UNIVERSIDADE FEDERAL DE ALFENAS

JOÃO LUIZ BALDIM ZANIN

BIOINFORMATICS APPLIED TO NATURAL PRODUCTS DISCOVERY PROCESSES: SYSTEMATIZATION, BIOSYNTHETIC EVIDENCES, AND ISOLATION OF PROMISING SPECIES.

Alfenas/MG
2016

JOÃO LUIZ BALDIM ZANIN

BIOINFORMATICS APPLIED TO NATURAL PRODUCTS DISCOVERY PROCESSES: SYSTEMATIZATION, BIOSYNTHETIC EVIDENCES, AND ISOLATION OF PROMISING SPECIES.

Thesis presented as part of all requirements for obtaining the PhD title in Chemistry at the Federal University of Alfenas. Field: Organic Chemistry – Natural Products. Advisor: Dr. Marisi Gomes Soares.

Alfenas/MG
2016

Dados Internacionais de Catalogação-na-Publicação (CIP)
Biblioteca Central da Universidade Federal de Alfenas

Zanin, João Luiz Baldim.
Bioinformatics applied to natural products discovery processes: systematization, biosynthetic evidences, and isolation of promising species / João Luiz Baldim Zanin. -- Alfenas/MG, 2016.
160 f.

Orientadora: Marisi Gomes Soares.
Tese (Doutorado em Química) - Universidade Federal de Alfenas, Alfenas, MG, 2016.
Bibliografia.

1. Biossíntese de Peptídeos Independentes de Ácido Nucleico. 2. Biologia Computacional. 3. Betaproteobacteria. 4. Produtos Biológicos. I. Soares, Marisi Gomes. II. Título.

CDD-547

JOÃO LUIZ BALDIM ZANIN

BIOINFORMATICS APPLIED TO NATURAL PRODUCTS DISCOVERY PROCESSES: SYSTEMATIZATION, BIOSYNTHETIC EVIDENCES, AND ISOLATION OF PROMISING SPECIES.

The examining board below approved the manuscript concerning the PhD title, presented as part of the requirements for obtaining a PhD in chemistry at the Federal University of Alfenas. Field: Organic Chemistry – Natural Products

Approved in: 06/05/2016

Prof. Dr. Angel Maurício Castro Gamero
Universidade Federal de Alfenas – Unifal

Profa. Dra. Ana Carolina Barbosa Padovan
Universidade Federal de Alfenas – Unifal

Profa. Dra. Luciana Gonzaga de Oliveira
Universidade Estadual de Campinas – Unicamp

Profa. Dra. Mônica Tallarico Pupo
Universidade de São Paulo - USP

Profa. Dra. Marisi Gomes Soares
Universidade Federal de Alfenas – Unifal

“I dedicate this work to those who believe life is a continuous quest to the real knowledge. Especially, I dedicate it to the people who made me believe.

“Dedico este trabalho a todos que acreditam que a vida é uma eterna busca à sabedoria. E especialmente para as pessoas que me fizeram acreditar.”

AGRADECIMENTOS

Em quatro anos pôde-se vivenciar inúmeras transições. E, se escrevo estes agradecimentos agora, é pelo fato de eu estar perto de mais uma delas. Foram quatro anos de aprendizagem e todas as possibilitades de evolução como ser humano. Todos os sentimentos envolvidos, bons e ruins. Não há como desvencilhar uns dos outros, mas, há como lidar com cada um deles rumo ao crescimento espiritual.

Eu tenho total certeza de que todas as pessoas que conseguiram me tocar intelectualmente, emocionalmente e profissionalmente neste período sentem a minha energia de gratidão perante cada uma delas. O envolvimento profissional torna-nos mais humanos, pois doamos um pouco de nós para outros também em evolução.

A energia que todos que compartilho carga genética depositam em mim é sublime. É uma razão forte para que eu fosse impulsionado em direção ao meu sonho. Curvo-me em direção à vocês para demonstrar meu respeito: Família.

A distância que causei para me encher de experiências e histórias, espero recompensar em momentos mais próximos. Anseio compartilhar essa "tábula meio-cheia" que cultivo diariamente com vocês: Amigos.

Como almejava antes de realizar o processo seletivo de doutorado, queria muito conhecer um pouco da ciência e cultura de lugares que são referência no nosso planeta. Para tanto, não posso deixar de agradecer uma das pessoas mais iluminadas cientificamente que o mundo pôs diante de mim, o Professor Roger G. Linington, da Universidade da California – Santa Cruz (UCSC). Os demais cientistas, do Linington Lab também merecem agradecimento. Vocês me mostraram que sou pequeno demais dentro de um mundo ainda menor perante as probabilidades do universo. Agradeço pelas lições de humildade, amizade e calor humano quando dei mais um passo em busca dos meus sonhos. Um ano em Santa Cruz na California era o que eu precisava para enxergar, e não apenas ver, detalhes sobre a nossa natureza humana. Não posso deixar de agradecer pela amizade, pois vocês foram/são/serão essenciais para minha evolução. E, claro, não posso deixar de lado o tamanho da carga intelectual que trouxe comigo. Quero deixar escrito aqui o quanto sou grato por ter conhecido, vivido e trabalhado com vocês. Obrigado.

Na UNIFAL, houveram inúmeras mudanças, inúmeros contratempos e acertos. Tudo proporcionou o direcionamento e aprendizado necessário para chegar até este ponto. As

pessoas que passaram pelo Laboratótio de Fitoquímica e Química Medicinal foram muitas, as que ficaram também. As que participaram efetivamente foram especiais. Mais que especiais. Serão eternamente amigos todos aqueles que também carregam as mesmas grandezas, incertezas, ideias e ideais. Não coloco aqui os nomes, pois são muitos, mas, cada um que ler, sentirá a energia devida perante suas ações e vivência.

Tive a oportunidade de ser orientado por uma mulher de valores sólidos. De igualdade em direitos sem interferências de gêneros ou estereótipos. Uma mulher que soube direcionar e libertar cientificamente, criticamente e outras demais abrangências. Devo agradecer muito todas as vezes que lutou por mim (e por nós, seus alunos e alunos de outros), defendendo o que é certo. Atitudes como essas contam mais que qualquer palavra de exemplo. Professora Marisi Gomes Soares, obrigado por doar-se de si em prol da minha evolução científica.

Neste contexto, fui orientado também, no momento da defesa desta tese, por professores brilhantes que dedicaram seu tempo trazendo detalhes essenciais para a consolidação desta etapa. Professores Ana Carolina Padovan, Angel Maurício Gamero, Luciana Gonzaga, Mônica Pupo e Marisi Soares, deixo meu muito obrigado pelas contribuições, elogios e críticas. Tudo foi muito bem absorvido e a conclusão desta etapa, na forma desta tese, se faz mais elaborada pelas palavras de cada um de vocês. Obrigado.

Espero, nessa estrada diária da evolução humana, pessoal, profissional e espiritual, que este título de doutor me torne mais humilde com meus semelhantes, mais profissional para meus iguais e que eu possa transmitir todo o conhecimento que adquiri e apaixonar pessoas pela ciência. Talvez este seja o significado, em suma, da imortalidade, fazendo pessoas realmente se apaixonarem pelo que fazem. Que eu consiga, também, entusiasmar alguém como me senti ao desenvolver este trabalho.

Agradeço pelas oportunidades e situações que vivi. Delas trouxe a maturidade para redigir as palavras que aqui escrevo, desejando que amanhã, ou em dez anos, que eu as leia de forma que não reconheça mais quem escreveu agora, pelo fato de ter conseguido amadurecer e evoluir mais. Senão, claro, tudo perderia o propósito.

De um modo geral, sou grato.

ACKNOWLEDGES

In four years it was possible to experience many transitions periods. And if I write down this acknolwledges righ now, is because I’m, in fact, next to one more of them. It took four years of learning and understanding for my evolution as a human being. All the feelings involved, both good and bad. There is no way of separating them from each other, but there is a way of dealing with each one of them toward spiritual growth.

I'm completely sure that all the people who could touch me intellectually, emotionally and professionally in this period will feel my gratitude. Professional and personal involvement makes us more human, because we donate some of us to others also in the evolution processes.

The energy those who share my genetic code deposited in me is sublime. It is a strong reason for me to be following all my dreams and make them come true. I bend over in front of each one of you to show my respect: Family.

The distance caused by me to fill myself with experiences and stories, I hope I could reward you all in closer moments ahead. I long to share this "half-full slate" that growing daily with you: Friends.

As I craved before making the PhD selection process, I wanted to know a little of the science and culture from reference places on our planet. Therefore, I have to thank one of the most enlightened scientifically person that the world has put in my life, Professor Roger G. Linington from the University of California - Santa Cruz (UCSC) as one of them. Other scientists, in the Linington Lab also deserve thanks. You have shown me that I am too small inside an even smaller world against the odds of the universe. I appreciate the lessons of humility, friendship and warmth when I took another step in pursuit of my dreams. A year in Santa Cruz, California was exactly what I needed to see further details of our human nature. All people I met there were essential to my journey. I must thank you all for the friendship. You were/are/will be essential to my development. And of course, I can not ignore the size of the intellectual charge that I brought back to my country. I want to leave written, forever, how grateful I am to have met, lived and worked with you. Thank you.

Here in Brazil, at UNIFAL, there were numerous changes, numerous setbacks and successes. It provided me the direction and the learning needed to get to this point. People who have participated in our daily basis, at the Laboratótio de Fitoquímica e Química

Medicinal. Those who really participated were actually special. More than special, in fact. For all those who also carry the same energy, uncertainties, goals and ideas: you guys will be forever friends of mine.

I had the opportunity to be mentored by a woman of strong values. Equality in rights without interference of gender of stereotypes. A woman who knew how to direct and free scientifically, critically and too many other scopes. I thank you very much for each and every time you fought for me (and for us, students in general), defending what is right. Attitudes like these count more than any example heard from a sentence. Marisi Gomes Soares, thank you for giving me a little bit of your essence in my scientific development.

In this sense, I also was guided, in the exact moment that I defended my thesis, by brilliant professors. They dedicate their time to read, to understand, and to give me essential details about my work in order to finish it in the best way. Ana Carolina Padovan, Angel Maurício Gamero, Luciana Gonzaga, Mônica Pupo e Marisi Soares, thank you very much for contribute, compliments and critics. Everything was absorved and my thesis is improved according to each one of your considerations.

I hope, in this daily road heading to the human, personal, professional and spiritual evolution, this PhD title could make me a humble and more professional to my peers. I hope that I can transmit all the knowledge acquired till this very moment making people fall in love with science as I did. Maybe, this could be the real meaning of immortality, making people find love in what they do.

I acknowledge for all opportunities and situations I've passed through. From them, I brought maturity to write down these words. I hope that, tomorrow, or in ten years, I could not recognize who is writing them right now, due to the fact I will reach another state of evolution higher than now. Otherwise, everything could lose sense...

In all ways as possible, I’m grateful.

RESUMO

Estratégias guiadas por genoma foram utilizadas a fim de examinar o potencial biossintético de microorganismos da classe Betaproteobacteria no âmbito de Produtos Naturais. Uma estratégia capaz de ser expandida para todos os tipos de microorganismos foi criada para estimar as reações enzimáticas das Peptideo Sintetases Não Ribossomais a fim de sistematizar a e analisar suas similaridades biossintéticas. Todas as bases de dados e software *user-friendly* foram adotadas a fim de tornar esta estratégia simples e mais abrangente. Elas foram NCBI, KEGG, NORINE, antiSMASH, Cystoscape, Gitools, MEGA e Clustal. Os resultados tornaram possível a criação de uma stratégia, chamada XPAIRT (e*XPA*ndable *I*dentification of amino acids in non*R*ibosomal peptides *T*endencies) correlacionando pares de peptídeos e seus genomas similares via Jaccard Index e filogenia. Neste contexto, espécies Betaproteobacteria mostraram sintetizar produtos naturais seguindo certa similaridade biossintética na montagem de monômeros para a construção do esqueleto peptídico. Subunidades estruturais tais como asp.ser e orn.ser foram amplamente encontradas. Essas similaridades foram correlacionados gerando índice de similaridade entre espécies e sua distribuição entre genomas semelhantes, que foram nomeados como contribuíntes. Quanto maior a identidade genômica de um cluster de gene biossintético para um produto natural de forma geral, maior a chance de um contribuínte expressar pares similares relativos ao cluster em questão. A partir de análises de contagem de clusteres de genes biossintéticos, pôde-se eleger microorganismos promissores para isolamento de amostras ambientais. Essas análises mostraram que espécies do gênero *Burkholderia* são as mais promissoras quando comparadas a todos os genomas disponíveis da subclasse Betaproteobacteria. Análises genômicas da espécie padrão do gênero, *Burkholderia thailandensis* mostraram que cromossomos 1 e 2, em comparação a uma cepa produtora de antibióticos padrão, *S. coelicolor*, não apresentarem mesmas informações para biossíntese de compostos, mas apresentam similaridades de classes, sendo elas, Terpenos, T1PKS, Bacteriocinas e Peptídeos Não Ribossomais. Todos os resultados não tiveram correlações com os clusteres de *S. coelicolor* evidenciando que *B. thailandensis* apresenta-se promissora para a descoberta de novos compostos. Como espécies do gênero *Burkholderia* foram o principal alvo neste trabalho, um método guiado por genoma foi desenvolvido para isolar tanto quanto possível cepas de amostras ambientais. O método levou em consideração as necessidades básicas de um microorganismo para sobreviver: a) o tipo de microbioma que os microorganismos de interesse se encontram, analizados através de resultados de metagenômica, b) resistência à antibióticos e metais, c) capacidade de metabolizar compostos com papel biológico, d) crescimento celular e nutrientes, e e) variações de pH e crescimento celular. Todas as análises foram cruzadas e os melhores candidatos à composição de meios de culturas celular específicos para o isolamento de microorganismos do gênero *Burkholderia* foram selecionados. A estratégia foi bem-sucedida para diversos tipos de amostras. Estes experimentos excepcionais demonstraram a eficácia na resolução de problemas químico-biológicos auxiliando a análise posterior de novos produtos naturais.

Palavras-chave: Peptídeos não ribossomais. Bioinformática. Betaproteobacteria. Produtos naturais.

ABSTRACT

Genome-guided strategies were applied to examine Betaproteobacteria species potential for the biosynthesis of nonribosomal peptides. A generalizable strategy was created to track similarities in enzymatic reactions of nonribosomal peptides synthetases in order to organize their capability of assembling monomers building the peptides backbones. Databases and user-friendly software were adopted making this strategy a comprehensive one. Databases and software adopted, as well as, NCBI, KEGG, NORINE, antiSMASH, Cystoscape, Gitools, MEGA e Clustal were used for this purpose. Betaproteobacteria species showed to possess biosynthetic similarities in assembling monomers for the peptide backbone of a nonribosomal peptide. These evidences were correlated giving similarities indexes between species and their distribution between similar genomes. Predictions were fragmented in several ways, for example, monomers, pairs and triads. Correlation analyses displayed that pairs it is the best way of tracking similarities. This result turned possible to create a strategy, named XPAIRT (e*XPA*ndable *I*dentification of amino acids in non*R*ibosomal peptides *T*endencies) correlating pairs of peptides and their similar genomes via Jaccard Index and phylogeny. Thought these investigations it was noticed that Betaproteobacteria species generally assemble asp.orn and orn.ser, mainly *Burkholderia* species, among other pairs of peptides. Further analysis showed that species from the genera *Burkholderia* are the most promising ones due to their Biosynthetic Gene Cluster counting for all available *Betaproteobacteria* genomes. These species were further analyzed and a standard strain, *Burkholderia thailandensis,* was used to the identification of intraspecific variation for their biosynthetic potential. A specific study on Biosynthetic Gene Cluster variation was proceeded for discovering disparities between chromosomes 1 and 2, and a standard antibiotic producer strain, *S. coelicolor*. Results showed that *B. thailandensis* have different possibilities for biosynthesizing natural products. Even thought, common classes of compounds such as, Terpenes, Bacteriocins, T1PKS and Nonribosomal Peptides were identified for all strains. As *Burkholderia* species were the main target in this work, a genome-guided method was developed for isolating as much strains as possible from environmental samples. This very method took into account the basic needs for a microorganism to survive: a) the type of microbiome that microorganisms of interest coexist, analyzed through metagenomics, b) resistance to antibiotics and metals, c) ability to metabolize compounds with biological role, d) cell growth related to different nutrients, and e) cell growth under pH variations. The strategy was successful for diverse types of samples. These exceptional experiments are part of a novel way of working with Natural Products, using genomic, bioinformatics and visual statistical analysis in order to access common characteristics and uniqueness of species guiding the search of medically relevant natural products.

Keywords: Nonribosomal peptides. Bioinformatics. Betaproteobacteria. Natural products.

IMAGE INDEX

TABLE INDEX

ABBREVIATION INDEX

ABBREVIATION		MEANING
%GC	-	Percentage of Guanine/Citosine
1492R	-	Reward Primer Starting at the position 1492 of the 16s gene
16s rRNA	-	A Component of he 30S Small Subunit of Prokaryotic Ribosomes
8F	-	Forward Primer Starting at the Position 8 of the 16s Gene
A	-	Adenylation Domain
Ac	-	Acriflavin
Ala	-	Alanine
antiSMASH	-	antibiotics & Secondary Metabolite Analysis SHell
ARDB	-	Antibiotic Resistance Database
Arg	-	Arginine
Asp	-	Aspartic Acid
ASTDR	-	Agency for Toxic Substances and Disease Registry
ATCC	-	American Type Culture Collection
B_04	-	Well B-04 the 96-well plate
B_09	-	Well B-09 the 96-well plate
B4_9	-	Combination of the Duplicates B_04 and B_09
Bac	-	Bacitracin
BGCs	-	Biosynthetic Gene Clusters
BKD	-	*Burkholderia* spp.
BPB	-	Betaproteobacteria
BSA	-	Burkholderia Selection Media A
BSB	-	Burkholderia Selection Media B
BSC	-	Burkholderia Selection Media C
BSD	-	Burkholderia Selection Media D
BSE	-	Burkholderia Selection Media E
BSF	-	Burkholderia Selection Media F
BSG	-	Burkholderia Selection Media G

BSH	-	Burkholderia Selection Media H
BSI	-	Burkholderia Selection Media I
BSJ	-	Burkholderia Selection Media J
BSX	-	Burkholderia Selection from A to J
C	-	Condensation Domains
C&N-Sources	-	Carbon & Nitrogen-Sources
CARD	-	The Comprehensive Antibiotic Resistance Database
Chr1	-	Chromosome 1
Chr2	-	Chromosome 2
Cys	-	Cysteine
dab	-	2,4-Diaminobutyric Acid
DHA	-	Dihydroxyacetone
DSMZ	-	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DNP	-	Dictionary of Natural Products
D-Rib	-	D-Ribose
D-Ser	-	D-Serine
Fac	-	Fusaric Acid
FDA	-	Food and Drug Administration
glu	-	Glutamate
gly	-	Glycine
HGLKS-NRPS-T1PKS	-	Mixed Pathways Derived from Cyanobacterial Heterocyst Glycolipid Biosynthetic Pathways
iBGCs	-	identified Biosynthetic Gene Clusters
JI	-	Jaccard Index
KEGG	-	Kyoto Enciclopedia of Genes and Genomes
leu	-	Leucine
L-Sor	-	L-Sorbose
MGA	-	Microorganisms from MetaGenomic Analysis
MOA	-	Mode of Action
mod	-	Modified Peptide
mRNA	-	messenger RNA
NB	-	Nutrient Broth
NCBI	-	National Center of Biotechnology Information

NCEs	-	New Chemical Entities
NMR	-	Nuclear Magnetic Ressonace
NORINE	-	A platform that includes a database of nonribosomal peptides together with tools for their analysis
NPs	-	Natural Products
NRP	-	Nonribosomal Peptide
nrp.nrp	-	Unassigned Pair of Predictions (POPs)
NRPS	-	Nonribosomal Peptide Synthetase
NRPS-PKS	-	Hybrid System Between Nonsirobomal Peptide Synthetase and Polyketide Synthase
NRPS-T1PKS	-	Tipe 1 PKS in a hybrid system with NRPS machinery
OD	-	Optical Density
ohmal	-	Hydroxy Malate
OHP	-	Hydroxy Proline
orn	-	Ornithine
PBS	-	Phospathe Buffer Solution
PCP	-	Peptidyl Carrier Protein
PCR	-	Polimerase Chain Reaction
pH	-	Potential Hydrogen Ion Concentration
pk	-	Unknown PKS Product
PKS	-	Polyketide Synthase
POCs	-	Pairs of Compounds
POPs	-	Pairs of Prediction
$R_{(aa)}$	-	Unidentified Peptide
$R_{(nrp)}$	-	Peptide Backbone Continuation
RPS	-	Ribosomal Peptide Synthetase
Ser	-	Serine
sp^2	-	Carbon Hybridization State
sp^3	-	Carbon Hybridization State
T	-	Thiolation Domain
Te	-	Thioesterification Domain
Thr	-	Threonine
UNT	-	Untargeted Microorganisms

val	-	Valine
XPAIRT	-	e*XPA*ndable *I*dentification of non*R*ibosomal peptides *T*endencies

CONTENTS

1 INTRODUCTION

Microorganisms have the ability of producing a large variety of Natural Products (NPs) known by their bioactivities (AMINI; TAVAZOIE, 2011; FELNAGLE et al., 2008; NEWMAN; CRAGG, 2012). These chemical entities belong to a wide spectrum of classes, and many of them show anticancer, antibiotic, immunosuppressant and cytostatic effects that are essential for the level of quality of life that humankind has reached (FELNAGLE et al., 2008; NEWMAN; CRAGG, 2012; SCHWARZER; FINKING; MARAHIEL, 2003).

Some of these NPs comprehend one of the most important classes of drugs in clinical use, the nonribosomal peptides (NRPs) (NEWMAN; CRAGG, 2012).

Several compounds will be showed with the purpose of exemplify structures and biological activities of NRPs from microorganisms.

The first well known NRPS-derived molecule is the **penicillin (1)**. Its discovery boosted new studies for novel antibiotics-like compouds due to the appearance of resistance mechanisms. These mechanisms are correlated to constantly increasing interest in discovering NCEs (BRANDT et al., 2014). Due to these resistance mechanisms, diverse structural modifications were proposed creating different classes of betalactams antibiotics.

The **cephalosporins**, also a betalactam compound, exemplified with **cephalosporin C (2)**, were elected as efficient against pathogenic microorganisms, overcoming momentarily the problem observed. **Cephalosporins** are active predominantly agaist Gram-positive bacteria, however, different generations increased its range of activity including Gram-negative bacteria (LLARRULL et al., 2010).

Examples of cyclic NRPs as **vancomycin (3)**, in clinical use for 60 years, are important for health standards reached nowadays. **Vancomycin** is active against Gram-positive aerobic cocci and bacilli. One interesting point about this antibiotic is related to its activity against methicillin-resistant *Staphylococcus aureus* (MRSA) and *S. epidermidis* (MRSE), among others pathogenic strains. **Vancomycin** is an option to patients with allergic reactions to penicillins and cephalosporins and remains the first-line agent against these two resistant strains and their correlated diseases, as well as, bacteremia, endocarditis, osteomyelitis and pneumonia (RUBINSTEIN; KEYNAN, 2014). The lipopeptide antibiotic **daptomycin (4)**, a cyclic NRP, discovered in the late 1980s by researchers at Eli Lilly and

Company had its development ceased due to adverse effects on skeletal muscle such as myalgia and myositis. In 2003, FDA approved this compound with a trade name **Cubicin** for people older than 18 years for Gram-positive bacteria only. Its MOA is related to bacterial cell membrane function (TALLY; DEBRUIN, 2000).

Another example, **bleomycin (5)**, a medication used to treat cancer, was discovered in 1962, isolated from *Streptomyces verticillus* and it is classified as one important medication in a basic health system by Wolrd Health Organization List of Essential Medicines (VITORIA, 2015). **Bleomycin** is a NRPS-PKS hybrid compound used for the threatment of ovarian, testicular, and cervical cancers (CHEN; STUBBE, 2005).

Another famous drug approved by the FDA, **ciclosporin (6, Figure 1)**, is also a NRPS-related compound. Its use is comprehended from organ transplantation preventing rejections to immunossupresion interfering with the activity and growth of T cells (MATSUDA; KOYASU, 2000). **Ciclosporin** was first isolated from the fungus *Tolypocladium inflatum* from soil samples in the late 1960s. Its structure contains 11 amino acids and ciclosporin is also on the WHO Model List of Essential Medicines.

These examples are not classifying compounds due their importance or economic value, the list is huge and here, they were used only with the purpose of illustrate how complex is their architecture. These NRPS-derived compounds present diverse structures, biological activities and complexity, sheding lights to their importance and their characteristic of being target specific designed, once they were optimized for millions of years by evolutionary pressures for their specific biological purpose (KOEHN; CARTER, 2005).

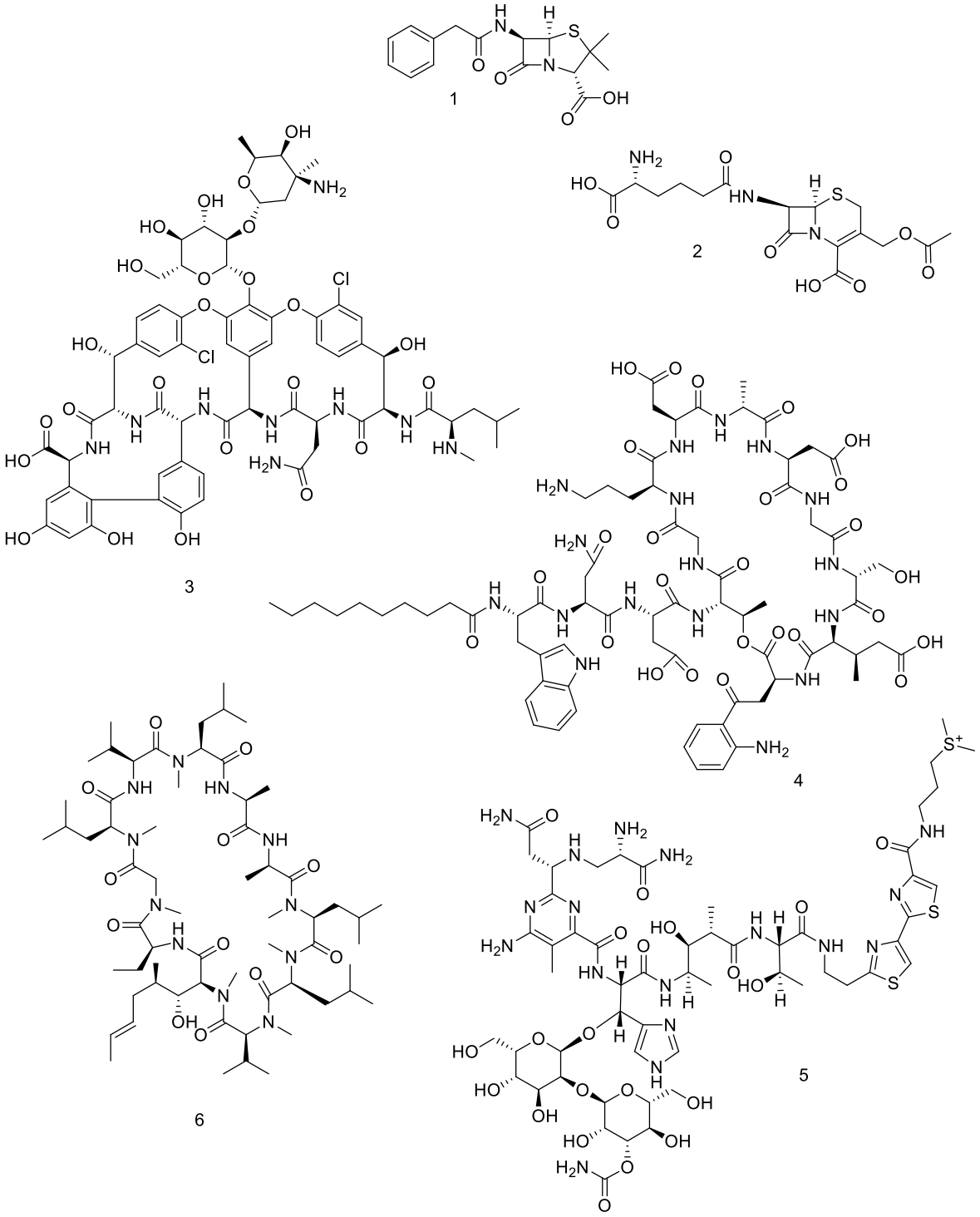

Figure 1- Classic NRPS-derived compounds.

Source: From the author.

These compounds are produced by microorganisms with diverse purpose, including killing competitive species. In the level of genes, whether they have information to produce bacitracins-like antibiotics, for example, they will present naturally, the bacitracin resistance cassette of genes, otherwise, it could cause damage to the producer cell (KANEHISA;

GOTO, 2000; LIU; POP, 2009; MEDEMA et al., 2011). This fact explains how resistance to antimicrobial agents is common in areas under no anthropogenic action since these genes are present in natural reservoirs of resistance in remote environment with no human contact. Moreover, these genes are similar to those encountered under anthropogenic action. It is against these sets of genes, their functions and evolutionary mutations that science have been struggling to improve human health conditions using NPs from diverse microorganisms (AMINI; TAVAZOIE, 2011).

Gradually, diverse scientific fields came together in order to understand and to develop consistently the knowledge about microcosms. For instance, Microbiology, Chemistry, Pharmacology, Genomic Science and Bioinformatics have turned possible to evolve rapidly the most assorted research fields. Thus, historical facts as well as the DNA discovery by Friedrich Miescher and the establishment of microbial culture techniques by Pasteur and Koch, provided the starting point to the level of knowledge existing nowadays. This allowed advances based on the most important discoveries of XX century, penicillin, by Alexander Fleming in 1928, cited above (AMINOV, 2010; SCHLEGEL, HANS G. JANNASCH, 1966). Nonetheless, after large using of penicillin, an effect until then unknown was noticed: the antimicrobial resistance (CASSIR; ROLAIN; BROUQUI, 2014). Since then, this effect started being studied with the aim of developing methods to overcome it (BRANDT et al., 2014).

After all crisis moments in antibiotics discovery, research in the field of NPs are re-emerging using state-of-the-art techniques providing novel strategies and rediscovering the nature's potential of producing biologically active compounds to overcome, in the case of antibiotics, the resistance effect. In this sense, different scientific approaches have provided several improvements to the logics of NPs characterization and biological activities profiles (HARVEY; EDRADA-EBEL; QUINN, 2015; KURITA; LININGTON, 2015; SCHULZE et al., 2013).

Some of these discoveries are made by using genome-mining and spectroscopy-guided strategies providing important new chemical entities (NCEs) via metabolomics approaches (HARVEY; EDRADA-EBEL; QUINN, 2015; NIKOLOULI; MOSSIALOS, 2012). Furthermore, the massive amount of data available along with the evolution of bioinformatics tools, and the increasing number of genomic sequences constantly uploaded allow the discovery of hidden potential of microorganisms for the production of NPs (CRAWFORD; CLARDY, 2012).

In microorganisms, the sets of genes responsible for NPs biosynthesis are chromosomally nearby genes called Biosynthetic Gene Clusters (BGCs) (MEDEMA; FISCHBACH, 2015). Identities between taxonomically close species have shown that their BGCs possess different levels of similarity in the way they produce compounds (CHALLIS; RAVEL; TOWNSEND, 2000; FORSETH et al., 2013; WANG et al., 2014). In this sense, the BGCs related to the production of NRPs and their hybrids are extensively studied due to their medical importance (FELNAGLE et al., 2008; NIKOLOULI; MOSSIALOS, 2012). These characteristics of genes, once structured, guides computational approaches that are applied to the search of novel NPs, including the NRPs-related NPs (MEDEMA; FISCHBACH, 2015; RICHARDS, 2015).

Recently, the potential of bacterial production of NPs was revealed (CIMERMANCIC et al., 2014), and the identified BGCs (called iBGCs in this work) of a large number of species showed that several genus own great likelihood for biosynthesize NPs.

Contradictorily, the total number of iBGCs indicates that there are still a lot to discover, since the number of known compounds to date is very inferior to the predicted capability of microorganisms in biosynthesizing them, implying that great potential remains undiscovered.

According to these evidences, the class Betaproteobacteria (BPB) has emerged as promising class of microorganisms regarding their real biosynthetic capability. Some BPB species considered as neglected producers of antibiotics, as well as, *Pantoea* sp., *Janthinobacterium* sp. and *Burkholderia* sp., provided diverse NRPS-related compounds after genome-mining strategies, presenting interesting biological activities (PIDOT et al., 2014).

These remarkable discoveries in the search of medically relevant NPs have motivated this present work to focus on the idiosyncrasies of iBGCs predictions for NRPs covering all available sequences of BPB species. A total of 359 BPB genomic sequences were studied. They were downloaded, classified and genome-mining analysis were proceeded, giving distribution of iBGCs types, identities of genes, and the prediction of core structures for NRPS-related products.

In order to systematize steps for discovering of NPs the use of bioinformatics tools was adopted creating libraries containing information to correlate species, their biosynthetic

capability, how to isolate species of interest by applying genomic-guided approaches and their features to guide the isolation of compounds.

In this work, all adopted software are classified as user-friendly hoping that this approach could comprehend the most diverse interests. Applications as Gitools (http://www.gitools.org), Cytoscape (http://www.cytoscape.org), and online databases as NCBI (http://www.ncbi.nlm.nih.gov), KEGG (http://www.genome.jp/kegg), AntiSMASH (http://antismash.secondarymetabolites.org), NORINE (http://bioinfo.lifl.fr/norine), CARD (http://arpcard.mcmaster.ca/), and ARDB (http://ardb.cbcb.umd.edu/index.html), among others have their successful place into diverse scientific fields providing rapid and efficient interpretation of huge amount of data from "omics" analysis. In this work, these tools have provided all information for understanding the capability of Betaproteobacteria species in biosynthesize NPs.

By using these tools, specifically for the purpose of discovering where are the greater number of promising BGCs and their products, was possible to choose what microorganisms could be isolated from environment allowing to develop genomic-guided methods to isolate them and predict their biosynthetic characteristics.

Thus, this analyzes demonstrate how bioinformatics tools are used to organize and to create libraries that could be used in order to aid the discovery of novel compounds, since the starting processes: isolation of promising species, and the prediction of biosynthetic capability of a given species, genus, order, or even family of microorganisms.

2 LITERATURE REVIEW

A brief literature review related to nonribosomal peptides, genome-guided analysis and betaproteobacteria species will be discussed ahead.

2.1 NONRIBOSOMAL PEPTIDES

NRPs are synthetized by Nonribosomal Peptide Synthetases (NRPSs) and are often hybrid with Polyketide Synthases (PKS) shaping the NRPS-PKS systems. These NRPSs are groups of specialized enzymes functionalized to activate, bind, and condense monomers building the most diverse types of peptide backbones (CHALLIS; NAISMITH, 2004; CHALLIS; RAVEL; TOWNSEND, 2000; WALSH; O'BRIEN; KHOSLA, 2013). The biological activities inherent to these compounds are constantly allied to some well-defined structural changing, for example, cyclization and/or posttranslational modifications (MCINTOSH; DONIA; SCHMIDT, 2009; PRABAKARAN et al., 2012; WALSH et al., 1997).

The NRPSs condensation domains are related to high level of monomers selectivity, although the so-called assembling line also accepts a specific set of variations in monomers selection (CHALLIS; RAVEL; TOWNSEND, 2000; RAUSCH et al., 2005; SCHAFFER; OTTEN, 2009). This event is described as substrate promiscuity or substrate flexibility and is often related to improvements in biological activities (XIE et al., 2014).

NRPS-like compounds extrapolates the possibilities of Ribosomal Peptides Synthetases (RPSs) in building peptide backbones. RPSs assemble only the twenty proteinogenic amino acids, while NRPSs machinery accepts different types of monomers as well as, non-proteinogenic amino acids, and fatty acids, for example, when hybrid with PKS.

The biosynthesis of NRPS-related NPs relies on monomer being incorporated into a oligomer (FISCHBACH; WALSH, 2006; WALSH, 2015) and its biosynthetic core contains a minimal of three domains (CHALLIS; NAISMITH, 2004). Besides the Peptidyl Carrier Protein (PCP) domain, basically present in each module, that is responsible for carrying the activated peptide to the following enzymatic module, these minimal NRPSs domains are

called chain initiation (adenylation domains (A), responsible for substrate specificity and activation), chain elongation (condensation domains (C), that catalyzes peptide bond formation between a new substrate and the peptide chain. Domains related to chain termination (associated to (T) domains that release the substrate to the cell, breaking the thioester bond) (CHALLIS; NAISMITH, 2004; FISCHBACH; WALSH, 2006). The size and number of domains in a given nonribosomal peptide could be recognized according to the number of monomers in its chemical structure (**Figure 2**) (KEATING; WALSH, 1999; MCINTOSH; DONIA; SCHMIDT, 2009).

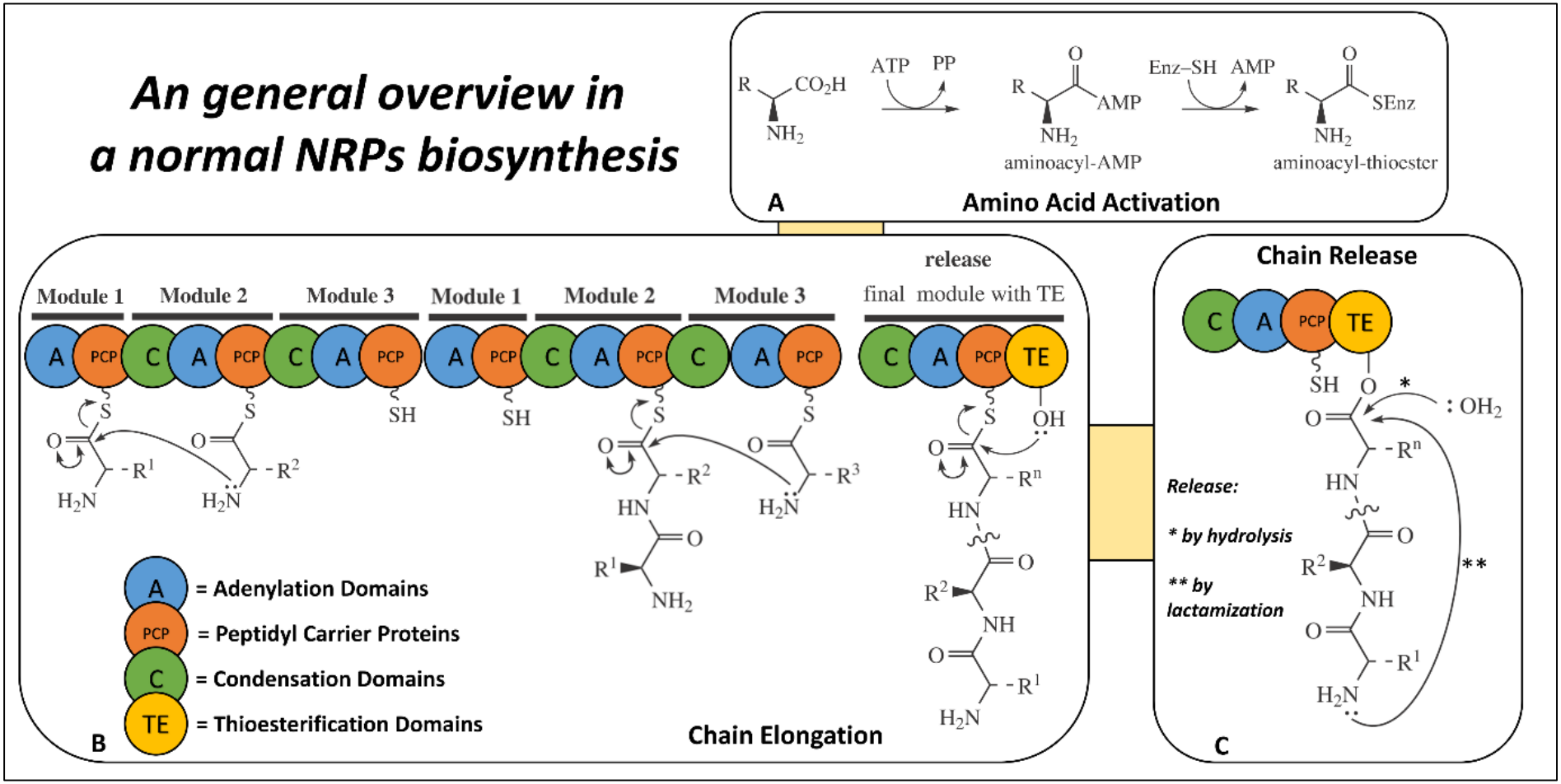

Figure 2- Biosynthesis of NPRS-like compounds. A) Activation of a monomer (peptide or organic acid). B) Growth of the peptide backbone. C) Finishing of the building peptide and two ways of releasing the molecule to the cytoplasm.

Source: From the author.

