distilled water. |
| LiAc/TE solution | For 100 mL, mix 10 mL 1M LiAc stock solution + 1 mL 1M Tris-HCl stock solution + 1 mL 1M EDTA stock solution + 88 mL distilled water. |

## 5. A.2. Microorganisms and culture conditions

Microorganisms used in this study are listed in Table 3. Molds (*Fusarium oxysporum*, *Galactomyces geotrichum*, *Botrytis cinerea* and *Rhizoctonia solani*) were gown on PDA (Sigma-Aldricht, St Louis, MO, U.S.A.). *Burkholderia ambifaria* AMMD wild-type (wt) and its mutants, and *Escherichia coli* were routinely grown at 37°C on Luria Bertani (LB) medium. *Burkholderia rhizoxinica* was grown on different media including Landy, TSB M9, MGY, bacto tryptic soy broth, 830 medium, nutrient broth and LB media. *Candida albicans* was grown on Sabouraud agar at 37°C and *S. cerevisiae* was grown on yeast-extract-peptone-dextrose at 30°C (Shanks et al. 2006). *Aeromonas hydrophila* was routinely grown in LB and *A. salmonicida* in nutrient medium.

Table 3: List of microorganisms used in the current study

| Strains | Relevant characteristics | Reference or source |
|---|---|---|
| *Burkholderia ambifaria* AMMD LMG 19182 | Wild type | LMG 19182 |
| *Burkholderia ambifaria* AMMD-Gm | Single-crossover mutant pMQ30-carrying $Gm^R$, $Suc^S$ | This study |
| *Burkholderia ambifaria* AMMD-Δbamb_6472 | Mutant with deletion of NRPS gene Bamb_6472; $Gm^S$; $Suc^R$ | This study |
| *B. ambifaria* AMMD $Or^-$ | Mutant with disruption of ornibactin cluster | This study |
| *B. ambifaria* AMMD $Cep^-$ | Mutant with disruption of cepaciachelin cluster | This study |
| *B. ambifaria* AMMD $Or^-Cep^-$ | Mutant with disruption of ornibactin and cepaciachelin clusters | This study |
| *Burkholderia rhizoxinica* HKI 454 | Wild type | DSM-19002 |
| *Burkholderia phymatum* STM815 | Wild type | DSM- 17167 |
| *Aeromonas hydrophila* subsp. *hydrophila* | Wild type | CIP |
| *A. hydrophila* ΔamoF | Mutant with disruption of NRPS gene AHA_2474 | This study |
| *Aeromonas hydrophila* ΔamoG | Crossover mutant with pMQ30AHA_2473 | This study |
| *Aeromonas salmonicida* subsp. *salmonicida* | Wild type | DSM-19634 |
| *Staphylococcus aureus* ATCC 25923 | Clinical isolate | ATCC 25923 |
| *Staphylococcus aureus* MethR ATCC 43300 | Isolated from human clinical | ATCC 43300 |
| *Listeria innocua ATCC51742* | Isolated from brain of cow | ProBioGEM lab stock |
| *B.subtilis BBG131* | Surfactin mono-producing | ProBioGEM lab stock |
| *Micrococcus luteus* | Wild type | ProBioGEM lab stock |
| *Fusarium oxysporum* F.sp | Phytopathgen | ProBioGEM lab stock |
| *Galactomyces geotrichum* MUCL2859 | Phytopathgen | ProBioGEM lab stock |
| *Botrytis cinerea* R16 | Phytopathgen | ProBioGEM lab stock |

| | | |
|---|---|---|
| *Rhizoctonia solani* S010-1 | Phytopathgen | ProBioGEM lab stock |
| *Candia albicans* ATCC10231 | Isolated from man with bronchmycosis | ATCC10231 |
| *Saccharomyces cerevisiae* | Wild type | ProBioGEM lab stock |
| *Saccharomyces cerevisiae* InvSc1 | Yeast strain for in vivo recombination *MAT*a/*MAT*α *leu*2/*leu*2 *trp*1-289/*trp*1-289 *ura*3-52/*ura*3-52 *his*3-Δ1/*his*3-Δ1 | Invitrogen |
| ***Escherichia coli*** | | |
| *Escherichia coli* ATCC 25922 | Clinical isolates | ATCC 25922 |
| WM3064 | Donor strain for conjugation | Phytopathology lab, Ghent university |
| WM3064-Bamb_6472 $Gm^R$ | Donor strain with pMQ30 Bamb_6472 $Gm^R$ | This study |
| WM3064 pMQ30 AHA_2473 $Gm^R$ | Donor strain with pMQ30 AHA_2473 $Gm^R$ | This study |
| DH5α | Host for cloning | ProBioGEM lab stock |
| HB101 | Helper strain for conjugation | ProBioGEM lab stock |
| JM109 | High-efficiency competent strain used to the amplification and the conservation of the different plasmids *recA1, endA1, gyrA96, thi, hsdR17 (rK–,mK+), relA1, supE44, Δ(lac-proAB), [F′, traD36, proAB, lacIqZΔM15]* | Promega |

## 5. B. Biological activity assays

### 5. B.1. Antifungal activity

Antifungal activity was demonstrated by contact antagonism assay. PDA plates were inoculated with 5 mm of agar plug of *F. oxysporum*, *G. geotrichum*, *B. cinerae*, and *R. solani* grown for two days on PDA. The plates were inoculated with bacterial culture and incubated at 25°C for 5-7 days. Control plates of fungi were also prepared and the results were scored by comparison to control plates.

### 5. B.2. Anti bacterial and yeast activity

Antibacterial activity was carried out by agar drop plate, disc diffusion, and agar well methods. Anti-yeast activity was performed by the agar drop plate method. Erlenmeyer flask containing 50 ml of different media including PDB, TSB, TSBM9, TSBgly, Landy, 830, and LB were inoculated with 0.5 mL of bacterial suspension containing about $10^8$ CFU $mL^{-1}$ and incubated 5 days under rotary shaking (160 rpm) at 37°C for *B. ambifaria* and 28°C for *B. rhizoxinica*. Then,

the culture was centrifuged at 10,000 g for 10 min at 4°C. The supernatant was concentrated 10 times via speed vacuum, and then 50 µL of this supernatant were deposited on 14 ml solid Sabouraud medium inoculated with target strains including *C. albicans* and *S. cerevisiae*, *M. luteus*, *L. innocua*, *S. aureus* MethR. After 24 h of incubation at 30°C, the activity was measured by the diameter of the zone of inhibition of target strains.

### 5. B.3. Haemolytic activity

The haemolytic activity was assessed at 30°C on blood agar. 50 µL of bacterial supernatant were spotted to the plates that were incubated for 48 h at 30°C. Positive results were indicted by the formation of a clear zone around the wells.

### 5. B.4. Siderophores activity

All free iron media used are listed in table 4. All the glassware were rinsed with 6M HCL then rinsed 10 times with Milli-Q distilled water. *Burkholderia* and *Aeromonas* strains were propagated three times in iron-deficient minimum medium (MM9) (Payne 1994) supplemented with 10 % of casamino acids previously treated, first with 3% 8-hydroxyquinoline in chloroform to remove contaminating iron and then with chloroform to remove remaining 8-hydroxyquinoline. Then, $MgCl_2$ 1 mM and $CaCl_2$ 0.1 mM were added. All glassware was rinsed with HCl 6 M. CAS (Chrome Azurol S) liquid and agar assays were performed in accordance with the original protocol (Schwyn and Neilands 1987). The color of CAS assay solution changed from blue to yellow when iron chelating agent is present in sample solutions. Production of siderophores was also demonstrated with liquid medium of MM9 and CAA inoculated with bacterial suspension of *B. ambifaria* AMMD and its mutants, *B. phymatum*, *A. hydrophila* and its mutants, and *A. salmonicida*. 0.5mL of filtered supernatant of 24h and 48h bacterial culture were mixed with 0.5mLof CAS regent. Colour change from blue to orange or yellow was noted and the production of siderophores was determined by measuring the absorbance at 620 nm and the difference between the first and last absorbance after 5 minutes was noted as ΔOD. MALDI-ToF Mass Spectrometry was also applied to detect the siderophores in supernatant.

