Contribution to unveiling the roles played by small non-coding RNAs in the biology and pathogenesis of Burkholderia cepacia complex bacteria

Soraia Isabel da Silva Guerreiro

Mestrado em Microbiologia Aplicada

Dissertação orientada por:
Professor Doutor Jorge Humberto Gomes Leitão (IST)
Professor Doutor Rogério Paulo de Andrade Tenreiro (FCUL)
ACKNOWLEDGMENTS

First, I would like to especially thank to Professor Jorge Leitão for his availability, optimism and exchange of scientific knowledge during this thesis, as well as for all the time spent correcting this manuscript. Thank you!

I also thank to Professor Isabel Sá-Correia for kindly receiving me in the BSRG group at Instituto Superior Técnico.

To Professor Rogério Tenreiro, I am thankful for accepting to be my internal supervisor.

I also would like to express my deepest gratitude and admiration to Dr. Sílvia Sousa, who was always there to help me. I recognize the amount of work I was and I thank you for all that you have done as well as for the critical thinking in fundamental scientific issues. Thank you very much, Sílvia. I will be eternally grateful to you.

To Joana Feliciano who was always available to answer my questions, even the dumiest ones, and to help me through this thesis. I am so thankful for all the talks we had, especially those to lift my spirit. Someday I want to be as strong as you are. No words can acknowledge you enough. I wish you all the best.

To Diana and Marília, I thank their incredible support and patience when I was a little grumpy. We made it, girls!

I would also like to thank to all the people of the 6th floor for the many laughs shared at lunch time and suggestions.

Last but not least, I have a very special thanks to my family. Mom and dad, thank you for your unconditional support, even though you are far away, and for helping me pursue my education. Big sis, thank you for all the support, encouragement, optimism and ears pulling throughout this year. Despite we were both in the same situation, when I was ready to throw the towel you were there to not let me to do it and I recognize it was also hard for you. Thank you for being the best sister and for giving me the best brother I could ask for. I also want to thank Sérgio for the patience, understanding and caring.
ABSTRACT

The development of high-throughput sequencing techniques and continued decrease of associated costs, together with advances in bioinformatics and increasing availability of powerful software for nucleotide and amino acid sequence analyses contributed to the exponential increase of available microbial genomes. These developments have unveiled several novel sequences presenting a size within the range of 50-500 nucleotides (nt) and encoding the now called small non-coding regulatory RNAs (sRNAs). These sRNAs are frequently encoded in intergenic regions, often partially overlapping with the 5' or 3' untranslated regions of the vicinal genes or annotated-opening reading frames (ORFs). The current known function for some sRNAs comprises the gene expression regulation through interference RNA mechanisms, mainly at the posttranscriptional level.

Bacterial sRNAs may regulate their targets (messenger RNAs) through positive or negative regulation mechanisms. To carry out a negative regulation, the sRNA base pairs with its target in the region containing the initiation codon and/or the ribosome-binding site (RBS), precluding the ribosome binding and consequently preventing the mRNA translation.

The *Burkholderia cepacia* (Bcc) complex comprises nowadays 20 validated and phylogenetically-related species of Gram-negative bacteria. These bacteria are phenotypically similar and genotypically distinct, being widely distributed in different ecological niches such as soil, water, plants, animals and in humans. Some Bcc bacteria have potential application as bioremediation and biocontrol/biopesticides agents due to their unusual metabolic abilities, which include xenobiotics metabolism, production of antifungal compounds and promotion of plant growth.

However, *Burkholderia* species have emerged in 1980s as important opportunistic pathogens, especially to cystic fibrosis (CF) patients. Bcc bacteria have gained the attention of medical and scientific community because they can cause in CF patients a rapid evolving clinical state known as the cepacia syndrome, which can cause a fast and necrotizing pneumonia, septicemia and ultimately results in the patient early death. Bcc bacteria produce a wide variety of virulence factors, possessing intrinsic resistance mechanisms to different antibiotics which lead to a difficult eradication. Over the last years, some Bcc species have also been recognized as emerging nosocomial pathogenic agents in hospitalized non-CF patients, particularly in cancer patients.

To identify potential genes involved in the Bcc bacteria virulence, researchers from iBB (Instituto de Bioengenharia e Biociências) prepared mutant libraries derived from *B. contaminans* IST408 and *B. cenocepacia* J2315 by random plasposon mutagenesis. This work aims to characterize a mutant derived from *B. cenocepacia* J2315, identified in an attenuated virulence screening using as model of infection the nematode *Caenorhabditis elegans*. This mutant, *B. cenocepacia* SJ2, contains the plasposon inserted in the intergenic region of *B. cenocepacia* J2315 chromosome 1, located between the coding gene of DNA gyrase subunit A (GyrA) enzyme and the coding gene of an outer membrane protein of the OmpA family.

