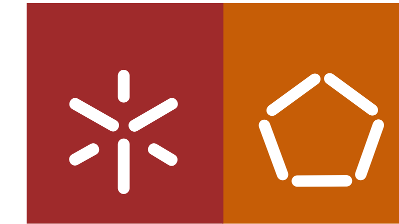
Characterization of gene or gene clusters responsible for the production of antimicrobial compounds in *Pseudoalteromonas atlantica*

Francisco Duarte da Cunha Ventura

UMinho | 2014

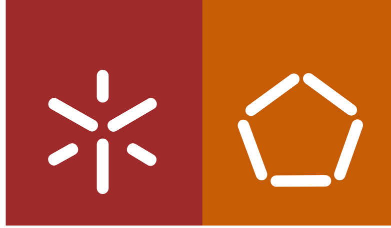


Universidade do Minho
Escola de Engenharia

Francisco Duarte da Cunha Ventura

Characterization of gene or gene clusters responsible for the production of antimicrobial compounds in *Pseudoalteromonas atlantica*

October 2014



Universidade do Minho
Escola de Engenharia

Francisco Duarte da Cunha Ventura

Characterization of gene or gene clusters responsible for the production of antimicrobial compounds in *Pseudoalteromonas atlantica*

Dissertation for the M.Sc. degree in Biomedical Engineering

Supervisors:
Maria João Vieira, PhD
Carina Almeida, PhD

October 2014

Nome: Francisco Duarte da Cunha Ventura
Endereço Electrónico: franciscoduarteventura@gmail.com
Cartão de Cidadão: 13973618
Título da dissertação: Characterization of gene or gene clusters responsible for the production of antimicrobial compounds in *Pseudoalteromonas atlantica*
Orientadora: Professora Doutora Maria João Vieira
Co-orientadora: Doutora Carina Almeida
Ano de Conclusão: 2014
Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica – Ramo Engenharia Clínica

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE/TRABALHO, APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, ___ / ___ / ______

Assinatura: ____

Acknowledgements

I would like to thank:

Professor Maria João Vieira for having me on board.

Carina Almeida, for all the supervision throughout the entire work. You have gone way beyond your duty. For always pointing me in the right direction, for all the time I made you lose, for the great support in the laboratory. I can't thank you enough! It's been an honor and a privilege to work with you for the past year!

Abstract

The intensive and unconscious use of antibiotics alongside with the decrease of investment in research of novel molecules (since the mid 1960s), has led to a rise of highly resistant bacteria. These microorganisms have been developing mechanisms that enable them to survive to the aggressions of classical antibiotics. Also, these self-defense mechanisms are easily transmitted between bacteria, which is a worrisome panorama.

It is necessary to get back to a proactive fight against these microorganisms. Living organisms have long proven to be a rich source for antimicrobial compounds due to their need to fight for their place in ecological niches. The marine environment is known to be prolific with microbial communities and thus a great diversity of bioactive compounds was already discovered.

In this thesis work, I searched for novel antimicrobial compounds in a marine bacteria, *Pseudoalteromonas atlantica*. A genome mining approach was followed. A search for clusters coding for secondary metabolites with antimicrobial characteristics was done using softwares such as AntiSmash, ClustScan, HHpred or BLASTx. This process resulted in the finding of 17 putative clusters coding for polyketide synthase and also 1 putative cluster coding for a bacteriocin.

To assess if *P. atlantica* is, in fact, capable of producing antimicrobial compounds, and in the positive case, to further enhance the production of such

compounds, a compound production optimization procedure was performed. Optimal culture conditions for production of antibiotic compounds in *P. atlantica* were met for Marine Broth culture medium at a temperature of 23ºC, a pH of 8 and a 120 rpm agitation. Bacteria-free spent media proved to inhibit the growth of *Escherichia coli* K12. Moreover, a bacteria-free spent media from cultures grown in the presence of a competitor (*E. coli K12*; *Staphylococcus aureus*; *Pseudomonas aeruginosa* or *Vibrio harveyi*) has also resulted in the inhibition of *Salmonella enteritidis*.

These results indicate that the *P. atlantica* genome might be a source for novel antimicrobial compounds. In fact, under the described conditions, *P. atlantica* was capable of producing antimicrobial molecules with a narrow activity spectrum.

Sumário

O uso intensivo e inconsciente de antibióticos e a quebra do investimento na investigação de novas moléculas (desde meados dos ano 60), levou ao aparecimento de bactérias altamente resistentes. Estes microorganismos têm desenvolvido mecanismos que os permite sobreviver às agressões provocadas pelos antibióticos clássicos. Além disso, estes mecanismos de auto-defesa são facilmente transmitidos entre diferentes espécies de bactérias, o que é um cenário muito preocupante.

É então necessário voltar a uma atitude proactiva na luta contra estes microorganismos. Os microorganismos já há muito provaram ser uma fonte rica em compostos antimicrobianos, devido à sua necessidade de lutar por um lugar nos respectivos nichos ecológicos. O ambiente marinho é conhecido por ser fértil em comunidades microbianas e portanto uma grande diversidade de compostos bioactivos foram já descobertos.

Nesta tese procurei por compostos antimicrobianos numa bactéria marinha, a *Pseudoalteromonas atlantica*. Seguiu-se uma estratégia de “genome mining”. Foi realizada uma procura de clusters de metabolitos secundários de cariz antimicrobiano, através do recurso a ferramentas como o AntiSmash, ClustScan, HHpred ou BLASTx. Todo este processo resultou na descoberta de 17 putativos clusters de poliquétidos sintetases e 1 putativo cluster de bacteriocina.

De forma a avaliar se a *P. atlantica* é capaz de produzir compostos antimicrobianos, e, em caso positivo, para aumentar a produção desses compostos, foi realizado um procedimento de optimização da produção do antimicrobiano. As condições óptimas para a produção de compostos antimicrobianos em *P. atlantica* revelaram ser o uso de meio marinho (Marine Broth), a uma temperatura de 23ºC, um pH de 8 e uma agitação de 120 rpm. O meio de cultura gasto desprovido de bactérias provou inibir o crescimento de *Escherichia coli* K12. Além disso, na presença de um competidor (*E. coli K12*; *Staphylococcus aureus*; *Pseudomonas aeruginosa* or *Vibrio harveyi*), o meio de cultura de *P. atlantica*, também resultou na inibição de *Salmonella enteritidis*.

Estes resultados indicam que o genoma da *P. atlantica* pode ser uma fonte de compostos antimicrobianos inauditos. De facto, nas condições referidas no presente trabalho, a *P. atlantica* foi capaz de produzir moléculas antimicrobianas, com um espectro de actividade aparentemente muito específico.

Table of contents

Acknowledgements iii
Abstract iv

Chapter I - Introduction: Antibiotics

1. Overview 3
2. A brief history 4
3. Mechanisms of action 6
4. The urgent need for new antimicrobials 10
5. Nature: A source for novel antimicrobials 13
   - 5.1. Polyketides 18
   - 5.2. Nonribosomal peptides 23
   - 5.3. Bacteriocins 25
6. Bioinformatic tools 27

Chapter II - Bioinformatic research

1. Overview 33
2. Identification of interesting regions 34
3. Clusters annotation and proteins identification 42
4. Discussion 51

Chapter III - Optimization of antimicrobial compounds production

1. Overview ... 57
2. Materials and methods ... 58
   - 2.1. Strains and cultures conditions optimization ... 58
     - 2.1.1. Effect of the growth medium ... 59
     - 2.1.2. Temperature, pH and agitation ... 60
     - 2.1.3. Addition of extracellular ATP ... 61
     - 2.1.4. Competitors ... 62
   - 2.2. Assessment of antimicrobial activity ... 64
   - 2.3. Determination of arbitrary units ... 66
3. Results and discussion ... 67
   - 3.1. Effect of culture media on cell growth and antimicrobial activity ... 67
   - 3.2. Effect of temperature, pH and agitation on cell growth and antimicrobial activity ... 70
   - 3.3. Effect of addition of extracelullar ATP ... 75
   - 3.4. Effect of competitors ... 76

Chapter IV - General conclusions and future work

1. General conclusions and future work ... 83

Chapter V - References and appendix

1. References ... 89
2. Appendix ... 96

List of figures and tables

Table 1 - Main antibiotic families and mechanisms of action. From Levy S. and Marshall B. (2004).Abstract 6

Figure 1 - New antibacterial agents approved in the US per 5-year period from 1983 to 2002. Adapted from Shlaes, D. et al (2004). 11

Figure 2 - Compounds with new antibacterial templates divided into development phases and their lead derivation source. Adapted from Butler M. and Cooper M. (2011). 13

Table 2 - Bioactivity of *Pseudoalteromonas* species. Adapted from Bowman (2007) 14

Figure 3 - Example of type I PKS (lovastatin), type II PKS (doxorubicin) and type III PKS (naringenin chalcone) From Hertweck (2009). 20

Table 3 - Examples of Polyketides and their bioactivity. Adapted from Pfeifer and Khosla (2001). 21

Figure 4 - Repartition of six main biological activities displayed by curated peptides in the Norine database (790 NRPs). From Caboche S. et al (2010). 24

Figure 5 - Parameters selected in AntiSmash. 34

Table 4 - Position of all clusters detected by AntiSmash 35

Figure 6 - Screenshot of ClustScan workspace. 36

Figure 7 - Script used for detection of nearby clusters 37

Table 5 - Position of all PKS genes and their main specifications given by ClustScan 38

Table 6 - Position of clusters detected in AntiSmash, ClustScan similarity and ClustScan E-value. 40

Figure 8 - Screenshot of CLC Sequence Viewer workspace. 42

Figure 9 - Screenshot of AntiSmash annotations. 43

Figure 10 - Screenshot of InterPro search. 44

Table 7 - PKS related ORFs detected near AntiSmash detection sites 45

FIgure 11 - Schematic representation of a hypothetical Polyketide Synthase in *P. atlantica* T6c composed by 4 distinct domains: AT - acyltransferase ; KS - ketosynthase ; KR - ketoreductase 47

Figure 12 - Graphic of homologous gene clusters, for PKS #8 cluster - AntiSmash. 48

Figure 13 - Schematic representation of a hypothetical bacteriocin in *P. atlantica* T6c. 49

Figure 14 - Graphic of homologous gene clusters, for bacteriocin cluster - AntiSmash. 50

Figure 15 - Representation of the culture conditions used for the competition assays. 62

Table 8 - Effect of culture media in P. atlantica growth and the production of inhibitory compounds. 68

Figure 16 - Inhibition halos in an *E.coli* K12 lawn. 69

Table 9 - Effect of culture media in the production of inhibitory compounds in *P. atlantica*. 70

Table 10 - Effect of pH on the production of inhibitory compounds in *P. atlantica*. 71

Table 11 - Effect of agitation on the production of inhibitory compounds in *P. atlantica*. 72

Figure 17 - Inhibitory effect of the filtered sterilized spent MB of *P. atlantica* in E. coli K12 lawns over time (A) **-** Arbitrary units over time (B) - Inhibition halos on E. coli K12 lawns, recorded after several dilutions of the spent medium. 73

Table 12 - Inhibition of targets by bacteria-free spent MB for optimized culture conditions. 74

Table 13 - Effect of competing species (E. coli K12, S.aureus, P. aeruginosa, V. harveyi) in dialysis membrane on production of antimicrobial compounds by P. atlantica. 77

Chapter I
Introduction: Antibiotics

1. Overview

Antibiotics are compounds that can either kill (bactericidals), or inhibit the growth (bacteriostatics) of bacteria. This categorization between bactericidals and bacteriostatics isn't as obvious as it may appear, since it depends on the drug type, concentration and on the bacterial species. They can be of natural source, when they are for instance produced by living organisms. Also, they can be produced by chemical synthesis or derive from a biological source.

Since their discovery, antibiotics have revolutionized medicine in many ways. Countless lives were saved by these compounds. Nevertheless, the use of antibiotics has been accompanied by a growing number of resistant microorganisms. It is feared that because of this, we may face a new era, like the preantibiotic one.

The study of antibiotics has had its ups and downs all over the years, but it has become of major interest in the last decade. It is important to understand how this all began, in order to understand where we stand nowadays.

2. A brief history

Sir Alexander Fleming set the beginning of antibiotics timeline. In 1929 Sir Fleming was conducting experiments with *Staphylococcus* variants. Plates containing *Staphylococcus* were left in contact with air and, by chance, they became contaminated by a fungus. Sir Fleming noticed that the areas around the mold became transparent, *Staphylococcus* were undergoing lysis. He concluded that the contaminant was a *Penicillum* mold that was producing a powerful bactericidal, the penicillin [1].

In 1940, years before the use of penicillin as a therapy, a bacterial penicillinase was identified. After its mass production, several strains showed the capacity to resist to penicillin. Just then, experiments were conducted so that the penicillin could be modified chemically, in order to prevent cleavage by penicillinases (β-lactamases) [2].

By 1944 Waksman, Feldman and Hinshaw discovered the streptomycin from *Streptomyces griseus*, a bacteria commonly found in the soil [2]. They found that streptomycin was effective against virulent human tubercle bacilli. This discovery led to a growing interest in soil bacteria, which later proved to be the main resource for the discovery of several antibiotics. Streptomycin was then used to cure tuberculosis, but in the mean time, strains of *Mycobacterium tuberculosis* resistant to the antibiotic arose during patient treatment [2].

As in penicillin and streptomycin, many powerful antibiotics were discovered or synthesized throughout the years, but then again, many resistant

strains appeared through a process that remained unknown until the mid 1950s [3]. In Japan, a country devastated by war at the time, an epidemic of *Shigella dysenteriae* turned to be particularly hard to irradiate, due to the growing number of resistant strains. Sulfonamide was no longer effective in 80% of the cases [3]. Kitamoto wrote about *S. dysenteriae* strains that could resist to four different antibiotics. Later on, it was proved that the resistance of these strains could be transferred to other *Enterobacteriaceae* simply requiring a cell-to-cell contact, indicating that bacterial conjugation was involved in the process [3].

In the mid 80s, Michael Syvanen proved that the uniformity of the gene code among nature would allow organisms to use genes transposed from organisms of different species (Horizontal Gene Transfer). This revealed to be of particular interest when one tries to understand the rapid bursts in the evolution of organisms, also helping to understand the way which pathogens gained resistance to several antibiotics [2].

Since Sir Fleming found penicillin in 1929 till now, the development of antibiotics has come up through different eras. It is safe to say that the majority of the antibiotics used nowadays derive from the ones discovered during the so called golden era, which took place between the 50s and the mid 60s. From then on came the lean years, or the innovation gap. A gap that lasted until 2000, when a decrease of the discovery rate was observed. The main approach for the development of novel drugs has been the modification of the molecules previously known [2].

3. Mechanisms of action

Antibiotics can be classified according to a different range of settings, such as their spectrum of activity, their chemical structure, but most commonly, antibiotics are categorized by their mechanism of action (Table 1). Each category, class or family aims to destroy or defunctionalize essential physiological or metabolic targets of the bacterial cell [4].

Table 1 - Main antibiotic families and mechanisms of action. From Levy S. and Marshall B. (2004).

Mechanism of action	Antibiotic families
Inhibition of cell wall synthesis	Penicillins; cephalosporins; daptomycin; monobactams; glycopeptides
Inhibition of protein synthesis	Tetracyclines; aminoglycosides; oxazolidonones; streptogramins; ketolides; macrolides; lincosamides
Inhibition of DNA synthesis	Fluoroquinolones
Competitive inhibition of folic acid synthesis	Sulfonamides; trimethroprim
Inhibition of RNA synthesis	Rifampin
Other	Metronidazole

- **Inhibition of cell wall synthesis**

Antibiotics that inhibit the cell wall synthesis act by interfering with the formation of the peptidoglycan wall present in bacteria. The process by which that occurs is what distinguishes the different families belonging to this mechanism of action. For example, β-Lactams like penicillins, carbapenems and cephalosporins act by inhibiting transpeptidase enzymes, whose function is to cross-link peptidoglycan chains that compose the bacterial cell wall, causing it to lyse [5].

Another way for an antibiotic to affect cell wall integrity is by binding with peptidoglycan units and by blocking transglycosylase (enzyme that adds disaccharide pentapeptides to extend the glycan strands of existing peptidoglycan molecules) and also blocking transpeptidase activity [5]. That is the case of a glycopeptides, such as vancomycin. Nevertheless, this family of antibiotics is ineffective against Gram-negative bacteria, due to their lower permeability, by contrast with β-Lactams, which are effective against both Gram-positive and Gram-negative bacteria. There are other ways to inhibit the cell wall synthesis or the cell wall integrity, besides the ones used by β-Lactams and glycopeptides, like affecting the transport of individual peptidoglycan (e.g. Bacitracin), for example [5].

- **Inhibition of protein synthesis**

Translation is a crucial process in biological lifeforms. It is a series of processes by which information coded in RNA is used to create proteins constituted by aminoacids. One of these processes includes the presence of a very important organelle - the ribosome. This organelle is divided into two

subunits. In the case of prokaryotes, there is a small subunit (**30s** - responsible for reading the mRNA) and the big subunit (**50s** - responsible for joining amino acids to the growing peptide chain). Antibiotics that act through this mechanism (inhibition of protein synthesis) are molecules that act by compromising the function of one of these two subunits. Therefore, they can be divided into 30s inhibitors or 50s inhibitors [6].

The 30s inhibitors like the tetracyclines act by compromising the function of the small ribosome subunit. Tetracycline, for example, blocks the aminoacyl-tRNA to the ribosome and therefore stops protein synthesis [5].

Antibiotics belonging to macrolides, lincosamides or oxazolidonones families are considered to be 50s inhibitors, since they act upon the big subunit of the bacterial ribosome. They stop protein synthesis by either blocking initiation or elongation of the translation process. Also, some act by hindering translocation of peptidyl-tRNA. Peptidyl is one of the binding sites in the ribosome for tRNA [5].

- **Inhibition of DNA synthesis**

Deoxyribonucleic acid or just DNA is an essential molecule of all living organisms. It would be correct to assume that a living organism would be killed if its DNA synthesis stopped. That is just how quinolones, a family of synthetic broad-spectrum antibiotics, work. Fluoroquinolone (a quinolone with a fluorine atom attached to the central ring system) targets the function of DNA-topoisomerase and DNA-gyrase complexes [5]. Topoisomerases and gyrases are enzymes that participate in DNA replication. By canceling their function, fluoroquinolones lead to bacterial cell death. Thus, they're considered

bactericidal. There are other families belonging to this mechanism of action and they target different components of the bacterial DNA machinery [5].

- **Competitive inhibition of folic acid synthesis**

Unlike humans, bacteria do not acquire folic acid through diet. Instead they synthesize it. It is by affecting this mechanism that antibiotics like the sulfonamide work. All cells require folate cofactors for the biosynthesis of diverse cellular components, like the formation of the so important amino acid methionine, crucial for starting protein translation. Sulfone inhibitors work by being analog of para-aminobenzoic acid (pABA), which is a required intermediate of bacterial synthesis of folate. Sulfonamides act as alternative substrates for dehydropteroate synthase, an enzyme involved in the folate pathway. By doing so, they inhibit the growth of the target microorganism, therefore they are considered to be bacteriostatical antibiotics [7].

- **Inhibition of RNA synthesis**

Like the DNA, ribonucleic acid or RNA plays an important role in expression of genes. Inhibition of RNA synthesis is a catastrophic event for prokaryotic nucleic acid metabolism. Drugs belonging to the rifamycin group inhibit or affect the normal behavior of RNA within bacterial cells. Rifampin, for example, hinders DNA transcription by inhibiting RNA polymerase, leading, in most cases, bacterial pathogens to death [5].

4. The urgent need for new antimicrobials

As stated by Hiroshi Nikaido in 2004, about 100 000 tons of antibiotics are produced per year [8]. Their intensive use has had a tremendous impact in worldwide microbial resistance. Because of this, many antibiotics that were effective many years ago, no longer produce a harmful effect in the same pathogens. Methicilin-resistant Staphylococcus aureus (MRSA) is a glaring example of a dangerous pathogen. It is resistant to methicilin and a wide range of antibiotics like aminoglycosides, macrolides, tetracycline, chloramphenicol, and lincosamides [8]. Since 2002 some strains of MRSA have also proved to be resistant to vancomicyn (the main strategy of eliminating MRSA until then) [9]. The rise of multidrug resistant bacteria, the scarcity of new classes of antibacterial drugs (Fig. 1) and the stagnation in antibiotics discovery is demanding new ways of controlling these very resistant pathogens [10, 11].