Specialized domains are present in different levels of modifications. Cyclic peptides into a peptide backbone and small subunits modifications are correlated to enzymatic processes that these amino acids or polyketides moieties undergo. These enzymes build domains similarly in both nonribosomal and ribosomal peptides (MCINTOSH; DONIA; SCHMIDT, 2009). Usually, there are three peptides that always present small modifications that confer biological activities to this class of NPs. Ser, Thr and Cys, are involved in these enzymatic processes and their final NPs often present oxazole, methyloxazole and thiazole rings, respectively (MCINTOSH; DONIA; SCHMIDT, 2009; WALSH; MALCOLMSON; YOUNG, 2012).

2.2 GENOME-GUIDE ANALYSIS APPLIED TO NATURAL PRODUCTS

Microbial genome mining is widely applied as a state-of-the-art approach to traditional methods for NPs discovery. NPs have their featured place into the chemical space, due to their medicinal importance. Lately, algorithms capable of further investigation of the hidden biosynthetic potential of microorganisms, lead to an enormous effort to rediscovering NPs (HARVEY; EDRADA-EBEL; QUINN, 2015). What was thought to be already well investigated and discovered, is currently, source of a vast reservoir of possibilities.

In the post-genomic era, diverse cases are related to success in using bioinformatics tools and genome-mining strategies to the discovery of NCEs from bacterial sources (PIDOT et al., 2014). In the beginning, the earliest *Streptomyces* genome sequence showed that genome mining and bioinformatics tools could transform compound discovery, as it is nowadays (MEDEMA; FISCHBACH, 2015).

Genome-mining studies lead to the discovery of **coelichelin (7)**, in the year 2000. Coelichelin is a peptide siderophore isolated from *Streptomyces coelicolor*. In that time, the BGC related to the production of this specific compound was identified from a partially sequenced genome (CHALLIS; RAVEL, 2000). Nowadays, the entire genome of *S. coelicolor* and innumerous others sequences of the most diverse microorganisms are available entirely or ongoing, due to the existence of high-throughput techniques of sequencing. In this sense, computational approaches made this process possible.

The discovery of **orfamide A (8)**, a NRPS-like antibiotic compound, from *Pseudomonas fluorescens* Pf-5 was carried out after genome-mining experiments adopting a “genomisotopic approach”, employing a combination of genomic sequence analysis and

isotope-guided fractionation of compounds biosynthesized by orphan gene clusters related to NRPS-like sets of genes (GROSS et al., 2007).

Examples of NPs that also act as antibiotics and had their discovery processes under genome-guided strategies are **salinilactam (9)**, from *Salinispora tropica* (UDWARY et al., 2007); **holomycin (10)**, from *Streptomyces clavuligerus* (LI; WALSH, 2010); **laspartomycin (11)**, from *Streptomyces viridochromeogenes* (WANG et al., 2011b); and **pyridomycin (12)**, from *Streptomyces pyridomyceticus* NRRL B-2517 (HUANG et al., 2011). These compounds, among others, were assigned to their chromosomes due to the characteristics of biosynthetic steps of well organized BGCs.

A recent example is the isolation of **teixobactin (13)**, a NRPS-related antibiotic able to kill Gram-positive and mycobacteria acting as lipid II antagonist. Teixobactin isolation was carried out after extensive genome-mining and isolated from *Eleftheria terrae* (LING et al., 2015). This antibiotic, known by its characteristic of killing pathogens without detectable resistance, has its peptide backbone built on D-amino acids (LING et al., 2015), damaging the pathogen at the level of cell wall biosynthesis. This discovery could be considered the most important finding in the new antibiotic era, confirming that the union between computational prediction and real production is very tight.

Phosphonate NPs also make part of genome-guided experiments, since they are ubiquitous in biological systems and present great pharmaceutical potential (3 of 20 isolated compounds have gone to commercialization). Their biological activities success is directly related to their ability of mimicking essential metabolites, including phosphate esthers (JU et al., 2015). Phosphonate NPs such as fosfomycin (Monurol®), fosmidomycin and phosphinothricin, although discovered through bioguided fractionation and originally isolated by assaying inhibition of bacterial growth, had guided the investigation of phosphonate-like BGCs due to their medical importance (JU; DOROGHAZI; METCALF, 2014). By mining genomes of 10,000 actinomycetes, Metcalf and collaborators discovered 19 new NPs in a four-year program. Compounds as **argolaphos A (14)** and **B (15)**, **hydroxyphosphonocystoximic acid (16)**, and **valinophos (17, Figure 3)** were isolated after genome-guided exploration (JU et al., 2015).

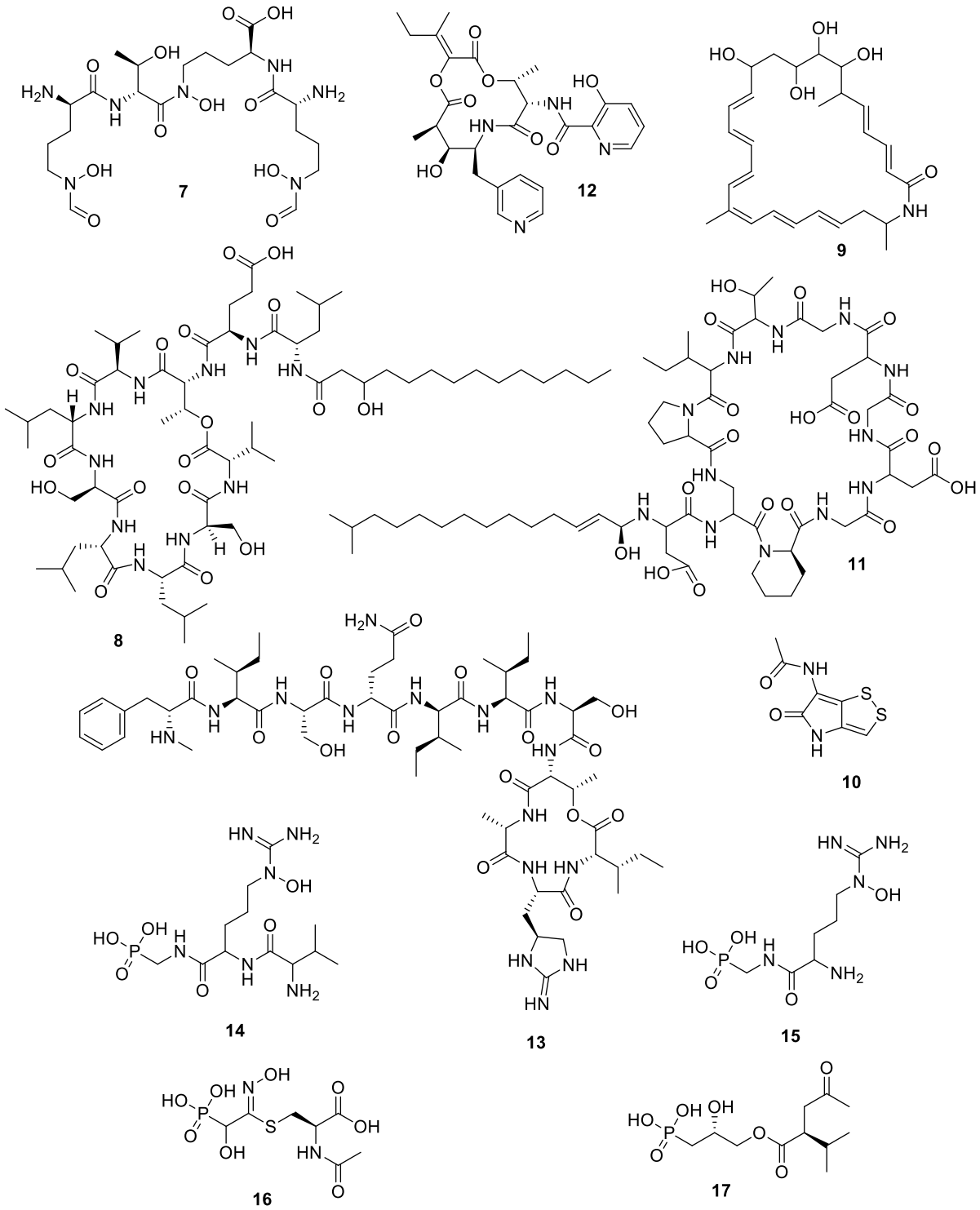

Figure 3- Compounds isolated after genome-guided analysis. These examples englobe NRPS-like, PKS-like and Phosphonate-like NPs.
Source: From the author.

Several reports relate massive genome mining-based datasets organized according to the capability of microorganisms in producing NPs. For each class, network profiles are created by assembling data to obtain valuable information based on genomic analysis. Currently, there are studies for Cyanobacteria (DITTMANN et al., 2015), Actinomycetes (DUNCAN et al.,

2014), a global view of microorganisms (CIMERMANCIC et al., 2014), and a remarkable investigation of NRPS synthetases and PKS synthases of microorganisms (WANG et al., 2014). These reports, when connected to metabolomics approaches, build a logic, fast and reliable mode of dereplication for novel extracts in the search of medically relevant NPs.

2.3 BETAPROTEOBACTERIA AND NATURAL PRODUCTS

NPs from Betaproteobacteria comprehend from small to big molecules. They are most of the time derived from NRPS, PKS, NRPS-PKS as a hybrid biosynthetic system, and some precursors of main compounds are known. The great majority of Betaproteobacteria species do not have NPs associated to them according to the literature. Although, species from the genus *Achromobacter* spp., *Alcaligenes* spp., *Burkholderia* spp., *Chromobacterium* spp., *Cupriavidus* spp., *Delftia* spp., *Janthinobacterium* spp., *Nitrossomonas* spp., *Ralstonia* spp., and *Thiobacillus* spp. have compounds correlated to their NPs producing capability.

The DNP database displays that compounds from *Achromobacter* spp. are related to micromolecules, when compared to other genera. In summary, chlorotryptophan derivatives, urocanic acid, cycloleucylprolyl are indexed. Due to the lack of biological activities data reported, their structures are not shown.

Compounds from Betaproteobacteria species present great diversity in biological activities and are classified in the most diverse structural classes. The compound **alcaligin (18)**, a macrocyclic dihydroxamate siderophore isolated from closely related species of *Alcaligenes* spp. and *Bordetella* spp, presented iron uptake increasing for *Bordetella bronschiseptica* and *B. pertussis* showing that this siderophore is crucial for strain growth (BRICKMAN et al., 1996). Other siderophore named **staphyloferrin B (19),** classified as a citrate-based polycarboxylate siderophore, is biosynthesized by *Ralstonia solanacerum* and *Ralstonia eutropha* (formely named *Alcaligenes eutrophus*), and utilized by *Staphylocuccus aureus* for iron uptake when colonizing the vertebrate host (MADSEN; JOHNSTONE; NOLAN, 2015). The synthesis of **staphyloferrin B** permitted to infer about biosynthetic mechanisms and molecular pathogenesis of this human pathogen. In the class of metallophores, the Magarvey group recently isolated a compound named **delftibactin (20)** presenting a unique ability of gold biomineralization (JOHNSTON et al., 2013). This compound has a typical NRPS-PKS

backbone composed by nonproteinogenic peptides including the presence of arginine, considered rare in NPs, in its structure.

An interesting compound from *Burkholderia thailandensis* 264, named as **bactobolin A (21)**, was recognized to have its biosynthetic starting related to *N*-acylhomoserine lactones. Bioinformatics experiments enabled to correlate bactobolin to its BGC, called *bta* and its preliminary assays results suggested that this compound possess a very potent antibiotic activity (CARR et al., 2011). The activity of bactobolin is very distinct while compared to previous antibiotics. Further investigation evidenced that a crystal structure of this compound was bound to a 70S ribosome site, providing its exact mode of action (MOA) (AMUNTS et al., 2015). Another antibiotic-like compound isolated from *Chromobacterium violaceum*, classified as a polyketide antibiotic, **aerocavin (22)**, presents activity against Gram-positive and Gram-negative bacteria such as several *Staphylococcus*, *Pseudomonas* and *Acinetobacter* species (SINGH et al., 1988). The first two species are known as the cause of hard treatment bacterial infection in hospitals due to their multidrug resistance characteristic.

From the genus *Burkholderia*, the Enacyloxins are unusual polyketide antibiotics produced by *B. ambifaria*. This species presents anti-Gram-negative activity even while testing other *Burkholderia* species. The production of **enacyloxin IIa (23)** was induced via a minimal culture media composition containing glycerol, a carbon source that promote antimicrobial compounds biosynthesis (MAHENTHIRALINGAM et al., 2011). Similar polyketide backbone, identified as **kalimantacin C (24, Figure 4)**, was isolated from *Alcaligenes* species display a strong antistaphylococcal effect. This discovery was taken in 1996, however, the last kalimantacin resistance report, in 2008, showed still no clinical resistance of *Staphylococcus* species under the treatment using this compound (MATTHEUS et al., 2010).

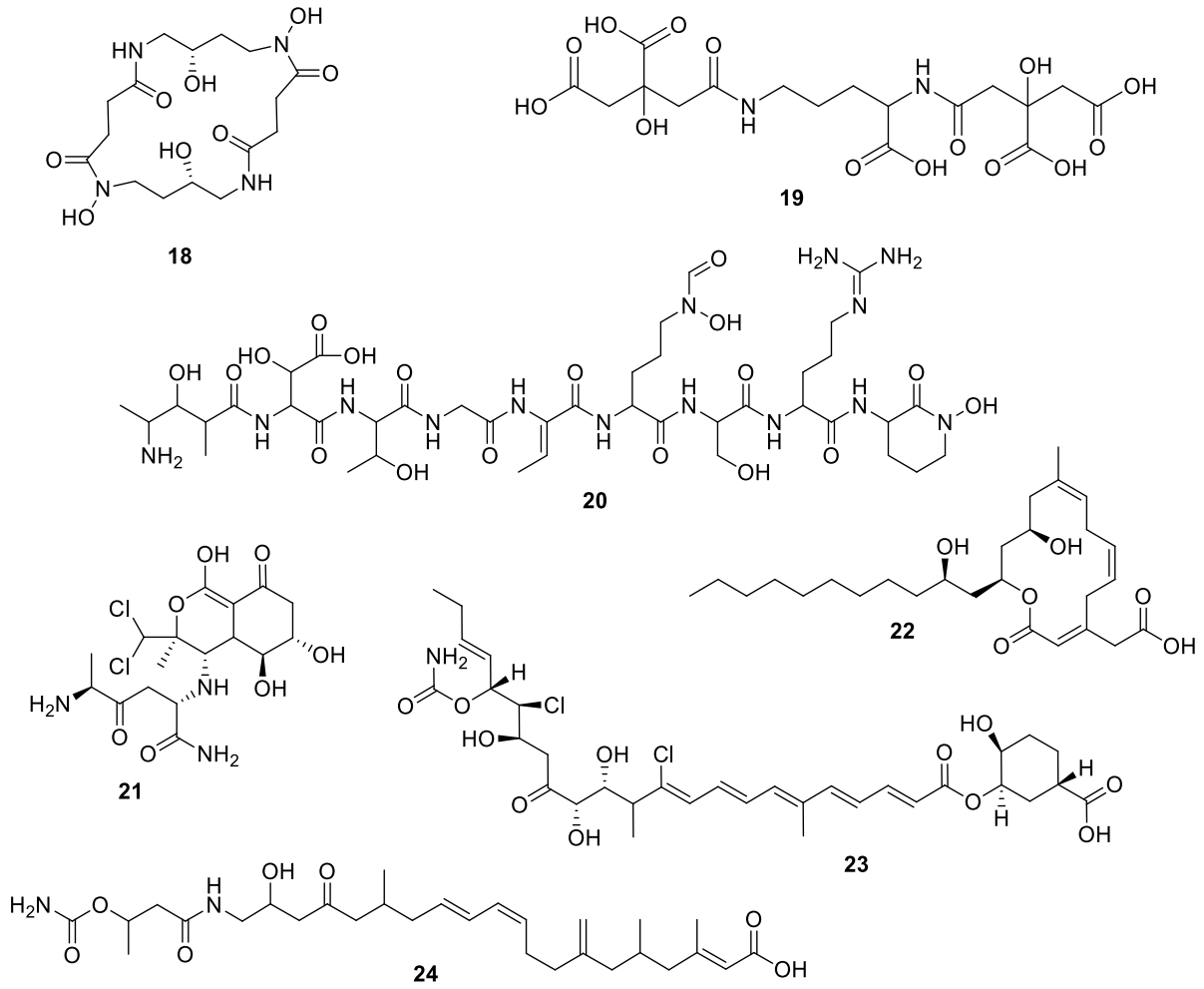

Figure 4- NPs from Betaproteobacteria species (Part 1).
Source: From the author.

An interesting NRPS-PKS-derived compound, called, **micacocidin (25)** was first reported in 1998 by Takeda group. Micacocidin is a metal-binding heterocyclic antibiotic first isolated from *Pseudomonas* species (KOBAYASHI et al., 1998). In the class Betaproteobacteria, *Ralstonia* species are known to produce this same compound. Some antibiotics even presenting strong activity do not have their MOA well established. In the case of **janthinocin (26)**, first discovered in the early 90's from *Jantinobacterium* species, its MOA remains unestablished (SULLIVAN et al., 1990). **Janthinocin** is a NRPS-derived NP that is considered an antibiotic from neglected bacterial sources (PIDOT et al., 2014). This discovery resurges after genomic era which also provided the genome-guided isolation of **jagaricin (27)**, another NRPS-PKS-derived biosynthesized via simple modules in NRPS-PKS hybrid machinery. **Jagaricin** exerts strong antifungal effects against human pathogens, as well as, *Candida albicans*, *Aspergillus fumigatus* and *Aspergillus terreus* in submicromolar concentrations (GRAUPNER et al., 2012).

Burkholdine (28), a linear lipopeptide and **occidiofungin (29)**, a cyclic glycol-lipopeptide also present antifungal activities. **Burkholdine** and **occidiofungin** displayed diverse and potent antifungal activities against several *Candida* species (EMRICK et al., 2013; KONNO et al., 2015).

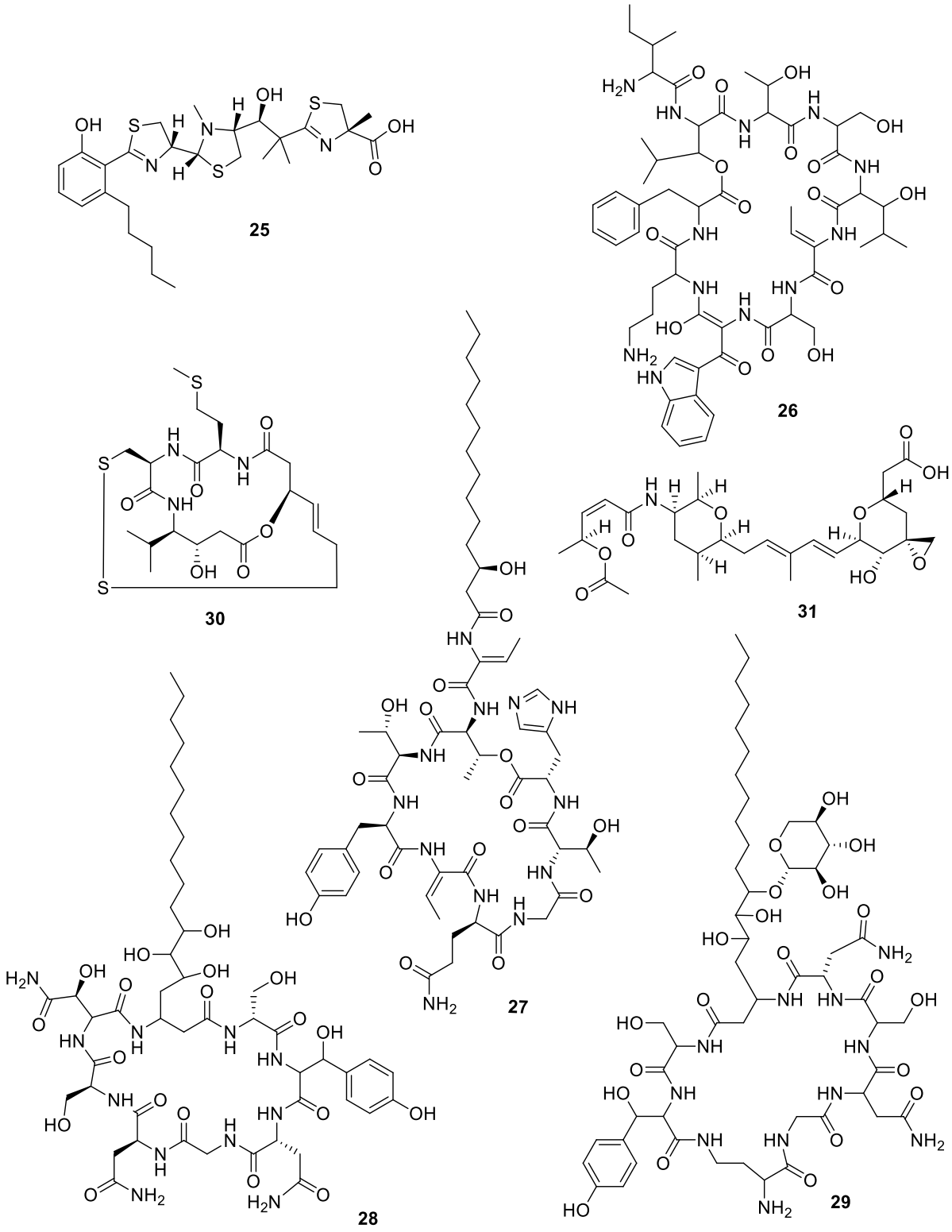

Figure 5- NPs from Betaproteobacteria species (Part 2).
Source: From the author.

Increasing the spectrum of biological activities of NPs from Betaproteobacteria species, the compound named **burkholdac (30)**, isolated from *Burkholderia* species, it is in clinical phase. **Burkholdac**, also known as **thailandepsin**, is an analogue of **FK228**, a FDA approved anticancer drug (FUKUI et al., 2014). The **thailanstatin (31, Figure 5)** also isolated from *Burkholderia* species presented *in vitro* evidences that its activity is correlated to mRNA splicing inhibition as potent as **spliceostatin A** (FR901464), one of its analogues (JAIN et al., 2013).

There are other examples that not make part of this review. However, the amount of compounds from Betaproteobacteria species are small next to the revealed potential of these microorganisms. In this brief overview, it was related diverse classes of compounds with biological role for this class os microorganisms. In this sense, it is worth to highlight that there is a great likelihood of discovering NCEs from Betaproteobacteria. Thus, systematizing their biosynthetic capability, the intraspecific variation of their BGCs and how to obtain them from environmental sources provide valuable information for NPs discovering processes.

3 OBJECTIVES

As a general objective, this thesis aimed to track and to discover similarities in enzymatic processes related to the production of nonribosomal peptides from Betaproteobacteria species, since these compounds have special place into the chemical space and this subclass of microorganisms presents high biosynthetic capability. This general objective could be divided in several specific objectives as follow:

a) To choose user-friendly applications in order to analyze all sets of data to be widely used for the most diverse interests;
b) To select all available sequences of microorganisms from the phylum Betaproteobacteria and investigate their potential as NPs producers and to create a library containing all features of each, including taxonomy, gene counting, clusters, type of compound, predicted core structure, etc., for easy accessing and interpretation;
c) To track biosynthetic evidences and to relate them to their genomic identity using a smart way of having information that will be used to correlate a specific microorganism to its similar ones;
d) To guide which genera of microorganisms will be firstly isolated according to their biosynthetic capabilities and to develop profiles of resistance, metabolism and nutrients (carbon and nitrogen sources) from microorganisms of interest and theoretical metagenomics analysis selecting the best candidates in a genome-guided strategy;
e) To investigate intraspecific variation of microorganisms of interest creating comparative profiles according to their genes, type of compounds and predicted core structure in order to guide and correlate to new discoveries.
f) To test all individual components in order to evaluate their behavior while acting together in the same culture media creating a computational genome-guided strategy for the isolation of selected microorganisms from environment;

4 RATIONALE

According to recent discoveries, the phylum Proteobacteria has been recognized to contain a substantial number of clusters encoding putatively novel NPs, but to date has not been well studied. Since the identification of NRPs-like compounds is still difficult given their complex architecture, this thesis aims to develop systematic methods to NPs discovery processes using state-of-the-art technologies.

Part of the phylum Proteobacteria, the class Betaproteobacteria is being always linked to remarkable discoveries in the field of novel compounds. In this sense, the aims of this whole work were to investigate a novel manner of working with Natural Products. Using bioinformatics tools, core questions were asked in order to make the search of novel compounds systematic. The key questions were:

a) Is there a way of measuring how promising Betaproteobacteria microorganisms are due to their biosynthetic capability? If yes, how could it be measured?;
b) How to select and how are taxonomically distributed the best candidates for studying?;
c) How different strains of a given candidate could produce chemical compounds? Are the possibilities likewise or are they discrepant?;
d) How different is the information in different chromosomes of a same strain?
e) How to reach these microorganisms in the microcosms and how to take advantage of their characteristics in order to isolate them?;

Since the genomics databases are increasing every day, a way of treating huge amount of data in order to solve chemical-biological problems is important given the current scenario. The elaboration of a technique able to track and correlate similar genomes according to their NPs biosynthetic capability and obtain these specimens from environment became substantial. This is a fact due to how arduous is the process of identification of novel NRPS-related NPs. Therefore, the creation of libraries correlating biosynthetic capability and genomic data englobe all purposes of this work in order to facilitate and guide the isolation of novel compounds.

5 EXPERIMENTAL SECTION

All the experiments developed for the three main analyses of this thesis will be explained ahead.

5.1 GENOME-GUIDED ANALYSIS OF BETAPROTEOBACTERIA BGCs: THE CREATION OF XPAIRT (e***XPA***ndable ***I***dentification of non***R***ibosomal peptides ***T***endencies).

Considering all literature contents, this experimental section focused on the fragmentation of NRPS-related structures creating a genomic-guided strategy. Fragmentations were made by pairs for predictions (pairs of predictions: POPs) and for known NPs (pairs of compounds: POCs) and the occurrence of each was investigated providing statistical and taxonomical visualization of their distribution in BPB species. **Figure 6** shows a step-by-step guide for XPAIRT analysis flow.

Consequently, calculating the likelihood of specific POPs showed to be conceivable in a given group of species. In this sense, XPAIRT revealed important biosynthetic evidences in BPB species. Considering the complicated architecture that NRPs can present, it is worth to highlight that the development of a simple technique able to track biosynthetic similarities, as proposed by XPAIRT, is substantial for aiding the structural elucidation of NCEs from metabolic profiles of promising species.

Furthermore, the extrapolation of XPAIRT is possible to any classes of microorganisms, making this guide a comprehensive strategy aiding the discovery of NRPS-related NPs. However, continuous improvement of XPAIRT dataset will give better matches, once the number of new sequences are often increasing and the genomic information of microorganisms is dramatically small when compared to the uncultured microbiome (MILSHTEYN; SCHNEIDER; BRADY, 2014).

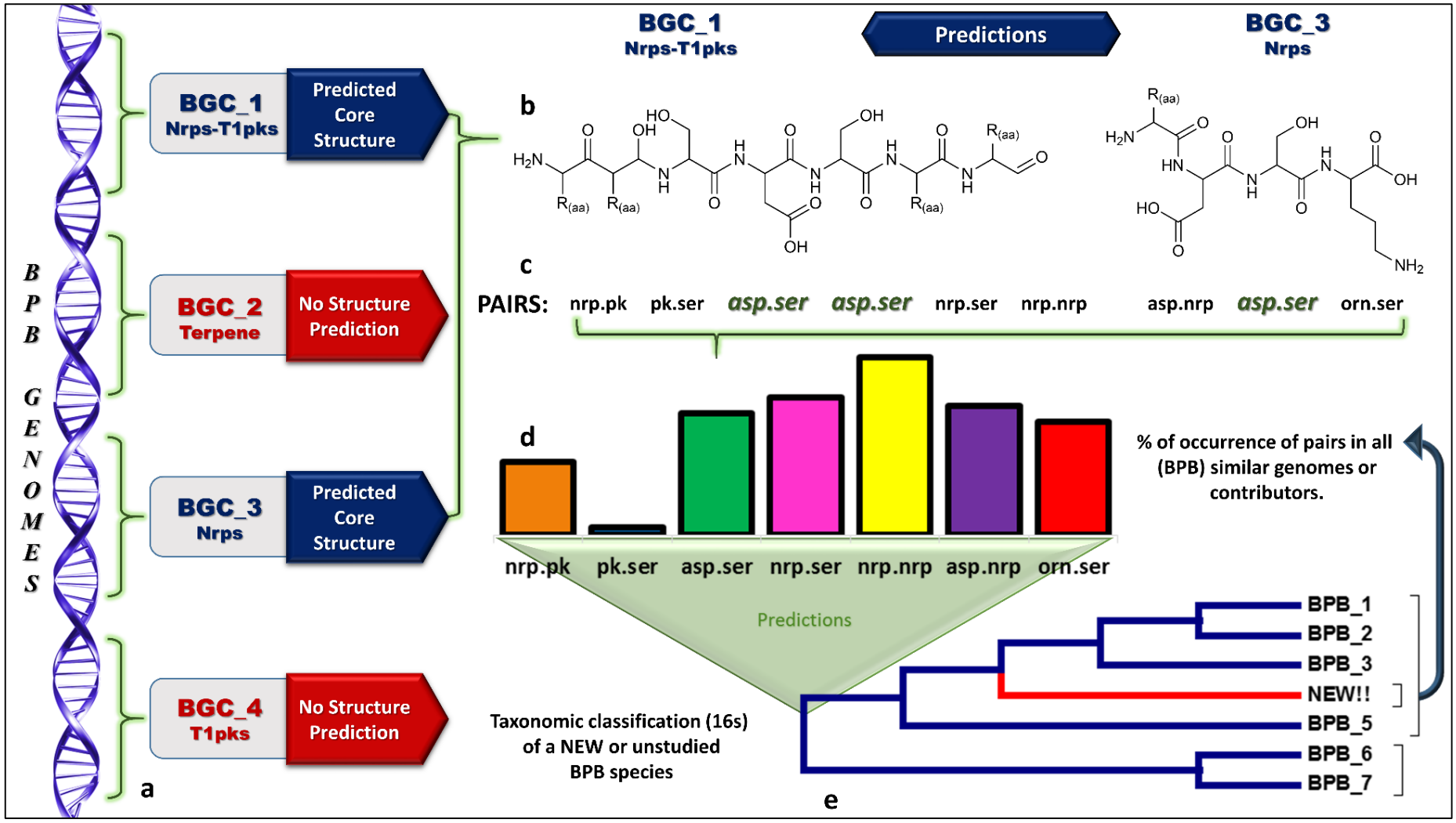

Figure 6- The scheme of XPAIRT workflow. The main steps related to A) analyzes of iBGCs products, B) the NRPS-related iBGCs), C) fragmentation (organization of pairs in logic sequences and alphabetical order, creating POPs), D) statistical analyzes (correlating producers in the most diverse taxa level: order, family, genus and species), and E) correlation to the closest taxonomic group are highlighted as the core steps of this strategy.

Source: From the author.

5.1.1 MATERIAL AND METHODS - XPAIRT CREATION

Material and methods adopted for the creationg and analysis of XPAIRT will be explained ahead.

5.1.1.1 Datasets

BPB genomic sequences were downloaded from the National Center of Biotechnology Information genome database (http://www.ncbi.nlm.nih.gov/genome). Sequence files were chosen according to the representative genome available. A complete final version of a main dataset is available in the end of this work at the section **Supplementary Files, Table 5**. Partially complete genomes were chosen according to their most complete sequences covering as much genus as possible of BPB. The dataset were created relating results from antiSMASH (BLIN et al., 2013; MEDEMA et al., 2011) algorithm and NCBI database.

5.1.1.2 BGCs Finder

BGCs predictions were made using antiSMASH 3.0 searching for NRPS-PKS detailed functional domain annotation, the core structure prediction, and gene cluster homologies. NRPS-PKS predicted results were elected according to the general consensus between NRPSprediction2, Stachelhaus code and Minowa. Only total consensus between two or more methods were considered as the predicted peptide. Divergent consensus for a single peptide was considered as **nrp** moiety (meaning that the R group could not be correctly assigned as proteinogenic or nonproteinogenic peptides, or an out of consensus moiety).

5.1.1.3 Rearranging NRPS predicted details by POPs

Predicted NRPS-PKS products were rearranged by pairs (POPs). Predicted core structures and known compounds had their peptide backbone fragmented and were used as standard comparisons. Two common structural abbreviations were adopted: "$R_{(nrp)}$": represents the continuation of the peptide backbone beyond the specific POP; "$R_{(aa)}$": represents the out of consensus moiety **nrp**. POPs and POCs had their peptides sequences normalized according to antiSMASH, KEGG (KANEHISA; GOTO, 2000) and Norine (CABOCHE et al., 2008). Detailed information was given for NRPS products, aiming the prediction of the most probable peptide backbones in BPB iBGCs. For both, POCs and POPs, all pairs had their three letters code for amino acids automatically converted in alphabetical order facilitating visual and statistical analysis (for example, predicted: val.ala; used: ala.val).

5.1.1.4 Illustrative scores for Delftibactin as a model compound

Identity values were analyzed and associated to their respective compounds. For each complete hit ($NRP_{(known)}$: ala.val; $Prediction_{(xpairt)}$: ala.val) the score was considered as 1 (100%); For each half complete hit containing an **nrp** moiety ($NRP_{(known)}$: cys.met; $Prediction_{(xpairt)}$: cys.nrp) the score was considered as 0.5 (50%), due to the lack on specificity of **nrp** moieties. An entirely non-recognized POP (nrp.nrp) or wrong POPs had their scores considered as 0 (0%).

5.1.1.5 Taxonomic Analysis

iBGCs sequences were downloaded as .fasta format and named according to their respective species and their related cluster number. iBGCs sequences were aligned and the dendrogram was built in MEGA6 (TAMURA et al., 2013). The analysis involved 113 nucleotide sequences aligned by CLUSTAL W using default parameters (LARKIN et al., 2007). The distribution of NRPS-related iBGCs was inferred using the Neighbor-Joining method (SAITOU; NEI, 1987a). Comparisons of the percentage of POPs expression in each

main branch were calculated with the purpose of differentiate iBGCs characteristics according to their phylogeny.

5.1.1.6 NRPS-related POPs and the Heatmap

Gitools v.2.2.2 (PEREZ-LLAMAS; LOPEZ-BIGAS, 2011) was used in order to provide integrative and visual analyses of data. Clustering and hierarchical calculations were made according to Jaccard Index (JI), clustering species to their similar level of similarity. The heatmap was constructed using species that could show at least one POP, even an out of consensus POP (**nrp.nrp**). The dataset was organized crossing microorganisms (and their taxonomic classification) to their POPs. Species and POPs results were overlapped and clustered according to a) scores for all pairs in NRPS-related iBGCs and b) Jaccard Index level of similarity via Euclidean distance method.

5.1.1.7 Network analysis

For visual analysis and graphical representation of networks, the software Cytoscape 3.2.1 was adopted. Statistical analysis for the network results were performed using the software Excel.

5.1.2 RESULTS AND DISCUSSION

Results and discussion related to XPAIRT experiments will be explained ahead.

5.1.2.1 iBGCs from BPB and their characteristics

About 1650 BGCs were identified in which 13.5% (237 iBGCs) are exclusively classified as NRPSs. Hybrid iBGCs, as well as, NRPS-T1PKS, HGLKS-NRPS-T1PKS,

corresponded to 28% of all iBGCs (**Figure 7, A**). In total, BPB species presented around 40% of their iBGCs classified as NRPs-related compounds.

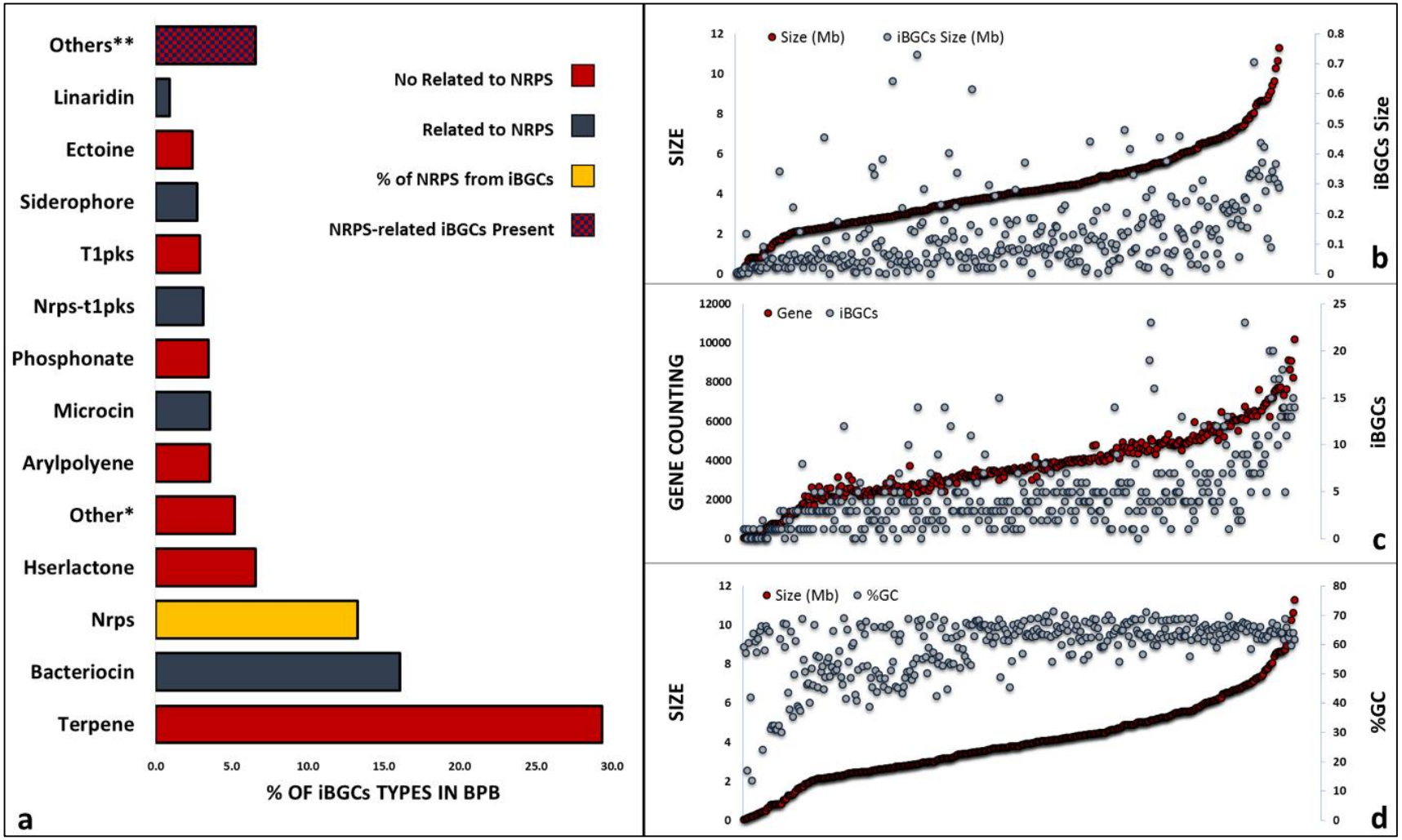

Figure 7- Characteristics of BPB genomic sequences according to their biosynthetic potential. A) Distribution of iBGCs classes in BPB. B) Gene size increasing connected to the magnitude of the genomic sequence for each species analyzed. C) Gene counting increasing related to the iBGCs number of each related species. D) Variation of %GC as genome size increases. *Other classes not identified by antiSMASH. **Other classes identified by antiSMASH grouped in order to facilitate visual analysis containing NRPS-related iBGCs.

Source: From the author.

As a tendency, the higher the gene counting, the higher the percentage of genomic information dedicated to the production of NPs in all sequences analyzed (**Figure 7, B**) and their distribution is widely spread. The incidence of hybrid pathways, besides of NRPS-PKS, occurred in very small percentage (3.24% of all iBGCs divided in: 23 NRPS-related iBGCs; and 34 non-NRPS-related iBGCs).

The average iBGCs size in BPB sequences remains around 4% of the entire genome. Species with larger replicons dedicate more genomic information to iBGCs while Genome Size rises (**Figure 7, C**). More than 60% of all sequences display 65% of GC content in their genomes (**Figure 7, D**). GC rich genomes are related to evolution questions, associated to stability of enzymes and protein structures and nutrients absorption (FOERSTNER et al., 2005; LASSALLE et al., 2015).

A large percentage of iBGCs was correlated to the production of Terpenes (496 iBGCs, 28.23%) and Bacteriocins (271 iBGCs, 15.42%), presenting diverse range of identity to already known BGCs of other species. Around 250 NRPS-related iBGCs did not have correlation to any known BGCs, suggesting that they could express novel compounds.

Apart of iBGCs homologous to other Proteobacteria species (**Table 1**), there were iBGCs related to taxonomically distant species. As an example, Bacillibactin (from Firmicutes; identity from 8 to 30%), Tubulysin (from Myxobacteria; identity of 6%), Laspartomycin (from Actinobacteria; identity from 6 to 28%), and Fuscachelin (from Actinobacteria; identity from 10 to 30%).

Table 1. Known compounds related to iBGCs from antiSMASH analysis in BPB showing the percentage of genes, which have identities, number of iBGCs associated, and their distribution within BPB Families.