Bioinformatic and Southern blot analyses allowed the identification of the plasposon insertion in the mutant *B. cenocepacia* SJ2. The plasposon is located 8 nucleotides (nt) after the predicted 5' UTR beginning of the *B. cenocepacia* J2315 OmpA-like protein. Therefore, the presence and functionality of the transcript *ompA* were evaluated by RT-PCR and Western blot, respectively. The results obtained indicated that the plasposon insertion did not affect *ompA* mRNA transcription or functionality.

Bioinformatic analyses also led to the identification of a possible opening reading frame (ORF) in the complementary strand (named ORF3) and encoding a 93 amino acids protein. The presence of this ORF was supported by Northern blot assays previously performed. These assays allowed the identification of
a transcript from the complementary strand. However, the amplification of complementary DNA ends (5’ and 3’ RACE) using specific primers for ORF3 region demonstrated that the transcript in study has a 5’ end with a lower size than the expected transcript of the potential protein (ORF3). A transcript containing approximately 179 base pairs and non-interrupted by the plasposon was identified. In the presence of these results, primers for the potential DNA sequence of ORF3 were designed and the results showed that this is the region interrupted by the plasposon.

Due to the absence of initiation and termination codons in the region closer to the 179 base pairs transcript, it was hypothesized that the transcript corresponds to a sRNA (MavA). In this work, the presence of two genetic elements, a sRNA and a protein, in the intergenic region of *B. cenocepacia* J2315 chromosome 1 is described. The coding sequences of both the sRNA and protein are overlapped in, at least, 100 nucleotides.

Bioinformatic analysis of the genetic elements showed that they are conserved among the *Burkholderia* genus. Moreover, the sequence of ORF3 protein also showed to be identical to the sequence of one protein encoded by an annotated ORF in the *B. multivorans* genome and to five proteins of *B. pseudomallei*.

The assays performed with the strain containing the ORF3 interrupted by the plasposon revealed the involvement of this protein in the resistance of *B. cenocepacia* J2315 strain to heat-shock stress (50 ºC), susceptibility to the detergent sodium dodecyl sulphate (SDS), biofilms formation, cellular hydrophobicity and antibiotics resistance (imipenem, ceftazidime and tetracycline). In general, these results suggest that this protein is likely involved in the maintenance of the outer membrane integrity and virulence of *B. cenocepacia* J2315 towards the nematode *C. elegans*.

MavA sRNA identified in the work herein presented was predicted to be a functional homologue of IstR-2 sRNA of *Escherichia coli* K-12 MG1655. Preliminary results of the MavA sRNA overexpression characterization indicate a possible direct or indirect role in the overexpression of ribosomal protein S12. Phenotypic analysis also showed the involvement in the swimming motility of *B. cenocepacia* J2315 strain and in resistance to thermal (50 ºC) stress. In addition, this sRNA is not involved in the virulence of *B. cenocepacia* J2315 towards the *C. elegans* nematode.

Overall, the results presented in this study contribute to a better knowledge of the intergenic region under study and to the identification of a putative ORF and a sRNA, contributing to a better understanding of the biology of Bcc bacteria and the role of sRNAs in the regulation of the expression of putative virulence factors. The identification of MavA sRNA and ORF3 protein highlight the importance of these studies in identifying genetic elements that might be exploited as targets for the development of effective treatments to the *B. cenocepacia* J2315 strain infections.

**Keywords**

*Burkholderia cepacia* complex; intergenic region; sRNAs; posttranscriptional regulation; ORF
RESUMO

O recente desenvolvimento de técnicas de elevado rendimento para sequenciar genomas e a diminuição contínua do custo associado, conjuntamente com os avanços na bioinformática e a disponibilidade de número crescente de ferramentas bioinformáticas para análise de sequências nucleotídicas e aminoácidas, tornou possível o aumento exponencial dos genomas microbianos disponíveis. Consequentemente, têm vindo a ser descobertos diversos pequenos transcritos com ação regulatória denominados de pequenos RNAs reguladores não codificantes (sRNAs), com tamanhos compreendidos entre 50-500 nucleótidos. Estes sRNAs têm sido identificados essencialmente em regiões intergénicas, junto a genes ou grelhas de leitura (ORFs) anotadas. A função conhecida de um grupo importante desses sRNAs consiste na regulação da expressão génica, principalmente ao nível pós-transcricional, por mecanismos de interferência de RNA.