Alongside the heavy health burden of this problem, also comes an economic burden that hasn't been correctly measured yet. As stated by WHO, "All costs for infections caused by resistant strains were consistently greater than those for infections caused by susceptible strains." That is mainly due not only to the increase in the intensity of care needed by infected patients, but also to an increase in the length of stay of those patients in the hospitals. Moreover, as expected, infections with resistant strains are associated with worse clinical outcomes [11].

The laborious legal procedures to approve new antimicrobials, especially in the United States, have delayed the introduction of new antibiotics and large pharmaceutical companies are fending themselves from this area of investigation [10]. This obviously results in a declining number of new antimicrobials in development. Although all of this may appear to be a dantesque scenario, there are some good news as well. State of the art techniques, like genomics, proteomics, structure based design, high-throughput screening and combinatorial chemistry are paving their own path into new antimicrobial compounds [10].

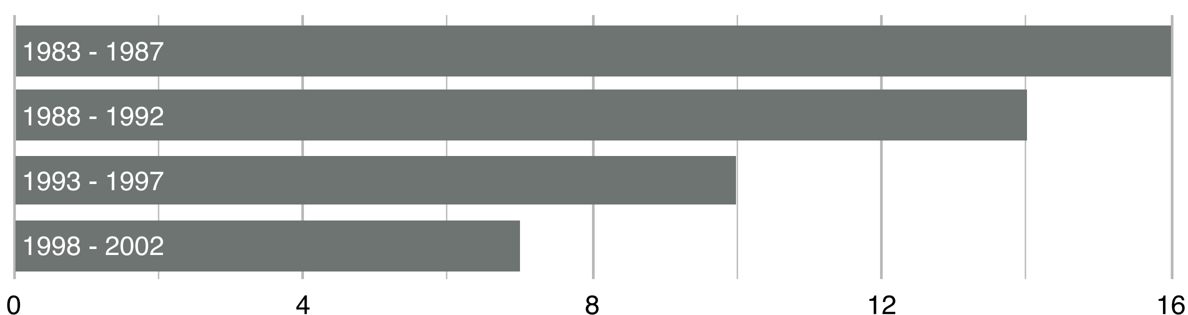


Figure 1 - New antibacterial agents approved in the US per 5-year period from 1983 to 2002. Adapted from Shlaes, D. et al (2004).

5. Nature: A source for novel antimicrobials

Today, there is an understanding that the method for finding new antibiotics used in the last 50 years is no longer effective. While 2nd, 3rd, etc. generations of antibiotics fail to deliver a continuous harm to pathogens along time, some researchers think that natural compounds might be the solution. Many antibiotics discovered until nowadays are secondary metabolites that some microorganisms naturally produce. Thinking that microorganisms have found their way against many pathogens long before they even become a known threat for humans, can be a successful contemplation [12].

Nowadays, efforts are being made in order to find novel antimicrobials from old sources, like *Streptomyces* [13]. Also, many are studying the capacity of other unstudied microorganisms (cyanobacteria or uncultured bacteria) to synthesize antimicrobials compounds. This is proving to be a successful strategy, since several novel scaffolds with antibiotic potential are being discovered. We can also think that many targets within pathogens haven't been explored yet (like the fatty acid synthesis) and if novel scaffolds can act upon these targets, a new era may be arriving. Naturally occurring antibiotics can benefit from the fact that they have features which are not present in libraries of synthetic drugs molecules, like those found via high-throughput screening. It is not surprising that rather than chemically synthesized, the majority of antibiotics in the last stages of clinical trials derive from natural sources (Fig. 2) [13].

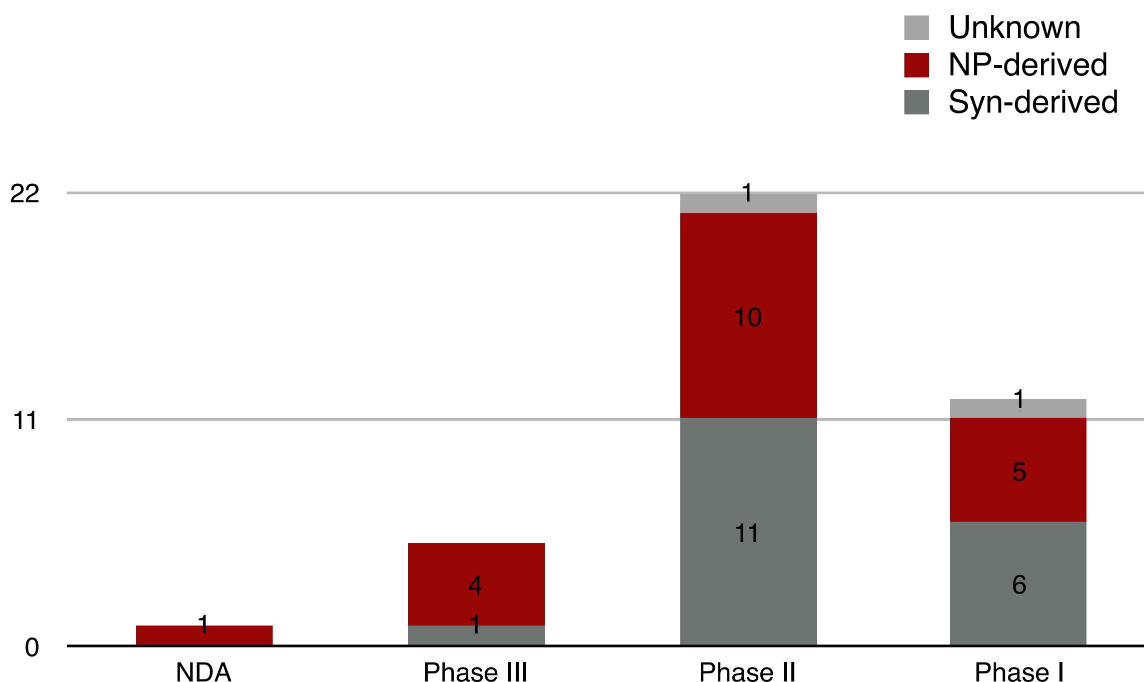


Figure 2 - Compounds with new antibacterial templates divided into development phases and their lead derivation source. Adapted from Butler M. and Cooper M. (2011).

The marine environment is known to be proliferous with microbial communities. Also, as many marine communities are highly specific to particular ecological niches (sponge or algae environments), there is a vast biodiversity among these microbial communities that inhabit seawater. This statement is useful when one tries to understand the implications of natural products discovery, such as naturally produced antibiotics [14].

It is estimated that about 0,5% to 6 % of the oceanic bacterioplancton belongs to the *Pseudoalteromonas* genus. Some species of this genus have

proved to be efficient in producing bioactive compounds, such as antimicrobials, anti-fouling or algicidals compounds [15, 16].

Table 2 summarizes the bioactivity of compounds produced by *Pseudoalteromonas* species. Although many bioactive compounds have already been detected, the synthesis of such compounds are quorum-sensitive. Thus, the ecosystem in which these species grow, influence the biosynthesis of these compounds. As many aspects of ecological networks remain unexplored, there is room for the discovery of novel natural products [15].

Table 2 - Bioactivity of Pseudoalteromonas species. Adapted from Bowman (2007).

Species	Source	Bioactive compounds	Bioactivity
P. aliena	seawater	Unknown compound(s)	Anti-tumorigenic activity - Ehrlich ascites carcinoma cell line inhibited
P. agarivorans	seawater, ascidians	Not Available	degrades algal polysaccharides
P. antarctica	seawater, sea-ice, muddy soils, sediment	Not Available	None observed polysaccharides, cold-active enzymes]
P. atlantica	seawater, marine alga	Not Available	May cause opportunistic disease in crabs [strong degrader of algal polysaccharides]
P. aurantia	surface of *Ulva lactuca*, seawater	Unknown compound(s)	Antimicrobial activity; inhibits settlement of invertebrate larvae
P. byunsanensis	tidal flat sediment	Not Available	—

Species	Source	Bioactive compounds	Bioactivity
P. carrageenovora	seawater, marine alga	Not Available	None observed degrader of algal polysaccharides]
P. citrea	seawater, mussels, ascidians, sponges	Unknown compound(s)	Inhibits settlement of invertebrate larvae; cytotoxic against sea urchin [algal polysaccharide degradation]
P. denitrificans	seawater	High molecular weight polyanionic substance; cycloprodigiosin HCl	Anti-tumorigenic activity; inhibits T-cell/lymphocyte proliferation; anti-malarial activity; induces settlement of sea urchin *Heliocidaris erythrogramma*
P. distincta	sponge	Not Available	
P. elyakovii	mussels, marine alga	Not Available	None observed
P. espejiana	seawater	Not Available	None observed
P. flavipulchra	seawater	Not Available	—
P. haloplanktis	seawater	Novel diketopiperazines	Probiotic benefits to shellfish; cold-active enzymes
P. issachenkonii	marine alga	Isatin; unknown reddish-brown compound	Anti-fungal activity; hemolytic
P. luteoviolacea	seawater, marine alga	Toxic antimicrobial protein; brominated pyrrole-containing compounds, 4-benzaldehyde; *n*-propyl-4-hydroxybenzoate	Antimicrobial activity; inhibits algal spore settlement; cytotoxic against sea urchin *Strongylocentrotus intermedius*; induces settlement of sea urchin *Heliocidaris erythrogramma*
P. maricoloris	sponges	Bromo-alterochromides A and B	Antibacterial activity; cytotoxicity against sea urchins

Species	Source	Bioactive compounds	Bioactivity
P. marina	tidal flat sediment	Not Available	—
P. mariniglutinosa	diatoms	Not Available	—
P. nigrifaciens	seawater, salted foods, mussels	Not Available	—
P. paragorgicola	sponge	Not Available	—
P. peptidolytica	seawater	Unknown compounds	Antimicrobial activity, hemolytic
P. phenolica	seawater	3,3',5,5'-tetra-bromo-2,2-biphenyldiol	Antimicrobial activity
P. piscicida (and related bacteria)	estuarine waters, fish samples	Toxic protein; possible yellow cyclic/acyclic brominated depsipeptide compounds; unknown anti-algal compound(s)	Antibacterial; algicidal activity; possible cytotoxicity [opportunistic fish pathogen; thrombolytic enzymes]
P. rubra	seawater	High molecular weight polyanionic substance; cycloprodigiosin HCl; rubrenoic acids	Antimicrobial activity; anti-tumorigenic activity; inhibits T-cell/lymphocyte proliferation; anti-malarial; bronchodilatatoric
P. ruthenica	shellfish	Unknown compounds	Antimicrobial activity
P. spongiae	sponge	Not Available	Strongly induces settlement of Hydroides elegans
P. tetraodonis	puffer fish	Tetrodotoxin	Neurotoxic effects
P. translucida	seawater	Not Available	—
P. tunicata	marine alga, tunicates	Unknown purple pigment; tambjamine-like alkaloid YP1; toxic protein AlpP; other unknown sunstances	Anti-fungal, anti-algal, antimicrobial, inhibits settlement of invertebrate larvae and algal spores; inhibits protists

Species	Source	Bioactive compounds	Bioactivity
P. ulvae	marine alga	Unknown substances	Inhibits invertebrate larval settlement and algal spore germination and settlement
P. undina	seawater, fish	Not Available	hemolytic; [probiotic benefits; possible opportunistic fish pathogen]

Because there are many strains belonging to the *Pseudoalteromonas* genus, many of them live in the wild under different environmental conditions, which leaves room for a vast spectrum of different bioactive compounds produced by different strains. Although some of these compounds were already described and tested, there is a believe that there is an enormous potential for the discovery of novel compounds [15, 16].

Among the naturally occurring antibiotics, polyketides and non-ribossomal peptides have gained their own status in antimicrobials research and production. Some of them are even known and widely used as antibiotics. Nevertheless, their synthase is yet to be fully understood. Besides, the ability for many microorganisms to produce these compounds is being tested in different species of bacteria and fungus [13].

5.1. Polyketides

Polyketides are small secondary metabolites that microorganisms produce. Their vast diversity in terms of structure and function is well known. They are produced by polyketide synthases, which are enzymatic assembly lines that determine the final structure of the produced polyketide. These metabolites can be of major interest in the clinical area (Table 3). For example, erythromycin A is a potent antibiotic used as a therapy against bacterial infections. Rapamycin is an immunosuppressant used in various surgery techniques. This is to say that polyketides may present very interesting characteristics. Some of them have already been useful, others may help us win the war against pathogens in the future. It is then important to try to understand how these powerful molecules are produced by some microorganisms [17].

Polyketides, are synthesized due to repetitive condensation reactions, in a process that is very similar to the synthesis of fatty acids. In these reactions that link carbon precursors, coenzyme A thioesters play an important role as they constitute the core of the molecule. However, polyketides are found in much more diverse structures than fatty acids. This diversity is of great usefulness since they also present different modes of action, thus can be used in different applications [18].

Polyketide synthases (PKS) are the enzymes of large dimension with specific catalytic domains that catalyze the referred condensations. Their core domains are ketosynthase (KS), acyltransferase (AT) and thiolation (T). PKS are

categorized into three classes (I, II and III) (Fig. 3) which differs slightly from fatty acid nomenclatures [18, 19].

- **Type I -** best exemplified by the PKS responsible for building the backbone of erythromycin (6-deoxyerythronolide B or 6-DEB). These PKS's are constituted by multidomains resembling type I fatty acid synthases (FAS).

- **Type II -** as one would imagine, this PKS resembles type II FAS. The growth of the polyketide is iterative and KS, AT and T domains are re-used during polyketide synthesis.

- **Type III -** They differ from other types by using an acyl-carrier-protein independent mechanism. Besides, they typically lack multiple catalytic domains.

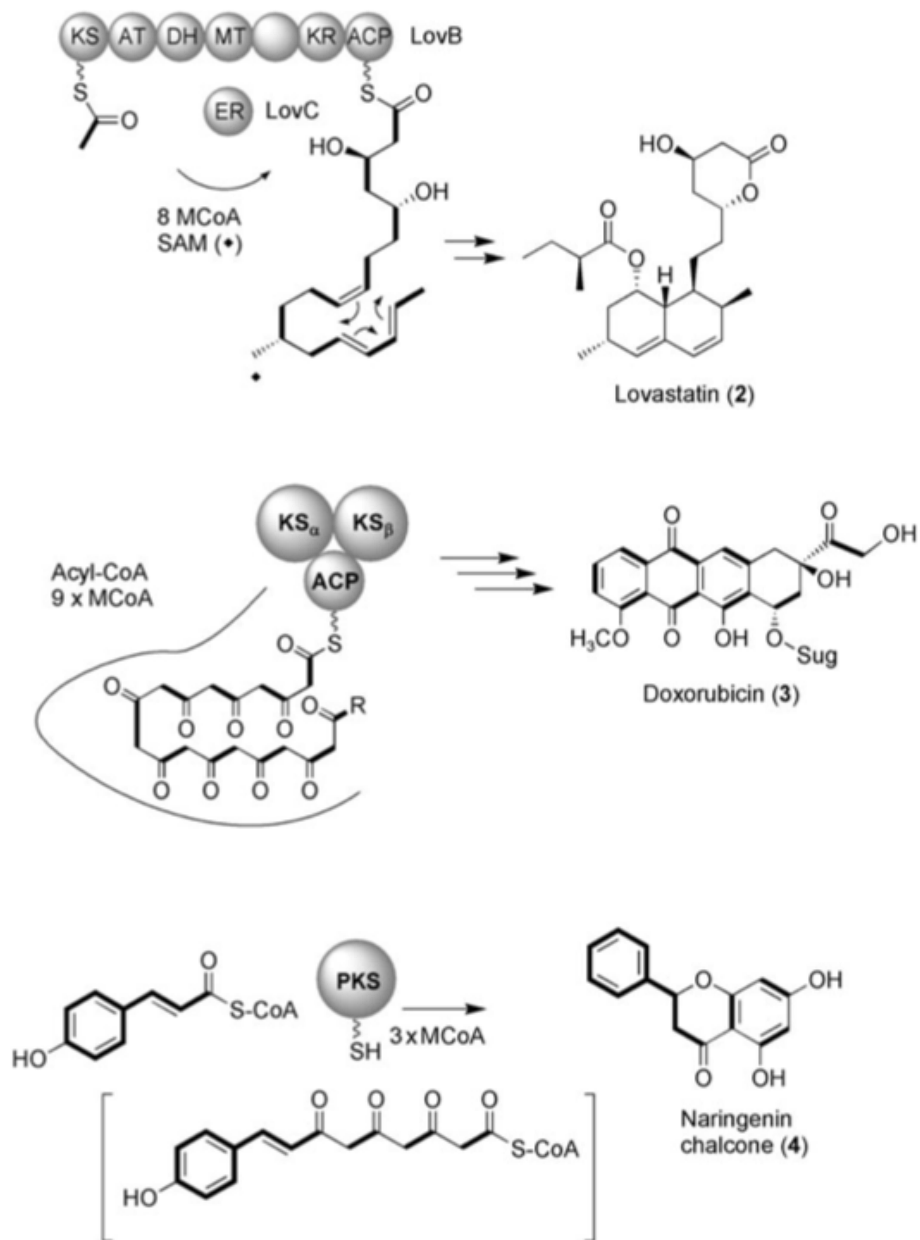


Figure 3 - Example of type I PKS (lovastatin), type II PKS (doxorubicin) and type III PKS (naringenin chalcone) From Hertweck (2009).

Table 3 - Examples of Polyketides and their bioactivity. Adapted from Pfeifer and Khosla (2001).

Polyketide	Bioactivity
Actinorhodin	Antibiotic
Doxorubicin	Antitumor agent
Erytrhomicin A	Antibiotic
Epothilone A	Anti-cancer agent
6-methylsalicilic acid	Antibiotic precursor
Lovastatin	Cholesterol-lowering agent

As polyketides are small metabolites hard to isolate, its correct characterization has been an hard quest for many researchers. Traditional methods used on the identification of polyketides usually involve phenotype screening, followed by the use of analytical chemistry techniques. For instance, Marinho and his colleagues [20] reported the presence of citreorosein (1), emodin (2), janthinone (3), citrinin (4), citrinin H1 (5) and dicitrinoln (6), six known polyketides produced by an endophytic fungi, *Penicillum herquei*. To evaluate the presence of these compounds, they have used classical methods of chromatography. They were identified by 1D and 2D Nuclear Magnetic Resonance spectroscopy (NMR) and Mass Spectrum analysis (MS), results were then compared to previous identifications of the referred compounds [20].

However, phenotype screening approaches are usually very time consuming and depend on the availability of large libraries of organisms. Besides, the production of some metabolites might not be induced under the testing conditions [21] and important compounds might be lost. Actually, the

advances in genomics and genome sequencing have shown that the bacteria potential to produce molecules of pharmacological interest has been greatly underestimated. Nowadays, the development of bioinformatic tools and the increasing update in genome databases are being helpful to finally understand the steps behind the synthesis of these small metabolites. It is now possible to follow a genome mining approach to identify regions with potential interest before proceeding with further laboratorial testing [21].

5.2. Nonribosomal peptides

Nonribosomal peptides are a class of potent antibiotics (like penicillin) and other important pharmaceuticals of great economic interest. These molecules are synthesized by a process that is independent of the ribosome and nucleic-acids, unlike the classical pathway for metabolites synthesis. They are assembled by nonribosomal peptide synthetases (NRPS), which are multimodular megaenzymes. Their biosynthesis relies not only in the 20 canonical amino acids, but also in some different building blocks, such as "d-configured and β-amino acids, methylated, glycosylated and phosphorylated residues, heterocyclic elements and even fatty acid building blocks" [22]. Due to this diversity of building blocks, there are generally a large number of active sites which are essential to the bioactive purposes of these compounds [22].