Table 4: List of iron free media used

| Media | Composition for 1L of medium |
|---|---|
| King's B (KB) | Proteose peptone 20 g, Glycerol 10 ml, $K2HPO_4$ 1.5 g, $MgSO_4$ 1.5 g, Bactor agar 15 g. |
| Casamino Acid (CAA) | Casamino acid 5 g, K2HPO4 1 g, and Chelex 100 Resin 5 g. pH was adjusted to 7. The medium was stirred during overnight at 4ºC. After autoclaving, MgCl2 was added till final concentration became 400 µM |
| Iron-deficient minimum medium (MM9) | Composed of MM9 medium supplemented with 10 % of casamino acids previously treated, first with 3% 8-hydroxyquinoline in chloroform to remove contaminating iron and then with chloroform to remove remaining 8-hydroxyquinoline. Then, $MgCl_2$ 1 mM and $CaCl_2$ 0.1 mM were added. |
| Chromeazurol (CAS) agar | 60.5 mg of CAS dissolved in 50 ml of deionised water, and mixed with 10 ml of a solution of iron (III) (1mmol $L^{-1}$ FeCl3 _ 6H2O, 10 mmol $L^{-1}$ HCl). While stirring, this solution was slowly mixed with 72.9 mg of hexadecyltrimethylammonium bromide (HDTMA) previously dissolved in 40 mL of water. The resulting dark blue solution was autoclaved, cooled to 50/60 1C and mixed with 900 ml of sterile MM9 containing 15 g $L^{-1}$ of agar |
| CAS liquid medium | Dissolved 4.3 g Piperazine in 30 ml water and adjusted the pH to 5.6 by addition of 6.25 ml concentrated HCl.<br>Dissolved 0.02 g Hexadecyl trimethyl ammonium bromide (HDTMA) in 50 ml dH2O in a separate flask<br>Added 7.5 ml CAS solution, 1.5 ml Fe solution, the piperazine solution and 2 mL 5-Sulfosalicylic acid while stirring and adjusted the volume to 100 ml |

### 5. B. 5. Surface tension measurement

To test whether *B. rhizoxinica* is able to decrease the surface tension through the production of mixed lipopeptides, bacterial strains were grown 72h in landy medium, cells were pelleted and culture supernatants were collected for analysis. Surface tension was measured by the ring method using a tensiometer (TD1, Lauda-Brinkmann) equipped with a Du Nouy ring. Briefly, the ring was placed just below the surface of 20mL of the liquid. Subsequently, the force to move this ring from the liquid phase to the air phase was determined. Surface tension values represent the average of three independent measurements performed at room temperature. A higher concentration in a test sample provides a lower surface tension until critical micelle concentration (CMC) is reached. For the calibration of the instrument, the surface tension of pure water was first measured. *B. subtilis* 131, surfactin mono-production, was also used as control.

### 5. B.6. Purification of culture supernatant

Culture supernatant especially for siderophore production were purified and extracted by C18 cartridges. The peptides absorbed on the C18 were then eluted in 8 mL of different concentrations of methanol or acetonitrile. The eluted peptides were dried by rotating vapour and speed vacuum followed by re-dissolving in 100 µL of methanol or acetonitrile. The final purified supernatant was then tested for further investigation including siderophore activity test, antimicrobial activity as well as LC-MS as well as MALDI-ToF Ms.

### 5. B.7. MALDI-ToF Mass Spectrometry

Mass spectrometry analysis was performed with MALDI-ToF, ToF-ToF, and MS/LC techniques. Supernatant was mixed with a matrix solution (10mg /ml cyano-4-hydroxycinnamic acid in 70% water, 30% acetonitrile, and 0.1% TFA). The samples were homogeneized on a Vortex and centrifuged at 5000 rpm. For classical analysis 1 µl of sample solution was spotted onto a MALDI-ToF MTP 384 target plate (Bruker Daltonik GmbH, Leipzig, Germany) according to the procedure of the dried-droplet preparation. Mass profiles experiments were analyzed with an Ultraflex MALDI-ToF/ToF mass spectrometer (Bruker, Bremen, Germany) equipped with a smartbeam laser. Samples were analyzed using an accelerating voltage of 25 kV and matrix suppression in deflexion mode at m/z 750. The laser power was set to just above the threshold of ionization (around 35%). Spectra were acquired in reflector positive mode in the range of 400 at 3000 Da. Each spectrum was the result of 2000 laser shots per m/z segment per sample delivered in 10 sets of 50 shots distributed in three different locations on the surface of the matrix spot. The instrument was externally calibrated in positive reflector mode using bradykinin [M+H]+ 757.3991, angiotensin II [M+H]+1046.5418, angiotensin I [M+H]+ 1296.6848, substance P [M+H]+ 1347.7354, bombesin [M+H]+ 1619.8223, and ACTH (1-17) [M+H]+ 2093.0862.

## 5. C. Molecular techniques

### 5. C.1. DNA Isolation

DNA was isolated using Wizard®Genomic DNA Purification kit from Promega. Overnight culture of *Burkholderia* and *Aeromonas* strains were grown in LB media. Bacterial cells were harvested by centrifugation and then re-suspended in 600 µL of Nuclei Lysis solution and incubated at 80°C for 5 minutes to lyse the cells. The RNA is digested by RNase. After

incubation at 37ºC for 1 h to degrade any RNA remaining, 200 µL of protein precipitation solution to precipitate protein matter, and the lysate was vortex vigorously for 20 seconds to mix the protein precipitation solution with the cell lysate, a high speed centrifugation was then used to remove proteins and polysaccharides. The cleaned lysate was transferred to a sterile 1.5 mL microtube containing 600 µL of isopropanol. After mixing, the tube was centrifuged for 2 minutes to collect the precipitated genomic DNA. The DNA pellet was washed with 600 µL of 70% ethanol and subjected to an additional centrifugation step to re-pellet the DNA. After removal of all the ethanol, the DNA pellet was air-dried for 10–15 minutes and dissolved in 100 µL of DNA rehydration solution and stored frozen at – 20ºC until use for further investigations. To examine the quality of the isolated DNA, 10 µL was loaded onto a 0.6% agarose gel. The concentrations were determined at 260 nm using NanoDrop (NanoDrop Lite spectrophotometer, Thermo Scientific, USA).

### 5. C.2. Agarose Gel electrophoresis

Agarose was made with Tris-Borate-EDTA (TBE) buffer. The molten agarose was supplemented with safe view dye (Gel Red; 5µL Gel Red/100 mL agarose gel) to enable the DNA samples to be subsequently visualized under UV illumination. A DNA size ladder (10 µL of 1 or 10 kb molecular ladder; Fermentas, Thermo Fisher Scientific, USA.) was also loaded and then, electrophoresis was performed at 100 volts for about 40-60 min in 0.5 x TBE buffer containing Gel Red (5 µLper 100 mL TBE). Then the agarose gel was viewed under UV light (Gel doc, Bio RAD) at 302 nm and images were taken using a gel documentation camera (BioRAD).

### 5. C.3. Polymerase chain reaction (PCR)

PCR Master Mix (2X) from Fermentas, USA, was used as a mixture of Taq DNA polymerase 0.05 units/ µL, reaction buffer, 4 mM MgCl2 and 0.4 mM of each dNTPs. The total volume of each reaction was 20 µL and contained 10 µL of master mix, 3µl sterile nuclease free $H_2O$, 4 µl of Q-Solution, 0.5 µL of each primer (0.5 pmol $\mu L^{-1}$ final concentration) and 2 µL of DNA (10 to 40 ng/ µL). PCR was carried out in a DNA Thermocycler (Labcycler, SensoQuest, Göttingen, Germany) with the following thermal cycles: Initial DNA denaturation at 94°C for 5min, DNA denaturing at 94°C for 1 min, primer annealing 50°C to 60°C for 30 sec. Time of elongation at 72°C depends on every fragment size and a final extension at 72°C for 10 min. PCR was

performed for 30 cycles. PCR products were evaluated by agarose gel electrophoresis as described above. All primers uses in this study are presented in table 5.