(continues…conclusion)

Compound	*iBGC Type*	*Identity (%)*	*iBGCs*	*Related BPB Families*
Arthrofactin	NRPS	From 2 to 30	7	*Burkholderiaceae; Chromobacteriaceae; Oxalobacteraceae; Unclassified.*
Coelichelin	NRPS	From 10 to 23	7	*Burkholderiaceae; Oxalobacteraceae.*
Cupriachelin	NRPS	From 1 to 71	41	*Alcaligenaceae; Burkholderiaceae; Comamonadaceae; Nitrosomonadaceae; Oxalobacteraceae; Rhodocyclaceae.*
Delftibactin	NRPS	From 9 to 100	26	*Alcaligenaceae; Burkholderiaceae; Burkholderiaceae; Chromobacteriaceae; Comamonadaceae; Nitrosomonadaceae; Oxalobacteraceae; Rhodocyclaceae; Unclassified.*
Malleobactin	NRPS	From 2 to 92	41	*Alcaligenaceae; Burkholderiaceae; Comamonadaceae; Nitrosomonadaceae; Oxalobacteraceae; Rhodocyclaceae.*
Paenibactin	NRPS	From 1 to 30	12	*Alcaligenaceae; Burkholderiaceae; Chromobacteriaceae; Oxalobacteraceae.*
Serobactins	NRPS	From 1 to 100	41	*Alcaligenaceae; Burkholderiaceae; Comamonadaceae; Nitrosomonadaceae; Oxalobacteraceae; Rhodocyclaceae.*

(continues…conclusion)

Compound	*iBGC Type*	*Identity (%)*	*iBGCs*	*Related BPB Families*
Turnerbactin	NRPS	From 15 to 76	7	*Burkholderiaceae; Chromobacteriaceae; Oxalobacteraceae.*
Vibriobactin	NRPS	From 14 to 23	10	*Alcaligenaceae; Burkholderiaceae; Oxalobacteraceae.*
Xenomamicins	NRPS	From 6 to 66	15	*Burkholderiaceae; Burkholderiaceae; Chromobacteriaceae; Oxalobacteraceae; Unclassified.*

5.1.2.2 XPAIRT workflow and validation steps

XPAIRT strategy follows a simple workflow: a) to organize iBGCs predictions of secondary metabolites; b) to fragment predicted structures in POPs; c) to calculate the percentage and distribution of each POP in BPB by similarity of structures and phylogeny.

To the validation of XPAIRT strategy, a two-step experiment was adopted. It was expected that when the identity of iBGCs were high, the probability of observing similar POPs should also be high ($\uparrow iBGCs\ Homology = \uparrow Correct\ POPs$), and lower levels of genomic identity between sets of iBGCs express different POPs. The two-step consisted in match results from predictions in both ways a) POCs to POPs and b) POPs to POCs.

In order to validate POCs to POPs, a network was created to correlate all known compounds and their levels of identities. The region called Cluster 1 (**Figure 8**), had their POCs organized according to their peptide backbones. Highly expressed POPs were similarly expressed to POCs of known BGCs identified in this work (orn.ser, predicted 14.47%, observed 13.64%; and asp.ser, predicted 15.79%, observed 12.12%).

Due to the number of proteinogenic and nonproteinogenic monomers are greater than 500 different subunits (WALSH; O'BRIEN; KHOSLA, 2013), the occurrence of pairs is related to low percentage values. Nonetheless, given the amount of combinations, small variations and distribution express their similarities. Other POPs of Cluster 1 also suggests that some out of consensus peptides (nrp) were linked to T1PKS-like moieties (pk). The higher percentage of PKS-related moieties was observed to aspartate in the form of asp.pk (1.32%, highlighted with **, **Figure 8**), suggesting that hybrid iBGCs in Cluster 1 is common (9.1% of all POPs). POCs as ala.asp and lys.orn were not predicted in POPs neither in this Cluster nor in the main dataset.

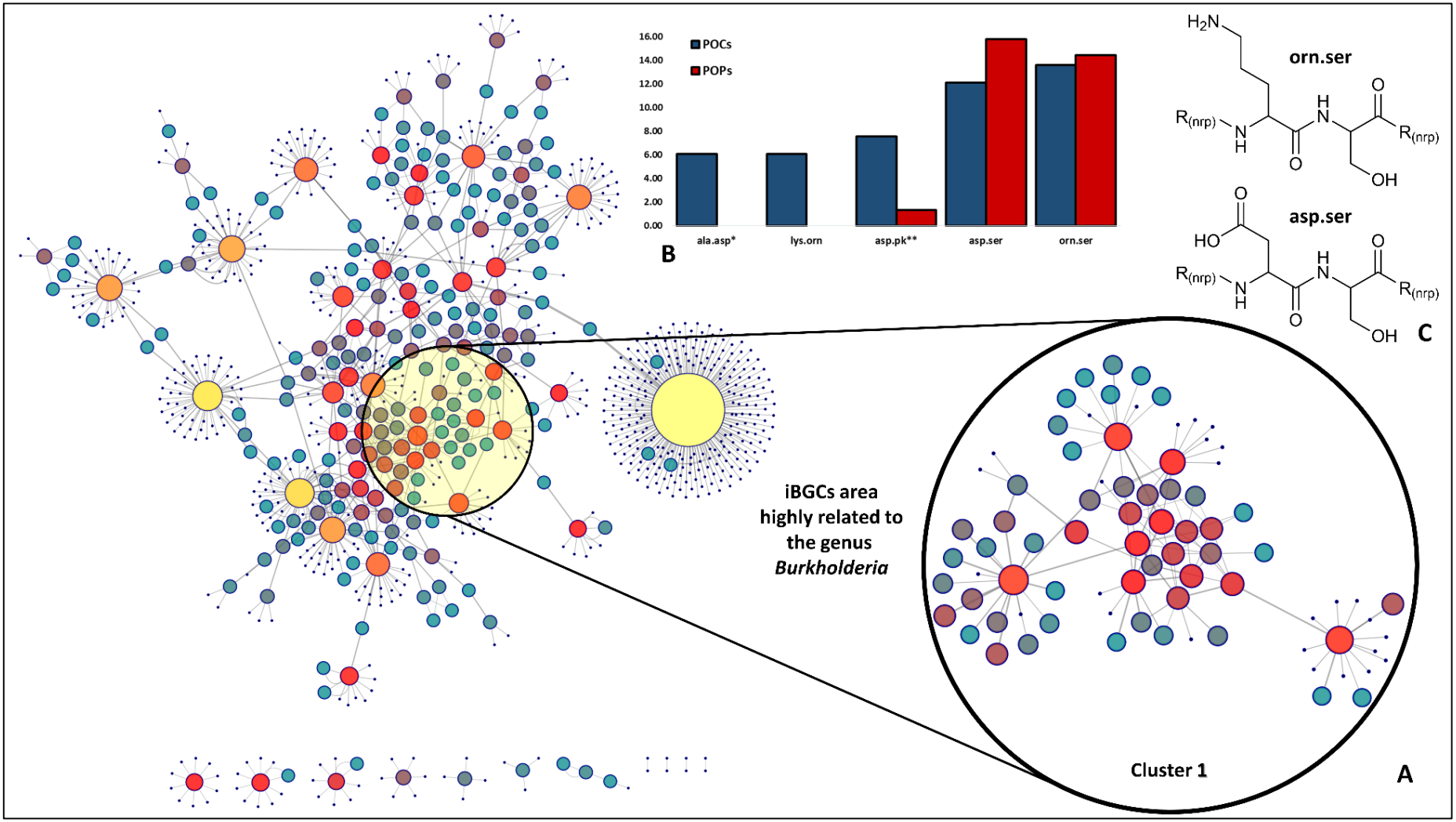

Figure 8- Network of known BGCs related to BPB species. A) The Cluster 1, a region highly related to *Burkholderia* species, and its first neighbors (compounds) selected, exported and clustered. B) Percentages of expression of each pair (POCs and POPs) from iBGCs. C) Pairs asp.ser and orn.ser are highlighted illustrating main results between predictions and real compounds. Calculations were made excluding degrees lower than 5 correlations. Network graph was built using Cytoscape 3.2.1. Nodes and color scale are related to their degree and clustered by using Spring-Electric Algorithm and Allegro Layout version 2.2.1. The biggest node is related to iBGCs with no identity to known BGC.

Source: From the author.

In order to validate POPs to POCs, the structure of Delftibactin, a metallophore related to gold biomineralization (JOHNSTON et al., 2013), was used as template. Its structure contains a classic NRPS-PKS hybrid product, and its monomers sequence was used as follows: $(pk_{(NH_2+ohmal)}+Asp_{(mod)}+Thr+Gly+Thr_{(mod)}+Orn_{(mod)}+Ser+Arg+Orn_{(cyclic)})$ (**Figure 9, A**). The fragmentation of Delftibactin peptide backbone was proceeded creating Delftibactin POCs, and all iBGCs products from the main dataset, related to the production of Delftibactin had their predicted structures fragmented by pairs, creating POPs.

Genome sequences related to Delftibacin iBGCs, were called contributors and organized in a cluster (**Figure 9, B**). The percentage of genomic identity for each iBGC relative to all contributors varied from 9 to 100%. The most noted POPs were orn.ser (10.9%), orn.thr (4.55%), asp.thr (8.18%). Moreover, the POP asp.ohmal reached a low score (1.82%) when compared to the other POPs (**Figure 9, C**). Furthermore, POCs of Delftibactin were correctly predicted when iBGCs presented higher levels of identity. The probability of finding correct POPs was directly proportional to the genomic identity level, since the decreasing in the level of identity led to lower scores and divergent POPs (**Figure 9, D**).

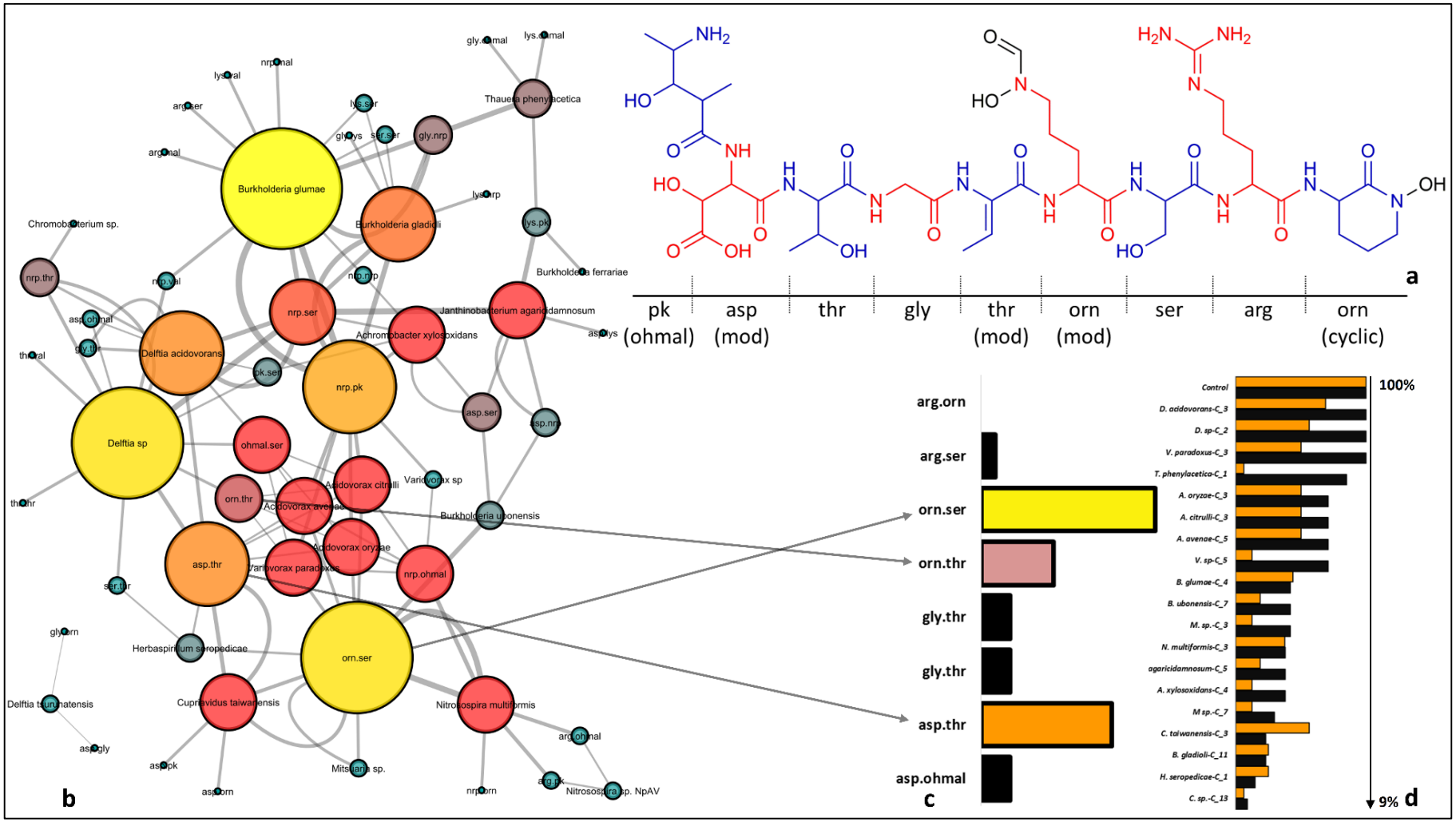

Figure 9- The Delftibactin contributors. A) The structure of Delftibactin fragmented relating the posttranslational modification sites ("mod" and "cyclic"). B) Contributors to the identification of POPs related to Delftibactin iBGCs. C) Relative percentage of each POPs and their percentage of occurrence. Pairs out of consensus with Delftibactin had lower incidence, were considered as having score 0 (zero), and lower degrees at Cytoscape. D) Relation between identity (black) and correct POPs (orange) associating the decreasing in identity and likelihood in finding POPs in a directly proportional manner.

Source: From the author.

The results of the two validation experiments, POCs to POPs (**Figure 8**) and POPs to POCs (**Figure 9**), were consistent, leading to further analyses correlation via phylogenic and similarity between distribution of POPs within genomic sequences from BPB species due to the NRPSs biosynthetic characteristics of assembling activated monomers for building peptides backbones.

In this sense, the biosynthesis of NRPS-related NPs relies on monomer being incorporated into a oligomer (FISCHBACH; WALSH, 2006; WALSH, 2015) and its biosynthetic core contains a minimal of three domains per module (CHALLIS; NAISMITH, 2004). These minimal NRPS domains are called chain initiation (adenylation domains (A), responsible for substrate specificity and activation), chain elongation (condensation domains (C), that catalyzes peptide bond formation between a new substrate and the peptide chain), and chain termination (related to (T) domains that covalently tethers the substrate to the respective enzyme, via a thioester bond) (CHALLIS; NAISMITH, 2004; FISCHBACH; WALSH, 2006). The size and number of domains in a given nonribosomal peptide could be recognized according to the number of monomers in its chemical structure (KEATING; WALSH, 1999; MCINTOSH; DONIA; SCHMIDT, 2009).

Different modifications can be present as the peptide undergo different enzymatic processes (MCINTOSH; DONIA; SCHMIDT, 2009) these processes are often related to cyclization and posttranslational modifications (SEO; LEE, 2004; WALSH et al., 1997). Structural modification of the finished peptide can also be present. These modifications are often related to Ser, Thr and Cys, creating oxazole, methyloxazole and thiazole rings, respectively (MCINTOSH; DONIA; SCHMIDT, 2009; WALSH; MALCOLMSON; YOUNG, 2012). Eventually, these heterocyclic moieties exchange the expected chemical shifts of POPs. However, they could be identified according to the NMR spectrum analyzes, by the occurrence of sp^2 carbons instead sp^3 and the lack of amide protons related to these three heterocyclic subunits (MCINTOSH; DONIA; SCHMIDT, 2009; WALSH; MALCOLMSON; YOUNG, 2012).

Assuming that peptide backbones of NRPs have complex architecture, their identification still is time consuming and difficult to assign when elucidating NCEs (YANG et al., 2015). Thus, XPAIRT, using only user-friendly software, tracks POPs variation in predicted metabolomics analysis, could be used to aid spectral data interpretation.

Currently, dereplication techniques have emerged as the most important steps when investigating microbial extracts in the search of novel NCEs by allowing the comparison to consistent compounds libraries avoiding re-isolation of previously identified compounds

(GAUDÊNCIO; PEREIRA, 2015; JOHNSTON et al., 2015; NG et al., 2009; YANG et al., 2015). Spectroscopic techniques are supported by a wide array of experiments based on genomic-mining strategies and the advances in proteomics and metabolomics fields have been strictly present at the level of detection of peptides NPs, habitually using LC/MS-based techniques (KERSTEN et al., 2011; MOHIMANI et al., 2014).

In this sense, XPAIRT respects the metabolic expertise of NRPSs of BPB species in linking known peptides in a paired fashion and do not underestimates their capability of incorporating different subunits (nonproteinogenic amino acids, pks moieties, etc.) in biosynthetic steps. In the case of novel species, that do not have their genome fully sequenced, the identification by 16s rRNA allocates it at the taxonomical level, and its POPs could be estimated according to their taxonomically related species, as proposed by XPAIRT.

The investigation of POPs in overlapped and clustered manners showed the most diverse indexes of similarity between BPB species. The essential iBGCs for the development of XPAIRT dataset contains single and hybrids NRPS-related pathways. Even with PKS products being part of the analyses they were only treated in a general way, due to the main focus in NRPSs products.

As an example, after overlapping and clustering all results, *B. caribensis* and *B. terrae* (Jaccard index, JI = 0.8) expressed the same pairs of peptides (asp.nrp, asp.ser, orn.ser) with only one out of consensus prediction associated to a NRPS-PKS moiety (nrp.redmal from *B. caribensis*). While investigating all POPs, regions with high Jaccard Index permitted to infer about similar expression of structural subunits between BPB species (**Figure 10**).

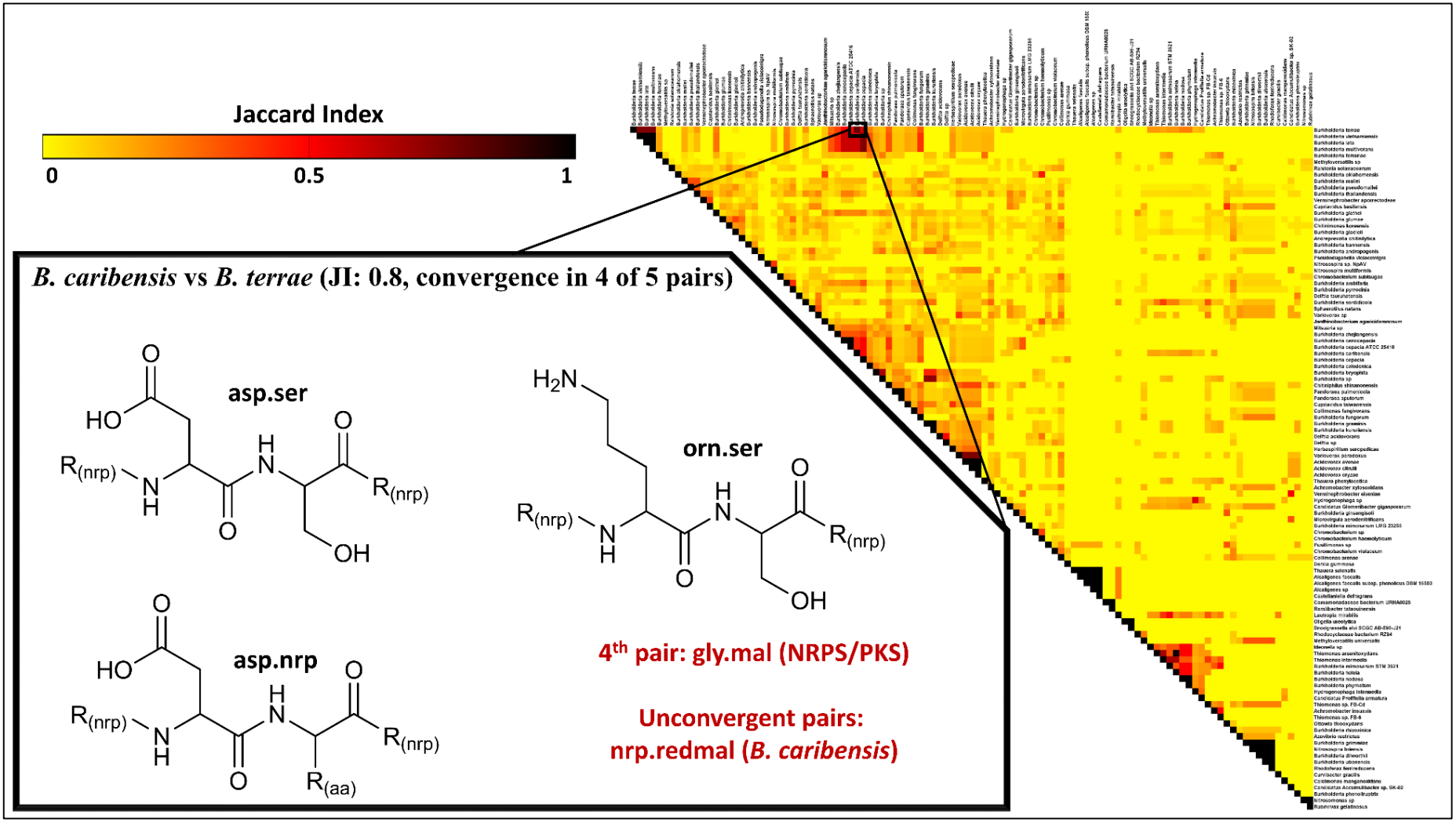

Figure 10- Heatmap comparing similarities of POPs in NRPS-related iBGCs of BPB species. One high Jaccard Index region was chosen selecting *B. caribensis* and *B. terrae*. Values closer to dark red and black mean that the POPs from different species are convergent, and when they are closer to yellow, divergent. The fourth convergent pair, gly.mal did not have its structure represented. The graph was constructed using Gitools 2.2.2 overlapping all POPs and their respective species. The clustering method adopted was Hierarchical applied to POPs. Settings of hierarchical clustering considered Euclidean distance and similarities were linked by default according to their average.

Source: From the author.

In order to attest that fragmentation by pairs was the best tactic, analyses were made by monomers and triads. It was observed that using monomers is useless, once there is no way of calculate variations of monomers linkage. On the other hand, the same strategy proceeded by triads of monomers is useful, although, the number of high Jaccard Index regions decreases drastically. As sequencing projects are still ongoing, in the future, a higher number of complete sequences will be available, making analysis by triads more efficient (**Figure 11**).

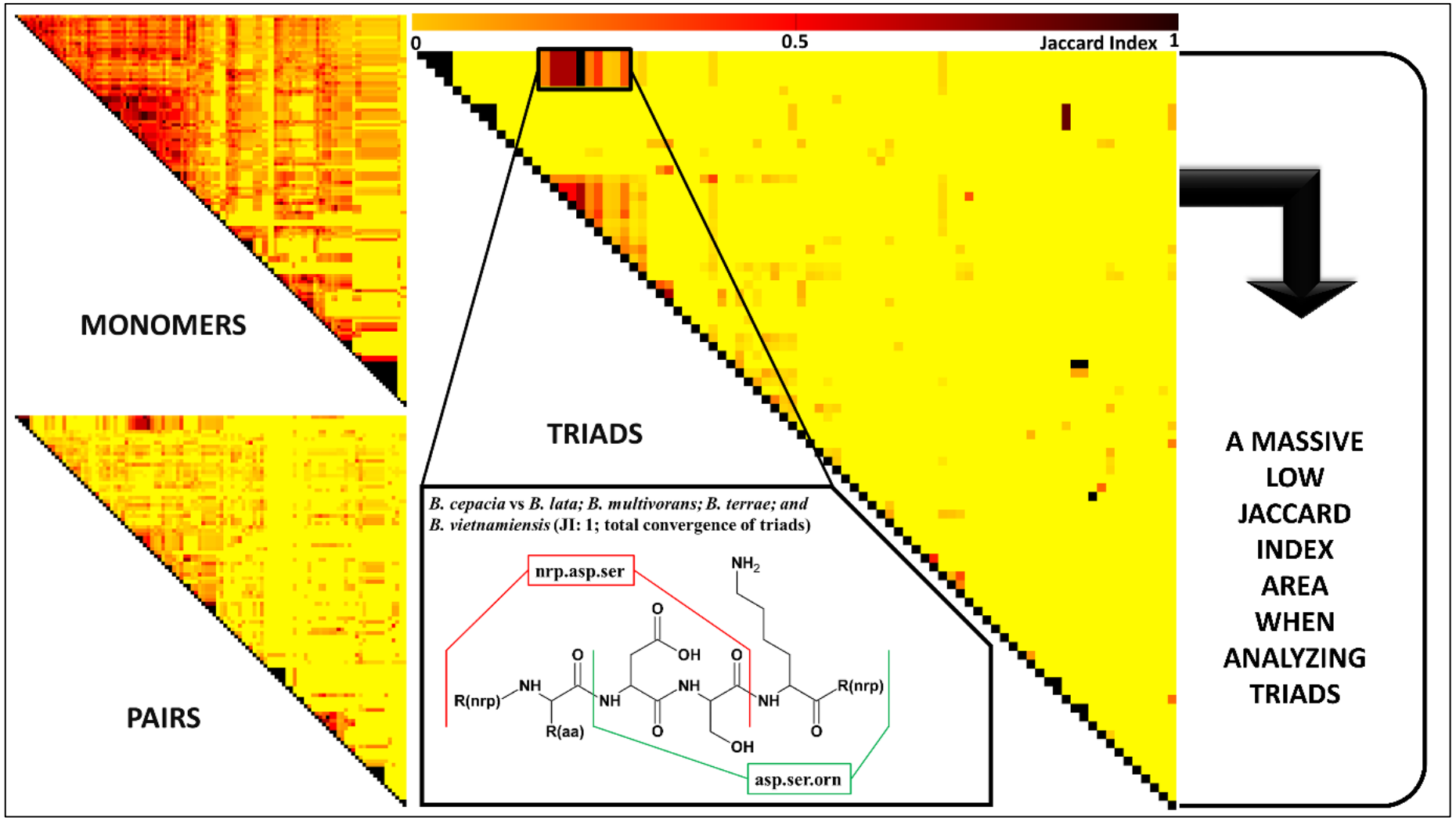

Figure 11- Heatmap comparing similarities of POPs in NRPS-related iBGCS of BPB species by TRIADS. One high Jaccard Index region was chosen selecting some *Burkholderia* species. Values closer to dark red and black mean that TRIADS from different species are convergent, and when they are closer to yellow, divergent. The graph was constructed using Gitools 2.2.2 overlapping all POPs and their respective species. The clustering method adopted was Hierarchical applied to TRIADS. Settings of hierarchical clustering considered Euclidean distance and similarities were linked by default according to their average.

Source: From the author.

The characteristic of NRPSs domains of working similarly in closely related species was previously noticed in some NRPS-PKS systems related to siderophores biosynthesis (KOEHN; CARTER, 2005) and this homologous enzymatic activities leads to the production of similar compounds (CHALLIS; RAVEL; TOWNSEND, 2000; FISCH, 2013; WANG et al., 2014). These evidences suggest that, in the nature, there are similarities in the way BGCs biosynthesize NRPs (WANG et al., 2014). Based on these evidences, 113 NRPS-related nucleotide sequences from iBGCs of BPB species identified by antiSMASH were aligned in a dendrogram with the purpose of investigating POPs expression at the genomic level.

The presence of iBGCs related to Malleobactin production (MiBIG code: BGC0000386_c1), were widely distributed into the dendrogram (**Figure 12**). This observation explains how different levels of identity and noncoding sites length infer in similar iBGCs alignment. The range of identities of Malleobactin related iBGCs happened from 14% to 92%, covering species from the genus *Burkholderia* (15 species) and *Collimonas* (1 species), and their average size around 53100aa. Their POPs are mostly associated to asp.ser (22.22%); orn.ser (20.37%); asp.nrp (14.81%); and asp.orn (7.41%), corresponding to about 65% of all POPs of these iBGCs. These POPs are expressed according to Malleobactin POCs. Although, the existence of non-Malleobactin POPs was noticed in iBGCs containing low level of identity to this NP.

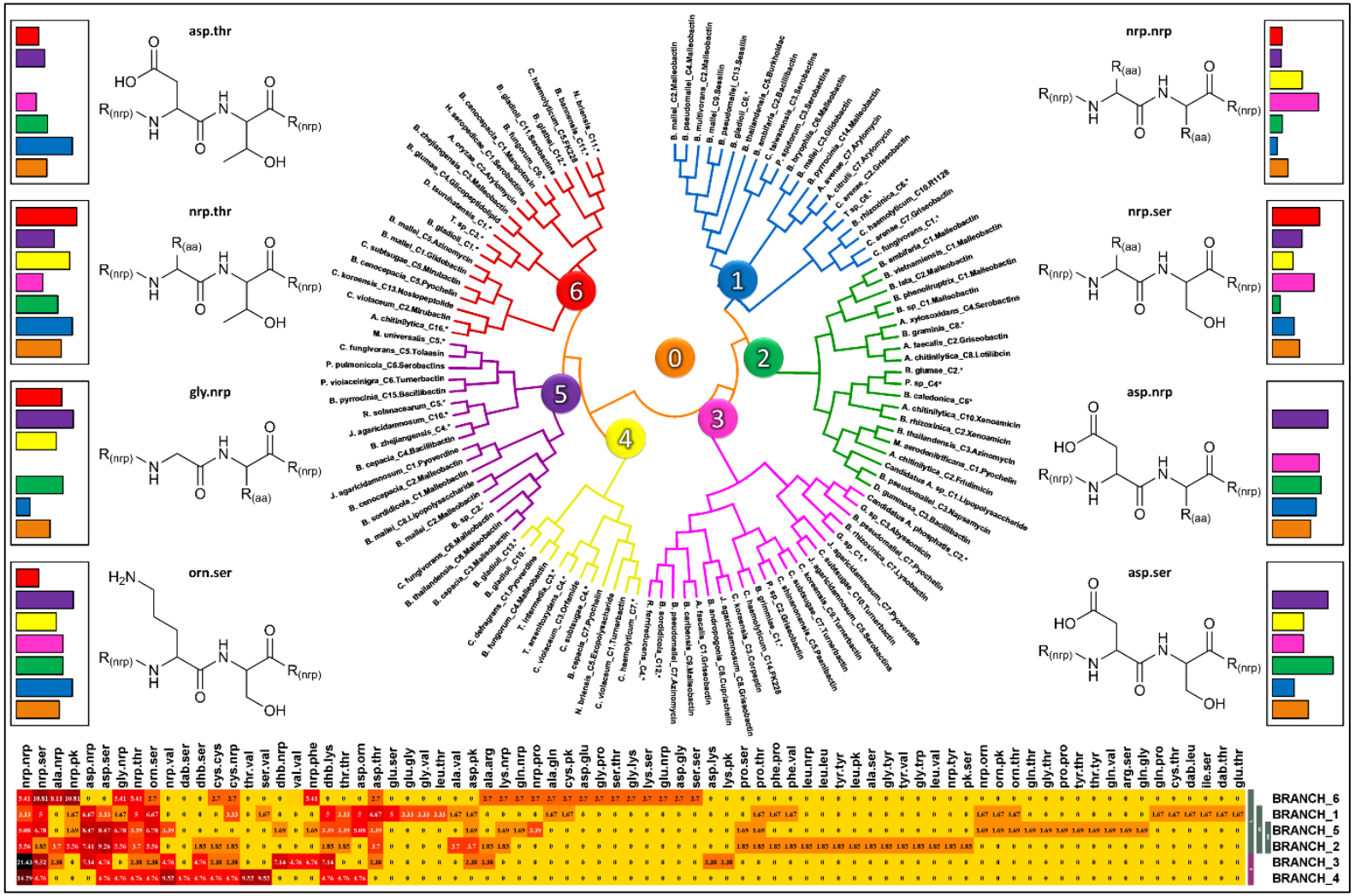

Figure 12- Distribution of POPs in 113 iBGCs sequences. Branches numbers represent main divisions of this dendrogram. The number zero (orange) represents the general percentage of occurrence of pairs in all the 113 nucleotide sequences. Each bar graph is associated directly to its respective structure. For each graph (from the bottom to the top), bar colors are associated to the branch color and their sizes are correlated to their percentage of occurrence. The absence of bars means that there is no occurrence of the respective POP on the respective branch. iBGCs sequences were aligned by Mega 6 with Clustal W algorithm, using Neighbor-Joining method. The clustered graph, was created by Gitools v.2.2.2 according to POPs. The bar graphs were made by using Excel. *iBGCs sequences that do not have their sequences associated to any known BGCs.

Source: From the author.

As observed in Delftibactin validation step, higher identities of iBGCs lead to correct matches between Malleobactin POPs and POCs. In this case, while exploring iBGCs presenting 92% of identity, non-Malleobactin POPS were inexistent. On the other hand, iBGCs presenting lower levels of identity showed the incidence of non-Malleobactin POPs (as well as, thr.thr; gly.nrp; asp.gly; asp.glu; arg.ser), suggesting that these iBGCs could synthesize different compounds (**Figure 13**).

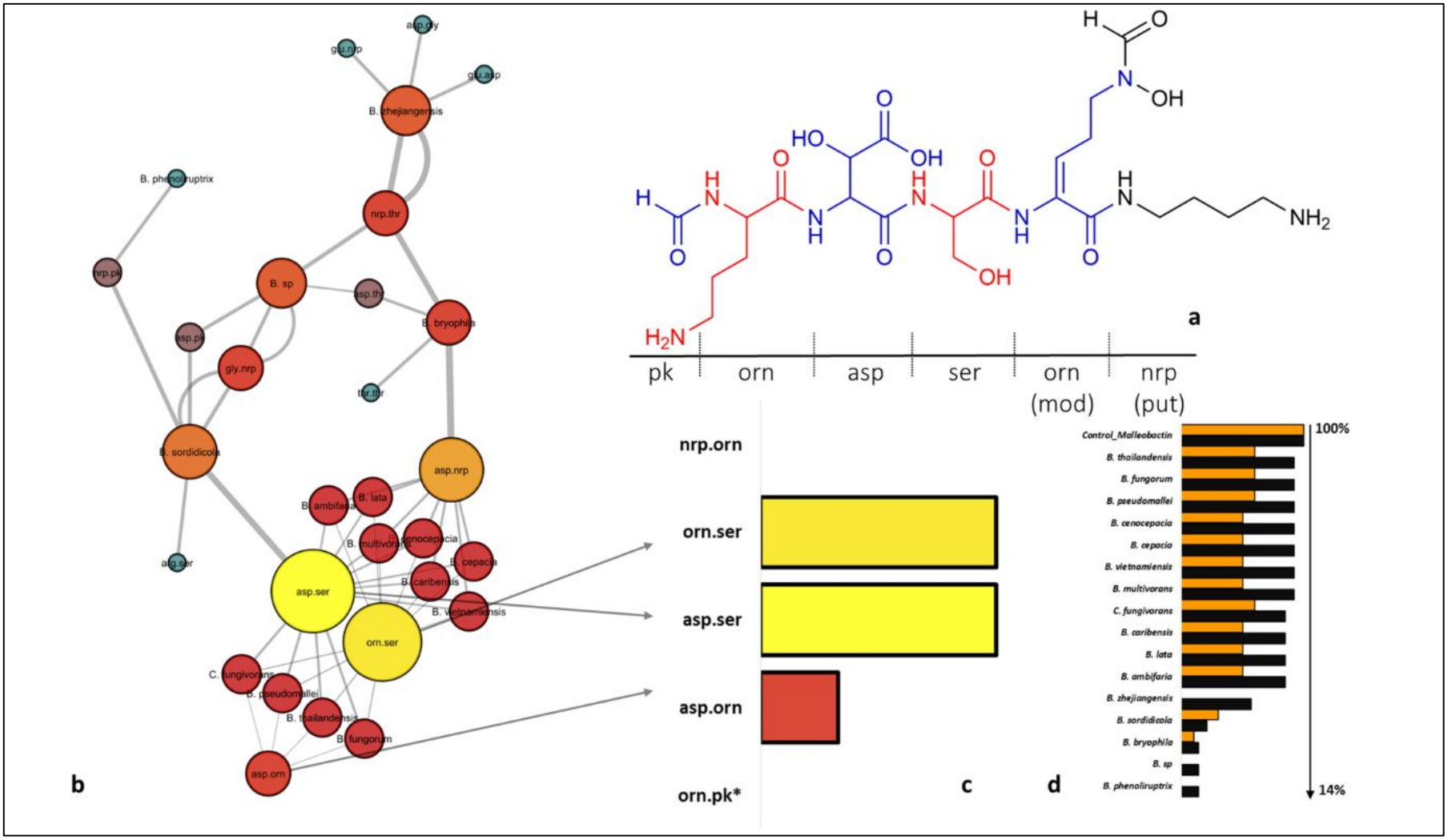

Figure 13- The Malleobactin contributors. A) The structure of Malleobactin fragmented relating the posttranslational modification sites ("mod" and "cyclic"). B) Contributors to the identification of POPs related to Malleobactin iBGCs. C) Relative percentage of each POPs and their percentage of occurrence. Pairs out of consensus with Malleobactin had lower incidence, were consider as having score 0 (zero), and lower degrees at Cytoscape. D) Relation between identity (black) and correct POPs (orange) associating the decreasing in identity and likelihood in finding POPs in a directly proportional manner.

Source: From the author.

Further investigation of the dendrogram revealed POPs specific to a given branch, for example, gln.pro, cys.thr, dab.leu, ile.ser, dab.thr, and glu.thr that are exclusive to Branch_1. However, the distribution of some POPs are spread into the dendrogram, as in the case of asp.ser, orn.ser. In addition, iBGCs sequences related to known compounds, as well as, Malleobactin, Serobactins and Sesillins were frequently grouped neighboring each other, although their occurred in different branches due to their level of identity to others iBGCs, related to the non-coding sites as identified for the biosynthesis of Malleobactin, in this case.

In order to illustrate the distribution of POPs, only the first eight presenting higher percentage of occurrence were displayed as structures on **Figure 12**. The clustered map crossing branches and POPs contains all the possibilities associated to the 113 iBGCs sequences. The clustered POPs also indicate that the distribution of NRPS-PKS hybrid machinery is common in BPB species (POPs as cys.pk; pk.ser; dab.leu, for example).

The out of consensus POP (nrp.nrp), that do not provide information about the exact monomers linked, was considered in all analyses. Relative high percentage of this POP was related to the Branch_3 and Branch_4, 21.43% and 14.29%, respectively. These values are supposed to decrease as much as the number of complete sequences of BPB species increases, given place to specific monomers. However, the occurrence of this POP indicates that there is the possibility of producing unusual NRPS-related compounds by BPB species, once the current methods are not able to identify with a high level of acceptance the correct peptide moiety.

This visual parameter shows important features, as well as, which group of iBGCs contain higher incidence of a given POP and the level of similarity to know BGCs into the dendrogram. Moreover, improving XPAIRT datasets is possible since genomic sequences are constantly deposited into genomic databases.

XPAIRT strategies were created as an effort to facilitate the elucidation of NRPs that were the main obstacle to the traditional NPs discovery process. The investigation of POPs similarities, associated to their homologous iBGCs and their contributors, is a simple manner of finding a fingerprint to support elucidation steps of NRPS-related structures. Hence, XPAIRT makes possible to identify these similarities in closely related microorganisms, as a non-costing strategy.

Finally, XPAIRT, as a genomic-based strategy guiding rapidly assignment of spectral data of BPB species, is possible to be built to any type of microorganisms classes that have their genomic information available in order to investigate their metabolomic capability.

5.1.4 CONCLUSION

Nonribosomal peptides have a featured position of the chemical space providing the most medically relevant compounds efficient to the clinical use by their biological activities. More than 500 building blocks, diverse architecture, peptides modifications, and the union with polyketide pathway increases exponentially the way these monomers are linked by these synthetases, turning arduous the process of structural elucidation.

Due to these features, XPAIRT showed that closely related microorganisms presents similarities at the biosynthetic level. The NRPSs related to Betaproteobacteria were clustered and several matches were encountered, as in the case of the two examples cited, Malleobactin and Delftibactin. POPs as asp.orn, asp.ser, orn.ser, etc, were expressed equally to POCs. They were correlated to all possible genomic sequences that presented different levels of identity to assembling these monomers in a paired manner. The use of clusterization methods made possible to correlate the distribution of pairs according to their phylogeny showing that species on the Branch_1 express exclusively glu.thr, dab.thr, ile.ser, cys.thr, and gln.pro. These pairs are not encountered in any other branch. The Jaccard Index experiments showed that the paired manner is the best way of analyzing these similarities and an abundant number of correlations provided an overview of all direct contributors.

In this sense, XPAIRT, as a genomic-guided strategy based on predicted metabolomics, is a simple, rapid and non-costing guide for quickly accession of probable structural moieties of NRPS-related compounds. These analyses facilitate the assignment of spectral data of novel NRPS-related compound candidates, by investigating trends in NRPS-related machinery at their taxonomic level. Finally, any efforts in aiding the discovery processes of novel NPs is essential in order to investigate the pharmacological potential of a NCE.

5.2 INVESTIGATION OF BIOSYNTHETIC CAPABILITY OF *Burkholderia thailandensis* STRAINS RELATED TO BGCs INTRASPECIFIC VARIATION.