NRP synthetases are modularly organized enzymes, which comprise multiple catalytic domains. Each module is responsible for adding one amino acid to the peptide and, therefore, their order in the chain influences the final product. The process of adding amino acids to the elongating chain continues until the final molecule is released by a thioesterase domain [23].

Norine is a database entirely dedicated to NRPs, from where it is possible to perform analysis of NRP-related peptides, like predicting functions. Biological activities presented by NRPs mainly range from immunomodulating, iron chelating, antibiotics, toxins, surfactants to anti-tumor (Fig. 4) [24].

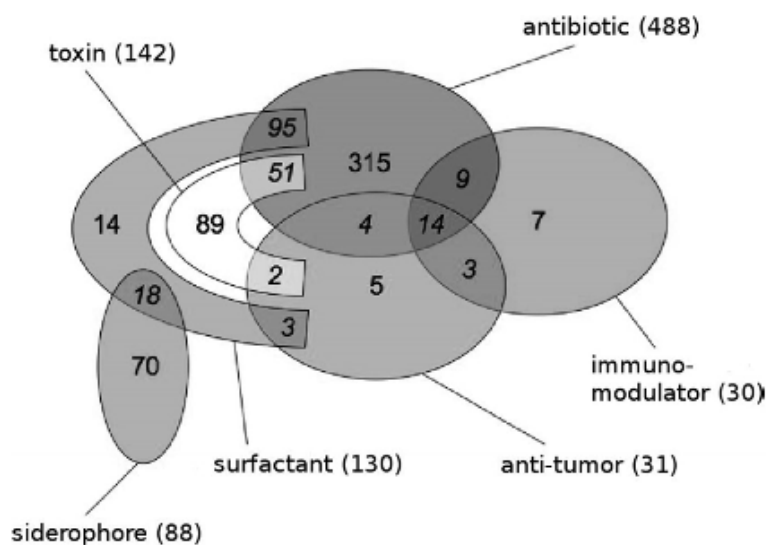


Figure 4 - Repartition of six main biological activities displayed by curated peptides in the Norine database (790 NRPs). From Caboche S. et al (2010).

5.3. Bacteriocins

Bacteriocins are antimicrobial peptides/proteins produced by bacteria that helps the producing bacteria to proliferate within an environment by eliminating other rival bacterial species that compete for the same environment. Unlike polyketides and non ribossomal peptides (NRP's), which are synthesized by non ribossomal pathways, bacteriocins follow the classical ribosomal pathway. Because of this, bacteriocins are structurally different from either PKS and NRP's, and thus some authors do not consider them as antibiotics. Also, unlike traditional antibiotics, bacteriocins usually restrict their activity to related species of the producing bacteria, and particularly to strains of the same species [25, 26].

As bacteriocins may exhibit significant potency against pathogens (some can even compromise antibiotic-resistant strains), they may be seen as a viable alternative to classic antibiotics and so help to solve the multi-resistant pathogens problem [27]. There are already many useful applications of these metabolites in some industries. For instance, bacteriocins are already used in the food industry to prevent the colonization by pathogens [25, 26]. On the other hand, as some pathogens produce their own bacteriocins that help them conquer unwanted places, such as the human nasopharynx (in the case of *Streptococcus pneumoniae*), they may also play a negative role by competing with commensal flora [26-28].

Bacteriocins include a very heterogeneous group of molecules, and so their classification is largely based on their molecular weight differences [29]. Their composition can consist of only 19 amino acids, while large bacteriocins

can have molecular weights up to 90 000 Da. Additionally, some small bacteriocins can also present some pos-translation modifications.

Generally, bacteriocins attack pathogens by compromising their membrane. They act by binding to cell's surface receptors that are recognized by that particular bacteriocin. In a microbial environment, there are usually three types of cells: the bacteriocinogenic (produce bacteriocins); the sensitive ones; the resistant ones. That diversity results in an ecological balance since each type has a stronger and weaker opponent. In spite of bacteriocin producers tend to kill strains belonging to their same species, there are some remarkable exceptions, for example, *E. coli* bacteriocins are proved to eliminate strains of the distant related *Hafnia alvei* [26].

As in polyketides, the growing number of genome databases is helping to improve the understanding of the genes or gene clusters involved in the production of bacteriocins, giving way for a regulation and functionalization of these powerful metabolites.

6. Bioinformatic tools

Although bioinformatics is a cross-disciplinary field that began to emerge in the mid 60's, it was only by the beginning of the last decade that it gained relevant advantage in microbiology. That is due to the massive amount of data generated by genomic research in the 90's. Since then, bioinformatic has proved to be an indispensable field of knowledge in order to store and organize genomic sequences or predict metabolic behaviors, to refer some of the immense applications [30].

One of the most approaches to predict a gene or protein function is data-mining or genome mining. It relies on the fact that different species may present a local similarity in some genes (closely related species present higher similarity). Using this technique, a researcher can search for patterns in databases of different species and if those patterns look alike with patterns that belong to a known sequence, this researcher finds evidence for what that sequence or gene might do [30]. Since 2000's, genome sequencing of various microorganisms has allowed many researchers to identify new genes and pathways for the production of powerful antibiotics. Genome mining of actinomycetes has already resulted in the finding of new natural products, i.e., strambomycins A-D [31].

The list below shortly describes some bioinformatic tools that can be useful for searching specific clusters/genes:

- **BLASTx**: BLAST stands for Basic Local Alignment Search Tools, and particularly, BLASTx searches proteins giving a translated nucleotide query.

- **AntiSmash**: It is a library with secondary metabolites produced by microbes which can be a powerful source of antibiotics and other pharmaceuticals. It uses a genome mining approach for biosynthetic clusters, to do so. Antismash can be used to find putative PKS, bacteriocins, lantibiotics, homo serine lactones or other types of secondary metabolites. AntiSmash is based on profile hidden Markov models (HMM) of genes that are specific for certain types of gene clusters. Profile HMM analysis complements standard pairwise search models used in BLAST, for example, for a better identification of gene clusters encoding secondary metabolites [32, 33].

- **ClustScan**: Or Cluster Scanner is a rapid, semi-automatic software for DNA sequences annotations. The main goal for the use of ClustScan is the discovery of novel biosynthetic gene clusters. ClustScan is particularly efficient in finding regions of a given genome that can possibly code interesting secondary metabolites, such as PKS, NRPS, immuno-supressants, etc. As in AntiSmash, ClustScan also uses HMM analysis to detect more accurately sequence alignments [34].

- **CLC Sequence Viewer**: Among a large amount of features, CLC sequence viewer was used as a genome browser. This software can be used to track

and retrieve sequences of the targets given by other tools, such as ClustScan.

- **HHpred**: It is an interactive server for protein homology. It allows to search for protein homology within a wide choice of databases. Various single queries sequences (retrieved, for instance, from CLC sequence viewer or from NCBI) can be run one by one in HHpred, in order to see similarities between these queries and proteins coded by other organisms.

- **InterPro**: It is a documentation resource for finding protein domains, families and functional sites of proteins. As each InterPro entry includes a description, annotations and relevant literature, it is a powerful tool for predicting a protein's derivatives and its function.

As these tools help in the prediction and understanding of the biosynthesis of interesting metabolites, they are of major relevance when one aims to discover and try to induce the production of novel interesting compounds.

Chapter II
Bioinformatic research

1. Overview

In order to evaluate if *Pseudoalteromonas atlantica* T6c possessed clusters with potential antibiotic activity, a genome mining approach was required. First, *P. atlantica T6c* genome was downloaded in raw format from NCBI database. A search for polyketides synthases (PKS), nonribosomal peptides synthases (NRPS) and bacteriocins was done using AntiSmash (an antibiotics and secondary metabolites analyzer). In addition to AntiSmash, results were confirmed by using a second bioinformatic tool, ClustScan (a powerful tool for scanning clusters).

CLC Sequence Viewer was then used as a resource for clusters annotation. First, each ORF within the genomic regions identified was copied and then pasted into BLASTx (Basic Local Alignment Search Tool), which searches proteins database using a translated nucleotide query. For each ORF within the interest region, its size, name, direction of translation and domains were captured.

Several software tools were used to search each protein's catalytic domains - HHpred, Interpro and BLAST. From all the clusters annotated, only one seems to be a completed conserved PKS, with all the domains present in most of the PKS found in the literature.

2. Identification of interesting regions

A search for PKSs in the genome of *P. atlantica* (accession number: NC_008228) was done using AntiSmash (**ANTI**biotics & **S**econdary **M**etabolite **A**nalysis **SH**ell). Because PKSs regions are usually of a great diversity or variability, in order to detect these clusters it is needed to select “Detect putative genes clusters based on PFAM domain probabilities” in search parameters. All the other parameters were set as default (Fig. 5).

+ Limit prediction to an input region
DNA of Eukaryotic origin
− Restrict which of the 24 supported secondary metabolite types to detect
all 24 cluster types
polyketides (type I) | polyketides (type II) | polyketides (type III)
heterocyst glycolipid-like polyketides | nonribosomal peptides | terpenes
lantibiotics | bacteriocins
beta-lactams | aminoglycosides / aminocyclitols | aminocoumarins
siderophores
ectoines | butyrolactones | indoles
nucleosides | phosphoglycolipids | melanins
oligosaccharides | furans | homoserine lactones
thiopeptides | phenazines | others
Detect putative gene clusters based on PFAM domain probabilities
smCOG analysis for functional prediction and phylogenetic analysis of genes
Gene Cluster Blast analysis
Subcluster Blast analysis
Whole-genome PFAM analysis

Figure 5 - Parameters selected in AntiSmash.

A fasta format file of *P. atlantica* T6c genome was once again retrieved from NCBI database and then loaded to the “Nucleotide input” area. AntiSmash searches resulted in the finding in 17 putative PKS clusters and a cluster coding for a bacteriocin (Table 4). Unfortunately, this effort didn’t result in the finding of any NRP cluster.

To maintain explicitness, bacteriocin cluster will be analyzed in a section ahead.

Table 4 - Position of all clusters detected by AntiSmash

Cluster	From	To
#1	240399	252932
#2	1133011	1138638
#3	1306704	1333830
#4	1591638	1596044
#5	1689023	1702894
#6	1716663	1730405
#7	2316179	2333946
#8	2561287	2572936
#9	2588568	2597466
#10	2799741	2809626
#11	3092399	3106725
#12	3541734	3557681
#13	3692166	3714982
#14	3841002	3874237

#15	3887354	3897050
#16	4408491	4423937
#17	4814303	4820705
#bacteriocin	3506713	3519819

In order to confirm these results and to better identify interesting regions, a similar approach was done using ClustScan. The *Pseudoalteromonas atlantica* T6c genome was imported to ClustScan. Then, a search for PKSs and NRPSs was requested. This action resulted in dozens of putative interesting genes. These genes appear sorted by their position in the genome (Fig. 6).

Relevant information is obtained in a ClustScan search for these genomic regions, such as the E-value (Expect value), which gives a good idea about the significance of a match. DNA and proteins coordinates, the size of the coded proteins and the sequence frame, are also some of the valuable information obtained from ClustScan. ClustScan was used in version 2.0.3.

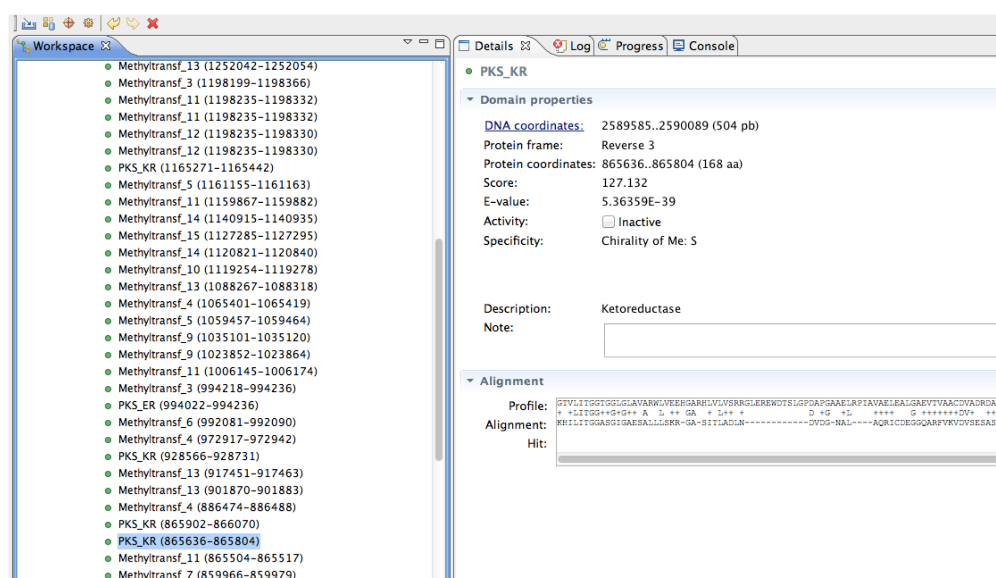


Figure 6 - Screenshot of ClustScan workspace.

DNA coordinates of genes given by this software were copied to lists of a python script (Fig. 7) that detects if coordinates of different genes are located within the same genomic area.

```
forward = [90414,271620,840903,1137894,1627875,1630572,2180463,2828073,3033876,3182778,4641642,4869999,
           3744511076502,2950479,3180975,3461127,4236369,4389030,4552839,5148300,
           252225,799500,2029677,2601018,2638161,3095406,3199524,3434229]
forward.sort()
forward_filtered=[forward[0]]

reverse = [490713,1085154,1595154,1722228,2318139,2570967,3905241,4057665,4595790,5167494,
           550062,881103,2206875,2568561,2569899,2570526,2615082,3312774,3602340,4416003,4522485,4637610,
           731472,1690671,2204289,2400804,2588787,2589585,2745321,3580353,3706116,3707997,5093046]
reverse.sort()
reverse_filtered=[reverse[0]]

for item in xrange(0,len(forward)-1):
    if (forward_filtered[-1]+15000)<forward[item+1]:
        forward_filtered.append(forward[item+1])

for item in xrange(0,len(reverse)-1):
    if (reverse_filtered[-1]+15000)<reverse[item+1]:
        reverse_filtered.append(reverse[item+1])

print len(forward)+len(reverse)
print len(forward_filtered)+len(reverse_filtered)
```

Figure 7 - Script used for detection of nearby clusters

This is important because two different genes don't always mean two different clusters. It just means that these two genes may belong to the same cluster and therefore could participate together in the biosynthesis of a polyketide. Results from running the script showed the presence of 8 superposed regions and therefore resulting in a final count of 53 PKS clusters, as clustscan detected 62 genes putatively belonging to PKS's (Table 5). These regions, where interesting genes are very close to one another are of major interest. Having the modular architecture of PKS's into account, different genes detected by clustscan may represent different modules of the same PKS.

Table 5 - Position of all PKS genes and their main specifications given by ClustScan.

Gene	From	To	E-value	Size (bp)
PKS_TE	90414	91092	5,1E-07	678
PKS_KR	252225	252708	1,1E-20	483
PKS_KR	271620	272130	5,5E-13	510
PKS_KR	374451	374949	6,5E-20	498
PKS_DH	490713	491214	9,9E-04	501
PKS_TE	550062	550713	2,1E-04	651
PKS_KR	731472	731982	2,7E-19	510
PKS_KS	799500	799959	2E-04	459
PKS_ACP	840903	841086	9E-04	183
PKS_KR	881103	881601	1,8E-29	498
PKS_KR	1076502	1077009	2,7E-23	507
PKS_KR	1085154	1085658	2,4E-26	504
PKS_KR	1137894	1138422	3,3E-33	528
PKS_KR	1595154	1595619	5,4E-21	465
PKS_KR	1627875	1628400	2,1E-15	525
PKS_ER	1630572	1631517	5,7E-19	945
PKS_KR	1690671	1691184	4,1E-24	513
PKS_KR	1722228	1722720	8,6E-23	492
PKS_AT	2029677	2030445	6,3E-04	768
PKS_ER	2180463	2181399	8,4E-22	936
PKS_ER	2204289	2204931	1,9E-06	642
PKS_KR	2206875	2207349	2,3E-19	474
PKS_KR	2318139	2318643	9,4E-35	504
PKS_KR	2400804	2401299	7,8E-27	495
PKS_KS	2568561	2569782	2,3E-41	1221

Gene	From	To	E-value	Size (bp)
PKS_ACP	2569899	2570100	5,7E-27	201
PKS_KR	2570526	2571018	4,1E-45	492
PKS_AT	2570967	2571972	5,5E-19	1005
PKS_KR	2588787	2589291	2,3E-33	504
PKS_KR	2589585	2590089	5,4E-39	504
PKS_TE	2601018	2601741	2,4E-05	723
PKS_ER	2615082	2616027	8,5E-20	945
PKS_KR	2638161	2638674	7,5E-23	513
PKS_AT	2745321	2746107	2,6E-04	786
PKS_KR	2828073	2828580	2,6E-32	507
PKS_TE	2950479	2951073	5E-04	594
PKS_KR	3033876	3034380	2E-10	504
PKS_KR	3095406	3095910	3,5E-28	504
PKS_DH	3180975	3181362	1,8E-04	387
PKS_KR	3182778	3183207	1,2E-10	429
PKS_ER	3199524	3200481	9,7E-15	957
PKS_KR	3312774	3313329	6,9E-24	555
PKS_ACP	3434229	3434406	9,5E-04	177
PKS_ER	3461127	3462114	3,9E-16	987
PKS_KR	3580353	3580860	6,4E-36	507
PKS_KR	3602340	3602886	4,6E-10	546
PKS_KR	3706116	3706611	2,9E-33	495
PKS_ACP	3707997	3708192	3,9E-06	195
PKS_ER	3905241	3905994	2,8E-04	753
PKS_KS	4057665	4058862	4,2E-30	1197
PKS_KR	4236369	4236873	1,6E-25	504
PKS_KR	4389030	4389540	6,1E-22	510

Gene	From	To	E-value	Size (bp)
PKS_KR	4416003	4416507	1,1E-24	504
PKS_AT	4522485	4523424	7,9E-04	939
PKS_KS	4552839	4553919	1,5E-09	1080
PKS_TE	4595790	4596426	2,7E-05	636
PKS_KR	4637610	4638126	6,6E-28	516
PKS_KS	4641642	4642215	8,4E-04	573
PKS_KR	4869999	4870473	1E-16	474
PKS_TE	5093046	5093772	1E-06	726
PKS_AT	5148300	5149071	5,7E-04	771
PKS_ACP	5167494	5167692	8,3E-04	198

From the 17 PKS clusters detected by AntiSmash, 11 were also detected by ClustScan (Table 6). Due to the fact that these two softwares work in a different manner (while ClustScan finds genes and allows the user to build clusters based on that information, AntiSmash provides the entire clusters found in the genome), the discrepancy between these two results was expected.

Table 6 - Position of clusters detected in AntiSmash, ClustScan similarity and ClustScan E-value.

#	DNA coordinates (From - To)		Detected by ClustScan	E-Value
1	240399	252932	✔	1,1E-20
2	1133011	1138638	✔	3,3E-33
3	1306704	1333830	✘	NA*
4	1591638	1596044	✔	5,4E-21
5	1689023	1702894	✔	4,1E-24

6	1716663	1730405	✔	8,6E-23
7	2316179	2333946	✔	9,4E-35
8	2561287	2572936	✔	4,1E-45
9	2588568	2597466	✔	5,4E-39
10	2799741	2809626	✘	NA
11	3092399	3106725	✔	3,5E-28
12	3541734	3557681	✘	NA
13	3692166	3714982	✘	2,9E-33
14	3841002	3874237	✘	NA
15	3887354	3897050	✘	NA
16	4408491	4423937	✔	1,1E-24
17	4814303	4820705	✘	NA

*NA - Not available

To further evaluate the potential of the selected clusters, it is necessary to analyze and annotate each ORF within the different clusters. This analysis will allow to understand which core domains are present in a specific cluster and which clusters might be functional.

(See appendix for more detailed information about the ORFs identified for each cluster.)