### 5. C.4. Restriction Enzyme Digest

For cloning PCR product, the vectors were prepared by cutting with a restriction endonuclease that cleaves DNA at specific sequences. The total volume of each reaction was 20 µL contained 15 µL of sterile nuclease free $H_2O$, 2 µL of restriction enzyme buffers (10X), 1 µL of DNA (0.5-1 µg/µL) and 2 µL of enzymes. The mixture was incubated at 37°C for 1.5 hour and analyzed on a 0.6% agarose gel.

### 5. C.5. Gel Extraction

Insert and vector DNA fragments is usually derived from restriction endonuclease digests and, thus, will be mixed with enzymes, other DNA fragments and salts that may inhibit the ligation reaction. Therefore, it is necessary to purify the DNA fragments prior to ligation. The DNA was purified from agarose gel according to Gel extraction protocol from Fermentas. After electrophoresis in agarose gel, an agarose block containing the fragment of interest is cut out by a scalpel blade. 1:1 volume of binding buffer was added to the gel slice and then incubated at 50-60°C for 10 min or until the gel slice is completely dissolved. The solubilized gel solution was subjected to GeneJET™ purification column and then washed with wash buffer previously diluted with ethanol. After centrifugation, the purified DNA was eluted in TE buffer and stored at -20°C.

Table 5: List of primers used in the current study

| **Primers** | **Primer sequences** **(5'-----> 3')** | Fragment length (bp) | Tm °C |
|---|---|---|---|
| Up6472-F | GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTGCGATCCAGTACCGCGACTAC | 1015 | 60 |
| Up6472-R | TCGGAAGGGAATAGGTCAGCTGATCGGTGACCAGTACGTT | | |
| Down6472-F | AACGTACTGGTCACCGATCAGCTGACCTATTCCCTTCCGA | 887 | 60 |
| Down6472-R | CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATTCCTTTTGCACAGGTTGACG | | |
| UPCepF | GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTGATCAAGGTACCCGTCGTGTT | 905 | 59 |
| UPCepR | CGGACGACGGATAACTGTCCGATGTCCTGTTTCGGCGC | | |
| DownCepF | GCGCCGAAACAGGACATCGGACAGTTATCCGTCGTCCG | 872 | 60 |
| DownCepR | CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATATCATGTGCTGTGCGACCG | | |
| UPOrnF | GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTGGCGCCTTTCCAGAACGAC | 832 | 59 |
| UPOrnR | GCGACATAGCCGATCAGTTGGCCCGTAACCGTTGATCAG | | |
| DownOrnF | CTGATCAACGGTTACGGGCCAACTGATCGGCTATGTCGC | 876 | 58 |
| DownOrnR | CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATCAACGGCGGATGAATCGTC | | |
| AhA1 | TTGTTGACCTATGGCGAGCTGGAG | 787 | 55 |
| AhA5 | GCACTCGGTCGGGCCGTACATG | | |
| AhA1 | TTGTTGACCTATGGCGAGCTGGAG | 1308 | 57 |
| AhA9 | CGGCACCATGTATTCGGGCAG | | |
| AhFo1 | GGGTGGACTGGCTGTTTGCTAC | 1360 | 60 |
| AhRev1 | CAACCGTGGCTCCCTCACTTTC | | |
| AhFo3 | CGCCTGAGCGAGACGAAATAC | 1317 | 58 |
| AhRev3 | CAAAGCCCTGCTCGAAGAACC | | |
| KmSalFo | GGGTTCGAAGTCGACGAG | 1045 | 59 |
| KmSalRev | AAAGGTCGACTTATTCAACAAAGC | | |
| UPAHA2473-F | GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTGTCAGTGAGGGCGGATCAGAT | 800 | 60 |
| UPAHA2473-R | AGTCCACTCCCCAGACCACCCGTAGCAGTTCCACAGC | | |
| DownAHA2472-F | GCTGTGGAACTGCTACGGGTGGTCTGGGGAGTGGACT | 827 | 60 |
| DownAHA2472-R | CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATCAATCCGTTTTCAGCCCAGT | | |
| 16srRNA_F | AGAGTTTGATC(A,C)TGGCTCAG | 1500 | 55 |
| 16srRNA_R | GG(A,C)TACCTTGTTACGA(T,C)TTC | | |

### 5. C.6. Plasmids Isolation

Plasmids were isolated using GeneJET™ Plasmid Miniprep Kit from Fermentas. Overnight culture of bacteria containing the plasmid was grown in a liquid media. Bacterial cells were harvested by centrifugation and then re-suspended in 250 µL of resuspension solution RNase and then treated with 300 µL of lysis solution. After that 250 µL of neutralization solution was added to bring the lysate pH back to normal. A high speed centrifugation was then used to separate the plasmid. Cleared lysate which contain plasmid was subjected to GeneJET™ purification column then washed with wash buffer previously diluted with ethanol. After centrifugation, the plasmid was eluted in TE buffer and stored at -20°C. Plasmids concentrations were measured using NanoDrop (NanoDrop Lite spectrophotometer, Thermo Scientific, USA).

### 5. C.7. Cloning

The purified fragments were cloned to pGEM-T Easy vector (Promega) or other linearized plasmids previously purified. For ligation reaction, the insert and vector were optimized with ratio 3:1with the following equation:

$$\frac{\textbf{Vector (ng) x size of insert (kb)}}{\textbf{size of vector (kb)}} \times \frac{\textbf{insert (3)}}{\textbf{vector molar ratio(1)}} = \textbf{ng of insert}$$

The ligation reaction volume was 10 µL contained 5 µL of 2X ligation buffer, 1 µL of digested Vector (50 ng), x µL of DNA insert, 1 µL of T4 DNA ligase. The final volume was completed to 10 µL with free DNA water. Ligation reactions were mixed and incubated overnight at 16°C for the maximum number of transformants.

### 5. C.8. Heat Shock Transformation

Transformation of plasmid DNA into *E. coli* using the heat shock method is a basic technique of molecular biology. It consists of inserting a ligation product into bacteria. The protocol we used describes the traditional method of transformation from Promega. 2 µL of each ligation reaction was added to 50 µL of competent cells. After incubation for 20 min in ice, a mixture of competent bacteria and ligation reaction placed at 42°C for 45 seconds (heat shock) and then placed back in ice for 2 min. SOC media was then added and the transformed cells were incubated at 37°C for 1.5 h min with agitation and then 100 µL of the resulting culture were

spread on LB plates containing the appropriate antibiotic marker or other selective marker as with pGEMTeasy, Amp 100 µg/mL, IPTG 200 µg/mL, and 20 µg/mL X-Gal. After incubation overnight at 37°C, blue/white screening was used for recombinants. Selected colonies were grown in LB with Amp 100 µg/mL and incubated overnight with agitation. The plasmids were extracted and digested to ascertain the success of transformation.

### 5. C.9. Electroporation

The bacterium was first streaked on LB agar plate and a single colony picked and grown in 10mL of LB medium at 37ºC for overnight with shaking. 2 mL of this culture was transferred to 100 mL of LB medium and incubated at 37ºC with vigorous shaking at 160 rpm till the cell density reached the absorbance of 0.5-1 at 600nm. Then the cells were collected by centrifugation and washed two times with 20 ml of MEB buffer or 10% ice-cold glycerol. Washed cells were resuspended in 100 mL of 10% ice-cold glycerol and transferred into a sterile eppendorf tube. The cell aliquots were either directly applied for electroporation or stored in -80ºC for future use. **2** µL (150 ng) of the plasmid was brought into the competence cells. Mix was placed on ice for 5 min and then transferred in a pre-chilled electroporation cuvette with precaution against bubbles and electroporated using Bio-RAD Gene Pulser set at 2.5 KV, 25 µF and 200 Ω. Electroporated cells were transferred into 900mL of LB containing 2% of glucose and placed on a rotary shaker at 37ºC for 1.5h at 150 rpm. This was followed by plating of cells on LB plates containing the appropriate antibiotic.