XPAIRT results made possible to elect the best candidates to be analyzed using genomic and bioinformatics tools according to the BGCs counting. The BGCs summing of *Burkholderia* species were far larger than all other species making them the target to be further studied in this work. In this sense, the standard microorganism *B. thailandensis* was used in order to evidencing NPs capability in its different strains.

Species of the genus *Burkholderia* have emerged as owing promising biosynthetic capability for diverse Natural Products (NPs). Recently, a remarkable study about the potential of microorganisms in biosynthesizing NPs pointed that Proteobacteria species present large number of Biosynthetic Gene Clusters (BGCs) (CIMERMANCIC et al., 2014). This global analysis included species from the genera *Pseudomonas* spp. and *Burkholderia* spp. as containing the majority of BGCs counting for Proteobacteria representatives. However, the great interest is related to *Burkholderia* spp., once *Pseudomonas* spp. are extensively studied. About the likelihood of producing novel NPs, *Burkholderia* genomes present, statistically, higher percentage of thiotemplate modular systems than those of bacilli, cyanobacteria, myxobacteria and fungi, and is only second to that of *Actinobacteria* (MINOWA; ARAKI; KANEHISA, 2007). These modular systems are related to the production of many classes of pharmaceutical compounds, including PKS- and NRPS-related products (DIMINIC et al., 2013). These evidences lead to the investigation of *B. thailandensis* strains consistent with their NPs biosynthetic potential using bioinformatics tools to track possible NPs changing in all available strains of this microorganism according to **Figure 14**.

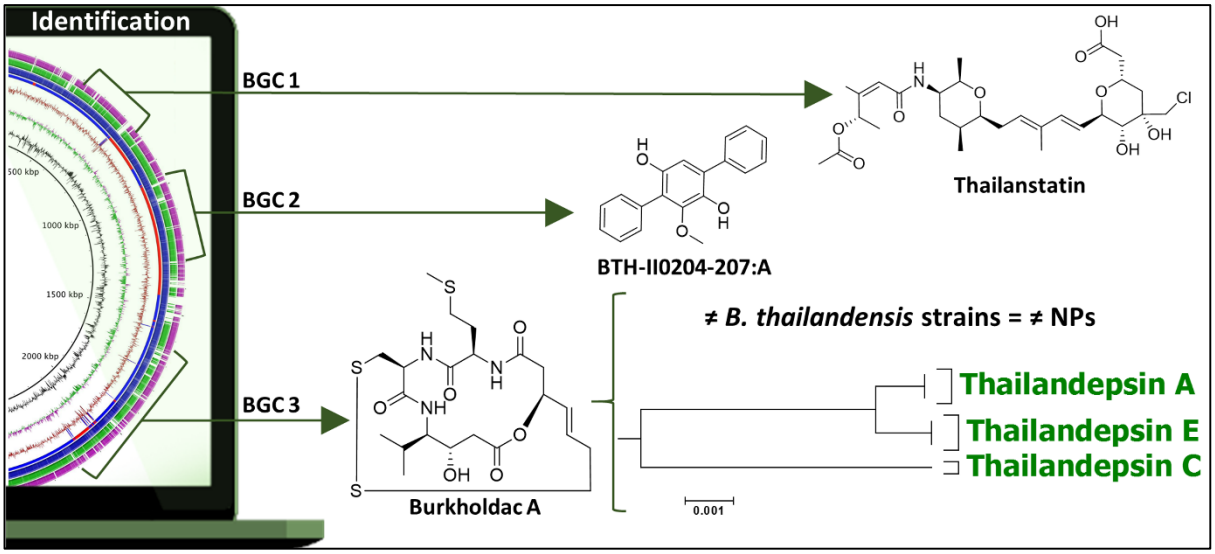

Figure 14- Replicons of *Burkholderia thailandensis* strains related to the production of three Natural Products. The BGCs related to Burkholdac (Thailandepsis) were further analyzed and their side chain differentiation were tracked according to their genomes explaining the side chains modifications for this NP.

Source: From the author.

According to the DMSZ database, the genus *Burkholderia* comprises more than 90 species (Accessed in March, 2016). These species inhabit the most diverse types of ecological niches such as soil, water, rhizosphere and plant surface (LUDOVIC et al., 2007).

NPs from *Burkholderia* spp. are structurally and functionally diverse, comprising benzoquinone; lactone; polyene compounds; lasso peptides; nonribosomal peptides; statins; and other polyketides. These compounds present important biological activities. In addition, some of these small molecules from *Burkholderia* sp. have entered as drug candidates to preclinical evaluation (WILSON; CHENG; KHABELE, 2012).

Burkholderia thailandensis (B. thailandensis) E264 presented a high level of similarity to the BGC BTH-II0204-207 from *Burkholderia pseudomallei* (*B. pseudomallei*) K96243 related to the production of betulinan/terferol analogues (EL-ELIMAT et al., 2013). Experiments indicates that the compound isolated from this BGC, **BTH-II0204-207:A** (**31**), is a potent PDE4 inhibitor (BIGGINS et al., 2011). It is worth to highlight that the first compound considered as PDE4 inhibitor to the treatment of chronic obstructive pulmonary disease was approved by US Food and Drug Administration (FDA) in 2011 (BIGGINS et al., 2011; TAEGTMEYER; LEUPPI; KULLAK-UBLICK, 2012). A specific BGC of *B. thailandensis* E264 also presents high level of identity to the BGC responsible to the production of class II lasso peptides of *E. coli*. Further studies with the strain E264 provided the discovery of a class II lasso peptide called Capstruin (structure not shown) that presents antimicrobial activities (KNAPPE et al., 2008). Species as *B. mallei* ATCC 23344, *B. pseudomallei* K96243 and *B.*

thailandensis E264 presented similar BGCs that encode hybrid polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) pathways presenting unusual domains that provided **malleilactone (33)** and **burkholderic acid (34)** (BIGGINS; TERNEI; BRADY, 2012; FRANKE; ISHIDA; HERTWECK, 2012). These compounds presented respectively, moderate activity against Gram-positive bacteria and weak cytotoxicity.(BIGGINS; TERNEI; BRADY, 2012; FRANKE; ISHIDA; HERTWECK, 2012). Genome-guided approaches led to the isolation of an interesting class of polyene amides named thailandamides from *B. thailandensis* E264, related to a hybrid 17-module *trans*-AT PKS-NRPS pathway. **Thailandamide lactone (35)** presented moderate antiproliferative activity against human tumor cell lines (ISHIDA et al., 2010; NGUYEN et al., 2008). The thailandepsins also isolated from *B. thailandensis* E264 are related to FK228 BGC in *Chromobacterium violaceum* (CHENG; YANG; MATTER, 2007). FK228 is an FDA approved anticancer drug related to the treatment of refractory cutaneous and peripheral T-cell lymphoma (KM. VANDERMOLEN, WILLIAM MCCULLOCH, CEDRIC J. PEARCE, 2011). The **thailandepsins A** (**30**) and **B** (**30**) also possess growth inhibition characteristics to different cancer cell lines, as well as, colon, melanoma and renal cancer cells (WANG et al., 2011a). However, the mechanism of action needs to be further studied.

Other interesting NPs from *B. thailandensis* MSMB43 are the **thailanstatins A (31)** and **B (37, Figure 15)**, belonging to the FR901464-family of microbial products that have a pyran ring heavily substituted with different groups in one end and an acetyl group at the other end. The biggest difference between **thailanstatins** and **FR901464** is the lack on the hydroxyl group and the presence of a carboxyl moiety, resulting in higher stability to **thailanstatins** (LIU et al., 2013). **Thailanstatins** inhibit mRNA splicing and are related to antiproliferative activities against human cancer cell lines (LIU et al., 2013). **Thailanstatins** compounds also showed therapeutic application against glaucoma due to modulation of glucocorticoid receptor splicing process (JAIN et al., 2013).

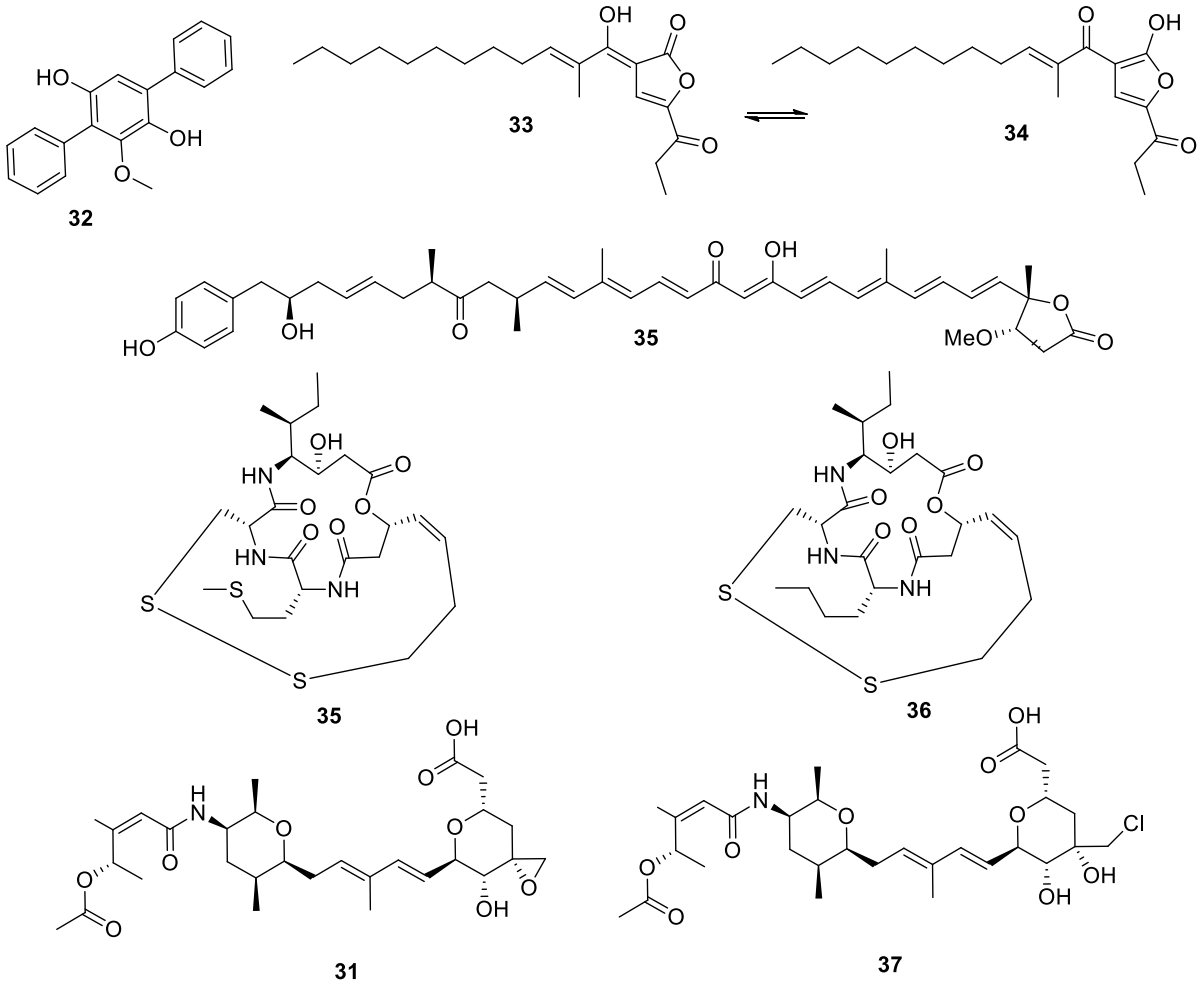

Figure 15- NPs isolated from *Burkholderia thailandensis* and their singular structures.
Source: From the author.

At a glance, genome-mining strategies are taking important role in the field of NPs discovering. Computational Biology and Genomics are changing the approach of NPs research by understanding specifically how nature produces compounds (MEDEMA et al., 2014). In this sense, species of the genus *Burkholderia* are providing an extensive number of NPs after genome-guided strategies applied. *B. thailandensis* has presenting an increasing interest due to its biosynthetic capability (ZHUO et al., 2012). Since the most studied strain is the *B. thailandensis* E264, all available genomic sequences related to this species were investigated in order to evaluate the potential of other *B. thailandensis* strains shedding lights to their biosynthetic potential, guiding NPs discovery process.

5.2.1 MATERIAL AND METHODS – BGCs INTRASPECIFIC VARIATION:

All material and methods related to the Intraspecific Variation experiments are showed ahead.

5.2.1.1 Dataset

Genomic information related to *B. thailandensis* strains were downloaded from the National Center of Biotechnology Information (NCBI) genome database (http://www.ncbi.nlm.nih.gov/genome). Chromosome I and II sequences were chosen in order to correlate the differentiation between them for all available sequences. All sequence files were downloaded according to the available data at NCBI on November 2015. All analyses used Chr1 results of an Actinobacteria representative *Streptomyces coelicolor* A3(2) (*S. coelicolor* A3(2)) as a control in order to compare similarities and uniqueness of *B. thailandensis* strains (Proteobacteria). *S. coelicolor* A3(2) is a well-studied antibiotic-producing bacterium (Accession Number: NC_003888.3).

5.2.1.2 BGCs finder and classification

BGCs predictions were made using antiSMASH 3.0 (WEBER et al., 2015). Investigations were made in order to search for, among other results, an overview in genomic information allowing detecting classes of compounds, level of genomic similarities to already known compounds, and their core structures, including the comparative gene cluster analysis. NRPS-PKS predicted results were elected according to the consensus between the classic comparisons made by the database using NRPSprediction2, Stachelhaus code and Minowa.

5.2.1.3 Network Profile

Datasets were submitted to Cytoscape 3.3.0 to the creation of a statistical visual correlation between Chr1 and Chr2 of *B. thailandensis* strains and the standard. The algorithm used in order to correlate results of chromosomes was Allegro Layout 2.2.3 with Fruchterman-Reingold Layout. Networks and clusters were formed for each step of analyses: a) chr vs type of compound and b) chr vs identity. Visual statistics were built correlating results to their respective degrees.

5.2.1.4 Jaccard Index Calculation

Jaccard Indexes (JI) were calculated in order to correlate the level of similarities between strains of *B. thailandensis* and compounds classes, structure and identity to known BGCs. Raw results were overlapped and clustered. The clustering settings followed hierarchical parameters, using values from Jaccard Index scores and Euclidean distance as standard method. Analyses were proceeded using the software Gitools 2.2.3.

5.2.1.5 Dendrogram Similarity

Taxonomic analysis about diversification of *B. thailandensis* BGCs was proceeded using sequences identified by antiSMASH results. BGCs sequences were downloaded as .fasta format and named according to their respective species, chromosome and type of compound homology. Nucleotide sequences were aligned with the purpose of comparing the distribution of BGCs of *B. thailandensis* strains and their identity levels at the phylogenetic level. The dendrogram was built in MEGA6 (TAMURA et al., 2013). The analysis involved 100 nucleotide sequences aligned by CLUSTAL W using default parameters (LARKIN et al., 2007). The distribution of BGCs classes was inferred using the Neighbor-Joining method (SAITOU; NEI, 1987b). Comparisons between higher identities levels and BGCs grouping were calculated with the purpose of differentiate BGCs characteristics according to their NPs most probable related structure.

5.2.2 RESULTS AND DISCUSSION

Most bacteria have one or two circular replicons that encode set of genes for the most diverse functions, including the production of NPs. In the case of genes encoding information of the biosynthesis of NPs , these sets of genes are called BGCs (MEDEMA; FISCHBACH, 2015). These BGCs encode diverse types of information related to enzymes, regulatory proteins and transporters that are essential to the biosynthetic machinery of a given metabolite. BGCs data also allows to mine genomes and identify sets of genes that participate in a specific biosynthetic pathway by computational analysis predicting their possible final products.

B. thailandensis is a model microorganism of the genus *Burkholderia* for the investigation of NPs biosynthetic capability. Due to its potential, the strain *B. thailandensis* E264 is well studied and diverse research groups has shown different methods to obtain NPs from it. In this sense, other *B. thailandensis* strains are also promising, since they share a significant level of DNA similarity.

Comparisons between Chr1, Chr2 of *B. thailandensis* strains along with *S. coelicolor* A3(2) and their related types of compounds are showed in **Figure 16, A**. Network results showed that both chromosomes of *B. thailandensis* putatively produce unique chemistry when compared to the standard.

The most probable class of compounds from these species are correlated to NRPS, TERPENE, and T1PKS. As usual, all chromosomes encode information to BGCs correlated to the production of BACTERIOCINS, that are proteinaceous toxins naturally produced by bacteria in order to colonize the environment in which they occur (BAKKAL et al., 2010).

Chr1 and Chr2 of *B. thailandensis* do not correlate themselves to their class of compounds other than the four classes cited above and Chr2 presents major diversity of classes. Due to evolutionary questions related to the production of specific compounds, these classes of compounds are different between replicons (**Figure 16, B**). In addition, independently of the four common classes of compounds correlated to all species, compounds linked to *S. coelicolor* A3(2) presented no direct correlation to Chr1 or Chr2 of *B. thailandensis* strains.

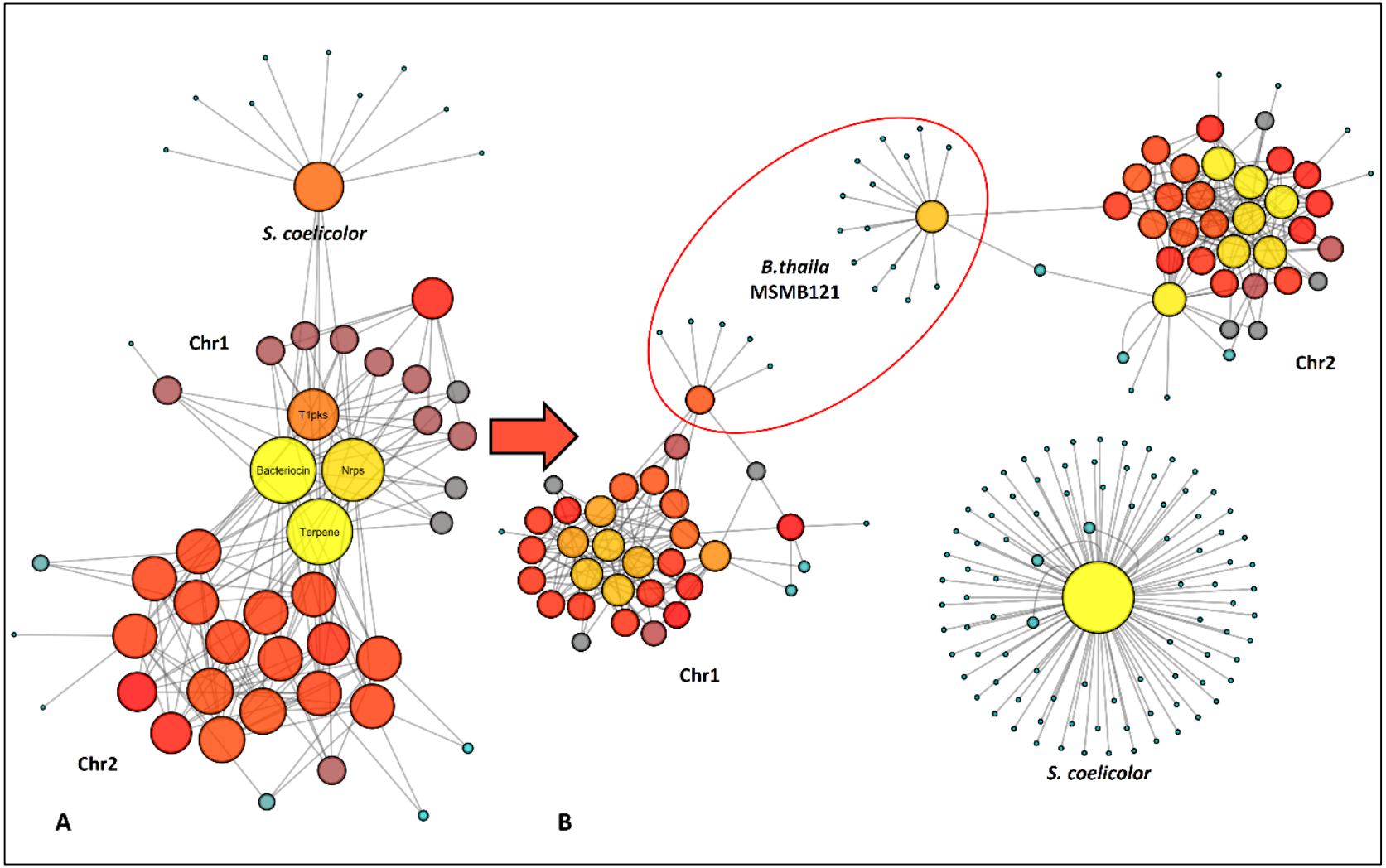

Figure 16- Network correlating replicons of *B. thailandensis* strains according to their classes of compounds and level of similarity to already known compounds. A) Type of compounds and their connection between Chr1, Chr2 and *S. coelicolor* A3(2). B) Identity between species investigated highlighting the most dissimilar strain. *S. coelicolor* A3(2), the standard strain show none correlation to *B. thailandensis* strains homologies.

Source: From the author.

Core structures produced by *B. thailandensis* strains do not present similarities when comparing Chr1 and Chr2. Their NRPSs machinery seems to not assembly the same monomers due to different levels of evolution between these two replicons. JI levels of core structures are also very small between Chr1 and Chr2 (**Figure 17**). However, when investigating one chromosome at time, they present genomic information pointing to the production of similar structures. In the case of different species, as reported in the literature, different substituents could be present in the same class of compounds, leading to different NPs including improvements of their biological activities (SCHAFFER; OTTEN, 2009; XIE et al., 2014). These results are well explained at **Figure 17**.

The BGCs associated to the Chr1 of *B. thailandensis* species generally encode information for NRPSs to assembly monomers such as alanine-arginine; cysteine-threonine and ornithine-aspartate-serine, while Chr2 assembles cysteine-cysteine; valine-glycine; and malate-aspartate most of the time. Chr2 of *B. thailandensis* strains also present higher level of genomic information related to hybrid pathways.

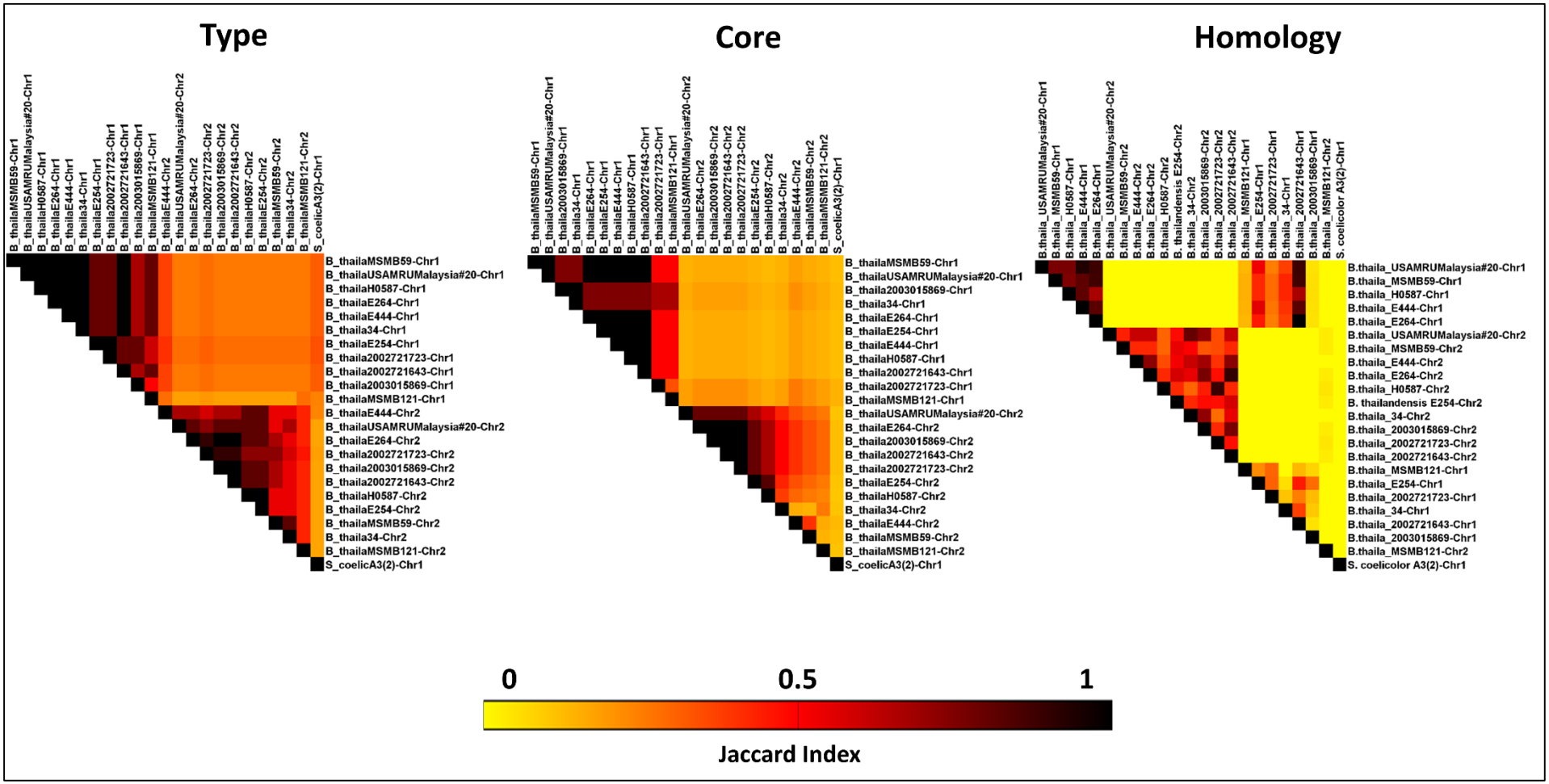

Figure 17- Overlapping Crh1 and Chr2 at three different levels of comparisons. Type, Core and Homology correlations are showed in Jaccard Index scores. Results are calculated based on comparisons with *S. coelicolor* A3(2) strain.

Source: From the author.

Lately, the genus *Burkholderia* has shown to produce differentiated NPs, even in known classes of compounds (PIDOT et al., 2014). Due to this characteristic, the use of *S. coelicolor* A3(2) as standard was successful in order to confirm uniqueness of *B. thailandensis* strains. The BGCs of *B. thailandensis* showed the lowest values of JI when compared to those of *S. coelicolor* A3(2). These values suggest that *B. thailandensis* species are highly different to the standard in the potential of biosynthesizing NPs.

In addition, core structures identified from Chr1 and Chr2 of *B. thailandensis* seems to not be originated from high levels of nucleotide identity to cluster themselves with similar scores, implying that they are independent in how they work in order to biosynthesize NPs.

One interesting detail about all analysis is that the strain MSMB121 presents higher levels of differentiation to all *Burkholderia* strains studied in this work. Chr1 and Chr2 of MSMB121 presented, respectively, 4 and 14 unique levels of identity to all other BGCs. These 18 levels of similarity do not present connection to other species in this work, suggesting that novel NPs from this strain could have structural moieties similar to their respective known NPs. Preliminarily, these observations suggest that *B. thailandensis* MSMB121 holds greater chances of biosynthesize novel NPs than other *B. thailandensis* strains. For comparisons, the second position in this analysis is *B. thailandensis* MSB59, containing 4 exclusive levels of similarity summing results of both chromosomes. Different chromosomal levels of similarity could imply directly in monomers flexibility leading to improvements in biosynthetic steps ending in different NPs (SCHAFFER; OTTEN, 2009; XIE et al., 2014). This could be explained observing different clusters related to the production of thailanstatins. Their different levels of similarity are related to thailanstatins-like compounds with different substituents or side chains.

Hierarchical analysis showed that the Chr2 of *B. thailandensis* MSMB121 and *S. coelicolor* A3(2) are grouped in the same sub-branch. This strain is the most similar to the standard between all *Burkholderia* strains studied in this work. Genomic comparisons between the standard and the strain MSMB121 would provide useful information about their NPs potential (**Figure 18**).

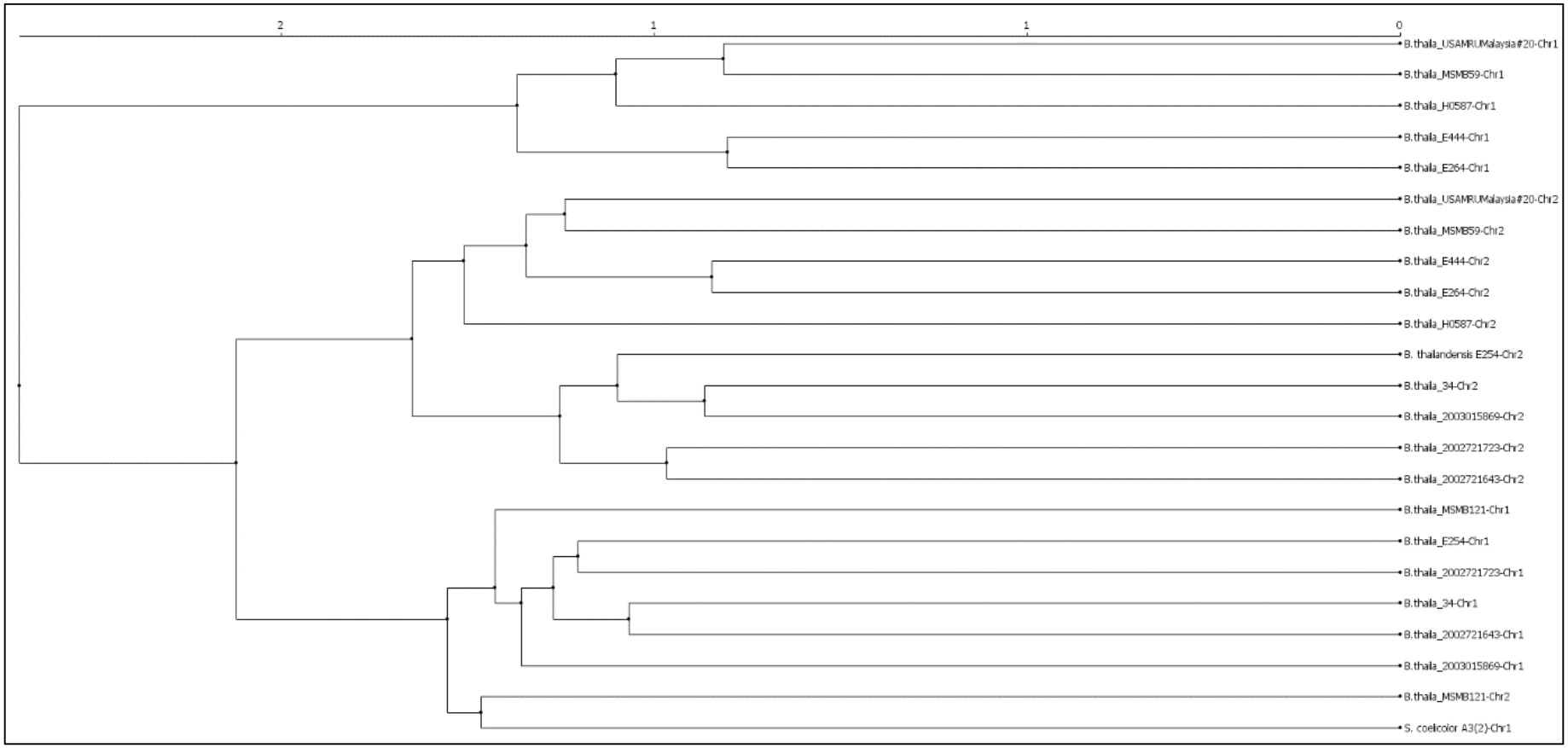

Figure 18- Hierarchical analysis of all chromosomes of *B. thailandensis* strains and *S. coelicolor* A3(2) according to their identities level to BGCs of already known compounds. This analysis was carried out by Gitools 2.2.3 using Euclidean distance. Results were linked by average of similarity and their scores. The strain *S. coelicolor* A3(2) was used as standard to compare identities. The last branch from the top to the bottom is grouping *B. thailandensis* MSMB121 and *S. coelicolor* A3(2) showing that this *Burkholderia* strain is the most similar strain according to their level of identity to BGCs of known compounds.

Source: From the author.

Results of hierarchical analyses using Euclidean distance further investigating NRPs- and PKs-related structures showed that the largest group expressing higher levels of similarity to the production of NPs is composed by the strains E444, E264, and H0587 (**Figure 19**).

They core structure is based on NRPS-related BGCs assembling alanine and arginine most of the time and presenting 4% of nucleotide identity to the BGC related to Azinomycin B. On the other hand, the strains grouped by their similarities in the Chr2 are the strains 2002721723, E264, 2002721643, and 2003015869. These strains are strongly correlated to the BGC that encodes information for the production of Malleobactin (identity greater than 90%), Pyochelin (identity of 100% for all strains) and Bactobolin (identity of 100% for all). These results confirmed that Chr1 and Chr2 of *B. thailandensis* strains are independent in the way they encode information for the biosynthesis of NPs and their genomes seems to be quite dissimilar to the standard.

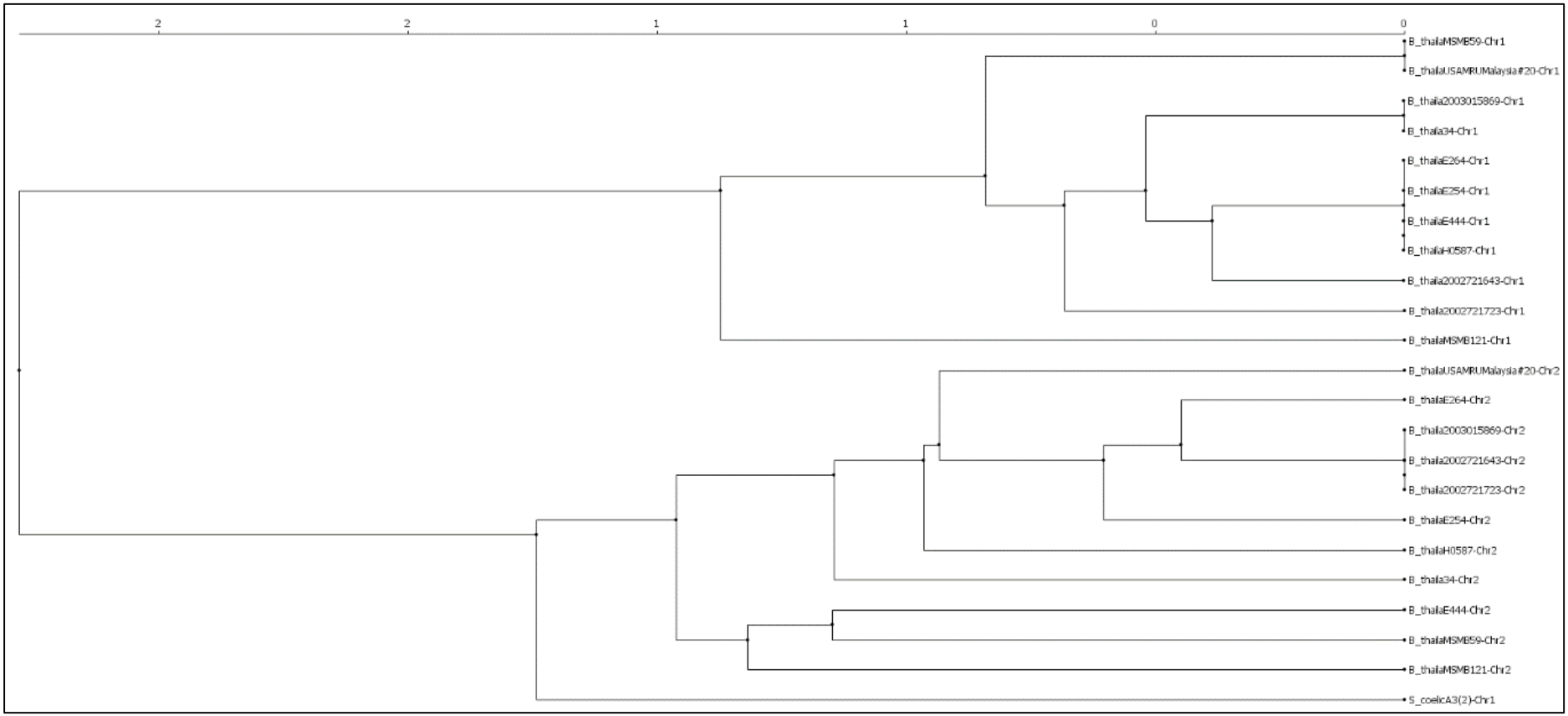

Figure 19- Hierarchical analysis of all chromosomes of *B. thailandensis* strains and *S. coelicolor* A3(2) correlation their cores structures. This analysis was carried out by Gitools 2.2.3 using Euclidean distance. Results were linked by average of similarity and their scores. The standard strain S. coelicolor A3(2) was added to this analysis for homologies comparisons. Groups of similar core structures producing strains are present explaining that their BGCs are similar at the genomic level.

Source: From the author.

Since genome-mining results showed that *B. thailandensis* strains differ in the level of similarity of their BGCs, all sequences classified as PKS- and NRPS-related compounds were further investigated at the genomic level.

BGCs of both chromosomes of *B. thailandensis* strains, as well as *S. coelicolor* A3(2) were aligned according to their nucleotide sequences. Divergent sequences expressing same core structures hold interesting features for the production of novel compounds due to the possibility of eliciting silent bacterial gene clusters after investigation of their biosynthetic power, resistance and metabolic profiles (SEYEDSAYAMDOST, 2014).

In the dendrogram, the differences in branch length of a specific subtree referring to a NP indicate the likelihood of biosynthesizing unique compounds to those already isolated, due to the variable levels of identity between BGCs. These small singularities in each group, are identified as possessing different sum of branch length explaining nucleotide modifications in all BGCs.

In the case of the subtree related to Malleilactone (**Figure 20**), there are two groups of BGCs, one (starting at position 1) containing two and other (starting at position 58) containing nine different BGCs. To the bigger subtree (position 58), there are four sequences presenting high level of DNA identity placed in the same sub-branch. On the other hand, the five other BGCs present dissimilar alignment. These evidences suggest that enzymes related to these BGCs could lead to different Malleilactone-related compounds in the biosynthetic steps. The same might occur to all other subtrees, tracking *B. thailandensis* strains potential in biosynthesize novel NPs.

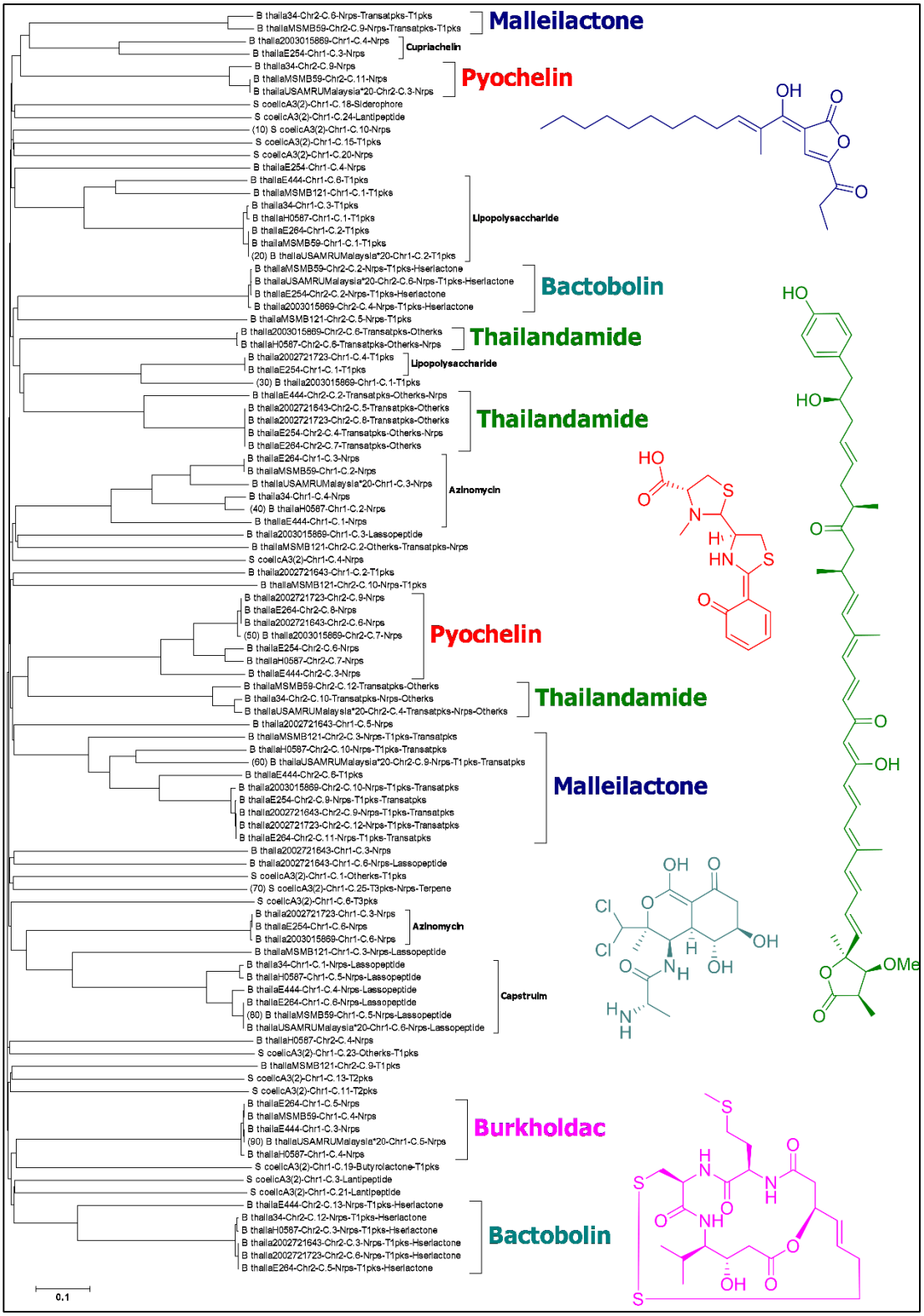

Figure 20- Distribution of *B. thailandensis* and *S. coelicolor* BGCs related to NRPS- and PKS-related compounds. Subtrees related to a specific compound were highlighted with brackets and each main compound was colorized according to its subtree. Differentiation in sum length relates the likelihood of different compounds based on the structure of the main compound. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. The analysis involved 100 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 923 positions in the final dataset.