3. Clusters annotation and proteins identification

Positions of interesting regions given by AntiSmash were then used to build a map of the most interesting clusters. Since PKS clusters are usually very large, ORF's within a ± 15 000 bp range of the target genes identified by AntiSmash and ClustScan were annotated. CLC Sequence Viewer (version 6.9) was used to locate these genes within *P. atlantica* genome (FIG. 8). To get to the target location, it is only needed to type the location desired (in bp) desired and then press the "find" button. This software also shows the direction of transcription of each ORF, and by using "selection" mode, one can click on the "arrows" and copy each ORF sequence for further analysis.

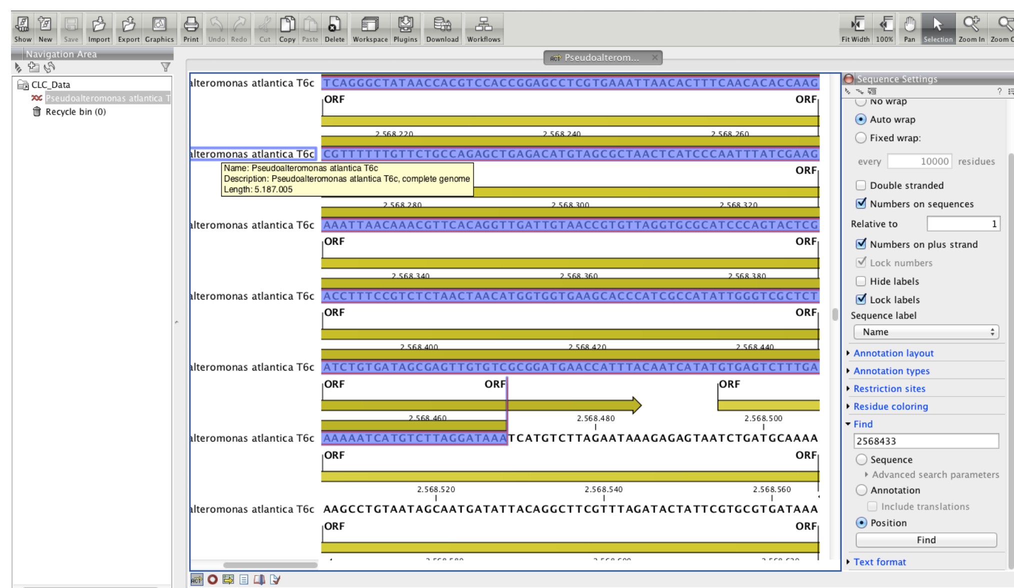


Figure 8 - Screenshot of CLC Sequence Viewer workspace.

Sequences of each ORF within the range selected were copied to the blastx of NCBI. This blast (basic local alignment search tool) allows to “Search protein database using a translated nucleotide query”. Default parameters with BLOSUM 62 (**BLO**cks of Amino Acid **SU**bstitution **M**atrix) were used. This lagging action resulted in several hits or sequences producing significant alignments. Hits with the best scores were considered for further studying. For example, from position 2568566 bp to 2569804 bp in *P. atlantica* genome, there is a 1239 bp long sequence that codes a 412 aa enzyme which is identified as a 3-oxoacyl-ACP synthase [Pseudoalteromonas atlantica], also known as Beta-ketoacyl-acyl-carrier-protein synthase I, typically involved in polyketide synthesis. [19]

Besides this synergy between CLC Sequence Viewer and blastx, AntiSmash is also of great utility when identifying proteins present in each cluster. By hovering the mouse over the arrows representing each ORF, a translated protein sequence appears with its identification as shown in a blastp (blast for proteins) of NCBI (Fig. 9). These annotations were confirmed with NCBI database.

Figure 9 - Screenshot of AntiSmash annotations.

In order to confirm if *P. atlantica* possesses the right means typically present in the biological machinery involved behind the production of compounds such as a polyketide or a bacteriocin, it is necessary to understand which role each of the identified proteins may be playing in the synthesis of these compounds. After the identification and annotation of all the proteins within the ranges of interest, an analysis was made using tools like InterPro and HHpred. InterPro is a tool that classifies proteins into families, predicts domains and identifies important sites (Fig. 10). HHpred detects protein homology in public databases and thus helps in the understanding of a protein function by presenting similar hits of known proteins.

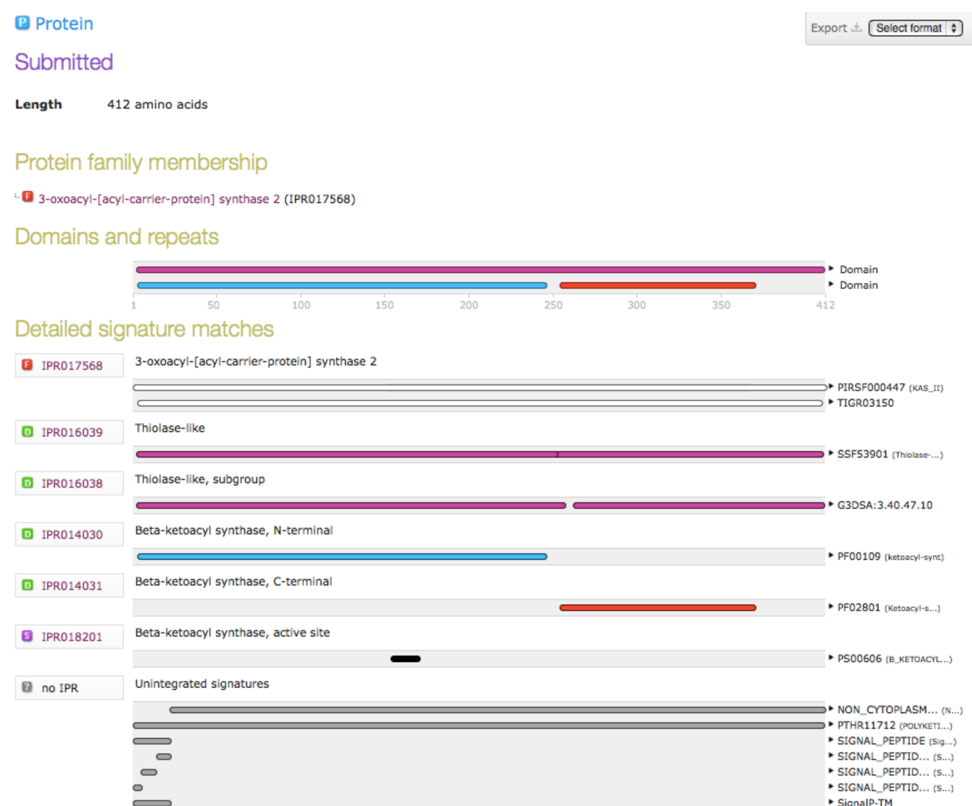


Figure 10 - Screenshot of InterPro search.

Biosynthesis of polyketides is catalyzed by a series of enzymes (PKS) that accomplish sequential decarboxylative condensations and reductive reactions in order to produce vast polyketide products. PKSs are characterized by having core domains that contain active sites [35]. These core domains are: AT (acyltransferase), ACP (acyl carrier protein) and KS (ketosynthase). A starter unit (acetyl) is loaded by the AT domain onto the KS domain, through a process that is mediated by ACP. Other domains, like KR (ketoreductase), ER (enoylreductase), DH (dehydratase) have been characterized as modifying domains. When present, these domains perform a modification of the initial carbonyl group, and therefore play a role in the structure of the final product. TE (thioesterase) domain catalyzes the release of the final product when it reaches its full length [35].

Table 7 shows the most glaring PKS-related ORFs for each annotated cluster.

Table 7 - PKS related ORFs detected near AntiSmash detection sites.

Cluster	ORF	Name	Predicted function
#1	Patl_0202	3-ketoacyl-CoA thiolase	Acyltransferase
	Patl_0209	short-chain dehydrogenase/reductase SDR	Ketoreductase
#2	Patl_0952	3-ketoacyl-ACP reductase	Ketoreductase
#3	Patl_1087	NAD-dependent epimerase/dehydratase	Dehydratase
	Patl_1093	acyltransferase 3	Acyltransferase
	Patl_1095	acyltransferase 3	Acyltransferase
	Patl_1096	acyltransferase 3	Acyltransferase
#4	Patl_1329	short-chain dehydrogenase/reductase SDR	Ketoreductase

Cluster	ORF	Name	Predicted function
#5	Patl_1046	3-oxoacyl-ACP reductase	Ketoreductase
#6	Patl_1432	short-chain dehydrogenase/reductase SDR	Ketoreductase
#7	Patl_1914	short-chain dehydrogenase/reductase SDR	Ketoreductase
#8	Patl_2120	beta-ketoacyl synthase	Ketosynthase
	Patl_2122	3-oxoacyl-(acyl-carrier-protein) reductase	Ketoreductase
	Patl_2123	malonyl CoA-acyl carrier protein transacylase	Acyltransferase
	Patl_2125	3-oxoacyl-(acyl-carrier-protein) synthase III	Ketosynthase
	Patl_2125	fatty acid/phospholipid synthesis protein PlsX	Acyltransferase
#9	Patl_2139	short-chain dehydrogenase/reductase SDR	Ketoreductase
#11	Patl_2552	short-chain dehydrogenase/reductase SDR	Ketoreductase
#12	Patl_2923	acetyl-CoA acetyltransferase	Acyltransferase
	Patl_2935	TesB family acyl-CoA thioesterase	Thioesterase
#13	Patl_3071	short-chain dehydrogenase/reductase SDR	Ketoreductase
#16	Patl_3661	3-oxoacyl-[acyl-carrier protein] reductase	Ketoreductase
#17	Patl_3994	aldo/keto reductase	Ketoreductase

The majority of the regions only presents one ORF with a PKS-related function or domain. A functional region should have ORFs with multiple domains, or multiple ORFs with PKS-related domains. This situation appears to happen only in one particular case, from Patl_2120 to Patl_2125.

- **PKS #8 CLUSTER**

Out of the 17 PKS clusters previously selected (Table 7 and appendix), one was identified as having all the core domains typically present in a PKS. From position 2568566 to 2574049 in the genome (Fig. 11), there is a transcripted cluster in reverse direction (complement) with possible two AT domains, one KS domains and one KR domain.

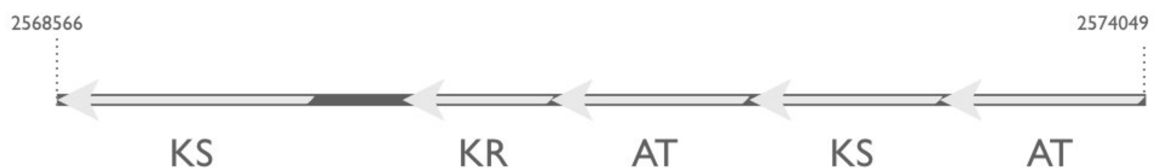


FIgure 11- Schematic representation of a hypothetical Polyketide Synthase in *P. atlantica* T6c composed by 4 distinct domains: AT - acyltransferase ; KS - ketosynthase ; KR - ketoreductase

The first protein of the cluster is a fatty acid/phospholipid synthesis protein PlsX (ORF Patl_2125), which has a 332 aa region referred in NCBI as a putative acyltransferase. Besides the InterPro scan for this protein reports a molecular function "transferase activity, transferring acyl groups other than amino-acyl groups".

The second protein in the chain is a beta-ketoacyl-ACP synthase. Biosynthesis of PKS pressuposes that the ketoacyl and ACP modules work together in the catalyzation of the chain elongation. [35]

After the beta-ketoacyl-ACP synthase there is a malonyl CoA-acyl carrier protein transacylase. InterPro detects an acyltransferase domain and reports that these domains are involved PKS synthesis, as previous literature stated. [35]

The next protein in this cluster is a 3-oxoacyl-(acyl-carrier-protein) reductase (from NCBI blastx). This protein, also known as beta-Ketoacyl reductase. Although a small number of ketoreductases has been deeply studied, and the mechanistic basis of their function is poorly understood, KR domains are typically present in PKS, as they catalyze reduction of 2- methyl-3-ketoacyl-ACP substrates. [36]

The last protein of this cluster is again a beta-ketoacyl synthase, one of the core domains of a PKS.

This PKS cluster presents relative similarity with clusters in the genome of other species (Fig. 12). Blastp over proteins of this cluster detected that *Methylomonas methanica*, *Edwarsiella ictaluri* and *Xenorhabdus nematophila* presented homology with a sequence set of six ORFs of the entire cluster detected by AntiSmash.

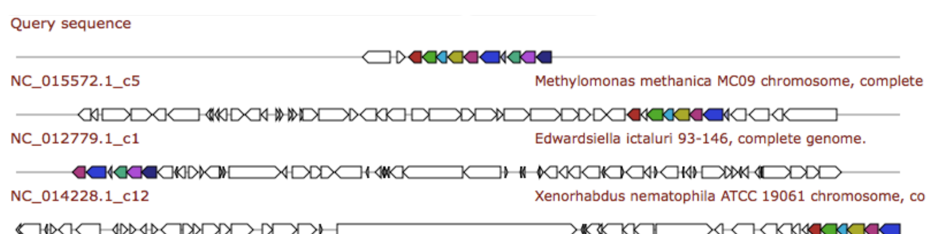


Figure 12 - Graphic of homologous gene clusters, for PKS #8 cluster - AntiSmash

Like in the case of *P. atlantica* there isn’t (yet) any reported proof that PKS are naturally produced by these bacteria.

- **BACTERIOCIN CLUSTER**

Detected by AntiSmash, this cluster (Fig. 13) is 13 106 bp long, and is constituted by 12 different proteins, two of them have a biosynthetic function.

Figure 13 - Schematic representation of a hypothetical bacteriocin in *P. atlantica* T6c.

The first protein of the cluster has a TonB-dependent receptor, and a TonB-dependent plug domains, which have a transport-related function.

The fourth protein in this cluster is detected by AntiSmash as having a biosynthetic role in the production of the bacteriocin. Surprisingly, a search (HHpred, AntiSmash, blastp) for this protein detects a beta-lactamase (typically antibiotic-resistance related). Nevertheless, BioGraph, a web tool that searches functional paths between different biomedical entities [37], detects a relation in the metabolic pathway of bacteriocins and Beta-lactamase-type transpeptidase fold proteins, just like the one present in this cluster.

The sixth protein in the cluster chain, catalogued as a “hypothetical protein” in NCBI database is identified by InterPro as a Xylose isomerase-like, TIM barrel domain, with none molecular function predicted. Although AntiSmash detects this protein as participatory in the biosynthesis of bacteriocins, there is no

evidence in the literature that links this kind of protein with the synthesis of bacteriocins.

Once again, AntiSmash presents an homology graphic. However, for this bacteriocin cluster, Blastp detected a lower similarity (less ORFs in the same position) between this cluster and sequences of other species (Fig. 14).

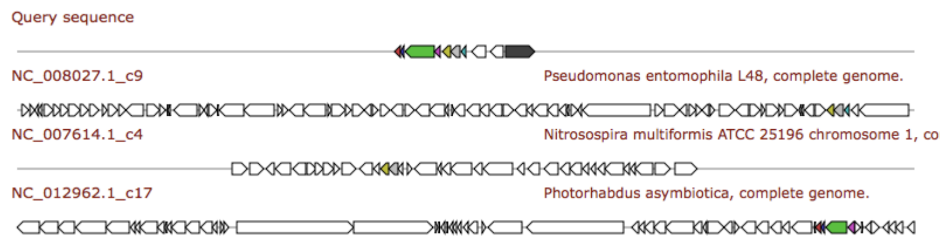


Figure 14 - Graphic of homologous gene clusters, for bacteriocin cluster - AntiSmash.

Once again, evidences that these bacteria produce bacteriocins are not available in the literature.

4. Discussion

The intense work over the analysis of the data retrieved by bioinformatic tools found in the literature is really impressive. The use of bioinformatic skills has led researchers to narrow genomic regions of interest and fully investigate over a smaller set of clusters.

Because *Streptomyces* is a genus known to be prolific in the production of secondary metabolites, many have already tried to detect PKS and NRPS clusters in genomes of different strains. For example, bioinformatic research using ClustScan and also AntiSmash has also resulted in the finding of 4 modular PKS clusters and 6 NRPS clusters in *Streptomyces tsukubaensis.* In this work over the *S. tsukubaensis* genome, ClustScan detected 60 putative KS domains, of which 38 were assigned to the referred clusters [38]. Moreover, genome analysis of *Streptomyces turgidiscabies* resulted in the detection of 17 PKS/NRPS clusters, using different bioinformatic tools (e.g. MiGAP - a microbial genome annotation software) [39].

Regarding *Pseudoalteromonas* species, the production of novel and powerful bioactive compounds among *Pseudoalteromonas* genus has already been recognized. Previous studies have already identified putative type I PKS or hybrid PKS/NRPS clusters in other *Pseudoalteromonas* species [40].

Although genome mining generally relies on the identification of gene clusters of known compounds such as NRPS or PKS, other important molecules were also found using this technique. For example, works over the genome of

Pseudoalteromonas tunicata led to the discovery of a 3-Formyl-Tyrosine, via searching for homologous ATP-grasp enzymes, which may possess antimicrobial properties [41].

AntiSmash has already been used to search clusters of antimicrobial interest. An AntiSmash analysis of the genome of *Pseudoalteromonas flavipulchra* resulted in the finding of four bacteriocin-type gene clusters, lantipeptide biosynthesis genes, four type I hybrid PKS/NRPS clusters and three NRPS clusters. Only one type I PKS gene was found in *P. flavipulchra* [42].

Although some of the state of the art techniques, like predicting compounds structure using bioinformatic tools, were not explored in this work, the number of antimicrobial clusters detected in *P. atlantica* T6c seems to be a sign that the genome of this strain might be prolific in regions responsible for antimicrobial compounds production [38, 43].

From the analysis of the results obtained here by bioinformatic research, it is presumable that if a PKS is really being translated through *P. atlantica* T6c genome, it must be type II PKS. That is because the PKS related proteins found in the genome, are monofunctional proteins, in contrast with the highly modular proteins typically present in type I PKSs. Also, an Acyl-Carrier-Protein (ACP) was found in various clusters, which eliminates the type III PKSs, since they are ACP independent. In spite of being shortly described, the presence of type II PKS clusters in marine bacteria has already been confirmed. However, no information is available for *Pseudoalteromonas*. BLAST searches over the genome *Streptomyces* and *Micromonospora* strains isolated from the soft coral tissue in the East China Sea detected the presence of type II PKS clusters [44]. Besides,

the occurrence of bacteriocin clusters in marine bacteria was already verified, as several gene clusters were found in many cyanobacteria strains. The identification of such clusters was done by searching for the best hits for proteins' sequences using BLASTp and Artemis (a genome browser and annotation tool), in a similar process used in this work [45]. Bacteriocin gene clusters normally comprise a region coding for an immunity protein against the bacteriocin itself [46], such protein was not detected through bioinformatic research in this work.

Regarding the most complete PKS cluster found (PKS #8 cluster), although it possesses the core domains of a PKS, it lacks a TE domain, which is essential to facilitate the release of the molecule from the enzyme. Nevertheless, it is possible that a particular domain may be positioned relatively far away (a few ORFs from the center of the cluster) [38], and thus escaping from the thorough analysis of that particular region (over the center of the cluster).

Chapter III

Optimization of antimicrobial compounds production

1. Overview

Many factors influence the efficiency of antibiotic production by living organisms. In this work, some culturing conditions were varied in order to understand their role in the biosynthesis of antimicrobial compounds. A one-factor-at-a-time method was used, each parameter was tested individually, which is a very time-consuming and expensive method, if large amount of trials are required. Nevertheless, it is an easy and simple method and that is why it is widely used in such experiments.

The following parameters were changed in the process of compound production optimization:

- culturing medium
- temperature
- agitation
- pH
- addition of extracellular ATP
- presence of competitors

The optimization of the referred parameters will be addressed in the next sections.

2. Materials and methods

2.1. Strains and culture conditions optimization

P. atlantica T6c (ATCC BAA-1087) was gently provided by Professor Anna Karls from University of Georgia - Department of Microbiology.