## 5. D. Bioinformatic tools

All the available bioinformatic tools including websites, softwares and database were used to analyze the NRPS of complete genomes of *Burkholderia* and *Aeromonas* strains.

### 5. D.1. National centre for Biotechnology Information (NCBI)

It is American National Center of Biotechnology Information advances science and health. It provides a series of databases including GenBank for DNA sequences and PubMed, a bibliographic database for the biomedical literature. All these databases are available online at http://www.ncbi.nlm.nih.gov/genbank/.

### 5. D.2. Basic Local Alignment Search Tool (Blast)

It is a tool used to compare the sequences of different proteins and DNA. Blast provides a powerful way to compare novel sequences with previously characterized genes. It is a useful method for rapid searching of nucleotide and protein databases.

### 5. D.3. Primer 3

It is a free online tool to design and analyze primers for PCR. The development of Primer3 and Primer3 web site was funded by Howard Hughes Medical Institute and by the National Institute of Health and the National Human Genome Research Institute. It is freely accessible at http://frodo.wi.mit.edu/primer3/.

### 5. D.4. NEB cutter

It is one of the services of the New England Bio Labs Insurance. This tool will take a DNA sequence and produce a comprehensive report of the restriction enzymes that will cleave the sequence. It provides a variety of outputs including restriction enzyme maps, theoretical digests and links into the restriction enzyme database. It is freely accessible at http://tools.neb.com/NEBcutter2/.

### 5. D.5. Amplifx software 1.4.4

Amplifx is a tool used to seek in a set of primers, those which can be use to amplify a particular target region and design strategies to screen recombinant clones by PCR. We can add a primer list to Amplifx to know the melting temperatures to design a PCR protocol in a thermocycler. All informations including TM, quality, and length, for a given primer, are automatically calculated by Amplifx.

### 5. D.6. Clone manager software

Clone manager software provides a comprehensive, fully integrated set of tools for primer design and analysis, graphic map drawing, cloning simulation, enzyme operations, global and local sequence alignments, similarity searches, and laboratory-sized sequence assembly projects. It is a quick and easy way to view or edit sequence files, find open reading frames, translate genes, find genes or text in files, or use primer or sequence phrase collections.

### 5. D.7. T-Coffe (ClustalW)

T-Coffe is a free online tool through the European Bioinformatics Institute (EBI) used to align multiple sequences and generate phylogenetic trees. It is free accessible at: http://www.ebi.ac.uk/Tools/msa/tcoffee/

### 5. D.8. NRPs bioinformatic tools

#### 5. D.8.1. Norine

Norine is a database dedicated to nonribosomal peptides (NRPs). It contains more than 1100 peptides and is freely accessible at http://bioinfo.lifl.fr/norine/. For each peptide, it provides its structure, biological activity, producing organisms and bibliographical references. It gives a complete computational tool for systematic study of NRPs in several species, and allows obtaining a better knowledge of these metabolic products.The databases can be queried in order to search for peptides through their annotations as well as through their monomeric structures. In the latter case, the user can specify the composition, the whole structure or a structural pattern (possibly including "undefined monomers") of the searched peptide.

#### 5. D.8.2. PKS- NRPS analysis

It is a project of the Institute of Genome Sciences University of Maryland School of Medicine (Baltimore, MD 21201), aimed to identify and predict the catalytic domain presented in NRPS and PKS genes. It is accessible on; http://nrps.igs.umaryland.edu/nrps/.

#### 5. D.8.3. SBSPKS

This database provides an interface to correlate chemical structures of NRPS/PKS products with the domains and modules in the corresponding polyketide syntheses (PKSs) or non-ribosomal peptide synthetases (NRPSs). The pictorial depictions combined with specific query interfaces facilitate easy identification of various domains and modules from a given polypeptide sequence. NRPS-PKS allows the user to extensively analyze and retrieve the sequence homology of various PKS or NRPS domains providing guidelines for carrying out domain and module swapping experiments. In addition it predicts the specificity of starter and extender precursor units of each potential adenylation domain or acyl transferase. It is accessible on:
http://www.nii.ac.in/~pksdb/sbspks/search_main_pks_nrps.html

### 5. D.8.4. AntiSMASH

The antibiotics and secondary metabolites analysis shell (antiSMASH) is a comprehensive pipeline for the automated mining of finished or draft genome data for the presence of secondary metabolites biosynthetic gene clusters. It is freely accessible at:
http://antismash.secondarymetabolites.org

### 5. D.8.5. NRPS predictor 2

It is a web server for predicting NRPS adenylation domain specificity based on four different levels. The first level, based on physic-chemical properties, the sequences is splitting in three different parts determining whether they are hydrophilic, hydrophobic- aromatic or hydrophobic-aliphatic. Then the sequences are grouped into large and small clusters. The final level predicts which substrate the given NRPS protein will bind. It is accessible on:
http://nrps.informatik.uni-tuebingen.de/Controller?cmd=SubmitJob

### 5. D.8.6. Natural Product Domain Seeker (Napdos)

It is a useful tool for the analysis and rapid detection of secondary metabolites genes. It is designed to detect and extract C- and KS domains from DNA or amino acids sequences. It is helpful to determine the C-domain types whether it is $C_{start}$, $^{D}C_{L}$, $^{L}C_{L}$ or dual C/E domains. The domains are identified based on sequences comparison to a broad set of manually curated reference genes from well characterized chemical pathways. It is accessible on: http://napdos.ucsd.edu/run_analysis.html

# References

The listed references below concern only the introduction, state of the art, discussions, and methodology as well as general conclusions.

Abuhammour W, Hasan RA, Rogers D (2006) Necrotizing fasciitis caused by *Aeromonas hydrophilia* in an immunocompetent child. Pediatr Emerg Care 22(1):48-51

Aizawa T, Ve NB, Vijarnsorn P, Nakajima M, Sunairi M (2010) *Burkholderia acidipaludis* sp. nov., aluminium-tolerant bacteria isolated from Chinese water chestnut (Eleocharis dulcis) growing in highly acidic swamps in South-East Asia. Int J syst Evol Microbiol 60(9):2036-2041 doi:10.1099/ijs.0.018283-0

Anand S, Prasad MR, Yadav G, Kumar N, Shehara J, Ansari MZ, Mohanty D (2010) SBSPKS: structure based sequence analysis of polyketide synthases. Nucl Acids Res:gkq340

Andrews SC, Robinson AK, Rodríguez-Quiñones F (2003) Bacterial iron homeostasis. FEMS Microbiol Rev 27(2-3):215-237 doi:10.1016/s0168-6445(03)00055-x

Bachmann BO, Ravel J (2009) Chapter 8 methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data Methods Enzymol. vol Volume 458. Academic Press, pp 181-217

Baichoo N, Helmann JD (2002) Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence. J Bacteriol 184(21):5826-5832 doi:10.1128/JB.184.21.5826-5832.2002

Baldani J, Caruso L, Baldani VL, Goi SR, Döbereiner J (1997) Recent advances in BNF with non-legume plants. Soil Biol Biochem 29(5):911-922 doi:10.1016/S0038-0717(96)00218-0

Balibar CJ, Vaillancourt FH, Walsh CT (2005) Generation of D Amino Acid Residues in Assembly of Arthrofactin by Dual Condensation/Epimerization Domains. Chem Biol 12(11):1189-1200 doi:10.1016/j.chembiol.2005.08.010

Baranašić D, Zucko J, Diminic J, Gacesa R, Long PF, Cullum J, Hranueli D, Starcevic A (2014) Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. J Ind Microbiol Biotechnol 41(2):461-467