Source: From the author.

The alignment of *S. coelicolor* BGCs are placed individually. When correlated to *B. thailandensis* strains, they are placed along with BGCs related to the production of Azinomycin. The highest level of identity encountered to *B. thailandensis* BGCs containing information to the production of Azinomycin was 4%, suggesting that the information present in these clusters mostly lead to the biosynthesis of other compounds.

These dissimilar branch length explains how compounds of the same class are biosynthesized by similar BGCs and the occurrence of different side chains or substituents. In some cases, there is the possibility of moieties presenting similar characteristics (polar or non-polar amino acids, for example) to be placed according to enzymatic steps involved in the biosynthesis of NPs (YANG et al., 2015). In the case of *B. thailandensis* MSM121, the range of substitution per site, related to each BGC, is larger than other strains. After aligned, the group of BGCs related to Malleilactone biosynthesis (position 58) presents the strain MSMB121 as possessing the larger differentiation compared to others in this subtree, followed by the strains H0587 and USAMRUMalasya*20. This could be observed in all *S. coelicolor* A3(2) BGCs, that present lager genomic differentiation compared to all *Burkholderia* strains. As genomes are strictly connected to biosynthetic pathways and the production of NPs, small identities (or higher levels of substitution per sites) in genomic information leads to the production of different NPs. Thus, results of network and phylogeny are direct correlated presenting no linkage to any level of similarity to BGCs related to the standard strain. These observations explain the reason of NPs from *Burkholderia* strains are very dissimilar to the standard suggesting that *B. thailandensis* is a good reservoir of novel NPs. The published work related to the intraspecific variation of BGCs of *B. thailandensis* strains can be accessed according to the following reference (BALDIM; SOARES, 2016).

5.2.3 CONCLUSION

Biosynthetic pathways are being strictly correlated to their genomic information in order to understand how diverse mechanisms are encoded, including the biosynthesis of NPs. Genomic-guided strategies are part of a new way of understanding how NPs are produced and how different they could be when investigating different strains of a given species based on their chromosomal composition.

In addition, experiments adopted in this work shed light to how NPs discovery processes are supported by stat-of-the-art techniques in order to give information that in the past was impossible to achieve making the process of discovery highly rationale. After analyzes of all *B. thailandensis* strains, it was possible to infer that there are great chances of isolating novel NPs using specific culture media due to their biosynthetic capability and the likelihood of side chains modifications of known NPs due to their phylogeny. Finally, the differentiation in their alignment revealed that similar NPs belonging to known classes often occur, increasing the potential of this species in biosynthesizing novel compounds.

5.3 GENOME-GUIDED ISOLATION OF PROMISING MICROORGANISMS ACCORDING TO THEIR BIOSYNTHETIC CAPABILITY.

After understanding how different is the capability of Betaproteobacteria strains in biosynthesize NPs, their characteristics, and which are the best genus to select and isolate from environmental sources in order to obtain novel NPs, a genomic-guided strategy was developed for *Burkholderia* species aiming to isolate them from environmental samples.

As related before, genomic studies revealed that resistance to antimicrobial agents is common in areas under anthropogenic action. Conversely, there are natural reservoirs of resistance in remote environment with no human contact, which can explain wild microorganisms having similar characteristics to those ones that science have been struggling to improve human health (ALLEN et al., 2010). This approach became important to develop a method to access *Burkholderia* species from environment, since genomic evidence found that this genus has a substantial number of biosynthetic gene clusters able of producing secondary metabolites and the resistance genes are naturally widely spread (ALLEN et al., 2010; CIMERMANCIC et al., 2014; LIU; CHENG, 2014; PIDOT et al., 2014).

The genus *Burkholderia*, discovered by Walter Bulkholder nearly 1950, has emerged as a new source of NPs and presents currently about 85 species according to the DSMZ-Germany (Bacterial Nomenclature Up-to-Date, Genus: Burkholderia; www.dsmz.de) and, after several analysis, different compound classes were discovered having their biological activities tested. However, there is a slight number of species studied in the chemistry point of view at present (LIU; CHENG, 2014).

Based on these discoveries and the fast development of online bioinformatics databases, nowadays is possible creating reliable and effective methods to isolate microorganisms, since there is massive interest in discovering new classes of NPs for the most diverse bioactivities (AMINI; TAVAZOIE, 2011; MEDINI et al., 2008). For that, a new isolation method highly selective for species of the genus *Burkholderia*, was developed, as a model method, using genomic, theoretical metagenomic and metabolic information to access its peculiarities at the microcosms level.

In this experimental section, all the restrictive conditions were found to avoid untargeted microorganisms and select only the target microorganisms (from the genus *Burkholderia*) using their basic characteristics as antibiotic and metal resistance, and their capability of metabolize compounds with biological role as carbon and nitrogen sources (C&N-

Sources). The method was compared to existing culture media (selective for one species only) (FRANCIS et al., 2006; KAWANISHI et al., 2011; MERK et al., 2001; VANLAERE et al., 2005, 2006; VERMIS; VANDAMME; NELIS, 2003; ZAID; BONASERA; BEER, 2012).

For the development of a highly specific isolation method, the seven most important constraints were controlled for avoiding drastically untargeted microorganisms to grow. They are 1) Antibiotic Resistance; 2) Metal Resistance; 3) Gene Similarities; 4) Amino Acids and 5) Metabolic Predictions as C&N-Sources; 6) Specific Nutrients and 7) pH as Basal Conditions. In addition, all of them correlated to species from metagenomic studies, for the first time with this purpose, reporting strains found in the same ecosystem in which the genus *Burkholderia* occurs. This strategy turned possible the development of a highly selective method using state-of-the-art techniques in a powerful and elegant way. The **Figure 21** explains step-by-step the workflow of the isolation technique.

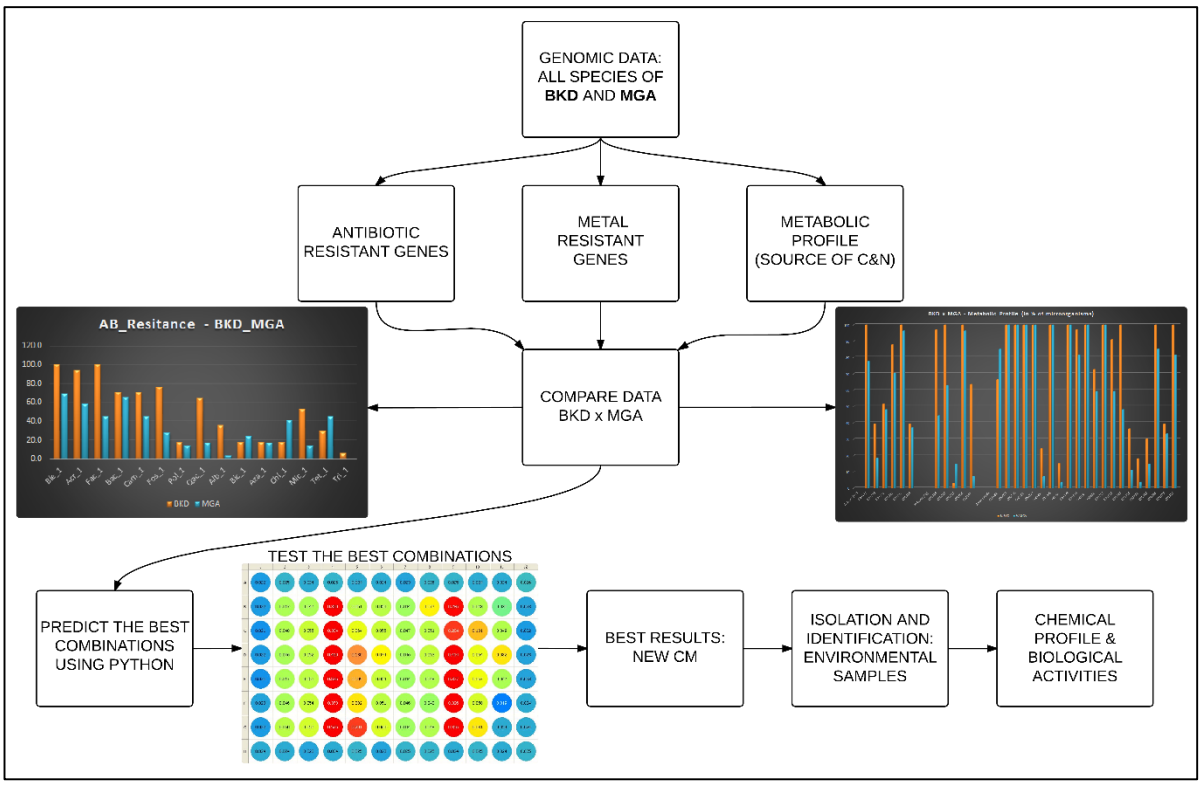

Figure 21- Workflow related to the strategy for the isolation of *Burkholderia* species from environment highlighting creation of resistance and metabolic profiles, predictions using python and test of all combinations in order to elect the best genome-guided candidates.
Source: From the author.

Thus, this strategy could show a novel way of creating a library of microorganisms in order to investigate their NPs biosynthetic potential.

5.3.1 MATERIAL AND METHODS - GENOME-GUIDED ISOLATION STRATEGY

All details related to the development of a Genome-Guided Isolation strategy will be showed ahead.

5.3.1.1 Antibiotics and Metals Resistance Profile

Based on genomic investigation of the genus *Burkholderia* (BKD) and theoretical Metagenomic Analysis (MGA), resistance profiles for antibiotics and metals were elaborated organizing BKD and MGA general characteristics. *In silico* profiles were built and organized in BKD and MGA groups. The antibiotic and metal resistance profiles were created by exploring individual genomic sequences for each putative resistance gene. Genome sequences were downloaded from NCBI database (www.ncbi.nlm.nih.gov).

5.3.1.2 C&N-Sources

Predictions were made to calculate the possibility of selected compounds in being converted to the target compound. Gram-negative microorganisms display at least sets of genes for Citrate Cycle, Glycolysis and Pentose Phosphate pathways (OSHIMA et al., 2004). *In silico* predictions using KEGG-PATHWAY database (www.genome.jp/kegg) turned possible to find the existence of one compound, common in all metabolic pathways cited above (KANEHISA et al., 2014). Interactions to Alpha-D-Glucose-6-Phosphate (KEGG Code: C00668) were related at some stage of these pathways and this compound was adopted for all conversions (**Figure 22**).

Thus, compounds selected were converted to Alpha-D-Glucose-6-Phosphate. This analysis considered species from BKD, MGA and compounds with biologic roles (http://www.genome.jp/kegg-bin/get_htext?br08001.keg). Standard procedures were adopted for a Cutoff length of up to 15 and only known enzymatic processes.

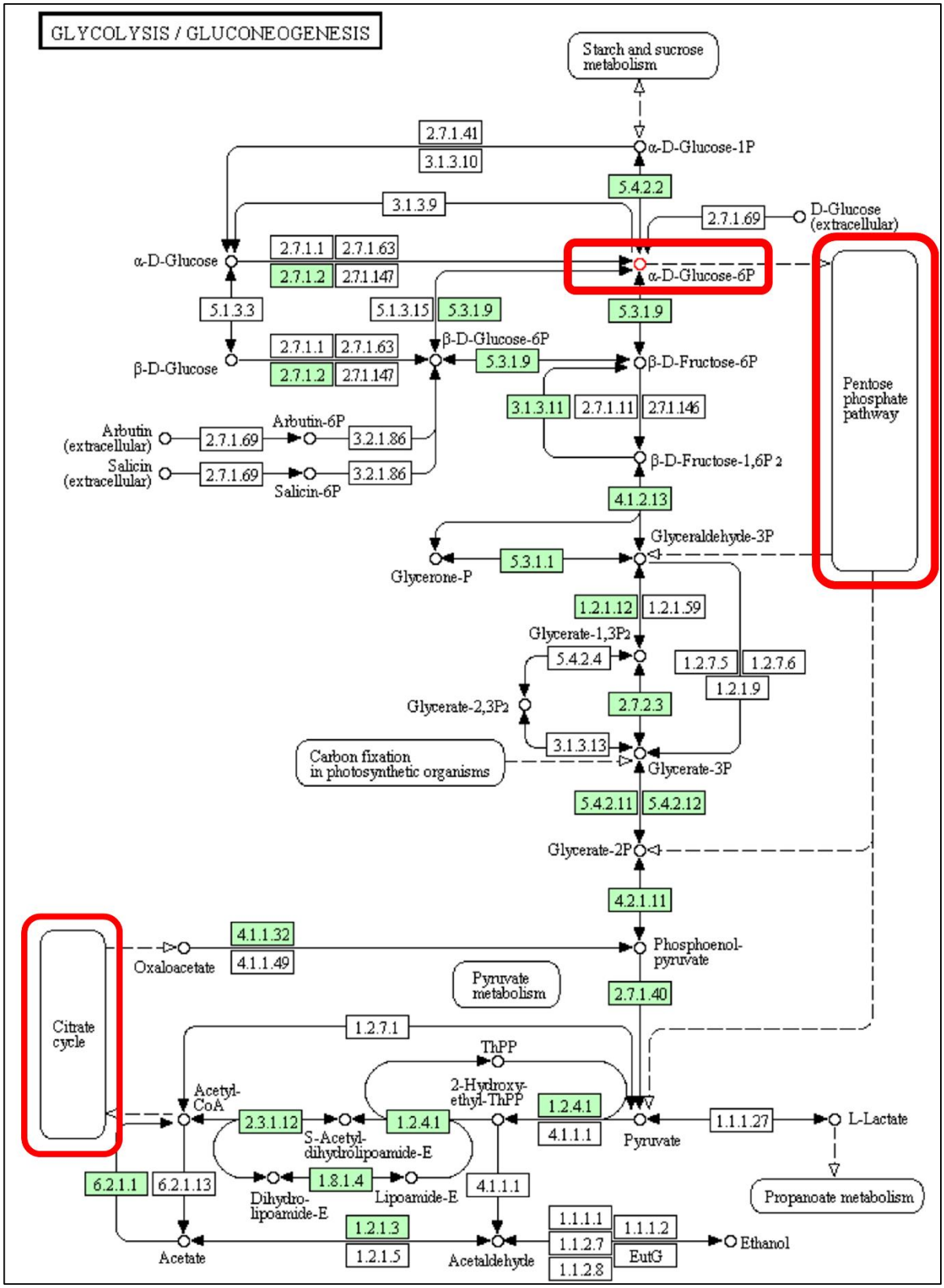

Figure 22- Metabolic pathways crossed for the correlation of Alpha-D-Glucose-6-Phosphate (KEGG Code: C00668) to the biosynthesis of compounds with biological role. Connections to other metabolic pathways and the target compound are highlighted in red.

Source: Adapted from KEGG (www.genome.jp/kegg/).

5.3.1.3 Theoretical Metagenomic Analysis

Genomic informations were investigated using representatives of bacterial communities based on metagenomic studies in which the genus *Burkholderia* was previously found (BELOVA; PANKRATOV; DEDYSH, 2006; WEISSKOPF; HELLER; EBERL, 2011). The species were analyzed for the creation of antibiotics and metals resistances profiles for MGA, as well as, the *in silico* bioconversion compounds profile. One representative species from each genus was chosen whether having both, fully genomic annotation and availability in KEGG-Pathway, for creating MGA profile. The **Table 2** contains all strains of microorganisms adopted in order to build comparative profiles.

Table 2. Selected microorganisms for further analyses of metal and antibiotic resistance and C&N sources.
(continues…conclusion)

Theoretical Metagenomic Analysis Microorganisms (MGA)		***Burkholderia species (BKD)***
Genus	***Representative species***	***Representative Species***
Stenotrophomonas	*Stenotrophomonas maltophilia K279*	*Burkholderia ambifaria **
Enterobacter	*Enterobacter cloacae 13047*	*Burkholderia cepacia GG4*
Bordetella	*Bordetella bonchiseptica 253*	*Burkholderia cenocepacia **
Chitinophaga	*Chitinophaga pinensis*	*Burkholderia lata (B. sp. 383)*
Herbaspirillum	*Herbaspirillum seropedicae smr1*	*Burkholderia multirovans**
Flavobacterium	*Flavobacterium johnsoniae*	*Burkholderia vietnamiensis G4*
Acinetobacter	*Acinetobacter baumannii AB 307-0294*	*Burkholderia thailandensis**
Mesornizobium	*Mesornizobium loti*	*Burkholderia gladioli BSR3*
Pantoea	*Pantoea Ananatis PA13*	*Burkholderia glumae BGR1*
Sphingomonas	*Sphingomonas wittichii RW1*	*Burkholderia mallei **
Rhizobium	*Rhizobium leguminosarum 3841*	*Burkholderia phenoliruptrix BR3459a*
Ralstonia	*Ralstonia solanacearum GMI 1000*	*Burkholderia phymatum STM815*
Microbacterium	*Microbacterium testaceum*	*Burkholderia phytofirmans PsJN*
Rhodopseudomonas	*Rhodopseudomonas palustris BisA53*	*Burkholderia pseudomallei**
Janthinobacterium	*Janthinobacterium sp. Marseille*	*Burkholderia rhizoxinica HKI 454*
Pectobacterium	*Pectobacterium carotovorumn PC1*	*Burkholderia sp.**
Methylobacterium	*Methylobacterium nodulans ORS 2060*	*Burkholderia xenovorans LB400*
Methylovorus	*Methylovorus glucosetrophus SIP3-4*	--
Catenulispora	*Catenulispora acidiphila 44928*	--

Table 2. Selected microorganisms for further analyses of metal and antibiotic resistance and C&N sources.
(continues…conclusion)

Theoretical Metagenomic Analysis Microorganisms (MGA)		***Burkholderia species (BKD)***
Genus	***Representative species***	***Representative Species***
Asticcacaulis	*Asticcacaulis excentricus*	--
Agrobacterium	*Agrobacterium radiobacter K84*	--
Variovorax	*Variovorax paradoxus B4*	--
Collimonas	*Collimonas fungivorans Ter331*	--
Azospirillum	*Azospirillum lipoferum 4B*	--
Acidobacterium	*Acidobacterium capsulatum*	--
Phelylobacterium	*Phenylobacterium zucineum HLK1*	--
Rhodanobacter	*Rhodanobacter denitrificans*	--
Rhizobium	*Rhizobium leguminosarum bv. viciae 3841*	--
Bradyrhizobium	*Bradyrhizobium japonicum USDA 6*	--

* Microorganisms presenting more than one representative were grouped and treated as one, due to the similarity in resistance and metabolic profiles.

5.3.1.4 Standard microorganism adopted from in silico analysis

Compounds resulting from genome-guided analysis and metabolism, BKD and MGA species were combined for the elaboration of a correlation graph. This strategy enabled choosing visually the standard microorganism to be used as a control for experimental procedures. To that, correlations were linked by using software Gephi® 0.8.2 beta.

5.3.1.5 Experimental analysis of in silico results

Previous samples obtained in for this work purpose, containing no BKD species, i.e. untargeted microorganisms (UNT), were enriched in liquid media and used as negative control. Compounds from genome-guided results were selected according to their greatest specificity to the genus *Burkholderia*. Then, BKD and UNT were submitted to cell growth in standard liquid media containing compounds from genome-guided strategy. Optical densities (OD) were investigated in time courses and the absorbance were obtained from PerkinElmer Envision Multilabel Plate Reader at 600 nm. In all experiments, T_0 was considered blank and the real

cell growth was measured by the difference of T_1 least T_0. Results were considered specific to BKD when $OD_{BKD} >> OD_{UNT}$ according to the equation 1, below:

$$\left(\boldsymbol{Abs_{BKD_{T1}} - Abs_{BKD_{T0}}}\right) - \left(\boldsymbol{Abs_{UNT_{T1}} - Abs_{UNT_{T0}}}\right) = \Delta \boldsymbol{Abs_{BKDvsUNT}}$$

Soil samples used as negative control were obtained in previous experiments, by using conditions already published on the literature, (KAWANISHI et al., 2011; VANLAERE et al., 2005; VERMIS et al., 2003; ZAID; BONASERA; BEER, 2012). These samples enclosed species as *Pseudomonas*, *Acrhomobacter*, *Ochrobactrum*, and *Bacillus*, confirmed by 16s rRNA gene amplification and identification and species with non-similar morphology to the genus *Burkholderia* when compared to *B. thailandensis* colonies morphology at the same conditions. In addition, using these culture media it was possible to isolated fungus species that were treated as contamination. Procedures for samples cell enrichment: 250 µL of soil supernatant from each sample were transferred to a new sterile Falcon tube containing 10 mL of sterile NB (pH 7) and incubated for 24 h. Negative control was taken by mixing all the overnighted cell culture of each sample onto a new sterile tube and incubating for 24 h. After incubation period, six aliquots of each were kept in cryogenic tubes containing 1 mL of Glycerol/Water (1:1 v/v) sterile solution and 1 mL of the cell culture maintained at -70 ^{o}C. The enriched cell solution was used as negative control for all experimental sets.

5.3.1.6 Samples Collecting

Different types of soil samples in which the genus *Burkholderia* is related and/or not were collected and analyzed (BALANDREAU et al., 2001; BELOVA; PANKRATOV; DEDYSH, 2006; BERGMARK et al., 2012; BRAGINA et al., 2013; ESTRADA-DE LOS SANTOS; BUSTILLOS-CRISTALES; CABALLERO-MELLADO, 2001; JACOBS et al., 2008; KOST et al., 2013; LEPLEUX et al., 2012; MARAVIĆ et al., 2012; PAUNGFOO-LONHIENNE et al., 2014). Collections field were Ocean Sediment (Panama); Bulk Soil (Santa Cruz; Randomly); Soil (5 cm from roots; soil surface; Santa Cruz); Rhizosphere (5 to 10 cm underneath the surface, strictly close to the roots; UCSC Farm; Santa Cruz). All the samples were kept in Falcon tubes containing standard PBS (pH 7.4) and frozen.

5.3.1.7 Cell Growth Standard Methodologies

Samples enrichment: 250 µL of soil supernatant were collected from environmental sources and added to a new assay tube containing the best basal condition, kept at 250 rpm, 25-27 ºC up until desirable OD reached.

For 96 well plates tests: 10 µL of each cell culture were transferred to each well containing 200 µL of the individual compounds, basal conditions or specific combinations. The plates were read using PerkinElmer Envision Multilabel Plate Reader at 600 nm, using Wallac Envision Manager software. Each sample was prepared in duplicates and the OD read in triplicates for each well.

All liquid media, solid media and microbiological glassware were autoclaved before using and/or maintained under UV light for 20 min.

5.3.1.8 Highly Selective Culture Media Elaboration

Procedures related to the test of all compounds and optimization steps will be discussed ahead.

a) Basal Conditions

Culture media with no antimicrobials, carbohydrates or amino acids from genome-guided strategy were called basal conditions. The ingredients used to make this step are listed as follows: a) Metals from genome-guided results as sulfate salts: $NiSO_4$; $CuSO_4$; b) Buffers: KH_2PO_4; K_2HPO_4; $NH_4H_2PO_4$; c) Nutrients: Yeast Extract; Peptone; Beef Extract; Nutrient Broth. All of them were grouped in order to provide as much unique combinations as possible, mandatorily containing $NiSO_4$ and $CuSO_4$ in optimum concentration. The combinations were tested to investigate cell growth of BKD against UNT.

b) Specific Conditions

Specific conditions were those in which antimicrobials, metals, carbohydrates and amino acids from genome-guided analysis were combined for the isolation of a great number of BKD. For that, a Python algorithm was created in order to correlate all possible combinations to avoid untargeted microorganisms. Combinations were named as *Burkholderia* Selection Media from BSA to BSJ, generally abbreviated as BSX conditions in this work. The strategy of this analysis is shown in **Figure 23**.

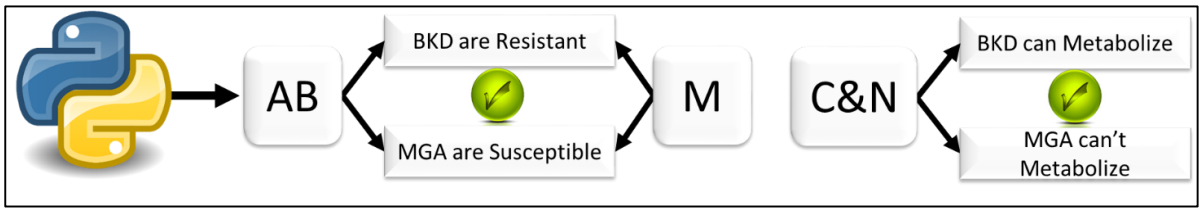

Figure 23- Python logics to create the algorithm in order to avoid MGA species.
Source: From the author.

c) pH Control

A range of pH variation (using HCl and NaOH as pH modifiers) was created for B4_9 conditions. Ingredients were mixed at the correct concentrations and pH was corrected using an electronic pHmeter. 200mL of all stock solutions were kept and poured in 96-well plates (250 uL per well) for further analysis. 10 uL of enriched cell culture containing BKD or UNT were poured into each well plate according to the pH Platemap. Timecourses analysis were proceeded in order to evaluate pH changes and cell growth of BKD and UNT.

5.3.1.9 Standard Conditions for Isolation of BKD

First, Basal Conditions with optimum pH were prepared and 10 mL transferred to each assay tube. Second, overnighted samples with considerable OD were streaked onto BSX conditions. Plates were analyzed daily for examination of any colony formation. After that, isolated colonies were streaked on new Basal Conditions plates plus Cycloheximide in order to analyze their purity. Morphological analyzes were made using *B. thailandensis* in the same BSX conditions as standard.

5.3.1.10 DNA Isolation, 16s rRNA Gene Amplification, Sequencing and Identification

DNA extraction from overnighted cell cultures was proceed following Wizard® Genomic DNA Purification Kit Protocol from Promega. After extraction, DNA was purified and submitted to PCR reactions using general primers 8F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′) with Platinum PCR High Fidelity Supermix (final volume of 50 µL). PCR was proceeded using Eppendorf Mastercycler Personal Thermocycler. Specifications: Denaturation: 95 ºC for 5 min; 35 cycles of denaturation at 95 ºC for 1 min; Annealing at 50 ºC for 1 min; Extension at 72 ºC for 10 min. Afterwards, agarose gel electrophoresis was proceeded (1% agarose gel in 1x TAE buffer) for confirmation. All PCR products were purified using QIAQuick PCR Purification Kit (Qiagen) and analyzed by Sequetech Corporation with the same primers cited above. Nucleotide sequences were treated with FinchTV 1.4.0 (Geospiza); assembled using PRABI-Doua (Pôle Rhône-Alpes de Bioinformatique Site Doua – CAP3 Sequence Assembly Program - http://doua.prabi.fr/software/cap3); Assembled sequences were submitted to Blastn (Database: Others (nr etc); Optimize for: Highly similar sequences (megablast); Algorithm parameters: Standard. The results presenting the lowest E-value and higher identity (higher max score, query cover, and ident) were used for the identification of all strains.

5.3.1.11 NPs production overview for Burkholderia species

44 genomes related to *Burkholderia* species were analyzed according to their biosynthetic capability in order to confirm previous results. This experiment was proceeded in order to understand how biosynthetic pathways are distributed within *Burkholderia* species compared to all BPB species and which kind of compounds they can mostly express from genome-mining investigation. As NRPS, PKS, homoserine lactones and phosphonate compounds are constantly related to important biological activities, *Burkholderia* species were examinated for the identification of their features. Sequences were downloaded from NCBI, their BGCs were identified and simply distributed onto a graph giving a general overview of classes.

5.3.2 RESULTS AND DISCUSSION

Results and discussion related to the creation of a specific genome-guided isolation strategy are presented ahead.

5.3.2.1 Antibiotics and Metals Resistance Profile (BKD vs MGA)

Only Representative Genome Sequences from NCBI Genome Database were used for building antibiotics and metals resistance profile. In microbial communities, species naturally contain genes associated to resistance to one or more antibiotics, i.e. to the antibiotic being produced and/or external stresses caused by another species, or even multifunctional gene systems encoding proteins known as efflux pumps (ALLEN et al., 2010). This investigation considered the existence of genes encoding resistance to specific antimicrobial compounds according to gene similarities and gene function in each genome sequence inspected from BKD and MGA and the results are shown in **Figure 24**:

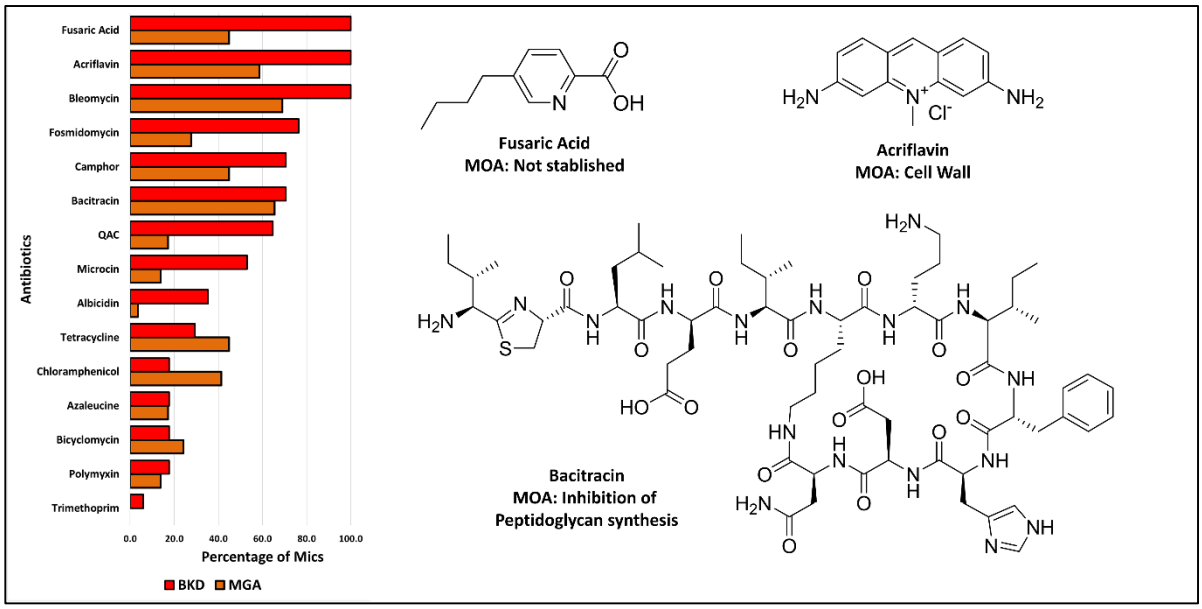

Figure 24- Genome-Guided Results of Antibiotic Resistance Profile. Results in percentage of species BKD vs MGA containing the same compound related to putative resistant genes from genomic annotations. Antibiotics structures are related to the best candidates according to their percentage, price and diversity of BKD species presenting the respective resistance genes.
Source: from the author.

Getting through the same examination, focusing on metals, it was possible to obtain the metal resistance profile. According to the genome annotations, and the normal occurrence of

metals in the environment, bacteria present genes encoding protection mechanisms against these elements, but in higher concentrations they show toxic effects on bacterial cells (HASSEN et al., 1998). Results showed that genes encoding proteins against metals toxicity in the environment are differently distributed among species and could be used with the purpose of building the highly specific isolation method (**Figure 25**):

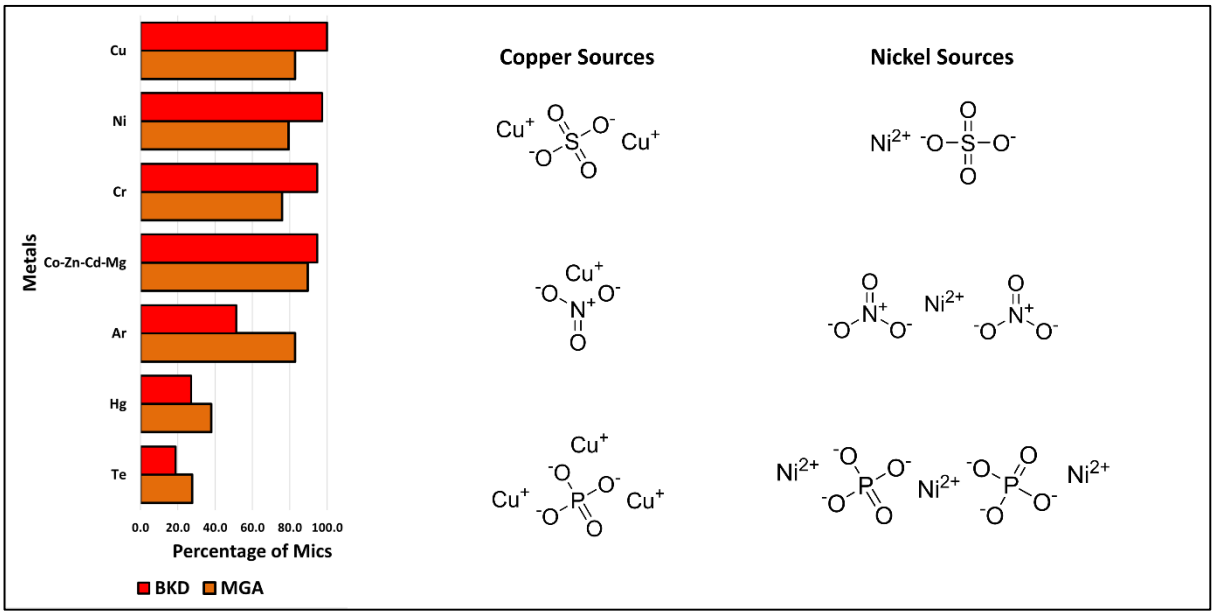

Figure 25- Genome-Guided Results of Metal Resistance Profile. Results in percentage of species BKD vs MGA containing the same compound related to putative resistant genes from genomic annotations.
Source: From the author.

These preliminary analyses showed that Bleomycin, Fosmidomycin, Fusaric Acid and Acriflavin are the best antibiotic for this method. Although, Bleomycin and Fosmidomycin are far expensive and Bacitracin, even presenting similar percentage between BKD and UNT, presents different BKD strains than other antibiotics. For metals, the number of resistance genes for BKD and MGA are quite similar. Therefore, an algorithm was created to obtain useful information.

5.3.2.2 In silico metabolic profile (BKD vs MGA)

For increasing the selectivity of this method, C&N-Sources sources were predicted as metabolizable or not in order to compare BKD and MGA species. According to *in silico*

conversions of compounds with biologic role, candidates were selected to build a theoretical profile displaying differences between their metabolic capabilities. The compounds analyzed were: a) Carbohydrates: Monosaccharides, Aldoses and Ketoses: b) Amino Acids: Other Amino Acids.

For MGA *in silico* metabolic investigation, representative strains were submitted to the same process using compounds showing values higher than zero for BKD. The **Figure 26** displays results for the metabolic profile between BKD and MGA representatives from *in silico* predictions:

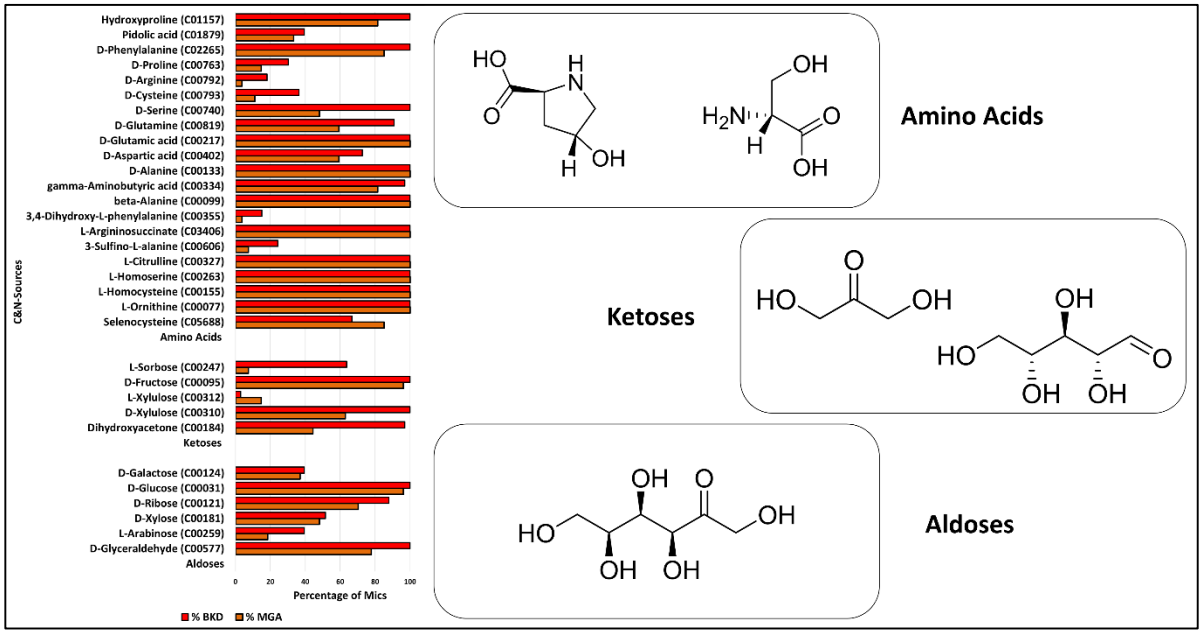

Figure 26- Comparison between percentages of BKD vs MGA able to metabolize specific compounds from in silico analysis using compounds with biological role. Codes are according to KEGG database.
Source: From the author.

5.3.2.3 Standard microorganism from in silico analysis

After elaboration of resistance and metabolic profiles, direct correlations (Source: microorganism – Target: compound) were established in a Nodes and Edges table and edited in Gephi 0.8.2®. *B. thailandensis*, elected as standard, have higher correlation to compounds common to BKD species (**Figure 27**). This strain, purchased from ATCC (ATCC_700388), was applied in all experimental procedures.

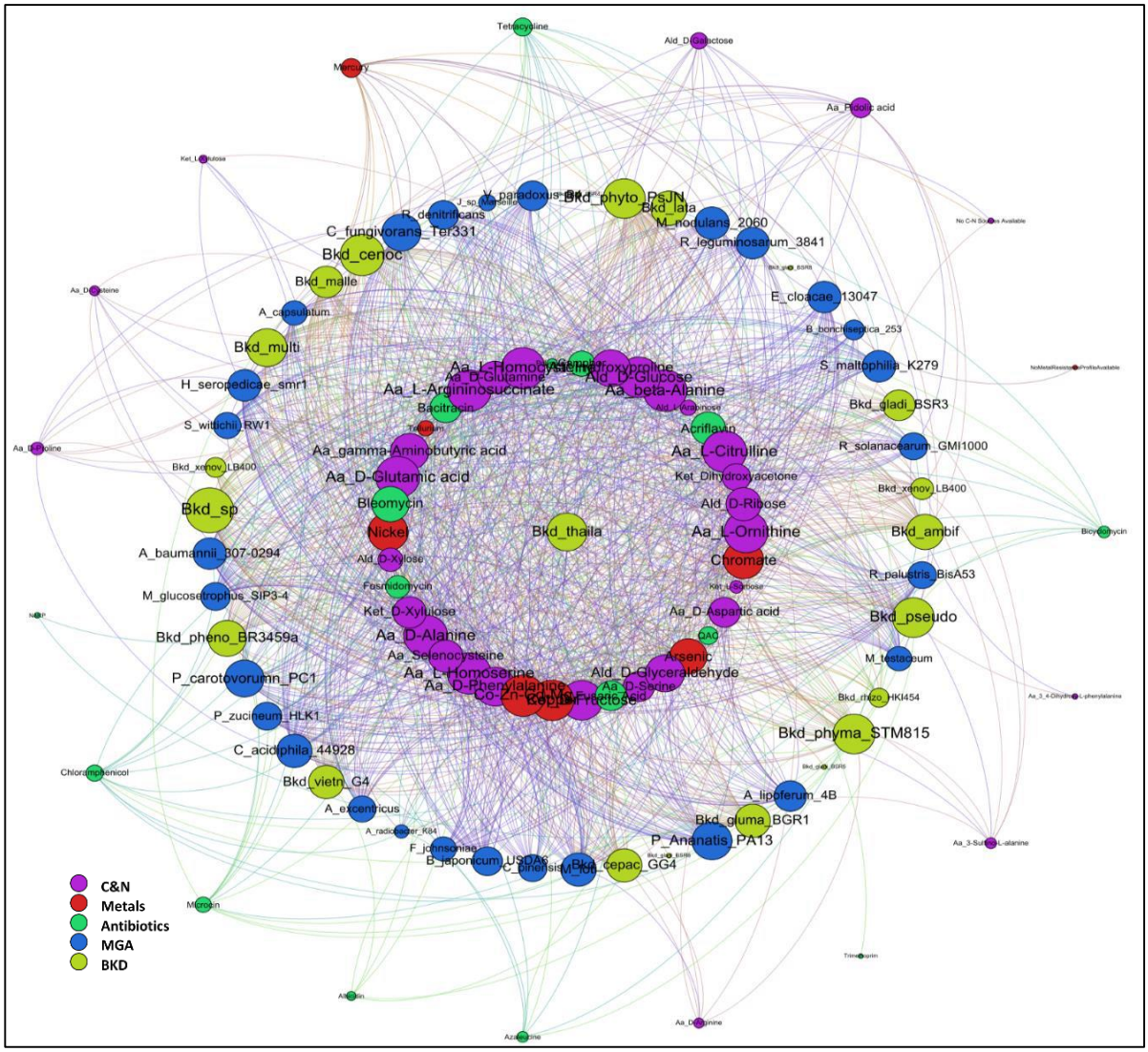

Figure 27- Visual profile from Gephi analysis showing (from the smaller to the bigger circle): a) *Burkholderia thailandensis* (Bkd_thaila) and its direct correlation with the best matches from genome-guided investigation. b) Compounds with direct correlation to Bkd_thaila. c) Microorganisms from BKD and MGA species, also correlated to the first compound circle; d) Compounds with no direct correlation to Bkd_thaila. The bigger the circle of each representative, the bigger the correlation between them. (Details on Gephy software: Layout: "Concentric Layout"; Distance: "750,00"; Node: "Bkd_thaila"; Speed: "7.0"; Coverage "2.5").