In order to evaluate if *P. atlantica* T6c produced antimicrobial compounds, and to test their activity spectrum, twelve target species, both reference strains and isolates, were assayed:

- *Escherichia coli* K12
- *Escherichia coli* K12 Δ*impA*
- *Staphylococcus aureus* CECT 229
- *Pseudomonas aeruginosa* PA01
- *Pseudomonas fluorescens* ATCC 27663
- *Salmonella enteritidis*
- *Listeria monocytogenes* CECT 4031
- *Enterococcus faecalis* V583
- *Bacillus subtilis*
- *Bacillus cereus*

- *Klebsiella pneumoniae* VK 089 RIFR
- *Vibrio harveyi*

E. coli K12 is a debilitated strain of *E. coli* commonly used in lab experimentations, because it does not normally colonize the human intestine [47]. The E. coli K12 Δ*impA* is a weaker strain that does not possess the *impA* gene, a gene that codes for an inner membrane protein [48] and thus it possesses a disrupted membrane.

All the bacteria isolates were retrieved from cryopreserved cultures. They were then grown overnight in a LB agar plate at 37° C and maintained at 4° C. *P. fluorescens* and *B. cereus,* which were grown overnight at 25° C.

2.1.1. Effect of the growth medium

In order to mimic the natural sea environment, in which *P. atlantica* inhabits, two culturing media related to this niche were assessed: DIFCO™ Marine Broth (MB) and Vatanen Nine Salts Solution (VNSS) [49]. Also, a Minimal Medium (MM) was tested, as it is common practice to optimize the production of antimicrobials in such medium, to determine optimal nutritional and culture conditions [50, 51]. This last medium was tested with four different and separate carbon sources: glucose, glycerol, lactose or galactose. At this stage, culturing

conditions were tried at 25ºC, 120 rpm and a pH of 7. Compositions of lab-made VNSS and MM media are shown below.

VNSS medium composition (all percentages are w/v): 0,1% peptone from soymeal; 0,05% yeast extract; 0,05% glucose; 0,5% soluble starch; 0,001% $FeSO_4 \cdot 7H_2O$; 0,001% $Na_sHPO_4 \cdot 2H_2O$; 1,76% NaCl; 0,147% Na_2SO_4; 0,008% $NaHCO_3$; 0,025% KCL; 0,004% KBr; 0,187% $MgCl_2 \cdot 6H_2O$; 0,041% $CaCl_2 \cdot 2H_2O$; 0,001% $SrCl_2 \cdot 6H_2O$ and 0,001% H_3BO_3.

MM composition (all percentages are w/v): 0,6% Na_2HPO_4 0,3% K_2HPO_4; 0,005% NH_4Cl; 0,006% $MgSO_4$; 0,75% $Na_2HPO_4 \cdot 2H_2O$; 0,55% $C_6H_{12}O_6 \cdot H_2O$ and 0,5% glucose / glycerol / lactose / galactose (only one of these carbon sources)

2.1.2. Temperature, pH and agitation

Once again, in order to mimic the natural habitat conditions of *P. atlantica,* and in accordance to the literature [14], temperatures between 20ºC and 30ºC were tested in a refrigerated incubator (Shel Lab®). These parameters are of utmost importance in the optimization of the production of antimicrobial compounds, as they affect organisms and have the ability to induce or inhibit the production of such compounds [51].

For the optimization of pH, MB medium and a temperature of 23ºC were used. In sea water, pH normally ranges between 8 and 8,5 [52]. Levels of pH between 5 and 9 were assayed.

For aeration optimization, MB medium, 23ºC, and a pH of 8 were used. Agitations of 100 to 150 RPM were then tested. Moreover, the dimensions of the flask and the media to flask size ratio are also important aspects to have into account, since they influence the aeration of the culture. Because of this, 250 mL of media were poured into 1000 mL sterile flasks. Cultures were grown during 120 hours.

2.1.3. Addition of extracellular ATP

As previous studies indicated that the addition of extracellular ATP enhances the production of antibiotics in *Streptomyces coelicolor* [53], the same was tried with *P. atlantica*. Ten µM (final concentration) of ATP were added to MB cultures at 23º C, 120 rpm agitation and pH of 8 in order to see if this factor influenced antimicrobial metabolites production. Cultures were grown during 120 hours.

2.1.4. Competitors

As the presence of a foreign (competitor) genus, or the metabolites produced by this competitor could affect antimicrobial metabolites production [15], three different configurations were assayed, as represented in Fig. 15.

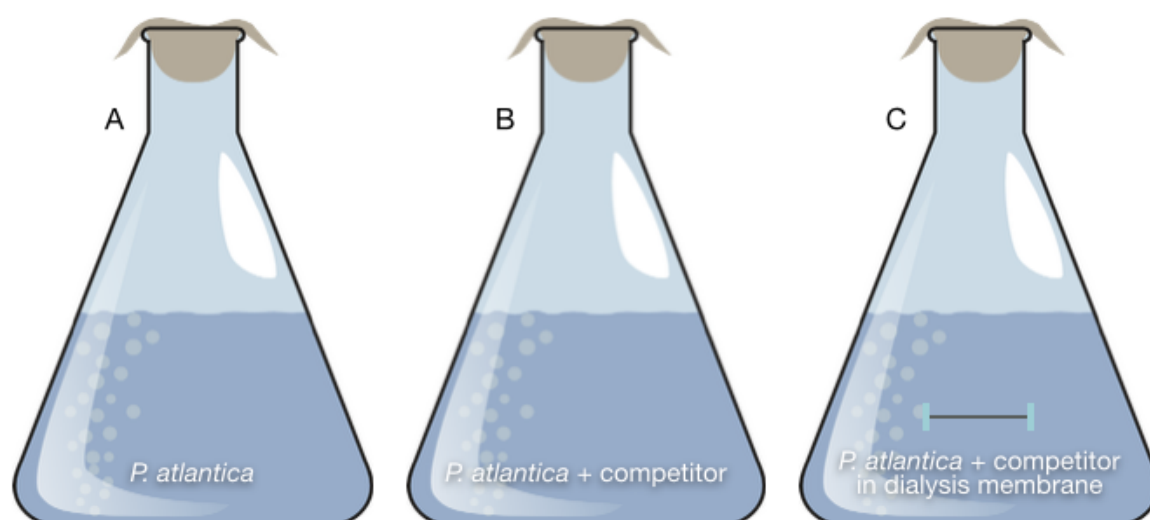


Figure 15 - Representation of the culture conditions used for the competition assays.

- **Fig. 15 - A** represents the simplest configuration. This configuration consists in a single culture of *P. atlantica* T6c with the selected conditions chosen in the previous steps in optimization phase; A small portion of biomass from a fresh culture of *P. atlantica* in marine agar was added aseptically to the Marine Broth medium. This flask served as a control condition.

- **Fig. 15 - B** represents a co-culture of *P. atlantica* T6c and *E. coli* K12. Three different sub-configurations (1:1, 10:1 and 100:1 ratios of *P. atlantica/E.coli*) were tested; One hundred μL from a pre-inoculum of *P. atlantica* grown overnight in Marine Broth were added to three eppendorfs cointaining 900 μL of phosphate buffered saline (PBS). After vortexing, the mix was poured onto the 1000 mL culture flasks containing 250 mL of Marine Broth. One hundred μL of a *E. coli* K12 pre-inoculum grown overnight in LB were added to an eppendorf containing 900 μL of PBS and three ten-fold dilutions were made and poured into the three different culture flasks.

- **Fig. 15 - C** represents a configuration in which a dialysis membrane (with a molecular weight cutoff of 14 000 Daltons) with a 2 mL culture of the competitor genus was put inside the culture flask. Four different sub-configurations were tested (variation of foreign genus inside membrane): *E. coli* K12, *P. aeruginosa*, *S. aureus*, *V. harveyi*; Ten centimeters of a dialysis membrane were cut using sterile scissors. One tip of this small portion was secured with a sterile clamp. Two milliliters of the foreign genus grown overnight at the respective conditions were added to the membrane with the help of a micropipette. The other tip was then secured with another sterile clamp and the apparatus was seamlessly put inside the large culture flask (as shown in Fig. 15 - C).

Cultures were incubated during 120 hours.

2.2. Assessment of antimicrobial activity

In order to measure the antimicrobial activity of the metabolites produced by *P. atlantica*, each 24 hours until a maximum of 120 hours of incubation, 10 mL of the culture were retrieved from the flasks under aseptic conditions. The spent media was then centrifuged at 15 000 RPM (refrigerated Sigma™ 3-16K) so that cell particles could deposit in the bottom of a falcon tube. The supernatant was then filtrated using a syringe filter with a 0,22 μm pore size (sterile Q-MAX® Syringe Filter from Frisenette) to further eliminate the presence of cells (obtaining bacteria-free spent media).

Target strains were grown overnight in LB. One hundred μL of the target cultures were shed into a Petri dish containing a thin layer of the respective media with solid agar. 4 mL of soft agar (0,5% agar) at 50° C were thereafter shed over the 100 μL culture. A slight shake of the Petri dish was made in order to improve an equitable distribution over the dish. This mix was allowed air dry for about 30 minutes. After that, 10 μL of the bacteria-free spent media and, when applicable, several dilutions were spotted into these bacterial lawns. These plates were incubated for 24h at 37ºC. Their ability to form an inhibition halo was recorded.

For experiments over medium, temperature, agitation and pH optimization, the antimicrobial activity was only assayed on *E. coli* K12 lawns. Only after the

optimization of the referred parameters, the entire array of strains was used to evaluate the compound activity spectrum.

2.3. Determination of arbitrary units

Arbitrary Units (AU) are a simple way of quantifying the ability of a compound to inhibit the growth of a target culture. They measure the strength of a compound. Arbitrary units are measured having the premise that the last dilution presenting inhibition has 1 compound unit. For example, if the last dilution presenting inhibition was a dilution of 1/16, then that sample would have 16 arbitrary units per mL [54].

3. Results and discussion

3.1. Effect of culture media on cell growth and antimicrobial activity

As culturing conditions may affect the antimicrobial capacity of *Pseudoalteromonas*, a set of default parameters found in literature was followed [55], varying the culture media to see what effect in cell growth and antimicrobial activity was being induced. Marine broth and the complex marine medium VNSS are culture media for the marine organisms and both are already linked to the successful production of bioactive secondary metabolites produced by *Pseudoalteromonas*. The Minimal Medium is a basic culturing medium usually used as a control in order to assay if a given nutrient is necessary for the growth of the culture or the production of a metabolite. [49, 56, 57]

P. atlantica showed no significant growth in MM medium in any of the chosen carbon sources, meaning that *P. atlantica* may be a demanding bacteria that requires some of the nutrients or amino-acids that are not present in such a poor medium. The standard approach is to culture in a Marine Minimal Medium (which contains a set micro and macronutrients required by the marine microbial consortia) [58], however, such medium was not tried. Otherwise, bacterial growth was observed in VNSS and MB media, which confirms that the richer marine

related media have a positive effect on *P. atlantica* growth. In fact, the use of Marine Broth and VNSS has already been related to successful growth of other strains of the *Pseudoalteromonas* genus [16, 59].

Table 8 summarizes the effect of chosen media in the bacterial growth and in the production of inhibitory compounds. As expected, the lack of growth verified with MM (with any of the carbon sources) led to an absence of effect on target. The strain tested (*E. coli* K12) grew over the drops of bacteria-free spent media of MM flasks. Not so expected, metabolites present in VNSS medium flasks were found innocuous, as an evident growth of *P. atlantica* was observed, but no harm for target variants was observed. In fact, the use of VNSS medium has already been linked to a successful production of antifouling agents [59], but any proof of the use of VNSS inducing the production of antimicrobial compounds in *Pseudoalteromonas* was found in the literature.

Table 8 - Effect of culture media in P. atlantica growth and the production of inhibitory compounds.

Media	Growth	Effect on target (*E. coli* K12)
MM - glucose	-	-
MM - glycerol	-	-
MM - lactose	-	-
MM - galactose	-	-
VNSS	+	-
MB	+	+

- No growth or no inhibition; + Growth or inhibition

On the other hand, bacteria-free spent media from MB flasks proved to inhibit the growth of target (*E. coli* K12) (FIG. 15), as clear inhibition halos were recorded. Marine Broth proved to be a medium where a notorious growth of *P. atlantica* was observed, with the formation of a thick biofilm on the walls of the flasks. From the three media tested, Marine Broth was the one where an antimicrobial activity was recorded. Hence, Marine Broth was the chosen medium for further assays. Other authors have recognized that the use of Marine Broth leads to a successful production of a potent antibiotic in *Pseudoalteromonas piscicida*, a compound that attacks the cell wall of MRSA pathogens [60]. Also, it was reported that a strain of *Pseudoalteromonas* closely related to *Pseudoalteromonas rubra* produces antimicrobial compounds of unknown structure with a broad spectral activity, when cultured in MB [61].

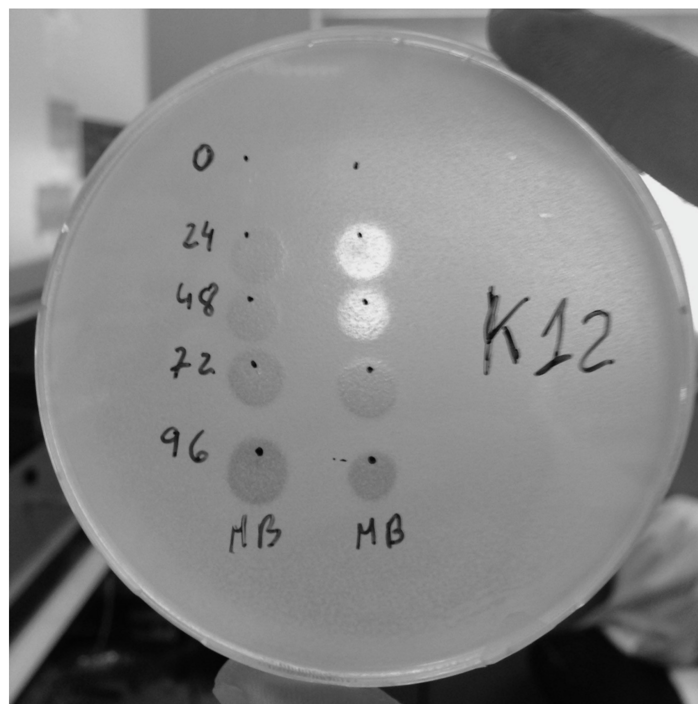


Figure 16 - Inhibition halos in an *E.coli* K12 lawn.

3.2. Effect of temperature, pH and agitation on cell growth and antimicrobial activity

The right set of culturing parameters can lead to a better production of some secondary metabolites. Parameters such as temperature, pH and agitation are of utmost importance when one tries to optimize the production of antimicrobial compounds. Living organisms have long proven to react to these exterior factors and thus they influence the production of such compounds [60, 62].

To test the effect of temperature, agitation and pH on cell growth, each parameter was tested individually. First, cultures of *P. atlantica* in Marine Broth were submitted to five different temperatures (20, 23, 25, 27, and 30ºC) (Table 9) at the standard pH medium (around 7,5).

Table 9 - Effect of culture media in the production of inhibitory compounds in P. atlantica.

Temperature	Inhibition
20ºC	+
23ºC	++
25ºC	++
27ºC	+
30ºC	+
+ Slight inhibition; ++ Good inhibition	

Antimicrobial activity was stronger for 23ºC and 25ºC, but a slight advantage was observed for the 23 ºC incubation, since a more vehement inhibition halo was recorded in the *E. coli* K12 lawns. Once again, the 23 ºC incubation temperature was elected for further analysis.

To test the pH effect, 5 different pHs were assayed in Marine Broth, ranging from pH 5 to 9 (Table 10). A clear advantage was attributed to the pH 8 (which is close to normal sea water levels) [52], since clearer inhibition halos were recorded. Previous works have already found a pH of around 7.5 to be ideal for the production of a target metabolite (a novel alginate-lyase) from *Pseudoalteromonas.* [63] Also, a pH of 7,5 was also found ideal for the production of antibiotics of unknown structure (not proteinaceus) by *Pseudoalteromonas piscicida.* The best conditions were also achieved in MB, but at a temperature of 37ºC [60].

Table 10 - Effect of pH on the production of inhibitory compounds in P. atlantica.

pH	Inhibition
5	-*
6	+
7	+
8	++
9	-

-* No growth; **+** Slight inhibition ; **++** Good inhibition

The same process was done to find which agitation provided the highest antimicrobial activity. Agitations varying from 80 rpm to 170 rpm (80, 100, 120, 150 and 170 rpm) were assayed (Table 11). Because there was no difference in antimicrobial activity nor bacterial growth, a 120 RPM agitation was chosen, according to previous works on *Pseudoalteromonas*, where rotations of about 120 rpm proved to be ideal for the production of secondary metabolites [55, 60].

Table 11 - Effect of agitation on the production of inhibitory compounds in P. atlantica.

Agitation (rpm)	Inhibition
80	++
100	++
120	++
150	++
170	++

+ Slight inhibition ; **++** Good inhibition

Also, previous experiments on how oxygen limitation influences the production of antimicrobial metabolites produced by *Pseudoalteromonas*, suggested that the antimicrobial activity is at a maximum for agitations up to 150 RPM. Antimicrobial activity was proved to be negligible for agitations greater than 250 RPM, which suggests that the RPM tried in this work were within the working window of agitations [64].

Antimicrobial assays on lawns of *E. coli* K12 (Fig 17 - B) for the best culture conditions were made in order to find the best time of incubation. (Fig. 17 - A) shows the arbitrary units (a compound strength measure) for antimicrobial assays on *E. coli* K12 of the bacteria-free spent media, from cultures of Marine Broth, at 23 ºC, 120 RPM and pH 8.

It is easily seen that the compound strength increases over time until 96 hours.

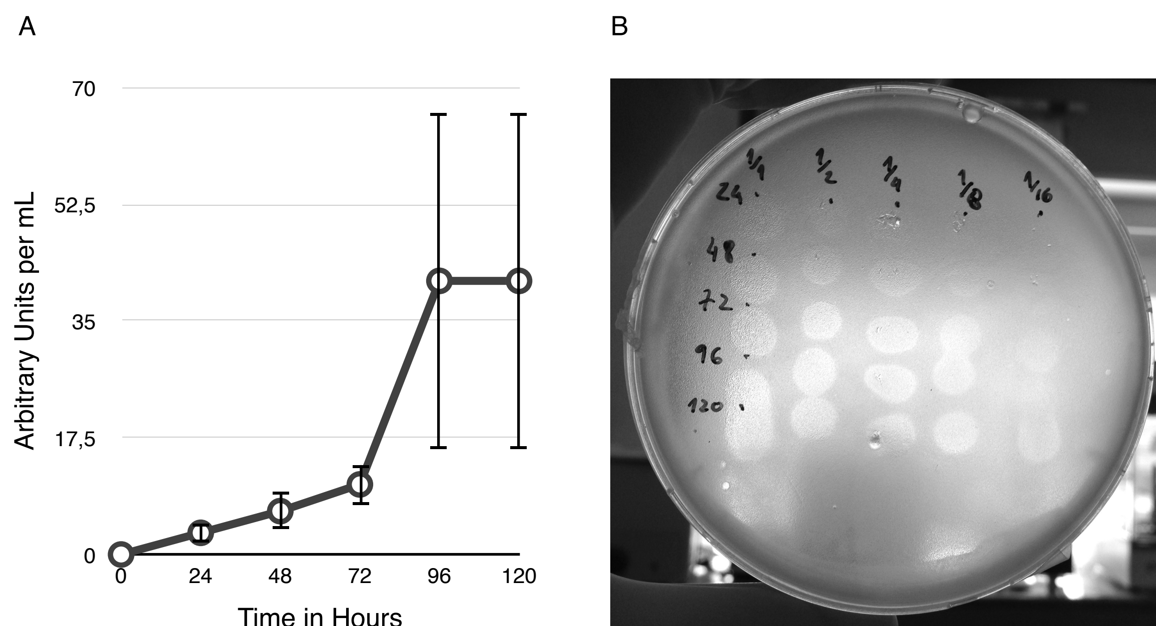


Figure 17 - Inhibitory effect of the filtered sterilized spent MB of *P. atlantica* in E. coli K12 lawns over time (A) **-** Arbitrary units over time (B) - Inhibition halos on E. coli K12 lawns, recorded after several dilutions of the spent medium.