Barelm I MJ-M, Taraz, K, Budzikiewicz H (1996) Cepaciachelin, a New Catecholate Siderophore from Burkholderia (Pseudomonas) cepacia. . Z Naturforsch 51C:627-630

Barghouthi S, Young R, Arceneaux JEL, Byers BR (1989a) Physiological control of amonabactin biosynthesis in *Aeromonas hydrophila*. Biol Metals 2(3):155-160 doi:10.1007/bf01142554

Barghouthi S, Young R, Olson MO, Arceneaux JE, Clem LW, Byers BR (1989b) Amonabactin, a novel tryptophan- or phenylalanine-containing phenolate siderophore in *Aeromonas hydrophila*. J Bacteriol 171(4):1811-1816

Beaz-Hidalgo R, Figueras M (2013) *Aeromonas* spp. whole genomes and virulence factors implicated in fish disease. J Fish Dis 36(4):371-388 doi:doi:10.1111/jfd.12025

Beiderbeck H, Taraz K, Meyer JM (1999) Revised structures of the pyoverdins from *Pseudomonas putida* CFBP 2461 and from *Pseudomonas fluorescens* CFBP 2392. Biometals 12(4):331-338 doi:10.1023/a:1009227520314

Bergendahl V, Linne U, Marahiel MA (2002) Mutational analysis of the C-domain in nonribosomal peptide synthesis. Eur J Biochem 269(2):620-629 doi:10.1046/j.0014-2956.2001.02691.x

Bok JW, Hoffmeister D, Maggio-Hall LA, Murillo R, Glasner JD, Keller NP (2006) Genomic mining for *Aspergillus* natural products. Chem Biol 13(1):31-37 doi:10.1016/j.chembiol.2005.10.008

Caballero-Mellado J, Martínez-Aguilar L, Paredes-Valdez G, Estrada-De Los Santos P (2004) *Burkholderia unamae* sp. nov., an N2-fixing rhizospheric and endophytic species. Int J Syst Evol Microbiol 54(4):1165-1172 doi:10.1099/ijs.0.02951-0

Caballero-Mellado J O-LJ, Estrada-de Los Santos P, Martínez-Aguilar L. (2007) The tomato rhizosphere, an environment rich in nitrogen-fixing *Burkholderia* species with capabilities of interest for agriculture and bioremediation. . Appl Environ Microbiol 73:5308-19 doi:10.1128/AEM.00324-07

Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. J Bacteriol 192(19):5143-5150 doi:10.1128/jb.00315-10

Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, Kucherov G (2008) NORINE: a database of nonribosomal peptides. Nucl Acids Res 36(suppl 1):D326-D331 doi:10.1093/nar/gkm792

Cane DE, Walsh CT, Khosla C (1998) Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. Sci 282(5386):63-68 doi:10.1126/science.282.5386.63

Caradec T, Pupin M, Vanvlassenbroeck A, Devignes M-D, Smaïl-Tabbone M, Jacques P, Leclère V (2014) Prediction of monomer isomery in Florine: A workflow dedicated to nonribosomal peptide discovery. PLoS ONE 9(1):e85667

Carlier AL, Eberl L (2012) The eroded genome of a *Psychotria* leaf symbiont: hypotheses about lifestyle and interactions with its plant host. Environ Microbiol 14(10):2757-2769 doi:10.1111/j.1462-2920.2012.02763.x

Cartwright DK, Chilton W, Benson DM (1995) Pyrrolnitrin and phenazine production by *Pseudomonas cepacia*, strain 5.5 B, a biocontrol agent of *Rhizoctonia solani*. Appl Microbiol Biotechnol 43(2):211-216

Castro-González R, Martínez-Aguilar L, Ramírez-Trujillo A, Estrada-de Los Santos P, Caballero-Mellado J (2011) High diversity of culturable *Burkholderia* species associated with sugarcane. Plant Soil 345(1-2):155-169 doi:10.1007/s11104-011-0768-0

Chain PSG, Denef VJ, Konstantinidis KT, Vergez LM, Agulló L, Reyes VL, Hauser L, Córdova M, Gómez L, González M, Land M, Lao V, Larimer F, LiPuma JJ, Mahenthiralingam E, Malfatti SA, Marx CJ, Parnell JJ, Ramette A, Richardson P, Seeger M, Smith D, Spilker T, Sul WJ, Tsoi TV, Ulrich LE, Zhulin IB, Tiedje JM (2006) *Burkholderia xenovoran*s LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. Proc Nati Acad Sci 103(42):15280-15287 doi:10.1073/pnas.0606924103

Challis GL (2005) A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. ChemBioChem 6(4):601-611 doi:10.1002/cbic.200400283

Challis GL, Ravel J (2000) Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. FEMS Microbiol Lett 187(2):111-114 doi:10.1111/j.1574-6968.2000.tb09145.x

Christian OE, Compton J, Christian KR, Mooberry SL, Valeriote FA, Crews P (2005) Using Jasplakinolide to Turn on Pathways that Enable the Isolation of New Chaetoglobosins from *Phomospis asparagi*. J Nat Prod 68(11):1592-1597 doi:10.1021/np050293f

Cornelis P (2010) Iron uptake and metabolism in *Pseudomonads*. Appl Microbiol Biotechnol 86(6):1637-1645 doi:10.1007/s00253-010-2550-2

Cornelis P, Matthijs S (2002) Diversity of siderophore-mediated iron uptake systems in *fluorescent pseudomonads*: not only pyoverdines. Environ microbiol 4(12):787-798

Corre C, Challis GL (2009) New natural product biosynthetic chemistry discovered by genome mining. Nat Prod Rep 26(8):977-986 doi:10.1039/b713024b

Cosmina P, Rodriguez F, de Ferra F, Grandi G, Perego M, Venema G, van Sinderen D (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol Microbiol 8(5):821-831 doi:10.1111/j.1365-2958.1993.tb01629.x

D'Aes J, Kieu NP, Léclère V, Tokarski C, Olorunleke FE, De Maeyer K, Jacques P, Höfte M, Ongena M (2014) To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain *Pseudomonas* CMR12a. Environ Microbiol 16(7):2282-2300 doi:10.1111/1462-2920.12462

Daher RK, Filion G, Tan SGE, Dallaire-Dufresne S, Paquet VE, Charette SJ (2011) Alteration of virulence factors and rearrangement of pAsa5 plasmid caused by the growth of *Aeromonas salmonicida* in stressful conditions. Vet Microbiol 152(3):353-360 doi:10.1016/j.vetmic.2011.04.034

Dallaire-Dufresne S, Tanaka KH, Trudel MV, Lafaille A, Charette SJ (2014) Virulence, genomic features, and plasticity of *Aeromonas salmonicida* subsp. *salmonicida*, the causative agent of fish furunculosis. Vet Microbiol 169(1):1-7 doi:10.1016/j.vetmic.2013.06.025

Daskalov H (2006) The importance of *Aeromonas hydrophila* in food safety. Food Control 17(6):474-483 doi:10.1016/j.foodcont.2005.02.009

de Bruijn I, de Kock MJ, Yang M, de Waard P, van Beek TA, Raaijmakers JM (2007) Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species. Mol Microbiol 63(2):417-428

Dieckmann R, Lee Y-O, van Liempt H, von Döhren H, Kleinkauf H (1995) Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. FEBS Lett 357(2):212-216 doi:10.1016/0014-5793(94)01342-X

Diminic J, Zucko J, Ruzic IT, Gacesa R, Hranueli D, Long PF, Cullum J, Starcevic A (2013) Databases of the thiotemplate modular systems (CSDB) and their *in silico* recombinants (r-CSDB). J Ind Microbiol Biotechnol 40(6):653-659

Du L, Chen M, Sánchez C, Shen B (2000a) An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. FEMS Microbiol Lett 189(2):171-175 doi:10.1111/j.1574-6968.2000.tb09225.x

Du L, Sánchez C, Chen M, Edwards DJ, Shen B (2000b) The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting

functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem Biol 7(8):623-642 doi:10.1016/s1074-5521(00)00011-9