Source: From the author.

5.3.2.4 Highly Selective Culture Media Elaboration

Optimization steps are discussed ahead in order to explain how the culture media was developed.

a) Individual Compound Analysis

Selected compounds from genome-guided analysis were tested in dilution series from 5 to 0.04 g.L^{-1}, using 96-well plates in duplicates for *B. thailandensis* (BKD standard) and UNT. Results were acquired by comparing OD variation (BKD vs UNT) from 0 h to 96 h. In this case, before elaboration of Basal Conditions, the Liquid Media reported for growing of *B. thailandensis* from ATCC (Medium 3: Nutrient Broth 8 g.L^{-1}) was used as standard. As expected results, the higher the difference between ODs of BKD vs UNT, the higher the specificity of one given compound for the highly specific isolation media. Results show that compounds from genomic annotation analysis are strongly specific to *Burkholderia* to the level of *in vitro* experiments.

Figure 28 displays the behavior of BKD and UNT when treated individually with compounds from genomic-guided analysis:

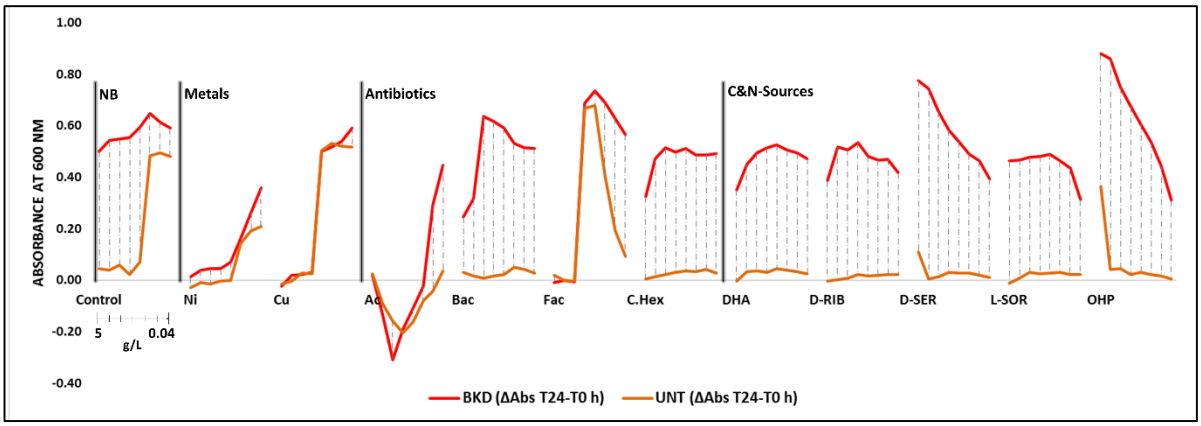

Figure 28- OD variations experiments for specific compounds from genomic-guided analysis. These results display the Control absorbance using only NB (Nutrient Broth); Metals as Sulfate Salts; Antibiotics: Ac (Acriflavin), Bac (Bacitracin), Fac (Fusaric Acid); C&N-Sources: DHA (Dihydroxiacetone); D-RIB (D-Ribose); D-SER (D-Serine); L-SOR (L-Sorbose); OHP (4-Hidroxy-L-Proline). All results are displayed as the difference between t24 (h) – t0 (h) for each point. Concentrations are showed in serial dilutions from 5.0 to 0.04 g/L from the left to the right for each compound. Values were compared in order to show the best concentration to use on the specific combinations.
Source: From the author.

Metals can be noxious in higher levels to some microorganisms, and in this case higher concentrations of Ni^{+2} and Cu^{+2} appeared to be harmful for both, BKD and UNT. However, in lower concentrations they were slightly specific to BKD, displaying higher OD and following results from *in silico* analysis. In accordance to the Agency for Toxic Substances and Disease Registry (ASTDR - http://www.atsdr.cdc.gov), Nickel and Copper can occur from 1 to 5000 and 1 to 70 mg/Kg of soil, respectively, and in high concentrations, they seemed to be cytotoxic, turning difficult cell growth. According to the antibiotics used Ac (MOA: cell wall), Bac

(MOA: synthesis of peptidoglycan inhibition) and Fac (MOA: not well understood) the best concentrations were chosen, and the results were similar to those found on *in silico* experiments. BKD presented higher OD than UNT when treated with Cycloheximide, used in order to avoid fungus contamination, which is, in a good perspective, helpful for increasing specificity to this method. Results showed that *B. thailandensis* happen to be resistant to its action. All the C&N-Sources tested were highly selective to BKD in our experiments. Individual analysis show that UNT did not grow well in the same condition as BKD.

b) Basal Conditions Analysis

Basal conditions were elaborated using antimicrobial-like salts $NiSO_4$ and $CuSO_4$, obtained from genome-guided analysis of metals resistance genes annotation in optimized concentrations according to the Individual Compounds Analyses, combined with Nutrients and Buffers Salts. Combinations containing Yeast Extract presented higher OD to BKD than to UNT. These conditions were analyzed in order to discover which ones had the best relationship OD vs Time, i.e. from time zero to the last hour of analysis. **Figure 29** shows results in duplicates for OD in all wells 4_9 (which presented higher OD) for all rows (from A to H using 96-well plates). In this experiment, higher values in the difference BKD-UNT were selected and the highest OD in the last hour was observed in order to elect the best Basal Condition. Combinations from duplicate wells B_04 and B_09, called B4_9 conditions, were adopted as common Basal Media.

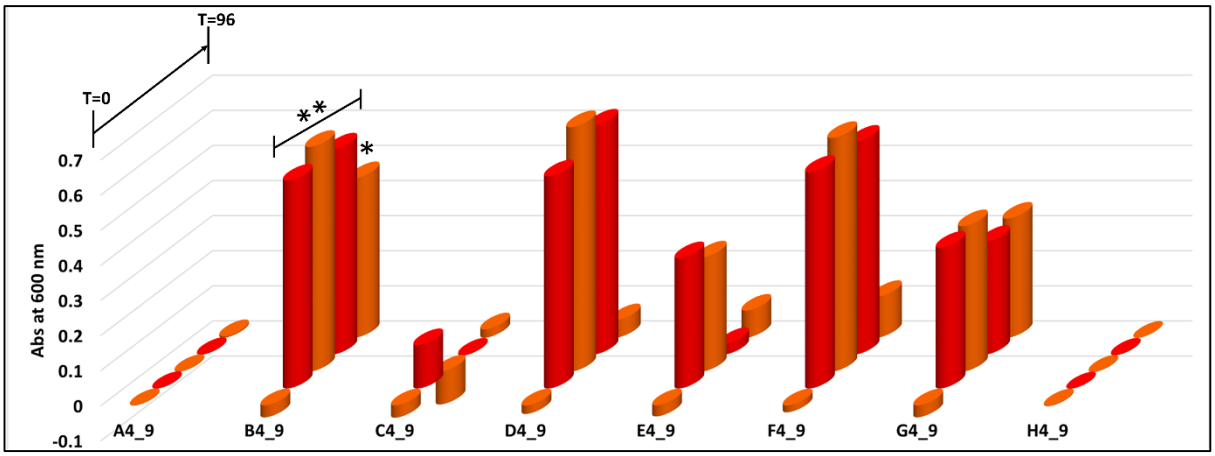

Figure 29- Time courses obtained from Basal Conditions Analysis of columns 4_9, using Yeast Extract as specific nutrient. These duplicates revealed highest OD values than all combinations present in the Table 1. *Higher OD for BKD-UNT in the last hour. ** Relative OD for the Basal Composition for the cells B4_9. The highest absorbance was taken in order to elect the common composition for all Specific Conditions.

Source: from the author.

c) pH Control

Considering pH variations, is possible to avoid unintended microorganisms to take advantage of the best specific combinations. Species from the genus *Burkholderia* are known to be more tolerant to low pH than non-*Burkholderia* species when existing in the same ecosystem (WEISSKOPF; HELLER; EBERL, 2011). *B. thailandensis* and UNT (*Pseudomonas* sp., *Chryseobacterium* sp., *Klebsiela* sp., *Cupriavidus* sp.) were submitted to several pH conditions (from 3 to 9 using HCl and NaOH as acidity modifiers) in time courses analysis, following some restrictions according to (ROBERTSON et al., 2010).

Bacterial densities can harbor up to 10^{10} bacterial cells and diversity can reach from four to fifty thousand species (RAYNAUD; NUNAN, 2014). Moreover, the major phylotype diversity in bacterial communities is given in pH between 6 and 8 (FIERER; JACKSON, 2006). Evidences suggests that *Burkholderia* sp. are often classified as plant-associated nitrogen fixers microorganisms and adaptive to dramatic changes in pH, being dominant at some acidic root systems (rhizosphere), and as one essential constraint a wide range of pH was tested to increase specificity to BKD species in this method (ROBERTSON et al., 2010; WEISSKOPF; HELLER; EBERL, 2011). In B4_9 conditions, was observed that lower pH (3-4) seems to interfere in cell growth of both *B. thailandensis* and UNT. When in pH 5-6 (highlighted as * on the graph), BKD exhibited higher OD than UNT species, in agreement to the literature wherein BKD species can survive even in severe environment conditions (BELOVA; PANKRATOV; DEDYSH, 2006; LEPLEUX et al., 2012) as can be seeing on **Figure 30**:

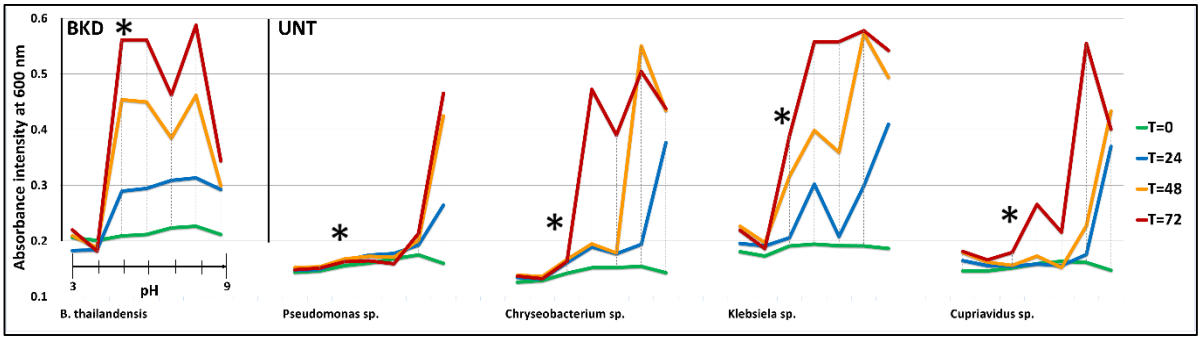

Figure 30- Cell growth profile showing absorbance for pH versus time. * Shows the best pH in which *B. thailandensis* displays higher absorbance than all of UNT microorganisms used as negative control. All experiments were made using B4_9 as basal conditions. Results were obtained from PerkinElmer Envision Multilabel Plate Reader at 600 nm.
Source: From the author.

Specific conditions were prepared using pH around 5.5 according to *B. thailandensis* results and used to reach higher specificity to BKD species.

d) Analysis of Specific Conditions

Constraints adopted in computational analysis allowed the maximum of 5 MGA as a limit for choosing combinations. In total, 60 specific culture media, comprising genome-guided Antibiotics, Metals and C&N-Sources were used in this step. Combinations allowed, theoretically, the isolation of up to 15 different species of the genus *Burkholderia* in diverse arrangements according to Table 3. Predicted results were tested using B4_9 and the optimum pH, and the OD of all specific culture media was compared between BKD and UNT.

Table 3. compounds combinations showing number of bkd and unt species. codes are according to computational codes.

(continues…conclusion)

Recipe code	***Combination***	***Bkd***	***BKD_roll***	***MGA_roll***
BSA	('Acriflavin', 'Bacitracin')-('Nickel', 'Copper')-('Aa_D-Serine',)	10	*['Bkd_cenoc', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_multi', 'Bkd_phyto_PsJN', 'Bkd_gladi_BSR3', 'Bkd_thaila', 'Bkd_cepac_GG4', 'Bkd_pseudo']*	*[]*
BSB	('Acriflavin', 'Bacitracin')-('Nickel', 'Copper')-('Ket_L-Sorbose', 'Aa_D-Serine')	10	*['Bkd_cenoc', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_multi', 'Bkd_phyto_PsJN', 'Bkd_gladi_BSR3', 'Bkd_thaila', 'Bkd_cepac_GG4', 'Bkd_pseudo']*	*[]*
BSC	('Acriflavin', 'Bacitracin')-('Nickel', 'Copper')-('Ket_L-Sorbose',)	6	*['Bkd_thaila', 'Bkd_sp', 'Bkd_ambif', 'Bkd_pseudo', 'Bkd_cenoc', 'Bkd_phyto_PsJN']*	*[]*
BSD	('Acriflavin', 'Fusaric Acid')-('Nickel', 'Copper')-('Aa_D-Serine',)	15	*['Bkd_xenov_LB400', 'Bkd_thaila', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_multi', 'Bkd_pseudo', 'Bkd_malle', 'Bkd_gladi_BSR3', 'Bkd_phyto_PsJN', 'Bkd_lata', 'Bkd_cenoc', 'Bkd_cepac_GG4', 'Bkd_pheno_BR3459a', 'Bkd_gluma_BGR1']*	*['P_Ananatis_PA13', 'C_fungivorans_Ter331']*

Table 3. compounds combinations showing number of bkd and unt species. codes are according to computational codes.

(continues…conclusion)

Recipe code	*Combination*	*Bkd*	*BKD_roll*	*MGA_roll*
BSE	('Acriflavin', 'Fusaric Acid')-('Nickel', 'Copper')-('Ket_L-Sorbose', 'Aa_D-Serine')	15	*['Bkd_xenov_LB400', 'Bkd_thaila', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_multi', 'Bkd_pseudo', 'Bkd_malle', 'Bkd_gladi_BSR3', 'Bkd_phyto_PsJN', 'Bkd_lata', 'Bkd_cenoc', 'Bkd_cepac_GG4', 'Bkd_pheno_BR3459a', 'Bkd_gluma_BGR1']*	*['P_Ananatis_PA13', 'C_fungivorans_Ter331']*
BSF	('Acriflavin', 'Bacitracin')-('Nickel', 'Copper')-('Ket_Dihydroxyacetone', 'Ket_L-Sorbose')	8	*['Bkd_cenoc', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_phyto_PsJN', 'Bkd_gladi_BSR3', 'Bkd_thaila', 'Bkd_pseudo']*	*['M_nodulans_2060', 'P_carotovorumn_PC1', 'M_glucosetrophus_SIP3-4']*
BSG	('Acriflavin', 'Bacitracin', 'Fusaric Acid')-('Nickel', 'Copper')-('Aa_Hydroxyproline', 'Aa_D-Serine')	10	*['Bkd_cenoc', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_multi', 'Bkd_phyto_PsJN', 'Bkd_gladi_BSR3', 'Bkd_thaila', 'Bkd_cepac_GG4', 'Bkd_pseudo']*	*['M_nodulans_2060', 'P_carotovorumn_PC1']*
BSH	('Bacitracin', 'Fusaric Acid')-('Nickel', 'Copper')-('Ket_L-Sorbose', 'Aa_Hydroxyproline')	10	*['Bkd_cenoc', 'Bkd_phyma_STM815', 'Bkd_sp', 'Bkd_ambif', 'Bkd_multi', 'Bkd_phyto_PsJN', 'Bkd_gladi_BSR3', 'Bkd_thaila', 'Bkd_cepac_GG4', 'Bkd_pseudo']*	*['M_nodulans_2060', 'P_carotovorumn_PC1', 'H_seropedicae_smr1']*
BSI	('Acriflavin', 'Fusaric Acid')-('Nickel', 'Copper')-('Ket_L-Sorbose',)	8	*['Bkd_thaila', 'Bkd_sp', 'Bkd_ambif', 'Bkd_pseudo', 'Bkd_malle', 'Bkd_lata', 'Bkd_cenoc', 'Bkd_phyto_PsJN']*	*['C_fungivorans_Ter331']*
BSJ	('Fusaric Acid',)-('Nickel', 'Copper')-('Ket_L-Sorbose',)	8	*['Bkd_thaila', 'Bkd_sp', 'Bkd_ambif', 'Bkd_pseudo', 'Bkd_malle', 'Bkd_lata', 'Bkd_cenoc', 'Bkd_phyto_PsJN']*	*['C_fungivorans_Ter331']*

These results provided the best composition for isolating BKD species from environment. **Figure 31** shows absorbance of each combination in relation to the difference of T=72 h – T=0 h to BKD vs UNT and the combinations elected in order to isolate BKD species from environment samples.

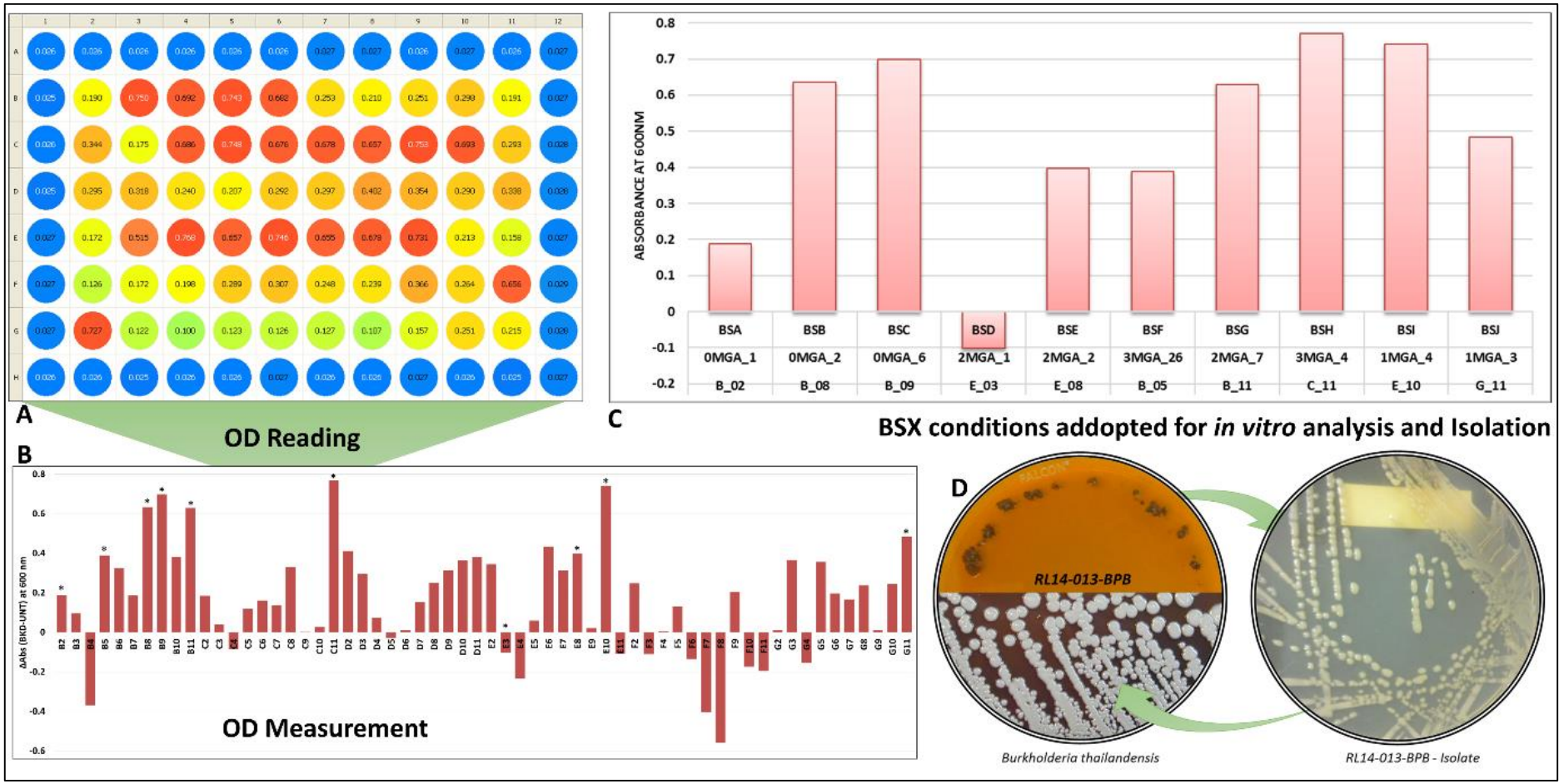

Figure 31- Relative absorbance of cell growth for each combination of Specific Culture Media comparing the difference of BKD-UNT (T72 least T0; according to equation 1). Results are organized according to the 96-well Plate-Map showing wells used. External wells were left empty. *Combinations elected for experimental analysis.

Source: From the author.

From these results, 5 combinations showing higher ΔAbs were chosen. In addition, 4 median ΔAbs and 1 negative ΔAbs were randomly selected in order to isolate BKD species from environmental samples, named as BSX (X = from A to J).

5.3.2.5 Strains Isolated

A two-step approach was taken for the isolation of *Burkholderia* species. First: Collected samples were enriched by using liquid media B4_9 in test tubes and left overnight up until displaying considerable OD; Second: BSX conditions, were made in Petri dishes with agar 2% and the overnighted cell culture were streaked by adding around 100 µL onto the plates. After this approach, all colony formation with similar morphology to *B. thailandensis* at BSX conditions were isolated, DNA extracted, 16s rRNA gene amplified and sent to Sequetech. Sequencing results were assembled, and similarity search was found using Blastn adopting the lower E-value for each blasted sequence according to **Table 4**.

Table 4. Results of all isolation experiments, culture media used and species identified using Blastn similarity search.

(continues…continuation…conclusion)

Experiments	***Samples Origin***	***Code****	***Species Name (Blastn)***
(1) Previous Culture Media From Literature	Bulk Soil.	RL14-002-BMS-A/IS	*Bacillus anthracis strain 2-Sj-2-2-26-M*
		RL13-114-BMF-B/IS	*Pseudomonas rhodesiae strain R2SsM3P1C14*
		RL13-115-BMF-A/RS	*Bacillus thuringiensis strain Lr7/2*
		RL14-003-BMS-B/RS	*Pseudomonas putida CFBP 4629*
		RL13-115-BMS-A/RS	*Pseudomonas brassicacearum NFM421*
		RL13-114-BMS-A/RS	*Pseudomonas brassicacearum strain Zy-2-1 16S*
		RL13-114-BMS-A/RS	*Pseudomonas brassicacearum NFM421*
		RL14-003-BMF-B/RS	*Achromobacter spanius strain MT3*
		RL13-115-BMS-B/RS	*Pseudomonas brassicacearum NFM421*
		RL14-003-BMS-B/RS	*Pseudomonas sp. QH11*
		RL13-115-BMM-A/RS	*Ochrobactrum intermedium strain 182I*
		RL14-003-BMM-A/IS	*Ochrobactrum sp. K3*
		RL13-114-BMB-D/IS	*Bacillus anthracis strain 2-Sj-2-2-26-M*
		RL14-003-BMV-A/IS	*Paenibacillus sp. Sd-7*
		RL13-111-BMF-A	*Bacillus sp. PG-5-9*
		RL13-111-BMM-A	*Pseudomonas stutzeri strain 28a42*
		RL13-111-BMS	*Bacillus anthracis strain 2-Sj-2-2-26-M*
(2) Samples	Ocean Sediment; Panama.	RL14-010-BPA-A	*Vibrio sp. Vibrio harveyi*
		RL14-011-BPA-A	*Vibrio sp. Vibrio harveyi*
		RL14-017-BPA-A	*Vibrio rotiferianus Vibrio harveyi*

Table 4. Results of all isolation experiments, culture media used and species identified using Blastn similarity search.

(continues…continuation…conclusion)

Experiments	*Samples Origin*	*Code**	*Species Name (Blastn)*
from Ocean, Panama		RL14-018-BPA-A	*Vibrio harveyi Vibrio sp.*
		RL14-019-BPA-A	*Vibrio sp. Vibrio harveyi*
		RL14-025-BPB-A	*Vibrio harveyi Vibrio sp.*
		RL14-026-BPB-A	*Vibrio sp. Vibrio harveyi*
		RL14-027-BPB-A	*Vibrio harveyi Vibrio sp.*
		RL14-028-BPB-A	*Vibrio sp. Vibrio harveyi*
		RL14-006-BPC-A	*Photobacterium sp.*
		RL14-009-BPD-A	*Vibrio sp. Vibrio harveyi*
		RL14-011-BPD-A	*Vibrio harveyi Vibrio sp.*
		RL14-017-BPE-A	*Vibrio harveyi Vibrio sp.*
		RL14-020-BPF-A	*Pseudomonas sp. /putida*
(3) Soil Samples.	5 cm from roots; soil surface.	RL14-048-BSA-A	*Pseudomonas sp. JC5*
		RL14-048-BSB-A	*Pseudomonas sp. R1SpM3P2C2*
		RL14-048-BSC-A	*Pseudomonas sp. R1SpM3P2C2*
		RL14-048-BSE-A	*Pseudomonas sp. R5SpM3P1C1*
		RL14-049-BSB-A	*Chryseobacterium jejuense strain JDG189*
		RL14-051-BSD-A	*Pseudomonas sp. UW4*
		RL14-052-BSD-A	*Pseudomonas sp. m1*
		RL14-052-BSE-A	*Klebsiella sp. SR55*
		RL14-052-BSF-A	*Klebsiella sp. SR55*
		RL14-052-BSF-B	*Pseudomonas moorei strain OR108*
		RL14-052-BSF-C	*Klebsiella sp. SR55*
		RL14-052-BSF-D	*Klebsiella sp. SR55*
		RL14-053-BSD-A	*Pseudomonas fluorescens strain S2*
		RL14-055-BSE-A	*Pseudomonas putida LMG 1246*
		RL14-057-BSB-A	*Pseudomonas sp. PALXIL01*
		RL14-059-BSA-A	*Cupriavidus sp. KSL5401-228*
		RL14-059-BSB-A	*Pseudomonas sp. 4097*
		RL14-059-BSD-A	*Cupriavidus sp. KSL5401-228*
		RL14-059-BSF-A	*Klebsiella oxytoca KONIH1*
(4) Rhizosphere One-Step B4_9, only.	Rhizosphere.	RL14-064-B49-A	*Pseudomonas putida strain*
		RL14-066-B49-A	*Pseudomonas salomonii*
		RL14-066-B49-B	*Serratia marcescens strain N-2*
		RL14-066-B49-C	*Serratia marcescens strain TC-1*
		RL14-066-B49-D	*Serratia marcescens WW4*
		RL14-066-B49-E	*Serratia marcescens WW4*
		RL14-066-B49-F	*Serratia marcescens strain TC-1*
		RL14-067-B49-A	*Pseudomonas oryzihabitans strain M-B1A*
		RL14-067-B49-B	*Pseudomonas oryzihabitans strain M-B1A*
		RL14-067-B49-C	*Serratia sp. R26(2012)*
		RL14-067-B49-E	*Serratia marcescens strain T18*
		RL14-068-B49-A	*Pseudomonas sp. JC11*
		RL14-068-B49-B	*Pseudomonas sp. C19*

Table 4. Results of all isolation experiments, culture media used and species identified using Blastn similarity search.

(continues…continuation…conclusion)

Experiments	*Samples Origin*	*Code**	*Species Name (Blastn)*
		RL14-068-B49-C	*Pseudomonas plecoglossicida strain NBFPALD_RAS144*
		RL14-068-B49-D	*Burkholderia gladioli strain CACua-73*
		RL14-068-B49-E	*Pseudomonas putida strain ATCC 17494*
		RL14-068-B49-F	*Serratia marcescens strain DSIP-2*
		RL14-068-B49-G	*Pseudomonas sp. SAUBS2-2*
		RL14-070-B49-A	*Pseudomonas putida strain ATCC 17494*
		RL14-071-B49-A	*Burkholderia gladioli KO-BH3(SUPP3014)*
		RL14-071-B49-B	*Burkholderia gladioli strain BgHL-01*
		RL14-071-B49-C	*Burkholderia gladioli strain BgHL-01*
		RL14-071-B49-D	*Burkholderia gladioli KO-BH3(SUPP3014)*
		RL14-072-B49-A	*Pseudomonas sp. VTAE174*
		RL14-072-B49-B	*Serratia sp. clone WRFC17*
		RL14-073-B49-A	*Pseudomonas oryzihabitans strain M-B1A*
		RL14-073-B49-B	*Pseudomonas salomonii strain +Y14*
		RL14-073-B49-C	*Pseudomonas oryzihabitans strain M-B1A*
		RL14-073-B49-D	*Pseudomonas oryzihabitans strain M-B1A*
		RL14-073-B49-E	*Pseudomonas putida strain CE1*
		RL14-073-B49-F	*Pseudomonas sp. GM12220*
		RL14-073-B49-G	*Pseudomonas oryzihabitans strain M-B1A*
		RL14-073-B49-H	*Pseudomonas sp. 2B1*
(5) Rhizosphere, Two-Step: B4_9/BSX.	5 to 10 cm underneath the surface, strictly close to the roots.	RL14-064-BSA-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSB-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSC-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSD-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSE-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSF-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSG-A	***Burkholderia sp. FSGSA12 16S ribosomal***
		RL14-064-BSH-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSI-A	***Burkholderia gladioli strain CACua-73 16S***
		RL14-064-BSJ-A	***Burkholderia sp. FSGSA12 16S ribosomal***
		RL14-065-BSA-A	***Burkholderia multivorans ATCC 17616***
		RL14-065-BSB-A	***Burkholderia gladioli strain 2002721590***
		RL14-065-BSC-A	***Burkholderia gladioli strain 2002721590***
		RL14-065-BSD-A	***Burkholderia sp. FSGSA12 16S***
		RL14-065-BSE-A	***Burkholderia gladioli strain 2002721590***
		RL14-065-BSF-A	***Burkholderia sp. FSGSA12 16S***
		RL14-065-BSG-A	***Burkholderia gladioli strain 2002721590***
		RL14-065-BSH-A	***Burkholderia sp. FSGSD3 16S***
		RL14-065-BSI-A	***Burkholderia gladioli Strain 2002721590***
		RL14-065-BSJ-A	***Burkholderia gladioli strain 2002721590***
		RL14-069-BSA-A	***Burkholderia multivorans ATCC 17616***
		RL14-069-BSB-A	***Burkholderia multivorans ATCC 17616***
		RL14-069-BSC-A	***Burkholderia multivorans ATCC 17616***

Table 4. Results of all isolation experiments, culture media used and species identified using Blastn similarity search.

(continues…continuation…conclusion)

Experiments	*Samples Origin*	*Code**	*Species Name (Blastn)*
		RL14-069-BSD-A	***Burkholderia multivorans ATCC 17616***
		RL14-069-BSE-A	***Burkholderia multivorans ATCC 17616***
		RL14-069-BSF-A	***Burkholderia sp. FSGSA12***
		RL14-069-BSG-A	***Burkholderia sp. W20 16S***
		RL14-069-BSH-A	***Burkholderia multivorans ATCC 17616***
		RL14-069-BSI-A	***Burkholderia multivorans ATCC 17616***
		RL14-069-BSJ-A	***Burkholderia sp. W20 16S***
		RL14-071-BSA-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSB-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSC-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSD-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSE-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSF-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSG-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSH-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSI-A	***Burkholderia gladioli SUPP3014***
		RL14-071-BSJ-A	***Burkholderia gladioli SUPP3014***

*Code adopted according to the lab procedures.

Plates were made using only B4_9 to confirm its efficacy by itself (one-step). According to the preliminary results, it was noticed that B4_9 condition allows isolation of only Proteobacteria species with no specificity, for instance, species from genus *Pseudomonas*, *Serratia* and *Burkholderia* (Experiment 4, Table 4). These results shows that Nickel (+2) and Copper (+2) salts increased selectivity for isolation of random Proteobacteria species. However, using the two-step approach, 100 % of BKD species were isolated from BSX conditions. This confirmed the necessity of a two-step method that showed to be highly specific for the purpose of its development. Analysis of resistance and metabolic profiles of BKD and MGA species, being part of the second step, provided Antimicrobials and C&N-Sources to increase specificity for the whole genus. In addition, BKD species are related to the production of bacteriocins (more than 80% of species shows Biosynthetic Gene Clusters for the production of these proteinaceous toxins). This class of compounds are synthesized by bacteria for avoiding closely related microorganisms to grow, suggesting that it turns difficult to grow non-*Burkholderia* species (FREY et al., 1996). Thus, the dominance of BKD species in some bacterial communities can also be explained by the production of these toxins, since there is no UNT isolated by using BSX combinations. Likewise, experimental results corroborate genome-guided strategy, confirming all *in silico* results.

5.3.2.6 Biosynthetic Pathways related to Burkholderia species.

In order to investigate features of *Burkholderia* species for biosynthesizing NPs, an overview in 44 genomes related to these species was proceeded. Strains used in this work were the following: *B. pseudomallei; B. Sp; B. mallei; B. thailandensis; B. cenocepacia; B. multivorans; B. glumae; B. ambifaria; B. cepacia; B. gladioli; B. kururiensis; B. mimosarum; B. dolosa; B. oklahomensis; B. phenoliruptrix; B. terrae; B. ubonensis; B. vietnamiensis; B. graminis; B. phytofirmans; B. xenovorans; B. phymatum; B. andropogonis; B. sprentiae; B. bryophila; B. caledonica; B. oxyphila; B. heleia; B. ginsengisoli; B. fungorum; B. ferrariae; B. bannensis; B. acidipaludis; B. pyrrocinia CH-67; B. nodosa; B. caribensis; B. sacchari; B. sordidicola; B. rhizoxinica; B. zhejiangensis; B. grimmiae; B. lata; B. dilworthii; B. glathei.*

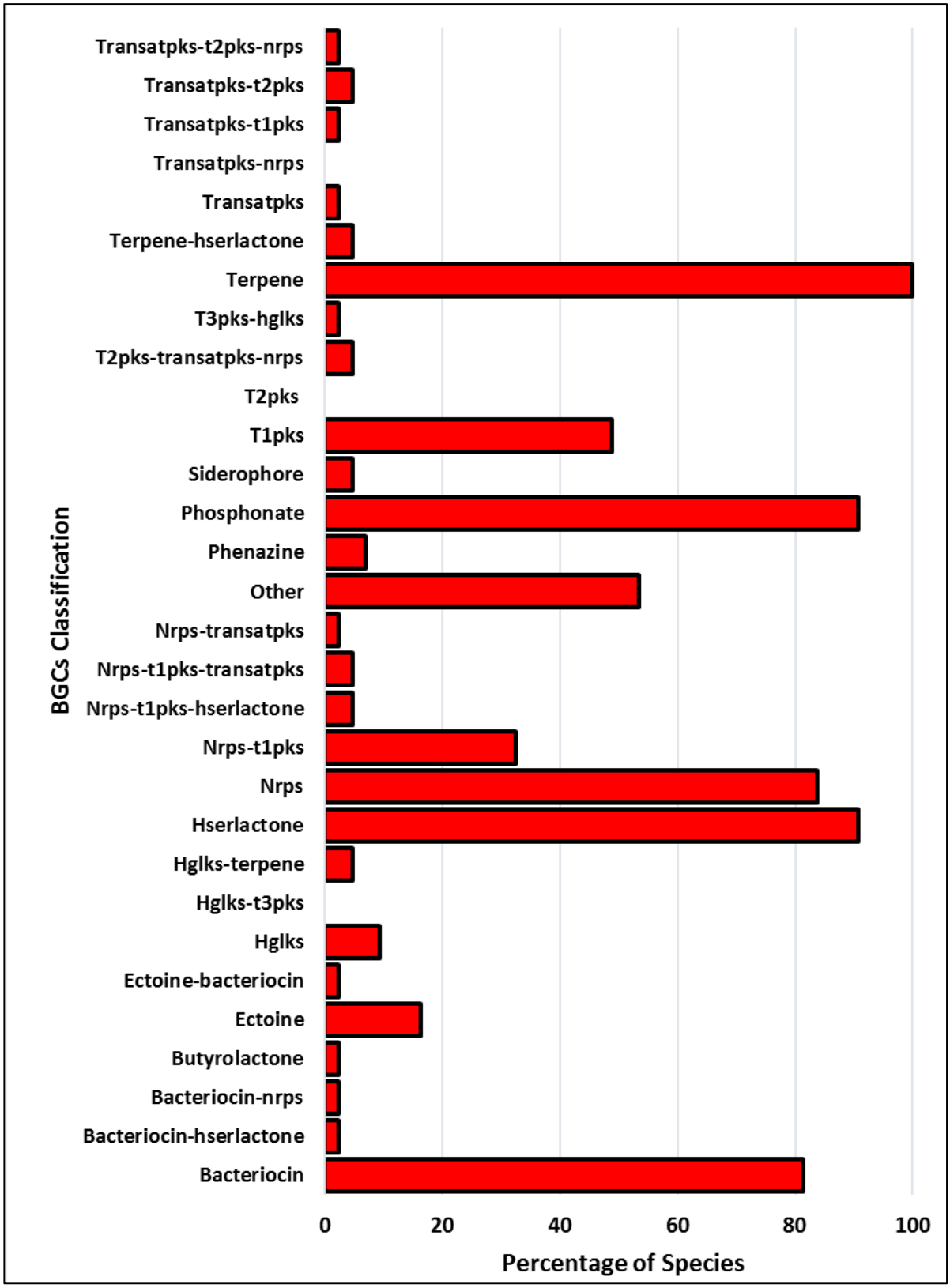

Figure 32- Classification of NPs identified for *Burkholderia* species and their percentage of occurrence in the investigated strains.

Source: From the author.

Classes of compounds as NRP, PK, phosphonates and homoserine lactone derivatives are considered as medically relevant NPs because of their importance into the chemical space. The percentage of microorganisms presenting sets of genes for the biosynthesis of NPs related to these compounds are high and techniques as those created in this work will make possible to awake bacterial gene clusters in order to isolate novel compounds (**Figure 32**). The incidence of siderophores is low, compared to other Betaproteobacteria species. Unknown classes of compounds present a considerable high presence in all genomes (>50% of all BKD genomes). This evidence suggests that novel pathways are involved in the way these microorganisms produce NPs and the discovery of novel compounds is promising.

5.3.3 CONCLUSION

The genomic-guided analysis of Resistance and Metabolic profiles along with optimizations made possible to isolate diverse species of the genus *Burkholderia*, the target of this work due to its NPs biosynthetic potential. In this method, created for isolation of species from the genus *Burkholderia*, it is possible to perceive how powerful these techniques are by getting information directly from genome sequences and bioinformatics tools. After diverse optimization, a generalizable method for the isolation of not only one species, but as much species from the same genus as possible, was created and the applications are indeed wide. In all experiments, it was noticed that all combinations also permitted the isolation of closely related microorganisms, as well as, *Pseudomonas* spp. and *Vibrio* spp., these species are also Proteobacteria representatives, indicating that in the absence of *Burkholderia* species, another closely related microorganisms could take advantage of all components into the culture media. The success of this strategy is noticed experiment after experiment leading to better culture media conditions. Factor as C&N-Sources, Antibiotic and Metal Resistance, Nutrients, pH and Theoretical Metagenomic Analysis are the key for making this strategy a comprehensive technique for the isolating of promising species in the point of view of their NPs biosynthetic capability.

6 FINAL CONCLUSION

Biosynthetic potential of microorganisms is larger than it was thought in the past. New techniques are exploring how the nature works and arranging it in very systematic answers. Genomic-guided analysis made an incredible advance for scientific investigation in microorganisms.