Os sRNAs bacterianos podem exercer nos seus alvos (RNAs mensageiros) uma regulação positiva ou negativa. No caso de ser exercida uma regulação negativa o sRNA estabelece com o seu alvo um emparelhamento na região que contém o codão de iniciação e/ou a região de ligação do ribossoma (RBS), tornando este local inacessível para a ligação do ribossoma e impedindo, assim, a tradução do mRNA.

O complexo *Burkholderia cepacia* (Bcc) é atualmente constituído por 20 espécies de bactérias Gram-negativas validadas e filogeneticamente próximas. As bactérias que compõem este grupo são fenotipicamente semelhantes e genotipicamente distintas, podendo ser isoladas de várias fontes. São exemplo de fontes o solo, a água, a rizosfera de plantas, animais e Humanos. Devido às suas capacidades metabólicas invulgares, algumas estirpes do Bcc apresentam um elevado potencial de aplicação ao nível do biocontrolo, bioremedição e agricultura, pois são capazes de metabolizar xenobióticos, produzir compostos com atividade antifúngica e promover o crescimento de plantas.

Contudo, na década de 80 as bactérias incluídas no género *Burkholderia*, incluindo as englobadas no complexo Bcc, emergiram como agentes patogénicos oportunistas em indivíduos com fibrose quística (FQ).

O facto de as bactérias deste complexo produzirem vários fatores de virulência, apresentarem mecanismos de resistência a um largo espetro de antibióticos sendo difíceis de erradicar e de causarem em doentes com FQ infeções por vezes acompanhadas por um estado clínico de evolução rápida que inclui pneumonia necrotizante e septicémia, levando à morte do doente – síndrome cepacia –, faz com que tenham merecido uma elevada atenção da comunidade médica e científica. Acresce que algumas espécies do Bcc têm vindo a ser reconhecidas nos últimos anos como sendo agentes patogénicos nosocomiais emergentes em doentes hospitalizados, principalmente em doentes oncológicos.

Com o objetivo de identificar possíveis genes envolvidos na virulência de bactérias do Bcc, investigadores do iBB (Instituto de Bioengenharia e Biociências) construíram bibliotecas de mutantes de *B. contaminans* IST408 e *B. cenocepacia* J2315 utilizando plasposões. As bibliotecas de mutantes foram rastreadas tendo como objetivo a identificação de mutantes que exibiam virulência atenuada, utilizando como modelo de infeção o nemátodeo *Caenorhabditis elegans*. O trabalho aqui apresentado teve como objetivo a caracterização de um mutante identificado no rastreio acima mencionado, contendo um plasposão inserido na região intergénica do cromossoma 1 de *B. cenocepacia* J2315, localizada entre o gene codificante da subunidade A da enzima ADN girase (GyrA) e o gene codificante de uma proteína da membrana externa do tipo A (OmpA).

Através de análise bioinformática e experimentais, foi possível localizar no mutante *B. cenocepacia* SJ2 a inserção do plasposão 8 nucleotídos (nt) após o início da região 5’ não traduzida (5’ UTR) do gene
que codifica a proteína do tipo OmpA. Neste sentido, foram avaliados a presença e a funcionalidade do transcrito ompA por ensaios de RT-PCR e Western blot, respectivamente. Os resultados obtidos confirmaram que a inserção do plasposão não afetou a transcrição nem a funcionalidade do transcrito correspondente ao gene ompA.

As análises bioinformáticas realizadas permitiram ainda a identificação de uma possível grelha de leitura aberta (ORF) na cadeia complementar da região intergénica (designada de ORF3), codificando para uma proteína de aproximadamente 93 aminoácidos. A confirmação da presença desta ORF foi suportada por ensaios de Northern blot realizados anteriormente e que permitiram a identificação de um transcrito a partir da cadeia complementar. No entanto, análises de amplificação em cadeia pela polimerase (PCR) das extremidades do ADN complementar (5’ e 3’ RACE) com sequências oligonucleotídicas iniciadoras específicas para a região da ORF3 demonstraram que o transcrito em causa possui uma extremidade 5’ menor que o esperado para o transcrito da possível proteína (ORF3), permitindo assim a identificação de um transcrito com cerca de 179 pares de bases que se verificou não ser interrompido pelo plasposão. Deste modo, tendo em conta os resultados aqui apresentados foram desenhadas sequências oligonucleotídicas iniciadoras específicas para a sequência de ADN da possível ORF3, tendo-se verificado que esta era a região interrompida pelo plasposão.