Besides the positive tests on both *E. coli* strains, the other targets were not harmed by this bacteria-free spent media (Table 12).

Further tests on the spectrum of activity of the antimicrobial compounds were performed on a set of gram-positive and gram-negative bacteria. The fact that only *E. coli* strains were affected by bacteria-free spent media, may be due to the fact that these two strains are weak variants of *E. coli* and thus are more easily affected. Besides, only direct samples of bacteria-free spent media were assayed. The concentration of the samples, for example via solvent extraction, could lead to a stronger and eventually broader inhibitory activity.

Table 12 - Inhibition of targets by bacteria-free spent MB for optimized culture conditions.

Strains	Inhibition
E. coli K12	++
E. coli K12 Δ*impA*	+
S. aureus	-
P. aeruginosa	-
P. fluorescens	-
S. enteritidis	-
L. monocytogenes	-
E. faecalis	-
B. subtilis	-
B. cereus	-
K. pneumoniae	-
V. harveyi	-
+ Slight inhibition ; **++** Good inhibition; **-** No inhibition	

3.3. Effect of addition of extracellular ATP

Evidences that the addition of extracellular ATP could promote the production of antimicrobial metabolites in *Pseudoalteromonas* are not available in the literature. However, this approach was experienced in *Streptomyces coelicolor* and resulted in an improved production of the known polyketide antibiotic actinorhodin [53]. The same approach was tried in this work.

The addition of extracellular ATP did not influence nor potentiate the production of antimicrobial compounds, since halos from cultures with extracellular ATP were not clearer than those from cultures without the ATP. Hence, the addition of extracellular ATP was ignored for further assays.

3.4. Effect of competitors

The battle for resources in a "tight economy" only leaves space for the strongest. It is known that the presence of competition induces the production of secondary metabolites such as antibiotics in some bacteria [65, 66]. It is then important to figure out if competition also affects the production of antimicrobial compounds in *Pseudoalteromonas atlantica.*

- **Co-cultures of *P. atlantica* and *E. coli* K12**

Co-cultures of *P. atlantica* and *E. coli* K12 in Marine Broth at 23 ºC, 120 RPM and pH 8 were tested for three different ratios (1:1, 10:1, 100:1) of *P. atlantica* : *E.coli* K12*.* Although cell growth was observed for the three configurations, antimicrobial activity was not registered. Bacteria-free spent media was spotted over lawns of all target variants (*E. coli*, *S. aureus*, *P. fluorescens*, *P. aeruginosa*, *S. enteritidis*, *L. monocytogenes*, *E. faecalis*, *B. subtilis*, *B. cereus*, *K. pneuomoniae* and *V. harveyi*), but in any of the plates an inhibition halo was perceptible nor evident. These results may be due to the fact that the used competitors may overgrow and therefore inhibit the growth of *P. atlantica* and thus there isn't production of antimicrobial compounds.

- **Cultures of *P. atlantica* and *E. coli* K12 / *V. harveyi* / *P. aeruginosa* / *S. aureus* in dialysis membrane**

To avoid the overgrowth of the competing species, a new strategy was used to confine and limit the growth of the competitors.

Competing species were placed inside of a dialysis membrane. They were then placed in a culture flask of *P. atlantica* with Marine Broth at 23 ºC, 120 RPM and pH 8. Significant cell growth was observed for all the flasks/competing species. The antimicrobial assays showed a great similarity for all the configurations, since all proved to inhibit the growth of *E. coli* K12 Δ*impA* and *S. enteritidis* (Table 13) for an incubation time superior to 72 hours. These results suggest that the metabolites produced by the competing species, which manage to cross the dialysis membrane due to their reduced size, could influence or induce the production of antimicrobial metabolites by *P. atlantica*.

Table 13 - Effect of competing species (E. coli K12, S.aureus, P. aeruginosa, V. harveyi) in dialysis membrane on production of antimicrobial compounds by P. atlantica.

Competing species / Targets	***E. coli*** **K12**	***S. aureus***	***P. aeruginosa***	***V. harveyi***
E. coli **K12**	-	-	-	-
E. coli **K12 Δ *imp A***	+	+	+	+
S. aureus	-	-	-	-
P. aeruginosa	-	-	-	-
P. fluorescens	-	-	-	-
S. enteritidis	+	+	+	+
L. monocytogenes	-	-	-	-
E. faecalis	-	-	-	-
B. subtilis	-	-	-	-
B. cereus	-	-	-	-
K. pneumoniae	-	-	-	-
V. harveyi	-	-	-	-

Because the first configuration (culture without competing species) managed to inhibit the growth of *E. coli* K12 and *E. coli* K12 Δ*impA*, while the dialysis membrane configuration inhibited the growth of *E. coli* K12 Δ*impA* and *S. enteritidis*, it is presumable that a different antimicrobial compound could be synthesized under different culture conditions. However, it was impossible to infer which metabolic switch might have been involved in the production of a different compound, and which compound is being produced under each particular conditions. Further chemical techniques would be necessary to isolate and characterize the compounds.

Chapter IV
General conclusions and future work

1. General conclusions and future work

The use of microorganisms as a source of powerful antibiotics has been recognized as a successful approach [2]. The search for species capable of producing metabolites with antibiotic activity has called the attention of many towards the marine environment [14, 15, 26]. The *Pseudoalteromonas* genus is known to colonize a vast range of habitats, which may be due to its capacity of producing powerful bioactive compounds that enable strains of this genus to conquer others' habitats. The high diversity of strains among this genus and the chemical diversity of the compounds already identified make the *Pseudoalteromonas* genus a very interesting target for the discovery of novel molecules [14, 15].

In this work, bioinformatic tools were used to detect regions of the *P. atlantica* genome with clusters of antimicrobial compounds, to further investigate these regions, detect ORFs and identify proteins domains and functions. Research over the genome of *P. atlantica* resulted in the finding of many clusters of PKSs. Although they all seem to miss one or more domains of a typical complete PKS, some of them could be working together in the synthesis of a single polyketide. Also, a PKS cluster possessing the core domains of a PKS was found and its ORFs were described. This cluster may in fact be an active producer of a polyketide. Besides the PKS clusters, a bacteriocin cluster was found. It is possible that these two very different metabolites could be

synthesized by *P. atlantica* under different culturing conditions, and thus explain, at some extent, some of the results obtained in laboratory testing. Nevertheless, at this point, there isn't any proof that supports such argument, since we have not further investigated the nature of the compounds.

Pseudoalteromonas atlantica is a marine bacterium that proved to inhibit the growth of other bacteria, under the right set of culturing parameters. While the VNSS or MM media have not induced the production of antimicrobial molecules, the MB has provided a clear inhibition of *E. coli* strains. Optimal conditions for antimicrobial production were found for cultures in Marine Broth, at a temperature of 23ºC and a pH of 8, at an agitation of 120 rpm). Also, *P. atlantica* showed different inhibition patterns in the presence of competing species, which might suggest that different conditions induce the production of different molecules.

It is possible to continue the work over *P. atlantica* antibacterial metabolites. A deeper investigation over the referred clusters could lead to the finding of more PKS domains working together in the assembly of a polyketide. The interruption of the identified clusters could be done to evaluate their involvement in the production of the detected antimicrobials. The purification, for instance by High-Performance Liquid Chromatography (HPLC), of the bacteria-free spent media would be a step forward in the definition of the antibacterial strength of the referred metabolites. Finally, if the purified molecule presented a potent effect, the work could move towards the characterization of the molecule structure.

Chapter V
References and appendix

1. References

1. Fleming, A., On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. influenzae. British Journal of Experimental Pathology, 1929. 10: p. 226-236.

2. Davies, J. and D. Davies, Origins and Evolution of Antibiotic Resistance. 2010. 74(3): p. 417-433.

3. Davies, J., Vicious Circles: Looking Back on Resistance Plasmids. Genetics, 1995: p. 1465-1468.

4. Levy, S.B. and B. Marshall, Antibacterial resistance worldwide: causes, challenges and responses. nature medicine, 2004. 10(12): p. 122-129.

5. Kohanski, M.A., D.J. Dwyer, and J.J. Collins, How antibiotics kill bacteria: from targets to networks. Nature Reviews Microbiology, 2010. 8: p. 423-435.

6. Alberts, B., et al., From RNA to Protein. Molecular Biology of the Cell. Vol. 4. 2002.

7. Bermingham, A. and J.P. Derrick, The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. BioEssays, 2002. 24(7): p. 637-648.

8. Nikaido, H., Multidrug Resistance in Bacteria. Annual Review of Biochemistry, 2009. 78(1): p. 119-146.

9. ´erichon, B.P. and P. Courvalin, VanA-Type Vancomycin-Resistant Staphylococcus aureus. ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, 2009. 53(11).

10. Shlaes, D.M., S.J. Projan, and J.E. Edwards, Antibiotic Discovery: State of the State. ASM News, 2004. 70: p. 270-280.

11. Organization, W.H., Antimicrobial Resistance: Global Report on Survaillance. 2014.

12. Hayashi, M.A.F., F.C. Bizerra, and P.I. Da Silva Junior, Antimicrobial compounds from natural sources. frontiers in MICROBIOLOGY, 2013. 4.

13. Clardy, J., M. Fischbach, and C. Walsh, New antibiotics from bacterial natural products. Nat Biotech, 2006. 24(12): p. 1541-1550.

14. Vynne, N.G., Bioactivity and phylogeny of the marine bacterial genus Pseudoalteromonas, in National Food Institute

Division of Industrial Food Research. 2011, Technical University of Denmark.

15. Bowman, J.P., Bioactive Compound Synthetic Capacity and Ecological Significance of Marine Bacterial Genus Pseudoalteromonas. Marine Drugs, 2007: p. 22.

16. Vynne, N.G., Bioactivity and phylogeny of the marine bacterial genus Pseudoalteromonas, in National Food Institute Division of Industrial Food Research. 2011, Technical University of Denmark.

17. Piasecki, S.K., Investigations into the biocatalytic potential of modular polyketide synthase ketoreductases. 2013, University of Texas.

18. Pfeifer, B. and C. Khosla, MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS : Biosynthesis of Polyketides in Heterologous Hosts. Vol. 65. 2001.

19. Chan, Y., et al., Biosynthesis of Polyketide Synthase Extender Units. National Institute of Health, 2009.

20. Marinho, A.M.R., et al., Active polyketides isolated from Penicillium herquei. Anais da Academia Brasileira de Ciências, 2013. 85: p. 909-912.

21. Arias, A.A.e., M. Craig, and P. Fickers, Science against microbial pathogens: communicating current research

and technological advances (Gram-positive antibiotic biosynthetic clusters: a review). Vol. 1. 2011.

22. Marahiel, M.A., Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. Journal of Peptide Science, 2009. 15(12): p. 799–807.

23. Sundlov, J.A., et al., Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. Chem Biol., 2013.

24. Caboche, S., et al., Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. JOURNAL OF BACTERIOLOGY, 2010. 192(19): p. 5143–5150.

25. Cotter, P., C. Hill, and P. Ross, Bacteriocins: developing innate immunity for food. Nature Reviews Microbiology, 2005. 3.

26. Bakkal, S., S. Robinson, and M. Riley, Bacteriocins of Aquatic Microorganisms and Their Potential Applications in the Seafood Industry. 2012.

27. Cotter, P., P. Ross, and C. Hill, Bacteriocins - a viable alternative to antibiotics? Nature Reviews Microbiology, 2013. 11.

28. Lux, T., et al., Diversity of Bacteriocins and Activity Spectrum in Streptococcus pneumoniae. Journal of Bacteriology, 2007. 189(21).

29. Klaenhammer, T.R., Genetics of bacteriocins produced by lactic acid bacteria. FEMS Microbiology Reviews, 1993. 12: p. 39–85.

30. Biology, N.R.C.U.B.o., Bioinformatics: Converting Data to Knowledge: Workshop Summary. 2000.

31. Scheffler, R.J., et al., Antimicrobials, drug discovery, and genome mining. Appl Microbiol Biotechnol, 2012.

32. Eddy, S.R., Profile Hidden Markov Models. Bioinformatics Review, 1998. 14(9): p. 755-763.

33. Blin, K., et al., antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res, 2013. 41(Web Server issue): p. W204-12.

34. Starcevic, A., et al., ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res, 2008. 36(21): p. 6882-92.

35. Yadav, G., R. Gokhale, and D. Mohanty, SEARCHPKS: a program for detection and analysis of polyketide synthase domains. Nucleic Acids Research, 2003. 31(13): p. 3654–3658.

36. You, Y.-O., C. Khosla, and D.E. Cane, Stereochemistry of Reductions Catalyzed by Methyl-Epimerizing Ketoreductase Domains of Polyketide Synthases. J Am Chem Soc, 2013.

37. Liekens, A.M., et al., BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation. Genome Biology, 2011.

38. Blažicˇ, M., et al., Annotation of the Modular Polyketide Synthase and Nonribosomal Peptide Synthetase Gene Clusters in the Genome of Streptomyces tsukubaensis NRRL18488. Applied and Environmental Microbiology, 2012. 78(23): p. 8183–8190.

39. Komaki, H., et al., Genome-wide survey of polyketide synthase and nonribosomal peptide synthetase gene clusters in Streptomyces turgidiscabies NBRC 16081. The Journal of General and Applied Microbiology, 2012. 58: p. 363-372.

40. Zhu, P., et al., Molecular Phylogeny and Modular Structure of Hybrid NRPS/PKS Gene Fragment of Pseudoalteromonas sp. NJ6-3-2 Isolated From Marine Sponge Hymeniacidon perleve. Journal of Microbiology and Biotechnology, 2009.

41. Blasiak, L.C. and J. Clardy, Discovery of 3-Formyl-Tyrosine Metabolites from Pseudoalteromonas tunicata through Heterologous Expression. American Chemical Society, 2009: p. 926-927.

42. Yu, M., et al., Genome analysis of Pseudoalteromonas flavipulchra JG1 reveals various survival advantages in marine environment. BMC Genomics, 2013.

43. Challis, G.L., Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology, 2008. 154: p. 1555–1569.

44. Sun, W., et al., Functional Gene-Guided Discovery of Type II Polyketides from Culturable Actinomycetes Associated with Soft Coral Scleronephthya sp. PLoS ONE, 2012. 7(8).

45. Wang, H., D.P. Fewer, and K. Sivonen, Genome Mining Demonstrates the Widespread Occurrence of Gene Clusters Encoding Bacteriocins in Cyanobacteria. PLoS ONE, 2011. 6(7).

46. Desriac, F., et al., Bacteriocin as Weapons in the Marine Animal-Associated Bacteria Warfare: Inventory and Potential Applications as an Aquaculture Probiotic. Marine Drugs, 2010. 8: p. 1153-1177.

47. Escherichia coli K-12 Derivatives Final Risk Assessment. 2012 [cited 2014; Available from: http://epa.gov/biotech_rule/pubs/fra/fra004.htm.

48. MINTZ, K.P. and P.M. FIVES-TAYLOR, impA, a Gene Coding for an Inner Membrane Protein, Influences Colonial Morphology of Actinobacillus actinomycetemcomitans. INFECTION AND IMMUNITY, 2000. 68(12): p. 6580–6586.

49. Rao, D., J.S. Webb, and S. Kjelleberg, Competitive Interactions in Mixed-Species Biofilms Containing the Marine Bacterium Pseudoalteromonas tunicata. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 2005. 71(4).

50. Neetia, D., et al., Optimization of the Production of the Lantibiotic Mutacin 1140 in Minimal Media. Process Biochem, 2011.

51. Jafarzade, M., et al., Influence of Culture Conditions and Medium Composition on theProduction of Antibacterial Compounds by Marine Serratia sp.WPRA3. Journal of Microbiology, 2013.

52. Kulthanan, K., P. Nuchkull, and S. Varothai, The pH of water from various sources: an overview for recommendation for patients with atopic dermatitis. Asia Pacific Association of Allergy, Asthma and Clinical Immunology, 2013.

53. Li, M., et al., Effects of extracellular ATP on the physiology of Streptomyces coelicolor A3(2). Federation of European Microbiological Societies, 2008.

54. GOEBEL, W., G. BARRY, and T. SHEDLOVSKY, The Production of Colicin K in Media Maintained at Constant pH. The Rockefeller Institute for Medical Research, 1956.

55. Jung, C., et al., Optimization of Pseudoalteromonas sp. JYBCL 1 Culture Conditions, Medium Composition and Extracellular β-agarase Activity. Biotechnology and Bioprocess Engineering, 2012.

56. Lodish, H., et al., Molecular Cell Biology. 2000.

57. Fehér, D., et al., Highly Brominated Antimicrobial Metabolites from a Marine Pseudoalteromonas sp. Journal of Natural Products, 2010. 73(11).

58. Kim, S.K., Marine Cosmeceuticals: Trends and Prospects. 2011: Taylor & Francis.

59. Dheilly, A., et al., Antibiofilm Activity of the Marine Bacterium Pseudoalteromonas sp. Strain 3J6. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 2010. 76(11): p. 3452–3461.

60. Darabpour, E., et al., ￼Isolation of a potent antibiotic producer bacterium, especially against MRSA, from northern region of the Persian Gulf. Bosnian Journal of Basic Medical Sciences, 2012: p. 108-121.

61. Chellarama, C., et al., Bioactive Peptides from Epibiotic Pseudoalteromonas Strain P1. APCBEE Procedia, 2012. 2: p. 37-42.

62. Song, Q., Y. Huang, and H. Yang, Optimization of Fermentation Conditions for Antibiotic Production by Actinomycetes YJ1 Strain against Sclerotinia sclerotiorum. Journal of Agricultural Science, 2012. 4(7).

63. Ma, Y., et al., Culture optimization and characterization of an alginate-lyase from Pseudoalteromonas sp. LJ1. 2009.

64. López, R., et al., OXYGEN LIMITATION FAVORS THE PRODUCTION OF PROTEIN WITH ANTIMICROBIAL ACTIVITY IN

PSEUDOALTEROMONAS SP. Brazilian Journal of Microbiology, 2012: p. 1206-1212.

65. Slattery, M., I. Rajbhandari, and K. Wesson, Competition-Mediated Antibiotic Induction in the Marine Bacterium Streptomyces tenjimariensis. Microbial Ecology, 2001. 41: p. 90-96.

66. Cornforth, D.M. and K.R. Foster, Competition sensing: the social side of bacterial stress responses. Nature Reviews Microbiology, 2013. 11.

67. Jung, C., et al., Optimization of Pseudoalteromonas sp. JYBCL 1 Culture Conditions, Medium Composition and Extracellular β-agarase Activity

. Biotechnology and Bioprocess Engineering, 2012.