Du L, Sánchez C, Shen B (2001) Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. Metabo Eng 3(1):78-95 doi:10.1006/mben.2000.0171

Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT (2000) The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. Proc Nati Acadof Sci 97(6):2509-2514 doi:10.1073/pnas.040572897

Ellis D, Gosai J, Emrick C, Heintz R, Romans L, Gordon D, Lu S-E, Austin F, Smith L (2012) Occidiofungin's chemical stability and In vitro potency against *Candida* species. Antimicrob Agents Chemother 56(2):765-769 doi:10.1128/aac.05231-11

Estrada-De Los Santos P, Bustillos-Cristales Ro, Caballero-Mellado J (2001) *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. App Environ Microbiol 67(6):2790-2798 doi:10.1128/AEM.67.6.2790-2798.2001

Finking R, Marahiel MA (2004) Biosynthesis of Nonribosomal Peptides. Annu Rev Microbiol 58(1):453-488 doi:doi:10.1146/annurev.micro.58.030603.123615

Fischbach MA, Walsh CT (2006) Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. Chem Rev 106(8):3468-3496 doi:10.1021/cr0503097

Flissi A, Dufresne Y, Michalik J, Tonon L, Janot S, Noé L, Jacques P, Leclère V, Pupin M (2016) Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. Nucl Acids Res doi:10.1093/nar/gkv1143

Gehring AM, Mori I, Walsh CT (1998) Reconstitution and Characterization of the *Escherichia coli* Enterobactin Synthetase from EntB, EntE, and EntF. Biochem 37(8):2648-2659 doi:10.1021/bi9726584

Gevers W, Kleinkauf H, Lipmann F (1968) The activation of amino acids for biosynthesis of gramicidin S. Pro Nati Acad Sci U S A 60(1):269-276

Gilad J (2007) *Burkholderia mallei* and *Burkholderia pseudomallei*: the causative micro-organisms of glanders and melioidosis. Recent Pat Anti-Infect Drug Discovery 2(3):233-241 doi:10.2174/157489107782497335

Gross H, Stockwell VO, Henkels MD, Nowak-Thompson B, Loper JE, Gerwick WH (2005) The Genomisotopic Approach: A Systematic Method to Isolate Products of Orphan Biosynthetic Gene Clusters. Chem Biol 14(1):53-63 doi:10.1016/j.chembiol.2006.11.007

Grünewald J, Marahiel MA (2006) Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. Microbiol Mol Biol Rev 70(1):121-146 doi:10.1128/mmbr.70.1.121-146.2006

Gudmundsdóttir BK, Björnsdóttir B (2007) Vaccination against atypical furunculosis and winter ulcer disease of fish. Vaccine 25(30):5512-5523 doi:10.1016/j.vaccine.2007.02.009

Haese A, Schubert M, Herrmann M, Zocher R (1993) Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. Mol Microbiol 7(6):905-914 doi:10.1111/j.1365-2958.1993.tb01181.x

Hammer PE, Burd W, Hill DS, Ligon JM, van Pée K-H (1999) Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains. FEMS Microbiol Lett 180(1):39-44 doi:10.1111/j.1574-6968.1999.tb08775.x

Hertweck C (2009) The biosynthetic logic of polyketide diversity. Angew Chem Int Edition 48(26):4688-4716 doi:10.1002/anie.200806121

Hiransuthikul N, Tantisiriwat W, Lertutsahakul K, Vibhagool A, Boonma P (2005) Skin and soft-tissue infections among tsunami survivors in southern Thailand. Clin Infect Dis 41(10):e93-e96 doi:10.1086/497372

Hubbard BK, Walsh CT (2003) Vancomycin Assembly: Nature's Way. Angew Chem Int Ed 42(7):730-765 doi:10.1002/anie.200390202

Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N (2012) DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucl Acids Res:gks1177

Isles A, Maclusky I, Corey M, Gold R, Prober C, Fleming P, Levison H (1984) *Pseudomonas cepacia* infection in cystic fibrosis: an emerging problem. J Pediatri 104(2):206-210

Janda JM, Abbott SL (2010) The Genus *Aeromonas*: Taxonomy, Pathogenicity, and Infection. Clin Microbiol Rev 23(1):35-73 doi:10.1128/cmr.00039-09

Jang JY, Yang SY, Kim YC, Lee CW, Park MS, Kim JC, Kim IS (2013) Identification of orfamide A as an insecticidal metabolite produced by *Pseudomonas protegens* F6. J Agric Food Chem 61(28):6786-6791 doi:10.1021/jf401218w

Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, Yang L, Zechel DL, Ma B, Magarvey NA (2015) An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. Nat Commun 6:8421 doi:10.1038/ncomms9421

Joseph S, Carnahan A, Brayton P, Fanning G, Almazan R, Drabick C, Trudo E, Colwell R (1991) *Aeromonas jandaei* and *Aeromonas veronii* dual infection of a human wound following aquatic exposure. J Clin microbiol 29(3):565-569

Kamra P, Gokhale RS, Mohanty D (2005) SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites. Nucleic Acids Research 33(suppl 2):W220-W225

Kang Y, Carlson R, Tharpe W, Schell MA (1998) Characterization of Genes Involved in Biosynthesis of a Novel Antibiotic from *Burkholderia cepacia* BC11 and Their Role in Biological Control of *Rhizoctonia solani*. Appl Environ Microbiol 64(10):3939-3947 doi:10 3939-3947

Keating TA, Ehmann DE, Kohli RM, Marshall CG, Trauger JW, Walsh CT (2001) Chain Termination Steps in Nonribosomal Peptide Synthetase Assembly Lines: Directed Acyl-S-Enzyme Breakdown in Antibiotic and Siderophore Biosynthesis. ChemBioChem 2(2):99-107 doi:10.1002/1439-7633(20010202)2:2<99::aid-cbic99>3.0.co;2-3

Khayatt BI, Overmars L, Siezen RJ, Francke C (2013) Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. PLoS ONE 8(4):e62136

Knudsen M, Søndergaard D, Tofting-Olesen C, Hansen FT, Brodersen DE, Pedersen CN (2015) Computational discovery of specificity-conferring sites in non-ribosomal peptide synthetases. Bioinf:btv600 doi:10.1093/bioinformatics/btv600

Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT (2001) Generality of Peptide Cyclization Catalyzed by Isolated Thioesterase Domains of Nonribosomal Peptide Synthetases†. Biochem 40(24):7099-7108 doi:10.1021/bi010036j

Kopp F, Marahiel MA (2007) Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. Curr Opin Biotechnol 18(6):513-520 doi:10.1016/j.copbio.2007.09.009

Krewulak KD, Vogel HJ (2008) Structural biology of bacterial iron uptake. Biochimica et Biophysica Acta (BBA)-Biomembranes 1778(9):1781-1804 doi:10.1016/j.bbamem.2007.07.026

Labby KJ, Watsula SG, Garneau-Tsodikova S (2015) Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. Nat Prod Rep 32(5):641-653 doi:10.1039/c4np00120f

Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT (1996) A new enzyme superfamily — the phosphopantetheinyl transferases. Chem Biol 3(11):923-936 doi:10.1016/s1074-5521(96)90181-7

Lange A, Sun H, Pilger J, Reinscheid UM, Gross H (2012) Predicting the Structure of Cyclic Lipopeptides by Bioinformatics: Structure Revision of Arthrofactin. ChemBioChem 13(18):2671-2675 doi:10.1002/cbic.201200532

Leclère V, Marti R, Béchet M, Fickers P, Jacques P (2006) The lipopeptides mycosubtilin and surfactin enhance spreading of *Bacillus subtilis* strains by their surface-active properties. Arch Microbiol 186(6):475-483 doi:10.1007/s00203-006-0163-z

Leclère V, Weber T, Jacques P, Pupin M (2016) Bioinformatics tools for the discovery of new nonribosomal peptides. In: Evans SB (ed) Nonribosomal Peptide and Polyketide Biosynthesis: Methods and Protocols. Springer New York, New York, NY, pp 209-232

Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH (2009) Automated genome mining for natural products. BMC Bioinf 10(1):1 doi:10.1186/1471-2105-10-185

Lim Y, Suh JW, Kim S, Hyun B, Kim C, Lee CH. (1994) Cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. II. Physico-chemical properties and structure lucidation. J Antibiot (Tokyo) 47:1406-16

Lin Z, Falkinham III JO, Tawfik KA, Jeffs P, Bray B, Dubay G, Cox JE, Schmidt EW (2012) Burkholdines from *Burkholderia ambifaria*: antifungal agents and possible virulence factors. J Nat Prod 75(9):1518-1523 doi:10.1021/np300108u

Lipmann F, Gevers, W., Kleinkauf, H., and Roskoski, R. Jr. (1971) Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. Adv Enzymol Relat Areas Mol Biol 35:1 – 34

Liu X-Y, Li C-X, Luo X-J, Lai Q-L, Xu J-H (2014) *Burkholderia* jiangsuensis sp. nov., a methyl parathion degrading bacterium, isolated from methyl parathion contaminated soil. Int J Syst Evol Microbiol 64(9):3247-3253 doi:10.1099/ijs.0.064444-0

Mahenthiralingam E, Baldwin A, Dowson CG (2008a) *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. J appl Microbiol 104(6):1539-1551 doi:10.1111/j.1365-2672.2007.03706.x

Mahenthiralingam E, Baldwin A, Dowson CG (2008b) Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. Journal of Applied Microbiology 104(6):1539-1551 doi:10.1111/j.1365-2672.2007.03706.x

Mahenthiralingam E, Urban TA, Goldberg JB. (2005) The multifarious, multireplicon *Burkholderia cepacia* complex. Nat Rev Microbiol 3:144-56 doi:10.1038/nrmicro1085

Marahiel MA, Stachelhaus T, Mootz HD (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. Chem Rev 97(7):2651-2674 doi:10.1021/cr960029e

Martin-Carnahan A, Joseph SW (2005) *Aeromonadales* ord. nov Bergey's Manual® of Syst Bacteriol. Springer, pp 556-587

Martínez P, Agulló L, Hernández M, Seeger M (2007) Chlorobenzoate inhibits growth and induces stress proteins in the PCB-degrading bacterium *Burkholderia xenovorans* LB400. Arch Microbiol 188(3):289-297

Massad G, Arceneaux JEL, Byers BR (1991) Acquisition of iron from host sources by mesophilic *Aeromonas* species. Microbiol 137(2):237-241 doi:doi:10.1099/00221287-137-2-237

Massad G, Arceneaux JEL, Byers BR (1994) Diversity of siderophore genes encoding biosynthesis of 2,3-dihydroxybenzoic acid in *Aeromonas* spp. Biometals 7(3):227-236 doi:10.1007/bf00149553

Mattos KA, Pádua VL, Romeiro A, Hallack LF, Neves BC, Ulisses TM, Barros CF, Todeschini AR, Previato JO, Mendonça-Previato L (2008) Endophytic colonization of rice (*Oryza sativa* L.) by the diazotrophic bacterium *Burkholderia kururiensis* and its ability to enhance plant growth. An Acad Bras Ciênc 80(3):477-493 doi:10.1590/S0001-37652008000300009

Mavrodi DV, Blankenfeldt W, Thomashow LS (2006) Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation*. Annu Rev Phytopathol 44:417-445 doi:10.1146/annurev.phyto.44.013106.145710

May JJ, Wendrich TM, Marahiel MA (2001) The dhb Operon of *Bacillus subtilis* Encodes the Biosynthetic Template for the Catecholic Siderophore 2,3-Dihydroxybenzoate-Glycine-Threonine Trimeric Ester Bacillibactin. J Biol Chem 276(10):7209-7217 doi:10.1074/jbc.M009140200

McClerren AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. Proc Nati Acad Sci U S A 103(46):17243-17248 doi:10.1073/pnas.0606088103

McNamara CJ, Leff LG (2004) Response of biofilm bacteria to dissolved organic matter from decomposing maple leaves. Microb Ecol 48(3):324-330

Meier JL, Burkart MD (2009) The chemical biology of modular biosynthetic enzymes. Chem Soc Rev 38(7):2012-2045 doi:10.1039/b805115c

Miché L, Faure D, Blot M, Cabanne-Giuli E, Balandreau J (2001) Detection and activity of insertion sequences in environmental strains of *Burkholderia*. Environ Microbiol 3(12):766-773 doi:10.1046/j.1462-2920.2001.00251.x

Miller DA, Walsh CT, Luo L (2001) C-Methyltransferase and Cyclization Domain Activity at the Intraprotein PK/NRP Switch Point of Yersiniabactin Synthetase. J Am Chem Soc 123(34):8434-8435 doi:10.1021/ja016398w

Mootz HD, Kessler N, Linne U, Eppelmann K, Schwarzer D, Marahiel MA (2002a) Decreasing the Ring Size of a Cyclic Nonribosomal Peptide Antibiotic by In-Frame Module Deletion in the Biosynthetic Genes. J Am Chem Soc 124(37):10980-10981 doi:10.1021/ja027276m

Mootz HD, Marahiel MA (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J Bacteriol 179(21):6843-6850

Mootz HD, Schwarzer D, Marahiel MA (2002b) Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. ChemBioChem 3(6):490-504 doi:10.1002/1439-7633(20020603)3:6<490::aid-cbic490>3.0.co;2-n

Moulin L, Munive A, Dreyfus B, Boivin-Masson C (2001) Nodulation of legumes by members of the β-subclass of *Proteobacteria*. Nat 411(6840):948-950

Nguyen T, Ishida K, Jenke-Kodama H, Dittmann E, Gurgui C, Hochmuth T, Taudien S, Platzer M, Hertweck C, Piel J (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat Biotech 26(2):225-233 doi:10.1038/nbt1379

Ongena M, Jacques P (2008) *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16(3):115-125 doi:10.1016/j.tim.2007.12.009

Partida-Martinez LP, Hertweck C (2007) A gene cluster encoding rhizoxin biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. ChemBioChem 8(1):41-45 doi:10.1002/cbic.200600393

Pauwelyn E, Huang C-J, Ongena M, Leclère V, Jacques P, Bleyaert P, Budzikiewicz H, Schäfer M, Höfte M (2013) New linear lipopeptides produced by *Pseudomonas cichorii* SF1-54 are involved in virulence, swarming motility, and biofilm formation. Mol Plant Microbe Interact 26(5):585-598 doi:10.1094/MPMI-11-12-0258-R

Payne SM (1994) Detection, isolation, and characterization of siderophores Methods Enzymol. vol 235. Academic Press, pp 329-344

Perin L, Martínez-Aguilar L, Castro-González R, Estrada-De Los Santos P, Cabellos-Avelar T, Guedes H, Reis V, Caballero-Mellado J (2006) Diazotrophic *Burkholderia* species associated with field-grown maize and sugarcane. App Enviro Microbiol 72(5):3103-3110 doi:10.1128/AEM.72.5.3103-3110.2006

Prieto C, García-Estrada C, Lorenzana D, Martín JF (2012) NRPSsp: non-ribosomal peptide synthase substrate predictor. Bioinf 28(3):426-427

Pupin M, Esmaeel Q, Flissi A, Dufresne Y, Jacques P, Leclère V (2015) Norine: A powerful resource for novel nonribosomal peptide discovery. Synth Syst Biotechnol doi:10.1016/j.synbio.2015.11.001

Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT (1998) Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. Biochem 37(6):1585-1595 doi:10.1021/bi9719861

Quan C, Li X, Wang J, Zheng W, Fan S (2009) Optimal control of a batch bioreactor for the production of a novel antifungal substance CF66I. Afr J Biotechnol 8(21)

Rausch C, Hoof I, Weber T, Wohlleben W, Huson D (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evo Biol 7(1):78 doi:10.1186/1471-2148-7-78