In this sense, strategies as showed in this whole work, walking through interdisciplinary scientific fields are the future of NPs research processes. All user-friendly bioinformatics tools adopted here, provided important biosynthetic aspects of Betaproteobacteria species, as well as, the assembling of specific monomers and their distribution within this class. Creating datasets like those of this work make the processes of discovering novel NPs extraordinarily systematic. Similar iBGCs will give evidences to NPs produced by novel microorganisms and vice-versa by tracking biosynthetic similarities of known sequences. Similar sequences could also be analyzed according to their biosynthetic potential revealing differences and similarities. These features could guide genetic engineering in producing genes to increase a given compound production range. Bioguided examinations proceeded in this work also shed lights to distribution of compouns within classes, families, order and genus. A more detailed investigation showed intraspecific characteristics of a given microorganisms, turning possible to elect the best strain of a given species in order to find new chemistry, as in the case of the strain *B. thailandensis* MSMB121, due to its levels of identity to several NPs.

After electing the most promising candidates for isolation, through XPAIRT results, a generalizable genome-guided method was developed making possible to isolate species of interest from environmental samples. Isolation results, for *Burkholderia* species, presented showed that this method is very powerful for increasing libraries of microorganisms, since the number of uncultured species is far largest than the number of known species. NPs discovery could take advantage of this type of strategies for adding into the chemical space important NCEs. The experiments executed in this work are, firstly, zero waist investigation. Secondly, they are totally directed to use as lower amount of chemicals as possible in order to maintain levels of environmental damage next to zero. This is possible due to targeting specifically a characteristic at time by analyzing genomic sequences. In this sense, is very necessary to create smart strategies like proposed in this work, once is possible to understand the genomic

machinery as the “code of life for biological processes” to the systematization of NPs discovery processes.

REFERENCES

ALLEN, H. K. et al. Call of the wild: antibiotic resistance genes in natural environments. **Nature Reviews Microbiology**, v. 8, n. 4, p. 251–9, abr. 2010.

AMINI, S.; TAVAZOIE, S. Antibiotics and the post-genome revolution. **Current Opinion in Microbiology**, v. 14, n. 5, p. 513–8, out. 2011.

AMINOV, R. I. A brief history of the antibiotic era: lessons learned and challenges for the future. **Frontiers in Microbiology**, v. 1, n. December, p. 134, jan. 2010.

AMUNTS, A. et al. Bactobolin A binds to a site on the 70S ribosome distinct from previously seen antibiotics. **Journal of Molecular Biology**, v. 427, n. 4, p. 753–755, 2015.

BAKKAL, S. et al. Role of bacteriocins in mediating interactions of bacterial isolates taken from cystic fibrosis patients. **Microbiology**, v. 156, n. 7, p. 2058–2067, 2010.

BALANDREAU, J. et al. Burkholderia cepacia genomovar III is a common plant-associated bacterium. **Applied and Environmental Microbiology**, v. 67, n. 2, p. 982–985, 2001.

BALDIM, J. L.; SOARES, M. G. An Insight Into the Intraspecific Variation of Biosynthetic Gene Clusters Between Strains of Burkholderia thailandensis spp. **Journal of the Brazilian Chemical Society**, v. 00, n. 00, p. 1–8, 2016.

BELOVA, S. E.; PANKRATOV, T. A.; DEDYSH, S. N. Bacteria of the genus Burkholderia as a typical component of the microbial community of sphagnum peat bogs. **Mikrobiologiia**, v. 75, n. 1, p. 110–117, 2006.

BERGMARK, L. et al. Assessment of the specificity of Burkholderia and Pseudomonas qPCR assays for detection of these genera in soil using 454 pyrosequencing. **FEMS Microbiology Letters**, 2012.

BIGGINS, J. B. et al. Metabolites from the induced expression of cryptic single operons found in the genome of burkholderia pseudomallei. **Journal of the American Chemical Society**, v. 133, n. 6, p. 1638–1641, 2011.

BIGGINS, J. B.; TERNEI, M. A; BRADY, S. F. Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of Burkholderia pseudomallei group pathogens John. **Journal of American Chemical Society**, v. 134, n. 32, p. 13192–13195, 2012.

BLIN, K. et al. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. **Nucleic Acids Research**, v. 41, n. Web Server issue, p. 204–212, 2013.

BRAGINA, A. et al. Vertical transmission explains the specific Burkholderia pattern in Sphagnum mosses at multi-geographic scale. **Frontiers in Microbiology**, v. 4, n. DEC, p. 1–10, 2013.

BRANDT, C. et al. The bigger picture: The history of antibiotics and antimicrobial resistance displayed by scientometric data. **International Journal of Antimicrobial Agents**, v. 44, n. 5, p. 424–430, 25 ago. 2014.

BRICKMAN, T. J. et al. Purification, spectroscopic analysis and biological activity of the macrocyclic dihydroxamate siderophore alcaligin produced by Bordetella pertussis and Bordetella bronchiseptica. **Biometals**, v. 9, n. 2, p. 191–203, 1996.

CABOCHE, S. et al. NORINE: A database of nonribosomal peptides. **Nucleic Acids Research**, v. 36, n. SUPPL. 1, p. 326–331, 2008.

CARR, G. et al. Sources of diversity in bactobolin biosynthesis by burkholderia thailandensis E264. **Organic Letters**, v. 13, n. 12, p. 3048–3051, 2011.

CASSIR, N.; ROLAIN, J.-M.; BROUQUI, P. A new strategy to fight antimicrobial resistance: the revival of old antibiotics. **Frontiers in Microbiology**, v. 5, n. October, p. 1–15, 20 out. 2014.

CHALLIS, G. L.; NAISMITH, J. H. Structural aspects of non-ribosomal peptide biosynthesis. **Current Opinion in Structural Biology**, v. 14, n. 6, p. 748–756, 2004.

CHALLIS, G. L.; RAVEL, J. Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. **FEMS Microbiology Letters**, v. 187, n. 2, p. 111–114, 2000.

CHALLIS, G. L.; RAVEL, J.; TOWNSEND, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. **Chemistry and Biology**, v. 7, n. 3, p. 211–224, 2000.

CHEN, J.; STUBBE, J. Bleomycins: towards better therapeutics. **Nature Reviews. Cancer**, v. 5, n. 2, p. 102–112, 2005.

CHENG, Y. Q.; YANG, M.; MATTER, A. M. Characterization of a gene cluster responsible for the biosynthesis of anticancer agent FK228 in Chromobacterium violaceum No. 968. **Applied and Environmental Microbiology**, v. 73, n. 11, p. 3460–3469, 2007.

CIMERMANCIC, P. et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. **Cell**, v. 158, n. 2, p. 412–421, jul. 2014.

CRAWFORD, J. M.; CLARDY, J. Microbial genome mining answers longstanding biosynthetic questions. **Proceedings of the National Academy of Sciences**, v. 109, n. 20, p. 7589–7590, 2012.

DIMINIC, J. et al. Databases of the thiotemplate modular systems (CSDB) and their in silico recombinants (r-CSDB). **Journal of Industrial Microbiology & Biotechnology**, v. 40, n. 6, p. 653–659, 2013.

DITTMANN, E. et al. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. **Trends in Microbiology**, v. 23, n. 10, p. 642–652, 2015.

DUNCAN, K. R. et al. Molecular Networking and Pattern-Based Genome Mining Improves Discovery of Biosynthetic Gene Clusters and their Products from Salinispora Species.

Chemistry and Biology, p. 460–471, 2014.

EL-ELIMAT, T. et al. Benzoquinones and terphenyl compounds as phosphodiesterase-4B inhibitors from a fungus of the order chaetothyriales (MSX 47445). **Journal of Natural Products**, v. 76, n. 3, p. 382–387, 2013.

EMRICK, D. et al. The antifungal occidiofungin triggers an apoptotic mechanism of cell death in yeast. **Journal of Natural Products**, v. 76, n. 5, p. 829–838, 2013.

ESTRADA-DE LOS SANTOS, P.; BUSTILLOS-CRISTALES, R.; CABALLERO-MELLADO, J. Burkholderia, a Genus Rich in Plant-Associated Nitrogen Fixers with Wide Environmental and Geographic Distribution. **Applied and Environmental Microbiology**, v. 67, n. 6, p. 2790–2798, 2001.

FELNAGLE, E. A. et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. **Molecular Pharmaceutics**, v. 5, n. 2, p. 191–211, 2008.

FIERER, N.; JACKSON, R. B. The diversity and biogeography of soil bacterial communities. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 3, p. 626–631, 2006.

FISCH, K. M. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. **RSC Advances**, v. 3, n. 40, p. 18228, 2013.

FISCHBACH, M. A.; WALSH, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. **Chemical Reviews**, v. 106, n. 8, p. 3468–3496, 2006.

FOERSTNER, K. U. et al. Environments shape the nucleotide composition of genomes. **EMBO Reports**, v. 6, n. 12, p. 1208–1213, 2005.

FORSETH, R. R. et al. Homologous NRPS-like gene clusters mediate redundant small-molecule biosynthesis in Aspergillus flavus. **Angewandte Chemie - International Edition**, v. 52, n. 5, p. 1590–1594, 2013.

FRANCIS, A. et al. An improved selective and differential medium for the isolation of Burkholderia pseudomallei from clinical specimens. **Diagnostic Microbiology and Infectious Disease**, v. 55, p. 95–99, 2006.

FRANKE, J.; ISHIDA, K.; HERTWECK, C. Genomics-Driven Discovery of Burkholderic Acid, a Noncanonical, Cryptic Polyketide from Human Pathogenic *Burkholderia* Species. **Angewandte Chemie International Edition**, v. 51, n. 46, p. 11611–11615, 2012.

FREY, P. et al. Bacteriocin typing of Burkholderia (Pseudomonas) solanacearum race 1 of the French West Indies and correlation with genomic variation of the pathogen. **Applied and Environmental Microbiology**, v. 62, n. 2, p. 473–479, 1996.

FUKUI, Y. et al. Total synthesis of burkholdacs A and B and 5,6,20-tri-epi-burkholdac A: HDAC inhibition and antiproliferative activity. **European Journal of Medicinal Chemistry**, v. 76, p. 301–313, 2014.

GAUDÊNCIO, S. P.; PEREIRA, F. Dereplication: racing to speed up the natural products

discovery process. **Natural Products Reports**, v. 32, n. 6, p. 779–810, 2015.

GRAUPNER, K. et al. Imaging mass spectrometry and genome mining reveal highly antifungal virulence factor of mushroom soft rot pathogen. **Angewandte Chemie - International Edition**, v. 51, n. 52, p. 13173–13177, 2012.

GROSS, H. et al. The Genomisotopic Approach: A Systematic Method to Isolate Products of Orphan Biosynthetic Gene Clusters. **Chemistry and Biology**, v. 14, n. 1, p. 53–63, 2007.

HARVEY, A. L.; EDRADA-EBEL, R.; QUINN, R. J. The re-emergence of natural products for drug discovery in the genomics era. **Nature Reviews Drug Discovery**, v. 14, n. 2, p. 111–129, 2015.

HASSEN, A. et al. Resistance of environmental bacteria to heavy metals. **Bioresource Technology**, v. 64, n. 1, p. 7–15, 1998.

HUANG, T. et al. Identification and characterization of the pyridomycin biosynthetic gene cluster of Streptomyces pyridomyceticus NRRL B-2517. **Journal of Biological Chemistry**, v. 286, n. 23, p. 20648–20657, 2011.

ISHIDA, K. et al. Induced biosynthesis of cryptic polyketide metabolites in a Burkholderia thailandensis quorum sensing mutant. **Journal of the American Chemical Society**, v. 132, n. 40, p. 13966–13968, 2010.

JACOBS, J. L. et al. Identification and onion pathogenicity of Burkholderia cepacia complex isolates from the onion rhizosphere and onion field soil. **Applied and Environmental Microbiology**, v. 74, n. 10, p. 3121–9, maio 2008.

JAIN, A. et al. Effects of thailanstatins on glucocorticoid response in trabecular meshwork and steroid-induced glaucoma. **Investigative Ophthalmology & Visual Science**, v. 54, n. 5, p. 3137–42, 2013.

JOHNSTON, C. W. et al. Gold biomineralization by a metallophore from a gold-associated microbe. **Nature Chemical Biology**, v. 9, n. 4, p. 241–3, 2013.

JOHNSTON, C. W. et al. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. **Nature Communications**, v. 6, p. 8421, 2015.

JU, K. S.; DOROGHAZI, J. R.; METCALF, W. W. Genomics-enabled discovery of phosphonate natural products and their biosynthetic pathways. **Journal of Industrial Microbiology and Biotechnology**, v. 41, n. 2, p. 345–356, 2014.

JU, K.-S. et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. **Proceedings of the National Academy of Sciences**, v. 112, n. 39, p. 12175–12180, 2015.

KANEHISA, M. et al. Data, information, knowledge and principle: back to metabolism in KEGG. **Nucleic Acids Research**, v. 42, n. Database issue, p. D199–205, jan. 2014.

KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.

KAWANISHI, T. et al. New detection systems of bacteria using highly selective media designed by SMART: Selective medium- design algorithm restricted by two constraints. **PLoS ONE**, v. 6, n. 1, 2011.

KEATING, T. A.; WALSH, C. T. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. **Current Opinion in Chemical Biology**, v. 3, n. 5, p. 598–606, 1999.

KERSTEN, R. D. et al. A mass spectrometry–guided genome mining approach for natural product peptidogenomics. **Nature Chemical Biology**, v. 7, n. 11, p. 794–802, 2011.

KM. VANDERMOLEN, WILLIAM MCCULLOCH, CEDRIC J. PEARCE, N. H. O. Romidepsin (Istodax, NSC 630176, FR901228, FK228, depsipeptide): a natural product recently approved for cutaneous T-cell lymphoma. **Journal of Antibiotics**, v. 64, n. 8, p. 525–531, 2011.

KNAPPE, T. A. et al. Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of Burkholderia thailandensis E264. **Journal of the American Chemical Society**, v. 130, n. 17, p. 11446–11454, 2008.

KOBAYASHI, S. et al. Micacocidin A, B and C, novel antimycoplasma agents from Pseudomonas sp. II. Structure elucidation. **Journal of Antibiotics**, v. 51, n. 3, p. 328–332, 1998.

KOEHN, F. E.; CARTER, G. T. The evolving role of natural products in drug discovery. **Nature Reviews. Drug Discovery**, v. 4, n. 3, p. 206–220, 2005.

KONNO, H. et al. Structure activity relationship study of burkholdine analogues toward simple antifungal agents. **Bioorganic & Medicinal Chemistry Letters**, v. 25, n. 16, p. 3199–3202, 2015.

KOST, T. et al. Oxalotrophy, a widespread trait of plant-associated Burkholderia species, is involved in successful root colonization of lupin and maize by Burkholderia phytofirmans. **Frontiers in Microbiology**, v. 4, n. JAN, 2013.

KURITA, K. L.; LININGTON, R. G. Connecting Phenotype and Chemotype: High-Content Discovery Strategies for Natural Products Research. **Journal of Natural Products**, p. 150302071352007, 2015.

LARKIN, M. A. et al. Clustal W and Clustal X version 2.0. **Bioinformatics**, v. 23, n. 21, p. 2947–2948, 2007.

LASSALLE, F. et al. GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. **PLOS Genetics**, v. 11, n. 2, p. e1004941, 2015.

LEPLEUX, C. et al. Correlation of the abundance of betaproteobacteria on mineral surfaces with mineral weathering in forest soils. **Applied and Environmental Microbiology**, v. 78, n. 19, p. 7114–7119, 2012.

LI, B.; WALSH, C. T. Identification of the gene cluster for the dithiolopyrrolone antibiotic holomycin in Streptomyces clavuligerus. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 46, p. 19731–5, 2010.

LING, L. L. et al. A new antibiotic kills pathogens without detectable resistance. **Nature**, v. 517, n. 7535, p. 455–459, 7 jan. 2015.

LIU, B.; POP, M. ARDB--Antibiotic Resistance Genes Database. **Nucleic Acids Research**, v. 37, n. Database issue, p. D443–7, jan. 2009.

LIU, X. et al. Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative agents from Burkholderia thailandensis MSMB43. **Journal of Natural Products**, v. 76, p. 685–693, 2013.

LIU, X.; CHENG, Y. Q. Genome-guided discovery of diverse natural products from Burkholderia sp. **Journal of Industrial Microbiology and Biotechnology**, 2014.

LLARRULL, L. I. et al. The future of the Betalactams. **Current Opinion in Microbiology**, v. 13, n. 5, p. 551–557, 2010.

LUDOVIC, V. et al. Burkholderia diversity and versatility: An investory of the extracellular products. **Journal of Microbiology and Biotechnology**, v. 17, n. 9, p. 1407–1429, 2007.

MADSEN, J. L. H.; JOHNSTONE, T. C.; NOLAN, E. M. Chemical Synthesis of Staphyloferrin B Affords Insight into the Molecular Structure, Iron Chelation, and Biological Activity of a Polycarboxylate Siderophore Deployed by the Human Pathogen Staphylococcus aureus. **Journal of the American Chemical Society**, v. 137, n. 28, p. 9117–9127, 2015.

MAHENTHIRALINGAM, E. et al. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic burkholderia ambifaria genomic island. **Chemistry and Biology**, v. 18, n. 5, p. 665–677, 2011.

MARAVIĆ, A. et al. Occurrence and antibiotic susceptiblitiy profiles of Burkholderia cepaciacomplex in Coastal Marine Environment. **International Journal of Environmental Health Research**, v. 22, n. 6, p. 531–542, 2012.

MATSUDA, S.; KOYASU, S. Mechanisms of action of cyclosporine. **Immunopharmacology**, v. 47, n. 2-3, p. 119–125, 2000.

MATTHEUS, W. et al. Isolation and Purification of a New Kalimantacin/Batumin-Related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. **Chemistry and Biology**, v. 17, n. 2, p. 149–159, 2010.

MCINTOSH, J. A; DONIA, M. S.; SCHMIDT, E. W. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. **Natural Product Reports**, v. 26, n. 4, p. 537–559, 2009.

MEDEMA, M. H. et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic Acids Research**, v. 39, n. Web Server issue, p. W339–46, jul. 2011.

MEDEMA, M. H. et al. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. **PLoS Computational Biology**, v. 10, n. 12, p. e1004016, 2014.

MEDEMA, M. H.; FISCHBACH, M. A. Computational approaches to natural product

discovery. **Nature Chemical Biology**, v. 11, n. 9, p. 639–648, 2015.

MEDINI, D. et al. Microbiology in the post-genomic era. **Nature Reviews. Microbiology**, v. 6, n. 6, p. 419–30, jun. 2008.

MERK, S. et al. Comparison of different methods for the isolation of Burkholderia cepacia DNA from pure cultures and waste water. **International Journal of Hygiene and Environmental Health**, v. 131, p. 127–131, 2001.

MILSHTEYN, A.; SCHNEIDER, J. S.; BRADY, S. F. Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. **Chemistry & Biology**, v. 21, n. 9, p. 1211–1223, 2014.

MINOWA, Y.; ARAKI, M.; KANEHISA, M. Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. **Journal of Molecular Biology**, v. 368, n. 5, p. 1500–1517, 2007.

MOHIMANI, H. et al. NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery. **Journal of Natural Products**, v. 77, p. 1902–1909, 2014.

NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. **Journal of Natural Products**, v. 75, n. 3, p. 311–35, 23 mar. 2012.

NG, J. et al. Dereplication and de novo sequencing of nonribosomal peptides. **Nat Methods**, v. 6, n. 8, p. 596–599, 2009.

NGUYEN, T. et al. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. **Nature Biotechnology**, v. 26, n. 2, p. 225–233, 2008.

NIKOLOULI, K.; MOSSIALOS, D. Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. **Biotechnology Letters**, v. 34, n. 8, p. 1393–1403, 2012.

OSHIMA, K. et al. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. **Nature Genetics**, v. 36, n. 1, p. 27–29, 2004.

PAUNGFOO-LONHIENNE, C. et al. A new species of Burkholderia isolated from sugarcane roots promotes plant growth. **Microbial Biotechnology**, v. 7, n. 2, p. 142–154, 2014.

PEREZ-LLAMAS, C.; LOPEZ-BIGAS, N. Gitools: Analysis and visualisation of genomic data using interactive heat-maps. **PLoS ONE**, v. 6, n. 5, 2011.

PIDOT, S. J. et al. Antibiotics from neglected bacterial sources. **International Journal of Medical Microbiology**, v. 304, n. 1, p. 14–22, 2014.

PRABAKARAN, S. et al. Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. **Wiley Interdisciplinary Reviews: Systems Biology and Medicine**, v. 4, n. 6, p. 565–583, 2012.

RAUSCH, C. et al. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). **Nucleic Acids**

Research, v. 33, n. 18, p. 5799–5808, 2005.

RAYNAUD, X.; NUNAN, N. Spatial ecology of bacteria at the microscale in soil. **PLoS ONE**, v. 9, n. 1, 2014.

RICHARDS, S. It ' s more than stamp collecting : how genome sequencing can unify biological research. **Trends in Genetics**, v. 31, n. 7, p. 411–421, 2015.

ROBERTSON, J. et al. The survival of Burkholderia pseudomallei in liquid media. **American Journal of Tropical Medicine and Hygiene**, v. 82, n. 1, p. 88–94, 2010.

RUBINSTEIN, E.; KEYNAN, Y. Vancomycin revisited - 60 years later. **Frontiers in Public Health**, v. 2, n. October, p. 217, 2014.

SAITOU, N.; NEI, M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees'. **Mol Biol Evol**, v. 4, n. 4, p. 406–425, 1987a.

SAITOU, N.; NEI, M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. **Mol Biol Evol**, v. 4, n. 4, p. 406–425, 1987b.

SCHAFFER, M. L.; OTTEN, L. G. Substrate flexibility of the adenylation reaction in the Tyrocidine non-ribosomal peptide synthetase. **Journal of Molecular Catalysis B: Enzymatic**, v. 59, n. 1-3, p. 140–144, 2009.

SCHLEGEL, HANS G. JANNASCH, H. W. Enrichment cultures. **Annual Review of Microbiology**, v. 21, p. 49–70, 1966.

SCHULZE, C. J. et al. "Function-first" lead discovery: mode of action profiling of natural product libraries using image-based screening. **Chemistry & Biology**, v. 20, n. 2, p. 285–95, 21 fev. 2013.

SCHWARZER, D.; FINKING, R.; MARAHIEL, M. A. Nonribosomal peptides: from genes to products. **Natural Product Reports**, v. 20, n. 3, p. 275–287, 2003.

SEO, J.; LEE, K.-J. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. **Journal of Biochemistry and Molecular Biology**, v. 37, n. 1, p. 35–44, 2004.

SEYEDSAYAMDOST, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. **Proceedings of the National Academy of Sciences**, v. 111, n. 20, p. 7266–7271, 2014.

SINGH, P. D. et al. Aerocavin, a new antibiotic produced by chromobacterium violaceum. **The Journal of Antibiotics**, p. 446–453, 1988.

SULLIVAN, J. O. et al. Janthomicins A, B and C, novel peptide lactone antibiotics produced by Janthinoba cterium lividum. **The Journal of Antibiotics**, v. XLIII, n. 8, p. 913–919, 1990.

TAEGTMEYER, A. B.; LEUPPI, J. D.; KULLAK-UBLICK, G. A. Roflumilast - A phosphodiesterase-4 inhibitor licensed for add-on therapy in severe COPD. **Swiss Medical Weekly**, v. 142, n. JULY, p. 1–9, 2012.

TALLY, F. P.; DEBRUIN, M. F. Development of daptomycin for Gram-positive infections. **Journal of Antimicrobial Chemotherapy**, v. 46, n. 4, p. 523–526, 2000.

TAMURA, K. et al. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725–2729, 2013.

UDWARY, D. W. et al. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 25, p. 10376–10381, 2007.

VANLAERE, E. et al. A novel strategy for the isolation and identification of environmental Burkholderia cepacia complex bacteria. **FEMS Microbiology Letters**, v. 249, n. 2, p. 303–7, 2005.

VANLAERE, E. et al. Growth in Stewart ' s medium is a simple , rapid and inexpensive screening tool for the identification of Burkholderia cepacia complex. **Journal of Cystic Fibrosis**, v. 5, p. 137–139, 2006.

VERMIS, K. et al. Isolation of Burkholderia cepacia complex genomovars from waters. **Systematic and Applied Microbiology**, v. 26, n. 4, p. 595–600, 2003.

VERMIS, K.; VANDAMME, P. A. R.; NELIS, H. J. Burkholderia cepacia complex genomovars: Utilization of carbon sources, susceptibility to antimicrobial agents and growth on selective media. **Journal of Applied Microbiology**, v. 95, n. 6, p. 1191–1199, 2003.

VITORIA, M. WHO Model list of essential medicines for children. n. March, p. 1–10, 2015.

WALSH, C. T. et al. Post-translational modification of polyketide and nonribosomal peptide synthases. **Current Opinion in Chemical Biology**, v. 1, n. 3, p. 309–315, 1997.

WALSH, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. **Nat. Prod. Rep.**, v. 00, p. 1–9, 2015.

WALSH, C. T.; MALCOLMSON, S. J.; YOUNG, T. S. Three ring posttranslational circuses: Insertion of oxazoles, thiazoles, and pyridines into protein-derived frameworks. **ACS Chemical Biology**, v. 7, p. 429–442, 2012.

WALSH, C. T.; O'BRIEN, R. V.; KHOSLA, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. **Angewandte Chemie International Edition**, v. 52, n. 28, p. 7098–7124, 2013.

WANG, C. et al. Thailandepsins: Bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. **Journal of Natural Products**, v. 74, p. 2031–2038, 2011a.

WANG, H. et al. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. **Proceedings of the National Academy of Sciences**, v. 111, n. 25, p. 9259–9264, 2014.

WANG, Y. et al. Molecular cloning and identification of the laspartomycin biosynthetic gene cluster from Streptomyces viridochromogenes. **Gene**, v. 483, n. 1-2, p. 11–21, 2011b.

WEBER, T. et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. **Nucleic Acids Research**, v. 43, n. W1, p. W237–W243, 2015.

WEISSKOPF, L.; HELLER, S.; EBERL, L. Burkholderia Species Are Major Inhabitants of White Lupin Cluster Roots Supplemental material Burkholderia Species Are Major Inhabitants of White Lupin Cluster Roots □ †. **Applied and Environmental Microbiology**, v. 77, n. 21, p. 7715–7720, 2011.

WILSON, A. J.; CHENG, Y.-Q.; KHABELE, D. Thailandepsins are new small molecule class I HDAC inhibitors with potent cytotoxic activity in ovarian cancer cells: a preclinical study of epigenetic ovarian cancer therapy. **Journal of Ovarian Research**, v. 5, n. 1, p. 12, 2012.

XIE, Y. et al. NRPS substrate promiscuity leads to more potent antitubercular sansanmycin analogues. **Journal of Natural Products**, v. 77, n. 7, p. 1744–1748, 2014.

YANG, L. et al. Exploration of Nonribosomal Peptide Families with an Automated Informatic Search Algorithm. **Chemistry & Biology**, v. 22, n. 9, p. 1259–1269, set. 2015.

ZAID, A. M.; BONASERA, J. M.; BEER, S. V. OEM — A new medium for rapid isolation of onion-pathogenic and onion-associated bacteria. **Journal of Microbiological Methods**, v. 91, n. 3, p. 520–526, 2012.

ZHUO, Y. et al. Revised Genome Sequence of Burkholderia thailandensis MSMB43 with Improved Annotation. **Journal of Bacteriology**, v. 194, n. 17, p. 4749–4750, 2012.