2. Appendix

#1

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	peptidase S9, prolyl oligopeptidase active site region	←	100	0,00E+00	231294	233234	1941	
2	hypothetical protein PatI_0195	←	100	0,00E+00	233297	234433	1137	
3	glutathione S-transferase-like protein	➡	100	9,00E-149	234572	235192	621	
4	hypothetical protein	←	100	2,00E-84	235224	235691	468	
5	hypothetical protein PatI_0198	←	100	0,00E+00	235660	236664	1005	
6	proline dipeptidase	←	100	0,00E+00	236751	238118	1368	
7	lytic transglycosylase	←	100	0,00E+00	238178	240124	1947	
8	multifunctional fatty acid oxidation complex subunit alpha	➡	100	0,00E+00	240363	242558	2196	
9	3-ketoacyl-CoA thiolase	➡	100	0,00E+00	242559	243740	1182	AT
10	TonB-dependent receptor, plug	←	100	0,00E+00	243936	246056	2121	
11	major facilitator superfamily transporter	➡	100	0,00E+00	246350	247648	1299	
12	hypothetical protein	➡	100	1,00E-113	247678	248220	543	
13	hypothetical protein PatI_0206	←	100	0,00E+00	248297	250078	1782	
14	twin-arginine translocation pathway signal protein	←	100	0,00E+00	250329	251312	984	
15	TetR family transcriptional regulator	←	100	2,00E-144	251449	252081	633	
16	short-chain dehydrogenase/reductase SDR	➡	100	8,00E-175	252090	252932	843	KR
17	hypothetical protein PatI_0210	➡	100	0,00E+00	253285	254175	891	
18	patatin	←	100	0,00E+00	254179	255222	1044	
19	alanine--glyoxylate aminotransferase	←	100	0,00E+00	255445	256740	1296	
20	hypothetical protein PatI_0213	➡	100	4,00E-132	256969	257700	732	
21	glyoxalase/bleomycin resistance protein/dioxygenase	←	100	2,00E-97	257682	258134	453	
22	adenylosuccinate synthetase	←	100	0,00E+00	258225	259526	1302	
23	ensor signal transduction histidine kinas	←	100	0,00E+00	259847	261367	1521	

#2

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	TonB-dependent receptor	←	100	0,00E+00	1120789	1123092	2303	
2	TonB-dependent receptor	➠	100	0,00E+00	1123519	1125987	2468	
3	peptidase M19	➠	100	0,00E+00	1125957	1127231	1274	
4	hypothetical protein Patl_0942	←	100	7,00E-130	1127325	1127897	572	
5	hypothetical protein Patl_0943	➠	99	5,00E-152	1128169	1128711	542	
6	rhodanese-like protein	←	100	1,00E-92	1128966	1129391	425	
7	hypothetical protein Patl_0945	←	100	5,00E-111	1129375	1130202	827	
8	beta-lactamase-like protein	➠	100	0,00E+00	1130244	1131113	869	
9	response regulator receiver protein	←	100	0,00E+00	1131124	1132773	1649	
10	methylmalonate-semialdehyde dehydrogenase	➠	100	0,00E+00	1132999	1134501	1502	
11	acyl-CoA dehydrogenase-like protein	➠	100	0,00E+00	1134471	1135670	1199	
12	enoyl-CoA hydratase/isomerase	➠	99	0,00E+00	1135761	1136894	1133	
13	3-hydroxyisobutyrate dehydrogenase	➠	100	0,00E+00	1136846	1137817	971	
14	3-ketoacyl-ACP reductase	➠	100	3,00E-155	1137865	1138638	773	KR
15	hypothetical protein Patl_0953	➠	100	0,00E+00	1138984	1140543	1559	
16	peptidoglycan binding domain-containing protein	➠	100	3,00E-139	1140676	1141278	602	
17	hypothetical protein Patl_0955	←	100	4,00E-56	1141279	1141581	302	
18	glutamine amidotransferase, class-II	←	100	0,00E+00	1141864	1142646	782	
19	hypothetical protein Patl_0957	➠	99	2,00E-90	1142775	1143269	494	
20	hypothetical protein Patl_0958	➠	100	0,00E+00	1143463	1145889	2426	
21	hypothetical protein Patl_0959	➠	100	0,00E+00	1145802	1146761	959	
22	hypothetical protein Patl_0960	➠	100	0,00E+00	1146762	1148213	1451	
23	serine/threonin protein kinase	←	100	0,00E+00	1148224	1150986	2762	
24	ECF subfamily RNA polymerase sigma-24 factor	➠	100	4,00E-137	1151044	1151646	602	
25	hypothetical protein Patl_0963	←	100	5,00E-74	1151732	1152079	347	
26	hypothetical protein Patl_0964	➠	100	0,00E+00	1152326	1153468	1142	

#3

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	hypothetical protein	←	100	0,00E+00	1305039	1306028	990	
2	NAD-dependent epimerase/ dehydratase	➡	100	0,00E+00	1306662	1307696	1035	DH
3	WecB/TagA/CpsF family glycosyl transferase	➡	99	0,00E+00	1307922	1308770	849	
4	group 1 glycosyl transferase	➡	100	0,00E+00	1308771	1309871	1101	
5	group 1 glycosyl transferase	➡	100	0,00E+00	1309872	1311002	1131	
6	exopolysaccharide production protein	➡	100	0,00E+00	1311008	1312306	1299	
7	glycosyl transferase family protein	➡	100	0,00E+00	1312297	1313250	954	
8	acyltransferase 3	←	100	0,00E+00	1313415	1314416	1002	AT
9	GumL protein	➡	99	0,00E+00	1315152	1315976	825	
10	acyltransferase 3	➡	100	0,00E+00	1316015	1317169	1155	AT
11	acyltransferase 3	←	100	0,00E+00	1317430	1319280	1851	AT
12	hypothetical protein Patl_1097	←	100	3,00E-159	1319374	1320060	687	
13	hypothetical protein	➡	100	0,00E+00	1320692	1321690	999	
14	hypothetical protein Patl_1099	➡	99	0,00E+00	1321961	1322884	924	
15	xanthan biosynthesis pyruvyltransferase GumL	➡	99	0,00E+00	1323322	1324146	825	
16	hypothetical protein Patl_1101	➡	100	0,00E+00	1324575	1326056	1482	
17	hypothetical protein Patl_1102	➡	100	0,00E+00	1326335	1328806	2472	
18	hypothetical protein Patl_1103	←	100	8,00E-99	1329011	1329448	438	
19	coproporphyrinogen III oxidase	←	100	0,00E+00	1329512	1330900	1389	
20	hypothetical protein	←	100	0,00E+00	1331029	1332498	1470	
21	phosphodiesterase I	←	100	0,00E+00	1332556	1333884	1329	
22	catalase	➡	100	0,00E+00	1334375	1336039	1665	

#4

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	riboflavin biosynthesis protein RibD	➠	99	,00E+0	1578875	1580002	1128	
2	riboflavin synthase subunit alpha	➠	100	,00E-15	1580137	1580811	675	
3	3,4-dihydroxy-2-butanone 4-phosphate synthase	➠	100	,00E+0	1580880	1582004	1125	
4	transcription antitermination protein NusB	➠	99	2,00E-94	1582731	1583153	423	
5	thiamine monophosphate kinase	➠	99	,00E+0	1583217	1584236	1020	
6	phosphatidylglycerophosphatase	➠	100	,00E-10	1584218	1584724	507	
7	hypothetical protein Patl_1318	←	100	,00E-10	1584812	1585270	459	
8	deoxyxylulose-5-phosphate synthase	←	100	,00E+0	1585355	1587265	1911	
9	polyprenyl synthetase	←	100	,00E+0	1587266	1588183	918	
10	flagellar motor protein MotB	➠	100	,00E-17	1589510	1590430	921	
11	putative lipoprotein	←	100	,00E-10	1590544	1591131	588	
12	hypothetical protein Patl_1325	←	100	,00E-11	1591110	1591637	528	
13	cyclopropane-fatty-acyl-phospholipid synthase	←	100	,00E+0	1591638	1592936	1299	
14	hypothetical protein	←	100	,00E+0	1592893	1593630	738	
15	amine oxidase	←	100	,00E+0	1593621	1594919	1299	
16	short-chain dehydrogenase/ Reductase SDR	←	100	,00E-16	1594909	1595640	732	KR
17	hypothetical protein Patl_1330	←	100	1,00E-97	1595622	1596047	426	
18	peptidase S16	➠	100	,00E-13	1596150	1596773	624	
19	RNA polymerase sigma factor	➠	100	,00E-14	1596769	1597443	675	
20	anti-ECF sigma factor ChrR	➠	100	,00E-16	1597430	1598107	678	
21	outer membrane biogenesis protein BamD	←	100	,00E-17	1598187	1598975	789	
22	RluA family pseudouridine synthase	➠	100	,00E+0	1599115	1600110	996	
23	hypothetical protein	➠	100	,00E+0	1600111	1600851	741	
24	protein disaggregation chaperone	➠	100	,00E+0	1600960	1603548	2589	
25	hypothetical protein	←	100	,00E+0	1603657	1606594	2938	
26	hypothetical protein Patl_1339	←	100	,00E-12	1606610	1607425	816	
27	ıcylglycerol--serine O-phosphatidyltraı	➠	100	,00E-15	1607744	1608577	834	

#5

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	glutaredoxin-like protein	→	100	2,00E-73	1681008	1681349	342	
2	BFD-like (2Fe-2S)-binding protein	←	100	5,00E-38	1681535	1681906	372	
3	alkyl hydroperoxide reductase	→	100	3,00E-45	1681945	1682601	657	
4	ribonuclease T	←	100	4,00E-156	1682716	1683357	642	
5	OmpA/MotB protein	→	100	0,00E+00	1683534	1684424	891	
6	glycerophosphodiester phosphodiesterase	←	100	0,00E+00	1684559	1685563	1005	
7	glyoxalase I	←	100	6,00E-88	1685585	1686070	486	
8	multidrug transporter	←	100	2,00E-61	1686071	1686448	378	
9	cellulase	←	100	0,00E+00	1686808	1688208	1401	
10	2-alkenal reductase	←	100	0,00E+00	1689023	1690057	1035	
11	3-oxoacyl-ACP reductase	←	100	0,00E+00	1690452	1691222	771	KR
12	acyl-CoA dehydrogenase	→	100	0,00E+00	1691916	1693760	1845	
13	hypothetical protein	←	100	1,00E-67	1694036	1694803	768	
14	hypothetical protein	←	100	9,00E-65	1694830	1695150	321	
15	hypothetical protein Patl_1411	←	99	4,00E-86	1695482	1695883	402	
16	XRE family transcriptional regulator	←	100	2,00E-67	1696542	1696880	339	
17	hypothetical protein	←	100	5,00E-173	1698607	1699413	807	
18	hypothetical protein	←	99	0,00E+00	1699407	1700396	990	
19	2-methylisocitrate lyase	→	100	0,00E+00	1702004	1702894	891	
20	methylcitrate synthase	→	100	0,00E+00	1703001	1704134	1134	
21	aconitate hydratase	→	100	0,00E+00	1704135	1706873	2739	
22	3-methylitaconate isomerase	→	100	0,00E+00	1706874	1708112	1239	
23	major facilitator superfamily transporter	→	100	0,00E+00	1708321	1709511	1191	
24	diguanylate cyclase	→	100	0,00E+00	1709596	1710717	1122	

#6

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	major facilitator superfamily transporter	➠	100	0,00E+00	1708321	1709511	1191	
2	diguanylate cyclase	➠	100	0,00E+00	1709596	1710717	1122	
3	NADH:flavin oxidoreductase	➠	100	0,00E+00	1711068	1712306	1239	
4	alcohol dehydrogenase GroES-like protein	➠	100	0,00E+00	1712442	1713539	1098	
5	ethanolamine ammonia lyase large subunit	←	100	1,00E-03	1713977	1714285	309	
6	ethanolamine ammonia lyase large subunit	➠	100	6,00E-57	1714223	1715626	1404	
7	ethanolamine ammonia-lyase small subunit	➠	100	0,00E+00	1715551	1716534	984	
8	hypothetical protein	←	100	0,00E+00	1716663	1718435	1773	
9	putative lipoprotein	←	100	0,00E+00	1718674	1720206	1533	
10	lyase -like protein	←	100	0,00E+00	1720632	1721957	1326	
11	short-chain dehydrogenase/reductase SDR	←	100	6,00E-151	172205	1722777	723	KR
12	hypothetical protein	➠	99	0,00E+00	1723048	1724295	1248	
13	glutamate synthase	➠	100	0,00E+00	1724342	1725979	1638	
14	glyoxalase	←	100	2,00E-84	1726151	1726546	396	
15	LysR family transcriptional regulator	←	100	0,00E+00	1726714	1727640	927	
16	quaternary ammonium transporter	➠	100	1,00E-58	1727820	1728155	336	
17	S-(hydroxymethyl)glutathione dehydrogenase	➠	100	0,00E+00	1728282	1729424	1143	
18	S-formylglutathione hydrolase	➠	99	0,00E+00	1729491	1730405	915	
19	LysR family transcriptional regulator	←	100	0,00E+00	1730483	1731466	984	
20	hypothetical protein	➠	100	0,00E+00	1731587	1732456	870	
21	hypothetical protein Patl_1442	←	99	2,00E-55	1732667	1733038	372	
22	TonB-dependent siderophore receptor	➠	100	0,00E+00	1733293	1735563	2271	
23	peptidase	➠	100	0,00E+00	1735640	1736914	1275	
24	short chain fatty acid transporter	←	100	0,00E+00	1737132	1738448	1317	

#7

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	agarase	➠	100	0,00E+00	2302140	2304242	2102	
2	beta-lactamase	➠	100	0,00E+00	2305159	2306280	1121	
3	thioredoxin domain-containing protein	➠	100	4,00E-74	2307294	2307635	341	
4	hypothetical protein Patl_1908	➠	99	0,00E+00	2307655	2309031	1376	
5	AraC family transcriptional regulator	←	100	0,00E+00	2309153	2309947	794	
6	major facilitator superfamily transporter	➠	99	0,00E+00	2310113	2311489	1376	
7	aldehyde dehydrogenase	➠	100	0,00E+00	2311680	2313143	1463	
8	peptidase M1, membrane alanine aminopeptidase	←	100	0,00E+00	2313351	2315954	2603	
9	AMP-dependent synthetase and ligase	←	100	0,00E+00	2316179	2317828	1649	
10	short-chain dehydrogenase/reductase SDR	←	100	1,00E-175	2317906	2318712	806	KR
11	haloalkane dehalogenase	←	100	0,00E+00	2318688	2319674	986	
12	3-hydroxyacyl-CoA dehydrogenase	➠	100	0,00E+00	2319863	2321980	2117	
13	acetyl-CoA acetyltransferase	➠	100	0,00E+00	2322000	2323196	1196	
14	acyl-CoA dehydrogenase-like protein	➠	100	0,00E+00	2323214	2325046	1832	
15	alpha/beta hydrolase fold protein	➠	100	0,00E+00	2325095	2326057	962	
16	diguanylate cyclase	←	100	0,00E+00	2326160	2327218	1058	
17	hypothetical protein Patl_1921	←	100	1,00E-75	2327535	2327981	446	
18	phytanoyl-CoA dioxygenase	←	100	0,00E+00	2328219	2329451	1232	
19	TetR family transcriptional regulator	➠	100	5,00E-163	2329482	2330186	704	
20	hypothetical protein Patl_1924	➠	100	9,00E-165	2330419	2331174	755	
21	hypothetical protein Patl_1925	➠	99	1,00E-92	2331254	2331676	422	
22	hypothetical protein Patl_1926	➠	100	0,00E+00	2331729	2332709	980	
23	major facilitator superfamily transporter	←	100	0,00E+00	2332720	2333970	1250	

#8

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	6-phosphogluconate dehydrogenase	←	100	0,00E+00	2555257	2556777	1520	
2	phage integrase	➠	100	0,00E+00	2557215	2558453	1238	
3	Pyrrolo-quinoline quinone	←	100	0,00E+00	2558550	2560994	2444	
4	alpha amylase	←	100	0,00E+00	2561287	2563095	1808	
5	hypothetical protein Patl_2114	➠	100	6,00E-106	2563461	2563985	524	
6	TatD family hydrolase	←	99	0,00E+00	2564158	2564940	782	
7	DNA-directed DNA polymerase	←	99	0,00E+00	2564994	2565968	974	
8	thymidylate kinase	←	99	3,00E-152	2565898	2566530	632	
9	aminodeoxychorismate lyase	←	100	0,00E+00	2566595	2567611	1016	
10	class IV aminotransferase	←	100	0,00E+00	2567618	2568469	851	
11	beta-ketoacyl synthase	←	99	0,00E+00	2568566	2569837	1271	KS
12	3-oxoacyl-(acyl-carrier-protein) reductase	←	100	1,00E-158	2570306	2571052	746	KR
13	malonyl CoA-acyl carrier protein transacylase	←	100	0,00E+00	2571058	2572032	974	AT
14	3-oxoacyl-(acyl-carrier-protein) synthase III	←	100	0,00E+00	2572049	2573008	959	KS
15	fatty acid/phospholipid synthesis protein PlsX	←	99	0,00E+00	2573012	2574049	1037	AT
16	hypothetical protein Patl_2127	←	100	3,00E-123	2574231	2574788	557	
17	maf protein	➠	100	5,00E-137	2574825	2575466	641	
18	dTDP-4-dehydrorhamnose reductase	←	99	0,00E+00	2575532	2576416	884	
19	dTDP-4-dehydrorhamnose 3,5-epimerase	←	100	1,00E-132	2576439	2576996	557	
20	glucose-1-phosphate thymidylyltransferase	←	100	0,00E+00	2577015	2577938	923	
21	dTDP-glucose 4,6-dehydratase	←	100	0,00E+00	2577911	2579008	1097	
22	ribonuclease	➠	100	0,00E+00	2580507	2583743	3236	
23	DEAD/DEAH box helicase	➠	100	0,00E+00	2583923	2585233	1310	
24	hypothetical protein Patl_2136	➠	99	0,00E+00	2585302	2587956	2654	

9

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	hypothetical protein Patl_2127	←	100	3,00E-123	2574231	2574788	557	
2	maf protein	➠	100	5,00E-137	2574825	2575466	641	
3	dTDP-4-dehydrorhamnose reductase	←	99	0,00E+00	2575532	2576416	884	
4	dTDP-4-dehydrorhamnose 3,5-epimerase	←	100	1,00E-132	2576439	2576996	557	
5	glucose-1-phosphate thymidylyltransferase	←	100	0,00E+00	2577015	2577938	923	
6	dTDP-glucose 4,6-dehydratase	←	100	0,00E+00	2577911	2579008	1097	
7	RluA family pseudouridine synthase	←	100	0,00E+00	2579054	2580052	998	
8	ribonuclease	➠	100	0,00E+00	2580507	2583743	3236	
9	DEAD/DEAH box helicase	➠	100	0,00E+00	2583923	2585233	1310	
10	hypothetical protein Patl_2136	➠	99	0,00E+00	2585302	2587956	2654	
11	hypothetical protein Patl_2138	➠	100	9,00E-59	2588133	2588471	338	
12	short chain dehydrogenase	←	100	2,00E-173	2588568	2589332	764	KR
14	acyl-CoA dehydrogenase-like protein	←	100	0,00E+00	2590282	2592189	1907	
15	AraC family transcriptional regulator	➠	99	0,00E+00	2592274	2593287	1013	
16	glycoside hydrolase	←	100	0,00E+00	2593859	2594176	317	
17	beta-galactosidase	➠	100	0,00E+00	2594233	2596209	1976	
18	electron transfer flavoprotein subunit alpha	←	100	0,00E+00	2596540	2597493	953	
19	electron transfer flavoprotein subunit beta	←	100	0,00E+00	2597466	2598281	815	
20	electron-transferring-flavoprotein dehydrogenase	➠	99	0,00E+00	2598617	2600293	1676	
21	histone-like nucleoid-structuring protein H-NS	➠	100	2,00E-72	2600460	2600876	416	
22	alpha/beta hydrolase fold protein	➠	100	8,00E-165	2600949	2601746	797	
23	flavodoxin FldA	➠	100	1,00E-122	2602051	2602599	548	
24	ferric uptake regulator	➠	100	3,00E-96	2602813	2603244	431	
25	molybdate ABC transporter inner membrane subunit	←	100	6,00E-123	2603549	2604247	698	
26	molybdenum ABC transporter periplasmic molybdate-binding protein	➠	100	9,00E-32	2604042	2604413	371	