Reith ME, Singh RK, Curtis B, Boyd JM, Bouevitch A, Kimball J, Munholland J, Murphy C, Sarty D, Williams J, Nash JHE, Johnson SC, Brown LL (2008) The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. BMC Genomics 9:427-427 doi:10.1186/1471-2164-9-427

Roongsawang N, Washio K, Morikawa M (2011) Diversity of Nonribosomal Peptide Synthetases Involved in the Biosynthesis of Lipopeptide Biosurfactants. Int J Mol Sci 12(1):141-172 doi:10.3390/ijms12010141

Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucl Acids Res 39(Web Server issue):W362-W367 doi:10.1093/nar/gkr323

Royer M, Koebnik R, Marguerettaz M, Barbe V, Robin G, Brin C, Carrere S, Gomez C, Hugelland M, Voller G, Noell J, Pieretti I, Rausch S, Verdier V, Poussier S, Rott P, Sussmuth R, Cociancich S (2013) Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. BMC Genomics 14(1):658 doi:10.1186/1471-2164-14-658

Schamfuß S, Neu TR, van der Meer JR, Tecon R, Harms H, Wick LY (2013) Impact of mycelia on the accessibility of fluorene to PAH-degrading bacteria. Enviro Sci Technol 47(13):6908-6915 doi:10.1021/es304378d

Schauwecker F, Pfennig F, Grammel N, Keller U (2000) Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. Chem Biol 7(4):287-297 doi:10.1016/s1074-5521(00)00103-4

Schuhmacher R, Stoppacher N, Zeilinger S (2007) Peptaibols of *Trichoderma atroviride*: screening, identification, and structure elucidation by liquid chromatography-tandem mass spectrometry. na

Schwarzer D, Finking R, Marahiel MA (2003) Nonribosomal peptides: from genes to products. Nat Prod Rep 20(3):275-287 doi:10.1039/b111145k

Schwarzer D, Marahiel MA (2001) Multimodular biocatalysts for natural product assembly. Naturwiss 88(3):93-101 doi:10.1007/s001140100211

Schwyn B, Neilands JB (1987) Universal chemical assay for the detection and determination of siderophores. Anal Biochem 160(1):47-56 doi:10.1016/0003-2697(87)90612-9

Sessitsch A, Coenye T, Sturz A, Vandamme P, Barka EA, Salles J, Van Elsas J, Faure D, Reiter B, Glick B (2005) *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. Int J Syst Evol Microbiol 55(3):1187-1192 doi:10.1099/ijs.0.63149-0

Shanks RMQ, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA (2006) *Saccharomyces cerevisiae*-based Molecular tool kit for manipulation of genes from Gram-negative bacteria. Appl Environ Microbiol 72(7):5027-5036 doi:10.1128/aem.00682-06

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7(2):285-295 doi:10.1016/S1367-5931(03)00020-6

Silakowski B, Nordsiek G, Kunze B, Blöcker H, Müller R (2000) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga151. Chem Biol 8(1):59-69 doi:10.1016/s1074-5521(00)00056-9

Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster Andrew LH, Wyatt MA, Magarvey NA (2015) Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucl Acids Res 43(20):9645-9662 doi:10.1093/nar/gkv1012

Stachelhaus T, Hüser A, Marahiel MA (1996) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. Chem Biol 3(11):913-921 doi:10.1016/S1074-5521(96)90180-5

Stachelhaus T, Marahiel MA (1995) Modular Structure of Peptide Synthetases Revealed by Dissection of the Multifunctional Enzyme GrsA. J Biol Chem 270(11):6163-6169 doi:10.1074/jbc.270.11.6163

Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA (1998) Peptide Bond Formation in Nonribosomal Peptide Biosynthesis: CATALYTIC ROLE OF THE CONDENSATION DOMAIN. J Biol Chem 273(35):22773-22781 doi:10.1074/jbc.273.35.22773

Stachelhaus T, Mootz HD, Marahiel MA (1999a) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chemistry & Biology 6(8):493-505 doi:10.1016/s1074-5521(99)80082-9

Stachelhaus T, Mootz HD, Marahiel MA (1999b) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol 6(8):493-505 doi:10.1016/s1074-5521(99)80082-9

Stachelhaus T, Walsh CT (2000) Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase†. Biochem 39(19):5775-5787 doi:10.1021/bi9929002

Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures. Nucl Acids Res 36(21):6882-6892

Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. Nat Prod Rep 18(4):380-416 doi:10.1039/a909079g

Štursová M, Žifčáková L, Leigh MB, Burgess R, Baldrian P (2012) Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. FEMS Microbiol Ecol 80(3):735-746 doi:10.1111/j.1574-6941.2012.01343.x

Suárez-Moreno ZR, Caballero-Mellado J, Coutinho BG, Mendonça-Previato L, James EK, Venturi V (2011) Common features of environmental and potentially beneficial plant-associated *Burkholderia*. Microb Ecol 63(2):249-266 doi:10.1007/s00248-011-9929-1

Tawfik K, Jeffs P, Bray B, Dubay G, Falkinham JO, Mesbah M, et al (2010) Burkholdines 1097 and 1229, potent antifungal peptides from *Burkholderia ambifari*a 2.2N. Org Lett 12:664-666 doi:10.1021/ol9029269

Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA (2000) Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. Chem Biol 7(10):753-764 doi:10.1016/s1074-5521(00)00021-1

Tseng CC, Bruner SD, Kohli RM, Marahiel MA, Walsh CT, Sieber SA (2002) Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. Biochem 41(45):13350-13359 doi:10.1021/bi026592a

Van Lanen SG, Shen B (2006) Microbial genomics for the improvement of natural product discovery. Curr Opin Microbiol 9(3):252-260 doi::10.1016/j.mib.2006.04.002

Vanlaere E, Baldwin A, Gevers D, Henry D, De Brandt E, LiPuma JJ, Mahenthiralingam E, Speert DP, Dowson C, Vandamme P (2009) Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. Int J Syst Evol Microbiol 59(1):102-111 doi:10.1099/ijs.0.001123-0

Vanlaere E, Sergeant K, Dawyndt P, Kallow W, Erhard M, Sutton H, Dare D, Devreese B, Samyn B, Vandamme P (2008) Matrix-assisted laser desorption ionisation-time-of of-flight mass spectrometry of intact cells allows rapid identification of *Burkholderia cepacia* complex. J Microbiol Methods 75(2):279-286 doi:10.1016/j.mimet.2008.06.016

Walsh CT (2008) The chemical versatility of natural-product assembly lines. Acc Chem Res 41(1):4-10 doi:10.1021/ar7000414

Weber G, Schörgendorfer K, Schneider-Scherzer E, Leitner E (1994) The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. Curr Genet 26(2):120-125 doi:10.1007/bf00313798

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller 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 Research 43(W1):W237-W243 doi:10.1093/nar/gkv437

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson D, Wohlleben W (2009) CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J Biotechnol 140(1):13-17 doi:10.1016/j.jbiotec.2009.01.007

Winsor GL, Khaira B, Van Rossum T, Lo R, Whiteside MD, Brinkman FSL (2008) The *Burkholderia* Genome Database: facilitating flexible queries and comparative analyses. Bioinf 24(23):2803-2804 doi:10.1093/bioinformatics/btn524

Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, Ezaki T, Arakawa M (1992) Proposal of *Burkholderia* gen. nov. and Transfer of Seven Species of the Genus *Pseudomonas* Homology Group II to the New Genus, with the Type Species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. Microbiol Immunol 36(12):1251-1275 doi:10.1111/j.1348-0421.1992.tb02129.x

Yadav G, Gokhale RS, Mohanty D (2003) Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. J Mol Biol 328(2):335-363 doi:10.1016/S0022-2836(03)00232-8

Zerikly M, Challis GL (2009) Strategies for the discovery of new natural products by genome mining. ChemBioChem 10(4):625-633 doi:10.1002/cbic.200800389

Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The Natural Product Domain Seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE 7(3):e34064