APPENDIX A

APPENDIX A – XPAIRT SUPPLEMENTARY FILES

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***ala.ala***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	3.5	68.1	3531	Nrps	53901	***ala.arg***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	53840	***ala.arg***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	53978	***ala.arg***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	53769	***ala.arg***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***ala.asn***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps	35744	***ala.cys***
Verminephrobacter aporrectodeae	*Burkholderiales*	*Comamonadaceae*	4.68	65.7	4654	Nrps-t1pks	21737	***ala.cys***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	34749	***ala.gln***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	63312	***ala.gln***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***ala.gly***
Verminephrobacter aporrectodeae	*Burkholderiales*	*Comamonadaceae*	4.68	65.7	4654	Nrps-t1pks	21737	***ala.mal***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	30889	***ala.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	3.91	68.1	3493	Nrps	65319	***ala.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	55272	***ala.nrp***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-hserlactone	79145	***ala.nrp***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps	35744	***ala.nrp***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***ala.nrp***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***ala.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia bannensis	*Burkholderiales*	*Burkholderiaceae*	8.65	64	6680	Nrps	46193	***ala.nrp***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	3.5	68.1	3531	Nrps	53901	***ala.nrp***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	53840	***ala.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	53978	***ala.nrp***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	53769	***ala.nrp***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-hserlactone	79145	***ala.nrp***
Verminephrobacter aporrectodeae	*Burkholderiales*	*Comamonadaceae*	4.68	65.7	4654	Other	3651	***ala.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	3.91	68.1	3493	Nrps-t1pks	67461	***ala.ohmal***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***ala.ohmal***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	88313	***ala.ohmal***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	88313	***ala.pk***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	30889	***ala.pk***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	55272	***ala.pk***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***ala.ser***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps	63420	***ala.ser***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***ala.ser***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***ala.val***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	63312	***ala.val***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	3.91	68.1	3493	Nrps-t1pks	67461	***ala.val***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps	63420	***ala.val***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***arg.mal***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Nitrosospira multiformis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.18	53.9	2895	Linaridin-nrps-t1pks	67398	***arg.ohmal***
Nitrosospira sp. NpAV	*Nitrosomonadales*	*Nitrosomonadaceae*	3.45	53.1	3252	Nrps-t1pks	30561	***arg.ohmal***
Nitrosospira multiformis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.18	53.9	2895	Linaridin-nrps-t1pks	67398	***arg.pk***
Nitrosospira sp. NpAV	*Nitrosomonadales*	*Nitrosomonadaceae*	3.45	53.1	3252	Nrps-t1pks	30561	***arg.pk***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***arg.ser***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	69464	***arg.ser***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***asn.cys***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Bacteriocin-nrps	36693	***asn.gly***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps-t1pks	40772	***asn.mal***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***asn.nrp***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***asn.nrp***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***asn.ohmal***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***asn.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***asn.ser***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps-t1pks	40772	***asn.ser***
Mitsuaria sp	*Burkholderiales*	*Unclassified*	6.66	67.8	???	Nrps	38533	***asp.asp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***asp.gln***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Nrps	38543	***asp.glu***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Nrps	38543	***asp.gly***
Delftia tsuruhatensis	*Burkholderiales*	*Comamonadaceae*	6.73	66.3	6065	Nrps	62938	***asp.gly***
Variovorax sp	*Burkholderiales*	*Comamonadaceae*	6.03	66.8	5650	Nrps	10761	***asp.gly***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	67286	***asp.lys***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***asp.mal***
Burkholderia ferrariae	*Burkholderiales*	*Burkholderiaceae*	7.94	64.8	6254	Nrps	17417	***asp.nrp***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***asp.nrp***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	67286	***asp.nrp***
Mitsuaria sp	*Burkholderiales*	*Unclassified*	6.66	67.8	???	Nrps	38533	***asp.nrp***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***asp.nrp***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***asp.nrp***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps	54705	***asp.nrp***
Burkholderia bryophila	*Burkholderiales*	*Burkholderiaceae*	7.38	61.9	6552	Nrps	57223	***asp.nrp***
Burkholderia caledonica	*Burkholderiales*	*Burkholderiaceae*	7.28	62	6289	Nrps	27523	***asp.nrp***
Burkholderia caribensis	*Burkholderiales*	*Burkholderiaceae*	9.42	62.5	9112	Nrps	54590	***asp.nrp***
Burkholderia cenocepacia	*Burkholderiales*	*Burkholderiaceae*	3.48	66.8	3165	Nrps	54657	***asp.nrp***
Burkholderia cepacia	*Burkholderiales*	*Burkholderiaceae*	8.95	65.9	7655	Nrps	54657	***asp.nrp***
Burkholderia cepacia ATCC 25416	*Burkholderiales*	*Alcaligenaceae*	8.61	66.6	772	Nrps	54696	***asp.nrp***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	47189	***asp.nrp***
Burkholderia lata	*Rhodocyclales*	*Rhodocyclaceae*	3.69	66.2	3369	Nrps	54660	***asp.nrp***
Burkholderia multivorans	*Burkholderiales*	*Burkholderiaceae*	3.45	66.9	3224	Nrps	54645	***asp.nrp***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.08	63.4	3543	Nrps	67079	***asp.nrp***
Burkholderia terrae	*Burkholderiales*	*Burkholderiaceae*	11.29	61.8	1029	Nrps	37155	***asp.nrp***
Burkholderia vietnamiensis	*Burkholderiales*	*Burkholderiaceae*	3.65	66.5	3326	Nrps	54723	***asp.nrp***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Arylpolyene	21238	***asp.nrp***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	67286	***asp.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***asp.nrp***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***asp.nrp***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***asp.ohmal***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***asp.ohmal***
Cupriavidus taiwanensis	*Burkholderiales*	*Burkholderiaceae*	2.5	67.9	2185	Nrps	67153	***asp.orn***
Burkholderia fungorum	*Burkholderiales*	*Burkholderiaceae*	8.7	61.8	7333	Nrps	54960	***asp.orn***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	3.5	68.1	3531	Nrps	55140	***asp.orn***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	55140	***asp.orn***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	55230	***asp.orn***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	54724	***asp.orn***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***asp.orn***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***asp.orn***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	54220	***asp.pk***
Burkholderia graminis	*Burkholderiales*	*Burkholderiaceae*	7.48	62.9	6788	Nrps	61960	***asp.pk***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	Nrps	64662	***asp.pk***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	69464	***asp.pk***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.06	63.7	3581	Nrps	62549	***asp.pk***
Cupriavidus taiwanensis	*Burkholderiales*	*Burkholderiaceae*	2.5	67.9	2185	Nrps	67153	***asp.pk***
Mitsuaria sp	*Burkholderiales*	*Unclassified*	6.66	67.8	???	Nrps	38533	***asp.pk***
Achromobacter xylosoxidans	*Burkholderiales*	*Alcaligenaceae*	7.01	66	6466	Nrps	70902	***asp.ser***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps	54705	***asp.ser***
Burkholderia caribensis	*Burkholderiales*	*Burkholderiaceae*	9.42	62.5	9112	Nrps	54590	***asp.ser***
Burkholderia cenocepacia	*Burkholderiales*	*Burkholderiaceae*	3.48	66.8	3165	Nrps	54657	***asp.ser***
Burkholderia cepacia	*Burkholderiales*	*Burkholderiaceae*	8.95	65.9	7655	Nrps	54657	***asp.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia cepacia ATCC 25416	*Burkholderiales*	*Alcaligenaceae*	8.61	66.6	772	Nrps	54696	***asp.ser***
Burkholderia fungorum	*Burkholderiales*	*Burkholderiaceae*	8.7	61.8	7333	Nrps	54960	***asp.ser***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	47189	***asp.ser***
Burkholderia lata	*Rhodocyclales*	*Rhodocyclaceae*	3.69	66.2	3369	Nrps	54660	***asp.ser***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	3.5	68.1	3531	Nrps	55140	***asp.ser***
Burkholderia multivorans	*Burkholderiales*	*Burkholderiaceae*	3.45	66.9	3224	Nrps	54645	***asp.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	55140	***asp.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps	11399	***asp.ser***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	69464	***asp.ser***
Burkholderia terrae	*Burkholderiales*	*Burkholderiaceae*	11.29	61.8	1029	Nrps	37155	***asp.ser***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	55230	***asp.ser***
Burkholderia vietnamiensis	*Burkholderiales*	*Burkholderiaceae*	3.65	66.5	3326	Nrps	54723	***asp.ser***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Arylpolyene	21238	***asp.ser***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	54724	***asp.ser***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	67286	***asp.ser***
Achromobacter xylosoxidans	*Burkholderiales*	*Alcaligenaceae*	7.01	66	6466	Nrps	70902	***asp.ser***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***asp.ser***
Thauera phenylacetica	*Rhodocyclales*	*Rhodocyclaceae*	5.02	68.6	4936	Nrps	8915	***asp.ser***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	54220	***asp.thr***
Burkholderia bannensis	*Burkholderiales*	*Burkholderiaceae*	8.65	64	6680	Nrps	41535	***asp.thr***
Burkholderia bryophila	*Burkholderiales*	*Burkholderiaceae*	7.38	61.9	6552	Nrps	57223	***asp.thr***
Burkholderia graminis	*Burkholderiales*	*Burkholderiaceae*	7.48	62.9	6788	Nrps	61960	***asp.thr***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	Nrps	64662	***asp.thr***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.08	63.4	3543	Nrps	67079	***asp.thr***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.06	63.7	3581	Nrps	62549	***asp.thr***
Cupriavidus taiwanensis	*Burkholderiales*	*Burkholderiaceae*	2.5	67.9	2185	Nrps	67153	***asp.thr***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***asp.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***asp.thr***
Herbaspirillum seropedicae	*Burkholderiales*	*Oxalobacteraceae*	5.51	63.4	4809	Nrps	54897	***asp.thr***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***asp.thr***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***asp.thr***
Acidovorax avenae	*Burkholderiales*	*Comamonadaceae*	5.48	68.8	4752	Nrps-t1pks	65516	***asp.thr***
Acidovorax citrulli	*Burkholderiales*	*Comamonadaceae*	5.35	68.5	4782	Nrps-t1pks	65471	***asp.thr***
Acidovorax oryzae	*Burkholderiales*	*Comamonadaceae*	5.53	68.7	4832	Nrps-t1pks	62251	***asp.thr***
Burkholderia bannensis	*Burkholderiales*	*Burkholderiaceae*	8.65	64	6680	Nrps	41535	***asp.thr***
Cupriavidus taiwanensis	*Burkholderiales*	*Burkholderiaceae*	2.5	67.9	2185	Nrps	67153	***asp.thr***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	65174	***asp.thr***
Comamonadaceae bacterium URHA0028	*Burkholderiales*	*Comamonadaceae*	4.9	68	4638	T1pks	50750	***ccmal.gly***
Ramlibacter tataouinensis	*Burkholderiales*	*Comamonadaceae*	4.07	70	6912	T1pks	50821	***ccmal.gly***
Burkholderia mimosarum LMG 23256	*Burkholderiales*	*Burkholderiaceae*	8.41	63.9	7524	T1pks	20067	***ccmal.nrp***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	T1pks	46221	***ccmal.nrp***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***ccmal.redmal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***ccmal.redmal***
Burkholderia cenocepacia	*Burkholderiales*	*Burkholderiaceae*	3	66.9	2633	Nrps	52782	***cys.cys***
Burkholderia cepacia ATCC 25416	*Burkholderiales*	*Alcaligenaceae*	8.61	66.6	772	Nrps	52824	***cys.cys***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	53900	***cys.cys***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Nrps	53288	***cys.cys***
Candidatus Glomeribacter gigasporarum	*Burkholderiales*	*Burkholderiaceae*	1.73	54.8	1815	Nrps-t1pks	54057	***cys.cys***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Microvirgula aerodenitrificans	*Neisseriales*	*Chromobacteriaceae*	4.04	64.2	3756	Nrps	52728	***cys.cys***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	3.72	67	3502	Nrps-t1pks	72756	***cys.cys***
Verminephrobacter aporrectodeae	*Burkholderiales*	*Comamonadaceae*	4.68	65.7	4654	Nrps	22290	***cys.cys***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***cys.gln***
Candidatus Glomeribacter gigasporarum	*Burkholderiales*	*Burkholderiaceae*	1.73	54.8	1815	Nrps-t1pks	54057	***cys.gly***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***cys.gly***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps-t1pks	55986	***cys.mal***
Burkholderia cenocepacia	*Burkholderiales*	*Burkholderiaceae*	3	66.9	2633	Nrps	52782	***cys.nrp***
Burkholderia ginsengisoli	*Burkholderiales*	*Burkholderiaceae*	6.54	63.3	5767	Nrps-t1pks	48997	***cys.nrp***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	85296	***cys.nrp***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	43612	***cys.nrp***
Burkholderia cepacia ATCC 25416	*Burkholderiales*	*Alcaligenaceae*	8.61	66.6	772	Nrps	52824	***cys.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	53900	***cys.nrp***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Nrps	53288	***cys.nrp***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	85296	***cys.nrp***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-hserlactone	62042	***cys.nrp***
Microvirgula aerodenitrificans	*Neisseriales*	*Chromobacteriaceae*	4.04	64.2	3756	Nrps	52728	***cys.nrp***
Verminephrobacter eiseniae	*Burkholderiales*	*Comamonadaceae*	5.57	65.3	4873	Nrps-t1pks	62690	***cys.ohmal***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	3.72	67	3502	Nrps-t1pks	72756	***cys.ohmal***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	85296	***cys.pk***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	31397	***cys.pk***
Chromobacterium sp	*Neisseriales*	*Chromobacteriaceae*	5.12	62.6	5113	Nrps	11607	***cys.pk***
Verminephrobacter aporrectodeae	*Burkholderiales*	*Comamonadaceae*	4.68	65.7	4654	Nrps-t1pks	21737	***cys.pk***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	43612	***cys.pro***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	85296	***cys.thr***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***cys.val***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***dab.leu***
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	59387	***dab.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***dab.thr***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps	46822	***dhb.lys***
Burkholderia cepacia	*Burkholderiales*	*Burkholderiaceae*	8.95	65.9	7655	Nrps	28108	***dhb.lys***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps	51981	***dhb.lys***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Nrps	46532	***dhb.lys***
Derxia gummosa	*Burkholderiales*	*Alcaligenaceae*	5.18	69.9	432	Nrps	62001	***dhb.lys***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	46591	***dhb.lys***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps	46894	***dhb.lys***
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	46903	***dhb.lys***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps	46894	***dhb.nrp***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Nrps	58393	***dhb.nrp***
Pusillimonas sp	*Burkholderiales*	*Alcaligenaceae*	3.88	56.9	3609	Nrps	55771	***dhb.nrp***
Alcaligenes faecalis	*Burkholderiales*	*Alcaligenaceae*	4.4	56.4	4068	Nrps	45833	***dhb.ser***
Alcaligenes faecalis subsp. phenolicus DSM 16503	*Burkholderiales*	*Alcaligenaceae*	4.25	56.4	3917	Nrps	45833	***dhb.ser***
Alcaligenes sp	*Burkholderiales*	*Alcaligenaceae*	4.27	56.6	3905	Nrps	25389	***dhb.ser***
Castellaniella defragrans	*Burkholderiales*	*Alcaligenaceae*	3.95	68.9	3667	Nrps	46128	***dhb.ser***
Lautropia mirabilis	*Burkholderiales*	*Burkholderiaceae*	3.14	65.6	2713	Nrps-t1pks	85431	***dhb.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Pusillimonas sp	*Burkholderiales*	*Alcaligenaceae*	3.88	56.9	3609	Nrps	55771	***dhb.ser***
Thauera selenatis	*Rhodocyclales*	*Rhodocyclaceae*	9.14	61	???	Nrps	30908	***dhb.ser***
Oligella ureolytica	*Burkholderiales*	*Alcaligenaceae*	2.67	44.6	2590	Nrps	7552	***dhb.thr***
Snodgrassella alvi SCGC AB-598-J21	*Burkholderiales*	*Comamonadaceae*	2.33	40.9	2485	Nrps	49314	***dhb.thr***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***gln.gly***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	47972	***gln.gly***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***gln.gly***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	34749	***gln.nrp***
Rhodocyclaceae bacterium RZ94	*Burkholderiales*	*Burkholderiaceae*	4.3	66.6	3958	Nrps	52335	***gln.nrp***
Methyloversatilis universalis	*Rhodocyclales*	*Rhodocyclaceae*	4.23	67	3981	Nrps	52089	***gln.nrp***
Rhodocyclaceae bacterium RZ94	*Burkholderiales*	*Burkholderiaceae*	4.3	66.6	3958	Nrps	52335	***gln.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	63312	***gln.pro***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***gln.thr***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	47972	***gln.thr***
Chromobacterium sp	*Neisseriales*	*Chromobacteriaceae*	5.12	62.6	5113	Nrps	6478	***gln.thr***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***gln.thr***
Methyloversatilis universalis	*Rhodocyclales*	*Rhodocyclaceae*	4.23	67	3981	Nrps	52089	***gln.val***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	Nrps	55703	***glu.gly***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***glu.gly***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Nrps	38543	***glu.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***glu.ser***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	Nrps	55703	***glu.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***glu.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***glu.thr***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***gly.gly***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***gly.lys***
Candidatus Glomeribacter gigasporarum	*Burkholderiales*	*Burkholderiaceae*	1.73	54.8	1815	Nrps-t1pks	54057	***gly.mal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***gly.mal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***gly.mal***
Burkholderia caribensis	*Burkholderiales*	*Burkholderiaceae*	9.42	62.5	9112	Nrps-t1pks	48072	***gly.mal***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	T2pks-transatpks-nrps	98014	***gly.mal***
Burkholderia heleia	*Burkholderiales*	*Burkholderiaceae*	8.01	64.6	7179	Nrps-t1pks	52248	***gly.mal***
Burkholderia mimosarum STM 3621	*Burkholderiales*	*Burkholderiaceae*	8.64	63.9	7762	Nrps-t1pks	57736	***gly.mal***
Burkholderia nodosa	*Burkholderiales*	*Burkholderiaceae*	9.62	64.1	8630	Nrps-t1pks	50358	***gly.mal***
Burkholderia phymatum	*Burkholderiales*	*Burkholderiaceae*	3.48	63	3158	Nrps-t1pks	50475	***gly.mal***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	T2pks-transatpks-nrps	100390	***gly.mal***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps-t1pks	52215	***gly.mal***
Burkholderia terrae	*Burkholderiales*	*Burkholderiaceae*	11.29	61.8	1029	Nrps-t1pks	50412	***gly.mal***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Transatpks-t2pks-nrps	108298	***gly.mal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Nrps-transatpks	66417	***gly.mal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***gly.mal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***gly.mal***
Ideonella sp	*Burkholderiales*	*Unclassified*	4.94	68.7	482	Nrps-t1pks	16907	***gly.mal***
Lautropia mirabilis	*Burkholderiales*	*Burkholderiaceae*	3.14	65.6	2713	Nrps-t1pks	85431	***gly.mal***
Thiomonas arsenitoxydans	*Burkholderiales*	*Burkholderiaceae*	3.74	63.8	3534	Nrps-t1pks	52413	***gly.mal***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Thiomonas intermedia	*Burkholderiales*	*Comamonadaceae*	3.4	63.9	3206	Nrps-t1pks	52414	***gly.mal***
Achromobacter insuavis	*Burkholderiales*	*Alcaligenaceae*	6.86	67.7	6152	T1pks	37090	***gly.nrp***
Burkholderia ferrariae	*Burkholderiales*	*Burkholderiaceae*	7.94	64.8	6254	Nrps	17417	***gly.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	T2pks-transatpks-nrps	98014	***gly.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***gly.nrp***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	53315	***gly.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***gly.nrp***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	Nrps	64662	***gly.nrp***
Burkholderia mimosarum STM 3621	*Burkholderiales*	*Burkholderiaceae*	8.64	63.9	7762	Nrps-t1pks	57736	***gly.nrp***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	69464	***gly.nrp***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.06	63.7	3581	Nrps	62549	***gly.nrp***
Burkholderia sp1	*Burkholderiales*	*Burkholderiaceae*	4.08	63.4	3543	Nrps	67079	***gly.nrp***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Bacteriocin-nrps	36693	***gly.nrp***
Lautropia mirabilis	*Burkholderiales*	*Burkholderiaceae*	3.14	65.6	2713	Nrps-t1pks	85431	***gly.nrp***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***gly.nrp***
Thauera phenylacetica	*Rhodocyclales*	*Rhodocyclaceae*	5.02	68.6	4936	Hserlactone-nrps-t1pks	51194	***gly.nrp***
Burkholderia ferrariae	*Burkholderiales*	*Burkholderiaceae*	7.94	64.8	6254	Nrps	17417	***gly.nrp***
Burkholderia ferrariae	*Burkholderiales*	*Burkholderiaceae*	7.94	64.8	6254	Nrps	17417	***gly.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***gly.nrp***
Burkholderia graminis	*Burkholderiales*	*Burkholderiaceae*	7.48	62.9	6788	Nrps	61960	***gly.nrp***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	Nrps	64662	***gly.nrp***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	69464	***gly.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.06	63.7	3581	Nrps	62549	***gly.nrp***
Burkholderia sp1	*Burkholderiales*	*Burkholderiaceae*	4.08	63.4	3543	Nrps	67079	***gly.nrp***
Thiomonas intermedia	*Burkholderiales*	*Comamonadaceae*	3.4	63.9	3206	Nrps	46497	***gly.nrp***
Thiomonas sp. FB-6	*Burkholderiales*	*Thiomonas*	4.2	70	3891	Nrps	25003	***gly.nrp***
Thiomonas sp. FB-Cd	*Burkholderiales*	*Thiomonas*	4.39	62.5	4036	Nrps	48865	***gly.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	3.91	68.1	3493	Nrps-t1pks	67461	***gly.ohmal***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***gly.ohmal***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	88313	***gly.ohmal***
Caldimonas manganoxidans	*Burkholderiales*	*Comamonadaceae*	3.53	66	3388	T1pks	51180	***gly.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Nrps-transatpks	66417	***gly.ohmal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***gly.ohmal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***gly.ohmal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***gly.ohmal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***gly.ohmal***
Thauera phenylacetica	*Rhodocyclales*	*Rhodocyclaceae*	5.02	68.6	4936	Hserlactone-nrps-t1pks	51194	***gly.ohmal***
Delftia tsuruhatensis	*Burkholderiales*	*Comamonadaceae*	6.73	66.3	6065	Nrps	62938	***gly.orn***
Burkholderia bannensis	*Burkholderiales*	*Burkholderiaceae*	8.65	64	6680	Nrps	41535	***gly.pk***
Caldimonas manganoxidans	*Burkholderiales*	*Comamonadaceae*	3.53	66	3388	Nrps	44643	***gly.pk***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	52889	***gly.pro***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***gly.redmal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***gly.redmal***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***gly.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia bannensis	*Burkholderiales*	*Burkholderiaceae*	8.65	64	6680	Nrps	41535	***gly.thr***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***gly.thr***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	47972	***gly.thr***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***gly.thr***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***gly.thr***
Burkholderia bannensis	*Burkholderiales*	*Burkholderiaceae*	8.65	64	6680	Nrps	41535	***gly.thr***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***gly.thr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***gly.trp***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***gly.tyr***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	53315	***gly.tyr***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	Nrps	55703	***gly.val***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***gly.val***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks-hserlactone	95292	***gly.val***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	88313	***gly.val***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	55998	***ile.ser***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	4.41	67.5	3964	Bacteriocin-nrps	64063	***leu.leu***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	58824	***leu.leu***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps	35508	***leu.leu***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps	35508	***leu.lys***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Nrps	41323	***leu.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	4.41	67.5	3964	Bacteriocin-nrps	64063	***leu.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	58824	***leu.nrp***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Nrps	41323	***leu.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	4.41	67.5	3964	Bacteriocin-nrps	64063	***leu.pk***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	58824	***leu.pk***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-bacteriocin	70214	***leu.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-bacteriocin	70214	***leu.ser***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Terpene	21777	***leu.thr***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	55998	***leu.thr***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***leu.thr***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	48991	***leu.val***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	Nrps-t1pks	56336	***lys.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks	64510	***lys.nrp***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps	51981	***lys.nrp***
Derxia gummosa	*Burkholderiales*	*Alcaligenaceae*	5.18	69.9	432	Nrps	62001	***lys.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***lys.nrp***
Thauera phenylacetica	*Rhodocyclales*	*Rhodocyclaceae*	5.02	68.6	4936	Hserlactone-nrps-t1pks	51194	***lys.ohmal***
Burkholderia ferrariae	*Burkholderiales*	*Burkholderiaceae*	7.94	64.8	6254	Nrps	7549	***lys.pk***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	Nrps-t1pks	56336	***lys.pk***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks	64510	***lys.pk***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	67286	***lys.pk***
Thauera phenylacetica	*Rhodocyclales*	*Rhodocyclaceae*	5.02	68.6	4936	Hserlactone-nrps-t1pks	51194	***lys.pk***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps	35508	***lys.ser***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***lys.ser***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***lys.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***lys.val***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Transatpks	27404	***mal.mal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.mal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.mal***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***mal.mal***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-transatpks-t1pks	75484	***mal.nrp***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Nrps-t1pks-transatpks	71171	***mal.nrp***
Candidatus Glomeribacter gigasporarum	*Burkholderiales*	*Burkholderiaceae*	1.73	54.8	1815	Nrps-t1pks	54057	***mal.nrp***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-t1pks-otherks	78169	***mal.nrp***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***mal.nrp***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps-t1pks	40772	***mal.nrp***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Otherks-nrps-t1pks	76953	***mal.nrp***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps-t1pks	53973	***mal.nrp***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps-t1pks	57263	***mal.nrp***
Pusillimonas sp	*Burkholderiales*	*Alcaligenaceae*	3.88	56.9	3609	Nrps-t1pks	66692	***mal.nrp***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***mal.nrp***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	52367	***mal.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Variovorax sp	*Burkholderiales*	*Comamonadaceae*	6.03	66.8	5650	Nrps-t1pks	52316	***mal.nrp***
Verminephrobacter aporrectodeae	*Burkholderiales*	*Comamonadaceae*	4.68	65.7	4654	Other	3651	***mal.nrp***
Burkholderia ginsengisoli	*Burkholderiales*	*Burkholderiaceae*	6.54	63.3	5767	Nrps-t1pks	48997	***mal.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	55272	***mal.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***mal.nrp***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	Nrps-t1pks	56336	***mal.nrp***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps-t1pks	57263	***mal.nrp***
Ottowia thiooxydans	*Burkholderiales*	*Comamonadaceae*	7.04	58.6	6382	T1pks-otherks	53483	***mal.nrp***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Transatpks	27404	***mal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Otherks-transatpks	67840	***mal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Otherks-transatpks	67840	***mal.ohmal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.ohmal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.ohmal***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Transatpks	27404	***mal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Otherks-transatpks	67840	***mal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Otherks-transatpks	67840	***mal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Otherks-transatpks	67840	***mal.ohmal***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Nrps-transatpks	66417	***mal.ohmal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.ohmal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.ohmal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.ohmal***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***mal.ohmal***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Transatpks-nrps	80567	***mal.pk***
Ottowia thiooxydans	*Burkholderiales*	*Comamonadaceae*	7.04	58.6	6382	T1pks-otherks	53483	***mal.pk***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps-t1pks	55986	***mal.pk***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***mal.pk***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	4.41	67.5	3964	Transatpks-t2pks	96075	***mal.pk***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	3.91	68.1	3493	Nrps-t1pks	58522	***mal.pk***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-transatpks-t1pks	75484	***mal.pk***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***mal.pk***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	T1pks	44727	***mal.pk***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Transatpks-nrps	118191	***mal.pk***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Nrps-t1pks-transatpks	71171	***mal.pk***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Transatpks-t2pks-nrps	108298	***mal.pk***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps-t1pks	53973	***mal.pk***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Ottowia thiooxydans	*Burkholderiales*	*Comamonadaceae*	7.04	58.6	6382	T1pks-otherks	53483	***mal.pk***
Pusillimonas sp	*Burkholderiales*	*Alcaligenaceae*	3.88	56.9	3609	Nrps-t1pks	66692	***mal.pk***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***mal.pk***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***mal.redmal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***mal.redmal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***mal.redmal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***mal.redmal***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***mal.redmal***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***mal.redmal***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***mal.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***mal.ser***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Transatpks-nrps	118191	***mal.ser***
Thiomonas sp. FB-Cd	*Burkholderiales*	*Thiomonas*	4.39	62.5	4036	Nrps-t1pks	51830	***mal.ser***
Curvibacter gracilis	*Burkholderiales*	*Comamonadaceae*	6.75	66	5529	Nrps-t1pks	47292	***mal.ser***
Lautropia mirabilis	*Burkholderiales*	*Burkholderiaceae*	3.14	65.6	2713	Nrps-t1pks	85431	***mal.ser***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***mal.thr***
Ideonella sp	*Burkholderiales*	*Unclassified*	4.94	68.7	482	Nrps-t1pks	16907	***mal.thr***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***mal.val***
Burkholderia heleia	*Burkholderiales*	*Burkholderiaceae*	8.01	64.6	7179	Nrps-t1pks	52248	***mal.val***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps-t1pks	52215	***mal.val***
Thiomonas arsenitoxydans	*Burkholderiales*	*Burkholderiaceae*	3.74	63.8	3534	Nrps-t1pks	52413	***mal.val***
Thiomonas intermedia	*Burkholderiales*	*Comamonadaceae*	3.4	63.9	3206	Nrps-t1pks	52414	***mal.val***
Thiomonas sp. FB-Cd	*Burkholderiales*	*Thiomonas*	4.39	62.5	4036	Nrps-t1pks	51830	***mal.val***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	52367	***mal.val***
Variovorax sp	*Burkholderiales*	*Comamonadaceae*	6.03	66.8	5650	Nrps-t1pks	52316	***mal.val***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.mmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.mmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.mmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.nrp***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.ohmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.ohmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.ohmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.ohmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.ohmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***mmal.ohmal***
Achromobacter xylosoxidans	*Burkholderiales*	*Alcaligenaceae*	7.01	66	6466	Nrps	70902	***nrp.nrp***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***nrp.nrp***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Nrps	41323	***nrp.nrp***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***nrp.nrp***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***nrp.nrp***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	30889	***nrp.nrp***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	54220	***nrp.nrp***
Burkholderia dilworthii	*Burkholderiales*	*Burkholderiaceae*	7.68	61.8	6931	Nrps	36271	***nrp.nrp***
Burkholderia fungorum	*Burkholderiales*	*Burkholderiaceae*	8.7	61.8	7333	Nrps	47379	***nrp.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	56949	***nrp.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	67282	***nrp.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	4.41	67.5	3964	Bacteriocin-nrps	64063	***nrp.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	56949	***nrp.nrp***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	4.41	67.5	3964	Bacteriocin-nrps	64063	***nrp.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***nrp.nrp***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	55272	***nrp.nrp***
Burkholderia grimmiae	*Burkholderiales*	*Burkholderiaceae*	6.7	63	6094	Nrps	47258	***nrp.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-transatpks-t1pks	75484	***nrp.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	58824	***nrp.nrp***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	58824	***nrp.nrp***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***nrp.nrp***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***nrp.nrp***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.nrp***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.nrp***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.nrp***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.nrp***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	47418	***nrp.nrp***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	47418	***nrp.nrp***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Nrps-t1pks-transatpks	71171	***nrp.nrp***
Burkholderia ubonensis	*Burkholderiales*	*Burkholderiaceae*	6.93	67.3	6762	Nrps	18217	***nrp.nrp***
Burkholderia ubonensis	*Burkholderiales*	*Burkholderiaceae*	6.93	67.3	6762	Nrps	18217	***nrp.nrp***
Burkholderia ubonensis	*Burkholderiales*	*Burkholderiaceae*	6.93	67.3	6762	Nrps	18217	***nrp.nrp***
Burkholderia ubonensis	*Burkholderiales*	*Burkholderiaceae*	6.93	67.3	6762	Nrps	18217	***nrp.nrp***
Burkholderia ubonensis	*Burkholderiales*	*Burkholderiaceae*	6.93	67.3	6762	Nrps	18217	***nrp.nrp***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Burkholderia ubonensis	*Burkholderiales*	*Burkholderiaceae*	6.93	67.3	6762	Nrps	18217	***nrp.nrp***
Candidatus Glomeribacter gigasporarum	*Burkholderiales*	*Burkholderiaceae*	1.73	54.8	1815	Nrps-t1pks	54057	***nrp.nrp***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-t1pks-otherks	78169	***nrp.nrp***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-hserlactone	62042	***nrp.nrp***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-hserlactone	62042	***nrp.nrp***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***nrp.nrp***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Transatpks-nrps	80567	***nrp.nrp***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Transatpks-nrps	80567	***nrp.nrp***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***nrp.nrp***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps-t1pks	53973	***nrp.nrp***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps-t1pks	53973	***nrp.nrp***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps-t1pks	53973	***nrp.nrp***
Cupriavidus basilensis	*Burkholderiales*	*Burkholderiaceae*	8.55	65.4	7600	Nrps-t1pks	53973	***nrp.nrp***
Methyloversatilis universalis	*Rhodocyclales*	*Rhodocyclaceae*	4.23	67	3981	Nrps	52089	***nrp.nrp***
Nitrosospira briensis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.19	53.2	2935	Nrps	25617	***nrp.nrp***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***nrp.nrp***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***nrp.nrp***
Rhodoferax ferrireducens	*Burkholderiales*	*Comamonadaceae*	4.71	59.9	4303	Nrps	52252	***nrp.nrp***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***nrp.nrp***
Thiomonas sp. FB-Cd	*Burkholderiales*	*Thiomonas*	4.39	62.5	4036	Nrps	48865	***nrp.nrp***
Variovorax sp	*Burkholderiales*	*Comamonadaceae*	6.03	66.8	5650	Nrps	6132	***nrp.nrp***
Acidovorax avenae	*Burkholderiales*	*Comamonadaceae*	5.48	68.8	4752	Nrps-t1pks	65516	***nrp.ohmal***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Acidovorax citrulli	*Burkholderiales*	*Comamonadaceae*	5.35	68.5	4782	Nrps-t1pks	65471	***nrp.ohmal***
Acidovorax oryzae	*Burkholderiales*	*Comamonadaceae*	5.53	68.7	4832	Nrps-t1pks	62251	***nrp.ohmal***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	2.33	69	2113	T1pks	48519	***nrp.ohmal***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks	64510	***nrp.ohmal***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-transatpks-t1pks	75484	***nrp.ohmal***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Nrps-t1pks-transatpks	71171	***nrp.ohmal***
Candidatus Accumulibacter sp. SK-02	*Unclassified*	*Unclassified*	4.36	61.3	3908	Nrps-t1pks	28168	***nrp.ohmal***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-t1pks-otherks	78169	***nrp.ohmal***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***nrp.ohmal***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Otherks-nrps-t1pks	76953	***nrp.ohmal***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***nrp.ohmal***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***nrp.ohmal***
Sphaerotilus natans	*Burkholderiales*	*Unclassified*	4.59	69.9	4236	Nrps-t1pks	72360	***nrp.ohmal***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	65174	***nrp.ohmal***
Variovorax sp	*Burkholderiales*	*Comamonadaceae*	6.03	66.8	5650	Nrps-t1pks	18568	***nrp.ohmal***
Azovibrio restrictus	*Rhodocyclales*	*Rhodocyclaceae*	4.02	65.4	3176	Transatpks-nrps	80285	***nrp.ohmal***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Otherks-nrps-t1pks	76953	***nrp.ohmal***
Microvirgula aerodenitrificans	*Neisseriales*	*Chromobacteriaceae*	4.04	64.2	3756	Nrps-t1pks	54386	***nrp.ohmal***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Nitrosospira multiformis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.18	53.9	2895	Linaridin-nrps-t1pks	67398	***nrp.ohmal***
Verminephrobacter eiseniae	*Burkholderiales*	*Comamonadaceae*	5.57	65.3	4873	Nrps-t1pks	62690	***nrp.ohmal***
Nitrosospira multiformis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.18	53.9	2895	Linaridin-nrps-t1pks	67398	***nrp.orn***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***nrp.orn***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***nrp.orn***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	7051	***nrp.phe***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	34749	***nrp.phe***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.phe***
Delftia tsuruhatensis	*Burkholderiales*	*Comamonadaceae*	6.73	66.3	6065	Nrps	65655	***nrp.phe***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	54561	***nrp.phe***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***nrp.phe***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.phe***
Achromobacter xylosoxidans	*Burkholderiales*	*Alcaligenaceae*	7.01	66	6466	Nrps	70902	***nrp.pk***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***nrp.pk***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	3.91	68.1	3493	Nrps-t1pks	58522	***nrp.pk***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***nrp.pk***
Burkholderia graminis	*Burkholderiales*	*Burkholderiaceae*	7.48	62.9	6788	Nrps	61960	***nrp.pk***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	T1pks	43920	***nrp.pk***
Burkholderia sordidicola	*Burkholderiales*	*Burkholderiaceae*	10.26	59.6	9080	Nrps	69464	***nrp.pk***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	85296	***nrp.pk***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	2.91	68.1	2369	Transatpks-t2pks-nrps	108298	***nrp.pk***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Pusillimonas sp	*Burkholderiales*	*Alcaligenaceae*	3.88	56.9	3609	Nrps-t1pks	66692	***nrp.pk***
Acidovorax avenae	*Burkholderiales*	*Comamonadaceae*	5.48	68.8	4752	Nrps-t1pks	65516	***nrp.pk***
Acidovorax citrulli	*Burkholderiales*	*Comamonadaceae*	5.35	68.5	4782	Nrps-t1pks	65471	***nrp.pk***
Acidovorax oryzae	*Burkholderiales*	*Comamonadaceae*	5.53	68.7	4832	Nrps-t1pks	62251	***nrp.pk***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***nrp.pk***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***nrp.pk***
Burkholderia phenoliruptrix	*Burkholderiales*	*Burkholderiaceae*	4.15	63.5	367	Nrps	65878	***nrp.pk***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	88313	***nrp.pk***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	52889	***nrp.pk***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***nrp.pk***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Transatpks-nrps	80567	***nrp.pk***
Delftia tsuruhatensis	*Burkholderiales*	*Comamonadaceae*	6.73	66.3	6065	Nrps	65655	***nrp.pk***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	65174	***nrp.pk***
Variovorax sp	*Burkholderiales*	*Comamonadaceae*	6.03	66.8	5650	Nrps-t1pks	18568	***nrp.pk***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	52889	***nrp.pro***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***nrp.pro***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***nrp.pro***
Achromobacter insuavis	*Burkholderiales*	*Alcaligenaceae*	6.86	67.7	6152	T1pks	37090	***nrp.redmal***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps-t1pks	40772	***nrp.redmal***
Burkholderia caribensis	*Burkholderiales*	*Burkholderiaceae*	9.42	62.5	9112	T1pks	46296	***nrp.redmal***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	40912	***nrp.ser***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	67282	***nrp.ser***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***nrp.ser***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***nrp.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	48702	***nrp.ser***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-t1pks-otherks	78169	***nrp.ser***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Nrps	48486	***nrp.ser***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***nrp.ser***
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	48519	***nrp.ser***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Otherks-nrps-t1pks	76953	***nrp.ser***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Transatpks-nrps	80567	***nrp.ser***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	55998	***nrp.ser***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***nrp.ser***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***nrp.ser***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***nrp.ser***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	54561	***nrp.ser***
Mitsuaria sp	*Burkholderiales*	*unclassifiedBurkholderiales*	6.66	67.8	???	Nrps	38533	***nrp.ser***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Nrps	58393	***nrp.ser***
Achromobacter xylosoxidans	*Burkholderiales*	*Alcaligenaceae*	7.01	66	6466	Nrps	70902	***nrp.ser***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps-t1pks	85431	***nrp.ser***
Burkholderia mimosarum LMG 23256	*Burkholderiales*	*Burkholderiaceae*	8.41	63.9	7524	Nrps	16388	***nrp.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***nrp.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	85332	***nrp.ser***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	51638	***nrp.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	28203	***nrp.ser***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-t1pks-otherks	78169	***nrp.ser***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	23907	***nrp.ser***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Terpene	21777	***nrp.ser***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Transatpks-nrps	80567	***nrp.ser***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***nrp.ser***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	47505	***nrp.ser***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	54561	***nrp.ser***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	67286	***nrp.ser***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***nrp.ser***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***nrp.thr***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	53315	***nrp.thr***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***nrp.thr***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***nrp.thr***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Nrps	38543	***nrp.thr***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	24855	***nrp.thr***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	23907	***nrp.thr***
Chromobacterium sp	*Neisseriales*	*Chromobacteriaceae*	5.12	62.6	5113	Nrps	7341	***nrp.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***nrp.thr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***nrp.thr***
Burkholderia andropogonis	*Burkholderiales*	*Burkholderiaceae*	6.18	58.9	5435	Nrps	54220	***nrp.thr***
Burkholderia bryophila	*Burkholderiales*	*Burkholderiaceae*	7.38	61.9	6552	Nrps	57223	***nrp.thr***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	56949	***nrp.thr***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	53315	***nrp.thr***
Burkholderia graminis	*Burkholderiales*	*Burkholderiaceae*	7.48	62.9	6788	Nrps	61960	***nrp.thr***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	Nrps	64662	***nrp.thr***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***nrp.thr***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.06	63.7	3581	Nrps	62549	***nrp.thr***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.08	63.4	3543	Nrps	67079	***nrp.thr***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Nrps	38543	***nrp.thr***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	47972	***nrp.thr***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	55998	***nrp.thr***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***nrp.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***nrp.thr***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***nrp.thr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	28459	***nrp.tyr***
Chromobacterium sp	*Neisseriales*	*Chromobacteriaceae*	5.12	62.6	5113	Nrps	9796	***nrp.tyr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***nrp.tyr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	28459	***nrp.tyr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	28459	***nrp.val***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	53315	***nrp.val***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-t1pks	88313	***nrp.val***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.val***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	43612	***nrp.val***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps	60906	***nrp.val***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***nrp.val***
Rhodocyclaceae bacterium RZ94	*Burkholderiales*	*Burkholderiaceae*	4.3	66.6	3958	Nrps	52335	***nrp.val***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	28459	***nrp.val***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	67282	***nrp.val***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***nrp.val***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***nrp.val***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-hserlactone	79145	***nrp.val***
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	59387	***nrp.val***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***nrp.val***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps-t1pks	57263	***nrp.val***
Pusillimonas sp	*Burkholderiales*	*Alcaligenaceae*	3.88	56.9	3609	Nrps-t1pks	66692	***nrp.val***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps	60906	***nrp.val***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Transatpks-t1pks	91132	***ohmal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Nrps-transatpks	66417	***ohmal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Nrps-transatpks	66417	***ohmal.ohmal***
Candidatus Profftella armatura	*Unclassified*	*Unclassified*	0.45	24.2	396	Nrps-transatpks	66417	***ohmal.ohmal***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***ohmal.ohmal***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***ohmal.ohmal***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-t1pks-otherks	78169	***ohmal.pk***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***ohmal.pk***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Otherks-nrps-t1pks	76953	***ohmal.pk***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***ohmal.pk***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Nitrosomonas sp	*Nitrosomonadales*	*Nitrosomonadaceae*	3.78	45.4	3521	T1pks-hglks	53030	***ohmal.pk***
Rubrivivax gelatinosus	*Burkholderiales*	*Rubrivivax*	5.04	71.2	4767	T1pks-hglks	53009	***ohmal.pk***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	3.72	67	3502	Nrps-t1pks	72756	***ohmal.pk***
Hydrogenophaga sp	*Burkholderiales*	*Comamonadaceae*	5.14	68.4	5009	Nrps-t1pks	89623	***ohmal.redmal***
Acidovorax avenae	*Burkholderiales*	*Comamonadaceae*	5.48	68.8	4752	Nrps-t1pks	65516	***ohmal.ser***
Acidovorax citrulli	*Burkholderiales*	*Comamonadaceae*	5.35	68.5	4782	Nrps-t1pks	65471	***ohmal.ser***
Acidovorax oryzae	*Burkholderiales*	*Comamonadaceae*	5.53	68.7	4832	Nrps-t1pks	62251	***ohmal.ser***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	65174	***ohmal.ser***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***ohmal.ser***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Otherks-nrps-t1pks	76953	***ohmal.ser***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***ohmal.ser***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***ohmal.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps-t1pks	64510	***ohmal.thr***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***ohmal.thr***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***ohmal.val***
Chromobacterium subtsugae	*Neisseriales*	*Chromobacteriaceae*	4.67	64.8	4336	Hglks-nrps-t1pks	77143	***ohmal.val***
Hydrogenophaga intermedia	*Burkholderiales*	*Comamonadaceae*	5.29	68.4	5087	Nrps-t1pks	95561	***ohmmal.redmal***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***orn.pk***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***orn.pk***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***orn.ser***
Cupriavidus taiwanensis	*Burkholderiales*	*Burkholderiaceae*	2.5	67.9	2185	Nrps	67153	***orn.ser***
Nitrosospira multiformis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.18	53.9	2895	Linaridin-nrps-t1pks	67398	***orn.ser***
Acidovorax avenae	*Burkholderiales*	*Comamonadaceae*	5.48	68.8	4752	Nrps-t1pks	65516	***orn.ser***
Acidovorax citrulli	*Burkholderiales*	*Comamonadaceae*	5.35	68.5	4782	Nrps-t1pks	65471	***orn.ser***
Acidovorax oryzae	*Burkholderiales*	*Comamonadaceae*	5.53	68.7	4832	Nrps-t1pks	62251	***orn.ser***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	Nrps	54705	***orn.ser***
Burkholderia caribensis	*Burkholderiales*	*Burkholderiaceae*	9.42	62.5	9112	Nrps	54590	***orn.ser***
Burkholderia cenocepacia	*Burkholderiales*	*Burkholderiaceae*	3.48	66.8	3165	Nrps	54657	***orn.ser***
Burkholderia cepacia	*Burkholderiales*	*Burkholderiaceae*	8.95	65.9	7655	Nrps	54657	***orn.ser***
Burkholderia cepacia ATCC 25416	*Burkholderiales*	*Alcaligenaceae*	8.61	66.6	772	Nrps	54696	***orn.ser***
Burkholderia fungorum	*Burkholderiales*	*Burkholderiaceae*	8.7	61.8	7333	Nrps	54960	***orn.ser***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	47189	***orn.ser***
Burkholderia lata	*Rhodocyclales*	*Rhodocyclaceae*	3.69	66.2	3369	Nrps	54660	***orn.ser***
Burkholderia mallei	*Burkholderiales*	*Burkholderiaceae*	3.5	68.1	3531	Nrps	55140	***orn.ser***
Burkholderia multivorans	*Burkholderiales*	*Burkholderiaceae*	3.45	66.9	3224	Nrps	54645	***orn.ser***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	4.07	67.7	3529	Nrps	55140	***orn.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps	11399	***orn.ser***
Burkholderia terrae	*Burkholderiales*	*Burkholderiaceae*	11.29	61.8	1029	Nrps	37155	***orn.ser***
Burkholderia thailandensis	*Burkholderiales*	*Burkholderiaceae*	3.81	67.3	3343	Nrps	55230	***orn.ser***
Burkholderia vietnamiensis	*Burkholderiales*	*Burkholderiaceae*	3.65	66.5	3326	Nrps	54723	***orn.ser***
Burkholderia zhejiangensis	*Burkholderiales*	*Burkholderiaceae*	7.77	62.7	7148	Arylpolyene	21238	***orn.ser***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	54724	***orn.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Cupriavidus taiwanensis	*Burkholderiales*	*Burkholderiaceae*	2.5	67.9	2185	Nrps	67153	***orn.ser***
Herbaspirillum seropedicae	*Burkholderiales*	*Oxalobacteraceae*	5.51	63.4	4809	Nrps	54897	***orn.ser***
Mitsuaria sp	*Burkholderiales*	*Unclassified*	6.66	67.8	???	Nrps	7095	***orn.ser***
Mitsuaria sp	*Burkholderiales*	*Unclassified*	6.66	67.8	???	Nrps	28988	***orn.ser***
Nitrosospira multiformis	*Nitrosomonadales*	*Nitrosomonadaceae*	3.18	53.9	2895	Linaridin-nrps-t1pks	67398	***orn.ser***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	65174	***orn.ser***
Acidovorax avenae	*Burkholderiales*	*Comamonadaceae*	5.48	68.8	4752	Nrps-t1pks	65516	***orn.thr***
Acidovorax citrulli	*Burkholderiales*	*Comamonadaceae*	5.35	68.5	4782	Nrps-t1pks	65471	***orn.thr***
Acidovorax oryzae	*Burkholderiales*	*Comamonadaceae*	5.53	68.7	4832	Nrps-t1pks	62251	***orn.thr***
Variovorax paradoxus	*Burkholderiales*	*Comamonadaceae*	5.63	67.6	5267	Nrps-t1pks	65174	***orn.thr***
Chitiniphilus shinanonensis	*Neisseriales*	*Chromobacteriaceae*	4.15	66.4	3647	Nrps-t1pks	48692	***orn.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***orn.thr***
Pandoraea pulmonicola	*Burkholderiales*	*Burkholderiaceae*	5.87	64.3	5020	Nrps	85645	***orn.thr***
Pandoraea sputorum	*Burkholderiales*	*Burkholderiaceae*	5.75	62.8	5044	Nrps	80996	***orn.thr***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	63312	***phe.pro***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps	63420	***phe.pro***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***phe.pro***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	63312	***phe.val***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps	63420	***phe.val***
Burkholderia ambifaria	*Burkholderiales*	*Burkholderiaceae*	3.56	66.9	3249	T1pks	63874	***pk.pk***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***pk.redmal***
Achromobacter xylosoxidans	*Burkholderiales*	*Alcaligenaceae*	7.01	66	6466	Nrps	70902	***pk.ser***
Delftia acidovorans	*Burkholderiales*	*Comamonadaceae*	6.77	66.5	6153	Nrps-t1pks	78539	***pk.ser***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***pk.ser***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***pro.pro***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps	63420	***pro.ser***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	29275	***pro.ser***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	43612	***pro.ser***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***pro.ser***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	63312	***pro.thr***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps	63420	***pro.thr***
Collimonas fungivorans	*Burkholderiales*	*Oxalobacteraceae*	5.19	59.6	4628	Nrps	66115	***pro.thr***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	43612	***pro.val***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***redmal.redmal***
Methyloversatilis sp	*Rhodocyclales*	*Rhodocyclaceae*	4.16	65.6	3763	Nrps-t1pks	102184	***redmal.val***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	83189	***ser.ser***
Burkholderia glumae	*Burkholderiales*	*Burkholderiaceae*	2.83	68.8	2286	Nrps-t1pks	107562	***ser.ser***
Burkholderia pyrrocinia	*Burkholderiales*	*Burkholderiaceae*	8.05	67.4	7218	Nrps-bacteriocin	70214	***ser.ser***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	28459	***ser.thr***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps	6200	***ser.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***ser.thr***
Burkholderia ferrariae	*Burkholderiales*	*Burkholderiaceae*	7.94	64.8	6254	Nrps	9020	***ser.thr***
Herbaspirillum seropedicae	*Burkholderiales*	*Oxalobacteraceae*	5.51	63.4	4809	Nrps	54897	***ser.thr***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-hserlactone	79145	***ser.tyr***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***ser.tyr***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	67282	***ser.val***
Chitinimonas koreensis	*Burkholderiales*	*Burkholderiaceae*	5.57	68.9	4534	Nrps-hserlactone	79145	***ser.val***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	*Order*	*Family*	*Size*	*GC%*	*Gene*	*Type*	*BGC Size*	*PAIRS*
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	59387	***ser.val***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps-t1pks	57263	***ser.val***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	51638	***ser.val***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Terpene	21777	***ser.val***
Pseudoduganella violaceinigra	*Burkholderiales*	*Oxalobacteraceae*	6.1	63	5697	Indole-transatpks-nrps	86260	***ser.val***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***ser.val***
Burkholderia bryophila	*Burkholderiales*	*Burkholderiaceae*	7.38	61.9	6552	Nrps	57223	***thr.thr***
Burkholderia gladioli	*Burkholderiales*	*Burkholderiaceae*	3.7	68.6	3006	Nrps	56949	***thr.thr***
Burkholderia graminis	*Burkholderiales*	*Burkholderiaceae*	7.48	62.9	6788	Nrps	61960	***thr.thr***
Burkholderia kururiensis	*Burkholderiales*	*Burkholderiaceae*	7.13	65	626	Nrps	64662	***thr.thr***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***thr.thr***
Burkholderia pseudomallei	*Burkholderiales*	*Burkholderiaceae*	3.17	68.5	2406	Nrps	81475	***thr.thr***
Burkholderia sp	*Burkholderiales*	*Burkholderiaceae*	4.08	63.4	3543	Nrps	67079	***thr.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***thr.thr***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***thr.thr***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***thr.tyr***
Chromobacterium haemolyticum	*Neisseriales*	*Chromobacteriaceae*	5.08	62.8	4571	Nrps	47972	***thr.tyr***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***thr.tyr***
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	59387	***thr.val***
Burkholderia oklahomensis	*Burkholderiales*	*Burkholderiaceae*	4.13	67	6427	Nrps	86121	***tyr.thr***
Janthinobacterium agaricidamnosum	*Burkholderiales*	*Oxalobacteraceae*	5.95	61.1	5184	Nrps	72606	***tyr.thr***
Ralstonia solanacearum	*Burkholderiales*	*Burkholderiaceae*	2.09	66.9	1683	Nrps-t1pks	78530	***tyr.thr***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***tyr.tyr***

Table 5. POPs table containing all information about Microorganisms, Taxonomy, Genome size and Percentage og GC, Gene counting, Typer of Compounds according to their biosynthetic pathway, the size of the BGC, and the expressed pair for each. The table is organized alphabetically according to PAIRS order.

(continues…continuation…conclusion)

Microorganism	***Order***	***Family***	***Size***	***GC%***	***Gene***	***Type***	***BGC Size***	***PAIRS***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	56865	***tyr.val***
Burkholderia glathei	*Burkholderiales*	*Burkholderiaceae*	8.64	64.4	7661	Nrps	53315	***tyr.val***
Andreprevotia chitinilytica	*Neisseriales*	*Chromobacteriaceae*	5.15	59.9	4500	Nrps	28459	***val.thr***
Chromobacterium violaceum	*Neisseriales*	*Chromobacteriaceae*	4.75	64.8	4378	Nrps	59387	***val.thr***
Collimonas arenae	*Burkholderiales*	*Oxalobacteraceae*	5.6	56.1	4971	Terpene	21777	***val.thr***
Delftia sp	*Burkholderiales*	*Comamonadaceae*	6.69	66.7	6029	Nrps-t1pks	84632	***val.thr***
Burkholderia rhizoxinica	*Burkholderiales*	*Burkholderiaceae*	2.76	61.2	2447	Nrps	74221	***val.val***
Glomeribacter sp. 1016415	*Burkholderiales*	*Burkholderiaceae*	2.52	48.8	2151	Nrps	46500	***val.val***