#10

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	phosphorylase kinase alphabeta	➠	100	0,00E+00	2787408	2790827	3420	
2	threonine transporter RhtB	←	100	5,00E-125	2791142	2791795	654	
3	LuxR family transcriptional regulator	←	100	5,00E-154	2791889	2792566	678	
4	ATPase	←	100	0,00E+00	2792533	2793963	1431	
5	TRAP dicarboxylate transporter subunit DctP	➠	100	0,00E+00	2794162	2795244	1083	
6	C4-dicarboxylate ABC transporter	➠	100	4,00E-115	2795355	2795906	552	
7	C4-dicarboxylate ABC transporter	➠	100	4,00E-115	2795904	2797196	1293	
8	hypothetical protein Patl_2295	➠	100	0,00E+00	2797206	2798246	1041	
9	hypothetical protein Patl_2296	➠	100	4,00E-85	2798017	2798730	714	
10	superoxide dismutase	←	100	1,00E-115	2798796	2799359	564	
11	hypothetical protein	➠	100	7,00E-90	2799729	2800136	408	
12	hypothetical protein	←	100	5,00E-89	2800762	2801229	468	
13	chemotaxis protein CheY	➠	100	2,00E-176	2801364	2802107	744	
14	sensor signal transduction histidine kinase	➠	100	0,00E+00	2802101	2803462	1362	
15	pyridine nucleotide-disulfide oxidoreductase	←	100	0,00E+00	2803538	2804797	1260	
16	2Fe-2S ferredoxin	←	100	2,00E-66	2804883	2805194	312	
17	cytochrome P450	←	100	0,00E+00	2805288	2806538	1251	
18	AraC family transcriptional regulator	←	99	0,00E+00	2806652	2807551	900	
19	histidine kinase	➠	100	0,00E+00	2807610	2808800	1191	
20	ATPase	➠	99	0,00E+00	2808712	2809626	915	
21	signal peptidase I	←	100	5,00E-148	2809739	2810437	699	
22	diguanylate cyclase	➠	99	0,00E+00	2810626	2812869	2244	
23	alcohol dehydrogenase GroES-like protein	➠	100	0,00E+00	2812850	2814148	1298	
24	LysR family transcriptional regulator	➠	99	0,00E+00	2814353	2815264	911	
25	response regulator receiver protein	←	100	0,00E+00	2815323	2816189	866	
26	dentin sialophosphoprotein	➠	100	3,00E-49	2816772	2818733	1961	
27	hypothetical protein Patl_2315	➠	100	2,00E-52	2818770	2819162	392	
28	peptidase C39, bacteriocin processing	➠	100	9,00E-154	2819170	2819853	683	

#11

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	periplasmic sensor hybrid histidine kinase	➠	99	0,00E+00	3084696	3086423	1728	
2	protein SapC protein	←	99	3,00E-164	3086533	3087279	747	
3	tryptophan halogenase	←	100	0,00E+00	3087311	3088897	1587	
4	TonB-denpendent receptor	←	100	0,00E+00	3089048	3091870	2823	
5	GntR family transcriptional regulator	←	99	2,00E-143	3092399	3093019	621	
6	uroporphyrinogen decarboxylase	➠	100	0,00E+00	3093130	3094311	1182	
7	molecular chaperone GroES	➠	100	0,00E+00	3094312	3095346	1035	
8	short-chain dehydrogenase/reductase SDR	➠	100	0,00E+00	3095340	3096137	798	KR
9	betaine aldehyde dehydrogenase	➠	100	0,00E+00	3096077	3097606	1530	
10	dethiobiotin synthase	←	100	0,00E+00	3097836	3098786	951	
11	SAM-dependent methlyltransferase	←	99	0,00E+00	3098767	3099753	987	
12	8-amino-7-oxononanoate synthase	←	100	0,00E+00	3098728	3099996	1269	
13	biotin synthase	←	100	0,00E+00	3100960	3102084	1125	
14	adenosylmethionine--8-amino-7-oxononanoate aminotransferase	➠	99	0,00E+00	3102105	3103448	1344	
15	hypothetical protein	←	99	0,00E+00	3103496	3104434	939	
16	invasion protein	←	100	8,00E-82	3104709	3105104	396	
17	metal-dependent phosphohydrolase	←	100	0,00E+00	3105172	3106830	1659	
18	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	←	100	0,00E+00	3106831	3107742	912	
19	MULTISPECIES: peptide chain release factor 1	←	100	0,00E+00	3107733	3108866	1134	
20	shikimate/quinate 5-dehydrogenase	←	100	0,00E+00	3108923	3110281	1359	
21	outer membrane lipoprotein LolB	➠	100	1,00E-150	3110272	3110940	669	
22	4-diphosphocytidyl-2C-methyl-D-erythritol kinase	➠	100	0,00E+00	3110853	3111785	933	
23	MULTISPECIES: ribose-phosphate pyrophosphokinase	➠	99	0,00E+00	3111877	3112893	1017	
24	mechanosensitive ion channel protein	➠	100	0,00E+00	3113026	3114120	1095	

#12

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	peptidase M48, Ste24p	➠	99	0,00E+00	3533970	3534986	1017	
2	hypothetical protein	←	100	2,00E-92	3535033	3535443	411	
3	glutaminyl-tRNA synthetase	←	100	0,00E+00	3536654	3538117	1464	
4	chemotaxis protein CheY	➠	100	0,00E+00	3539793	3541418	1626	
5	kynureninase	←	100	0,00E+00	3541734	3542981	1248	
6	lipase	←	100	0,00E+00	3542978	3544036	1059	
7	acetyl-CoA acetyltransferase	←	100	0,00E+00	3544078	3545295	1218	AT
8	MerR family transcriptional regulator	➠	100	4,00E-77	3545515	3545937	423	
9	isovaleryl-CoA dehydrogenase	➠	100	0,00E+00	3545919	3547103	1185	
10	methylcrotonoyl-CoA carboxylase	➠	99	0,00E+00	3547104	3548801	1698	
11	enoyl-CoA hydratase/isomerase	➠	100	0,00E+00	3548802	3549635	834	
12	carbamoyl-phosphate synthase subunit L	➠	100	0,00E+00	3549620	3551659	2040	
13	hydroxymethylglutaryl-CoA lyase	➠	100	0,00E+00	3551644	3552573	930	
14	hypothetical protein	←	100	9,00E-82	3552613	3553044	432	
15	polyketide cyclase	←	100	2,00E-127	3553169	3553720	552	
16	glucose-1-phosphate adenylyltransferase	➠	100	0,00E+00	3554101	3555423	1323	
17	alpha/beta hydrolase	←	99	0,00E+00	3555489	3556355	867	
18	MFS transporter	←	100	0,00E+00	3556356	3557696	1341	
19	TesB family acyl-CoA thioesterase	➠	100	0,00E+00	3557926	3558750	825	TE
20	histidine kinase	←	100	0,00E+00	3558857	3559753	897	
21	hypothetical protein Patl_2937	←	100	1,00E-172	3560531	3561328	798	
22	flagellar motor protein MotB	←	100	2,00E-111	3561298	3561861	564	
23	translocation protein TolB	←	100	0,00E+00	3561905	3563284	1380	
24	hypothetical protein	←	100	5,00E-70	3563260	3564189	930	

#13

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	hypothetical protein Patl_3056	←	99	2,00E-173	3690757	3691467	38222	
2	formyl transferase-like protein	←	100	2,00E-168	3691468	3692199	38366	
3	type 12 methyltransferase	←	100	1,00E-154	3692166	3692801	38496	
4	DegT/DnrJ/EryC1/StrS aminotransferase	←	100	0,00E+00	3692981	3694093	38707	
5	C-methyltransferase	←	100	0,00E+00	3694090	3695331	38942	
6	dTDP-4-dehydrorhamnose 3,5-epimerase	←	99	3,00E-134	3695307	3695888	39119	
7	CDP-glucose 4,6-dehydratase	←	100	0,00E+00	3695867	3696949	39281	
8	glucose-1-phosphate cytidylyltransferase	←	100	0,00E+00	3696925	3697722	39464	
9	glycosyl transferase family protein	←	100	0,00E+00	3697689	3699008	39669	
10	glycosyl transferase family protein	←	100	0,00E+00	3699060	3699929	39899	
11	glycosyl transferase family protein	←	100	0,00E+00	3700060	3701337	40139	
12	type 11 methyltransferase	←	99	0,00E+00	3701279	3702415	40369	
13	capsule polysaccharide biosynthesis	←	100	0,00E+00	3702385	3703725	40611	
14	UDP-galactopyranose mutase	←	100	0,00E+00	3703688	3704863	40855	
15	glycosyl transferase family protein	←	100	0,00E+00	3704864	3705823	41068	
16	short-chain dehydrogenase/reductase SDR	←	100	5,00E-138	3705903	3706643	41254	
17	hypothetical protein Patl_3072	←	100	0,00E+00	3706621	3708048	41467	
18	hypothetical protein Patl_3074	←	100	2,00E-150	3708230	3708841	41707	
19	spore coat polysaccharide biosynthesis protein glycosyltransferase-like protein	←	100	0,00E+00	3708832	3709977	41881	
20	N-acylneuraminate-9-phosphate synthase	←	100	0,00E+00	3709943	3711043	42098	
21	hypothetical protein Patl_3077	←	100	0,00E+00	3711049	3712164	42321	
22	acylneuraminate cytidylyltransferase	←	100	3,00E-172	3712127	3712828	42495	
23	DegT/DnrJ/EryC1/StrS aminotransferase	←	100	0,00E+00	3712839	3713996	42683	
24	polysaccharide biosynthesis protein CapD	←	100	0,00E+00	3713984	3714988	42897	
25	hypothetical protein Patl_3081	←	99	4,00E-54	3715114	3715452	43056	
26	flagellar protein FliS	←	100	6,00E-100	3715419	3715865	43128	
27	flagellar hook-associated 2-like protein	←	100	0,00E+00	3715870	3717339	43321	
28	flagellar protein FlaG protein	←	100	7,00E-98	3717355	3717798	43515	
29	flagellin-like protein	←	100	0,00E+00	3717919	3718803	43672	
30	hypothetical protein Patl_3086	→	99	5,00E-71	3718738	3719160	43789	
31	flagellin-like protein	←	100	0,00E+00	3720147	3720998	44114	
32	flagellin-like protein	←	100	0,00E+00	3721373	3722206	44358	

#14

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	putative signal transduction protein	➠	99	0,00E+00	3837469	3839271	1803	
2	isocitrate dehydrogenase	➠	100	0,00E+00	3839444	3840454	1011	
3	competence protein ComEA	➠	100	1,00E-53	3840598	3840921	324	
4	UTP--glucose-1-phosphate uridylyltransferase subunit GalU	←	100	0,00E+00	3841002	3841907	906	
5	UDP-glucose 6-dehydrogenase	←	100	0,00E+00	3841987	3843375	1389	
6	hypothetical protein	➠	100	8,00E-08	3843656	3844024	369	
7	hypothetical protein Patl_3202	➠	100	1,00E-164	3844025	3844786	762	
8	hypothetical protein Patl_3203	➠	100	1,00E-180	3844802	3845548	747	
9	hypothetical protein Patl_3204	➠	100	0,00E+00	3845470	3847674	2205	
10	rhamnosyltransferase	←	99	0,00E+00	3847863	3848735	873	
11	glycosyl transferase	←	100	0,00E+00	3848726	3849667	942	
12	glycosyl transferase family protein	←	100	0,00E+00	3849839	3851605	1767	
13	hypothetical protein Patl_3208	←	99	0,00E+00	3851369	3852490	1122	
14	hypothetical protein Patl_3209	←	100	0,00E+00	3852928	3854127	1200	
15	hypothetical protein Patl_3210	←	100	0,00E+00	3854192	3855307	1116	
16	glycerol-3-phosphate cytidylyltransferase	←	100	6,00E-93	3855288	3855701	414	
17	coagulation factor 5/8 type-like protein	←	100	0,00E+00	3855702	3857006	1305	
18	hypothetical protein Patl_3213	←	100	0,00E+00	3857007	3858977	1971	
19	hypothetical protein Patl_3215	←	99	0,00E+00	3860382	3861800	1419	
20	glycosyl transferase family 1	←	99	0,00E+00	3861081	3862355	1275	
21	glycosyl transferase family protein	←	99	0,00E+00	3863042	3864496	1455	
22	glycosyl transferase family protein	➠	100	0,00E+00	3864629	3867340	2712	
23	hypothetical protein Patl_3219	➠	100	0,00E+00	3867363	3868361	999	
24	group 1 glycosyl transferase	←	100	0,00E+00	3868384	3870429	2046	
25	hypothetical protein Patl_3221	←	100	0,00E+00	3870405	3871316	912	
26	hypothetical protein Patl_3222	←	99	0,00E+00	3871373	3872362	990	
27	hypothetical protein Patl_3223	➠	100	0,00E+00	3872446	3873228	783	
28	hypothetical protein Patl_3224	←	100	0,00E+00	3873350	3874252	903	
29	hypothetical protein Patl_3225	←	100	0,00E+00	3874320	3875684	1365	

#15

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	prepilin-type cleavage/methylation protein	←	100	3,00E-144	3877013	3877672	660	
2	prepilin-type cleavage/methylation protein	←	100	1,00E-78	3877673	3878113	441	
3	prepilin-type cleavage/methylation protein	←	100	8,00E-67	3878076	3878525	450	
4	type II secretion system protein G	←	100	9,00E-83	3878506	3878952	447	
5	secretion system protein	←	99	0,00E+00	3879049	3880278	1230	
6	type II secretion system protein	←	100	0,00E+00	3880236	3882023	1788	
7	hypothetical protein Patl_3233	➠	100	0,00E+00	3882072	3883031	960	
8	hypothetical protein Patl_3234	➠	100	2,00E-93	3882977	3883609	633	
9	hypothetical protein Patl_3235	➠	100	2,00E-90	3883584	3884063	480	
10	type II and III secretion system protein	➠	100	0,00E+00	3884103	3886241	2139	
11	hypothetical protein Patl_3237	←	100	9,00E-104	3886380	3886904	525	
12	lipopolysaccharide biosynthesis protein	←	100	0,00E+00	3887354	3888844	1491	
13	ABC transporter ATP-binding protein	←	100	4,00E-149	3888808	3889509	702	
14	hypothetical protein Patl_3237	←	100	9,00E-104	3889455	3890255	801	
15	sugar transporter	←	100	0,00E+00	3890259	3892031	1773	
16	hypothetical protein	←	99	1,00E-83	3892019	3892411	393	
17	undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphatetransferase	←	100	0,00E+00	3892486	3893568	1083	
18	amino acid/peptide transporter	➠	100	0,00E+00	3894046	3895593	1548	
19	hypothetical protein Patl_3245	➠	100	7,00E-36	3895715	3896029	315	
20	UDP-galactose-4-epimerase	➠	100	0,00E+00	3895962	3897050	1089	
21	VacJ-like lipoprotein	←	99	5,00E-171	3897176	3897952	777	
22	hypothetical protein Patl_3248	←	99	0,00E+00	3897940	3899241	1302	
23	cytochrome C biogenesis protein	←	100	7,00E-105	3899238	3899687	450	
24	thiol:disulfide interchange protein DsbE	←	100	3,00E-135	3899684	3900319	636	
25	heme lyase subunit CcmF	←	100	0,00E+00	3900253	3902295	2043	
26	cytochrome C biogenesis protein CcdA	←	100	9,00E-114	3902288	3902839	552	
27	heme exporter protein CcmC	←	100	3,00E-177	3903019	3903963	945	

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
28	heme exporter protein CcmB	←	100	7,00E-111	3903765	3904583	819	
29	heme exporter protein A	←	100	5,00E-135	3904430	3905083	654	
30	hypothetical protein Patl_3257	➠	100	0,00E+00	3905663	3908299	2637	

#16

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	TonB-denpendent receptor	➠	100	0,00E+00	4399367	4402414	3048	
2	heparinase II/III-like protein	➠	100	0,00E+00	4402676	4404937	2262	
3	cation transporter	←	100	0,00E+00	4405076	4406245	1170	
4	hypothetical protein Patl_3653	➠	100	0,00E+00	4406194	4406658	465	
5	NAD(P) transhydrogenase subunit beta	←	100	0,00E+00	4406182	4407579	1398	
6	NAD(P) transhydrogenase subunit alpha	←	100	5,00E-49	4408210	4408536	327	
7	NAD(P) transhydrogenase subunit alpha	←	100	0,00E+00	4408491	4409741	1251	
8	PfkB protein	←	100	0,00E+00	4409816	4410772	957	
9	MFS transporter	←	95	0,00E+00	4410804	4412159	1356	
10	alginate lyase	←	100	0,00E+00	4412160	4414472	2313	
11	GntR family transcriptional regulator	➠	100	5,00E-158	4414918	4415685	768	
12	3-oxoacyl-[acyl-carrier protein] reductase	←	98	4,00E-160	4415783	4416553	771	
13	monooxygenase	←	100	0,00E+00	4416901	4418097	1197	
14	SAM-dependent methyltransferase	←	100	0,00E+00	4418284	4420533	2250	
15	hypothetical protein Patl_3664	←	100	0,00E+00	4420470	4421612	1143	
16	CorA-like Mg2+ transporter protein	←	100	0,00E+00	4421790	4422827	1038	
17	ribosomal protein S6 modification protein	➠	100	0,00E+00	4422996	4423937	942	
18	hypothetical protein Patl_3667	➠	100	1,00E-104	4423913	4424458	546	
19	hypothetical protein Patl_3668	➠	100	2,00E-129	4424640	4425263	624	
20	diguanylate cyclase	←	100	0,00E+00	4425328	4426284	957	
21	TetR family transcriptional regulator	➠	100	5,00E-143	4426749	4427411	663	
22	mechanosensitive ion channel protein	➠	100	9,00E-165	4427701	4428546	846	
23	aldehyde oxidoreductase	←	100	0,00E+00	4428761	4430062	1302	

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
24	XRE family transcriptional regulator	➠	100	0,00E+00	4430067	4430984	918	
25	glycosyl transferase	←	100	0,00E+00	4430970	4432739	1770	

#17

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	OmpA/MotB protein	➠	100	0,00E+00	4801960	4803729	1770	
2	small GTP-binding protein	←	100	1,00E-118	4803741	4804250	510	
3	diguanylate phosphodiesterase	➠	100	0,00E+00	4804244	4806403	2160	
4	acyl-CoA dehydrogenase	←	100	0,00E+00	4806360	4807553	1194	
5	aldo/keto reductase	←	100	0,00E+00	4807609	4808661	1053	
6	hypothetical protein Patl_3995	←	100	2,00E-96	4808715	4809146	432	
7	elongation factor Tu	←	100	0,00E+00	4809741	4810928	1188	
8	pantothenate kinase	←	99	2,00E-158	4811511	4812251	741	
9	bifunctional biotin--[acetyl-CoA-carboxylase] synthetase/biotin operon repressor	←	100	0,00E+00	4812235	4813269	1035	
10	UDP-N-acetylenolpyruvoylglucosamine reductase	←	99	0,00E+00	4813218	4814294	1077	
11	protein-tyrosine kinase	←	100	0,00E+00	4814303	4816462	2160	
12	sugar ABC transporter substrate-binding protein	←	100	2,00E-116	4816432	4817019	588	
13	polysaccharide biosynthesis protein CapD	←	100	0,00E+00	4817020	4818894	1875	
14	aminotransferase	←	100	0,00E+00	4818864	4820069	1206	
15	sugar transferase	←	100	1,00E-133	4820695	4821411	717	
16	glycosyl transferase family 1	←	100	0,00E+00	4821308	4822498	1191	
17	group 1 glycosyl transferase	←	100	0,00E+00	4822489	4823751	1263	
18	hypothetical protein Patl_4008	←	100	0,00E+00	4823667	4825046	1380	
19	group 1 glycosyl transferase	←	100	0,00E+00	4825009	4826118	1110	

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
20	hypothetical protein Patl_4010	←	100	0,00E+00	4826009	4827310	1302	
21	hypothetical protein Patl_4011	←	100	2,00E-154	4827808	4828584	777	
22	carbonic anhydrase	←	100	3,00E-160	4830363	4831028	666	

#bacteriocin

ORFs	Name	⇆	Ident %	e-value	From	To	Size	Domains
1	recombinase RecR	←	100	5,00E-144	3506713	3507315	603	
2	hypothetical protein Patl_2896	←	100	9,00E-52	3507354	3507710	357	
3	ATPase AAA	←	100	0,00E+00	3507711	3510464	2754	
4	adenine phosphoribosyltransferase	←	99	2,00E-123	3510465	3511025	561	
5	conserved hypothetical protein	←	100	9,00E-172	3511242	3511991	750	
6	hypothetical protein Patl_2901	←	100	2,00E-24	3512956	3513528	573	
7	beta-lactamase	←	100	0,00E+00	3513966	3515288	1323	
8	anhydro-N-acetylmuramic acid kinase	←	100	0,00E+00	3515568	3516876	1309	
9	TonB-denpendent receptor	←	100	2,00E-95	3516945	3517634	690	