Devido à ausência de codões de iniciação e de terminação na região próxima do transcrito de 179 pares de base, colocou-se a hipótese deste transcrito corresponder a outro elemento genético, nomeadamente um sRNA, que se denominou de MavA. Neste trabalho é descrita a existência de dois elementos genéticos na região intergénica do cromossoma 1 de B. cenocepacia J2315, um sRNA e um gene que codifica uma proteína, sendo que as sequências codificantes de ambos encontram-se parcialmente sobrepostas em, pelo menos, 100 nucleotídos.

As análises bioinformáticas realizadas dos elementos genéticos permitiram a sua identificação como sendo ambos conservados no génere Burkholderia. Além disso, demonstraram também que a proteína ORF3 é idêntica à sequência de uma proteína codificada por uma ORF anotada no genoma de B. multivorans e à de cinco proteínas de B. pseudomallei.

Ensaios realizados com a estirpe contendo o plasposão a interromper a ORF3 demonstraram que esta proteína está envolvida na resistência da estirpe B. cenocepacia J2315 a choque térmico (50 ºC), susceptibilidade à presença do detergente dodecil sulfato de sódio (SDS), capacidade para formação de biofilmes e hidrofobicidade das células, bem como na resistência a antibióticos (imipeneme, ceftazidima e tetraciclina). De um modo geral, estes resultados apontam para que esta proteína esteja envolvida na manutenção da integridade da membrana externa e virulência da estirpe B. cenocepacia J2315 para o nemático C. elegans.

O sRNA MavA identificado neste trabalho foi bioinformaticamente previsto como tendo como seu homólogo funcional o sRNA IstR-2 de Escherichia coli K-12 MG1655. Os resultados preliminares da caracterização do sRNA MavA através da sua sobreexpressão demonstram um possível envolvimento direto ou indireto na sobreexpressão da proteína ribossomal S12. As análises fenotípicas mostraram ainda o seu envolvimento na mobilidade da estirpe B. cenocepacia J2315 por swimming e na resistência ao choque térmico (50 ºC). Os resultados obtidos, baseados em ensaios de morte lenta do nemático C. elegans sugerem que este sRNA não está envolvido na virulência de B. cenocepacia J2315 neste modelo animal de infeção.

Os resultados obtidos no presente trabalho contribuem para um conhecimento mais aprofundado da região intergénica estudada, tendo permitido identificar 2 elementos genéticos parcialmente sobrepostos que codificam uma proteína e um sRNA. Este trabalho constitui assim um contributo para o melhor...
conhecimento da biologia das bactérias do complexo Bcc e do papel dos sRNAs na regulação da expressão de fatores de virulência. A identificação do sRNA MavA e da proteína ORF3 realça a importância deste tipo de estudos na identificação de elementos genéticos que possam ser explorados como alvos para o desenvolvimento de estratégias terapêuticas mais eficazes com vista ao tratamento das infeções causadas pela estirpe *B. cenocepacia* J2315.

**Palavras-chave**

Complexo *Burkholderia cepacia*; região intergênica; pequenos RNAs regulatórios não codificantes; regulação pós-transcricional; grelha de leitura aberta
# TABLE OF CONTENTS

Acknowledgments .......................................................... I

Abstract ..................................................................... III

Resumo ....................................................................... V

Table of contents ........................................................... IX

List of figures ................................................................ XIII

List of tables .................................................................. XIV

Abbreviations ................................................................. XV

1. Introduction .................................................................. 1

1.1 Cells harbour different classes of RNAs .................. 1

1.2 Bacterial small non-coding RNAs (sRNAs) ............ 1

1.3 General regulatory RNAs classes ............................. 2

1.3.1 Regulatory 5’ UTR elements ............................... 2

1.3.2 cis-encoded antisense sRNAs .............................. 3

1.3.3 Trans-encoded sRNAs ........................................ 4

1.3.3.1 mRNAs 3’ regions are rich reservoirs for sRNAs .. 5

1.3.4 Regulatory sRNAs of protein activity .................. 6

1.4 Cell advantages of regulating with sRNAs ............... 7

1.5 *Burkholderia cepacia* complex (Bcc) – overview ..... 7

1.5.1 Bcc species as opportunistnic pathogens in Humans .. 9

1.5.2 Major virulence factors of the Bcc ....................... 10

1.5.3 Identification of novel Bcc virulence factors .......... 11

2. MATERIALS AND METHODS ................................. 13

2.1 Bacterial strains, plasmids, nematode and culture conditions .................................................. 13

2.2 Molecular biology techniques ................................. 14

2.2.1 Extraction and purification of genomic and plasmid DNA ...................................................... 14

2.2.2 Construction of a plasmid expressing the MavA sRNA .......................................................... 14

2.2.3 Insertion of foreign DNA in bacterial cells .......... 14

2.2.3.1 Heat-shock transformation of *E. coli* cells .......... 14

2.2.3.2 Triparental mating of *B. cenocepacia* cells .......... 15

2.2.4 DNA sequence analysis ........................................ 15
2.2.5 Extraction and purification of total RNA 15
2.2.6 Reverse transcription-polymerase chain reaction (RT-PCR) 16
2.2.7 Rapid amplification of cDNA ends (RACE) 16
2.2.8 Analysis of total soluble proteins by Tricine-SDS-PAGE 17
2.2.9 Western blot experiments 17

2.3 Nematode killing assays 18
2.3.1 Maintenance of C. elegans and egg preparation 18
2.3.2 C. elegans killing assays 18

2.4 Biofilm formation and cell surface hydrophobicity assays 18
2.4.1 Biofilm formation ability 18
2.4.2 Bacterial adhesion to hexadecane 19

2.5 Motility assays 19
2.5.1 Swimming assays 19
2.5.2 Twitching motility assays 19

2.6 Stress susceptibility experiments 19
2.6.1 Heat stress susceptibility 19
2.6.2 Anionic detergent shock experiments 20

2.7 Antibiotic susceptibility testing 20

2.8 Statistical analysis 20

2.9 Bioinformatics analysis 21
2.9.1 MavA sRNA bioinformatics analysis 21
2.9.2 Hypothetic ORF bioinformatics analysis 21

3. RESULTS AND DISCUSSION 23

3.1 Characterization of the B. cenocepacia J2315 chromosome 1 intergenic region flanked by genes gyrA and ompA 23

3.2 Biocomputational analysis and phenotypic characterization of the hypothetical protein encoded by ORF3 26
3.2.1 The hypothetical protein ORF3 is conserved among Burkholderia and has high homology with annotated ORFs 26
3.2.2 B. cenocepacia J2315 ORF3 does not impair growth ability 28
3.2.3 B. cenocepacia SJ2 mutant strain forms thicker biofilms and has an increased cell surface hydrophobicity 29
3.2.4 The twitching motility of B. cenocepacia SJ2 strain is apparently affected 30
3.2.5 B. cenocepacia SJ2 strain shows increased mortality at 50 °C and cell lysis rate in the presence of SDS 30
3.2.6 The B. cenocepacia SJ2 strain is more susceptible to ceftazidime, imipenem and tetracycline 32
3.2.7 B. cenocepacia SJ2 mutant strain exhibits decreased ability to kill the nematode C. elegans 32
3.3 Biocomputational analysis and phenotypic characterization of MavA sRNA

3.3.1 MavA sRNA is specific of the *Burkholderia* genus and has one functional homologue

3.3.2 MavA putative targets in *B. cenocepacia* J2315 genome

3.3.3 MavA overexpression leads to S12 protein overexpression

3.3.4 MavA overexpression does not affect *B. cenocepacia* J2315 ability to grow in minimal medium

3.3.5 MavA overexpression affects *B. cenocepacia* J2315 resistance to heat-shock

3.3.6 MavA overexpression does not affect biofilm formation

3.3.7 MavA sRNA overexpression does not affect the ability of *B. cenocepacia* J2315 to kill the nematode *C. elegans* significantly

4. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

References

Annex
**LIST OF FIGURES**

**Figure 1.1** Schematic representation of the three levels of gene expression regulation involving regulatory RNAs.  
**Figure 1.2** Regulatory mechanisms employed by *cis*-encoded antisense sRNAs are based on two configurations.  
**Figure 1.3** Regulatory mechanisms employed by *trans*-encoded sRNAs.  
**Figure 1.4** Major mechanisms of action of the RNA-binding protein Hfq.  
**Figure 1.5** sRNAs from bacterial 3’ UTRs.  
**Figure 3.1** Detection by RT-PCR of the transcript from the ORF putatively encoding an OmpA-like protein, located in *B. cenocepacia* J2315 chromosome 1 (A) and assessment of the *ompA* mRNA functionality by Western blot (B).  
**Figure 3.2** ORF prediction.  
**Figure 3.3** Analysis of 5’ and 3’ Rapid Amplification of cDNA Ends (RACE) in 2% (w/v) agarose gel electrophoresis.  
**Figure 3.4** Detection by RT-PCR of the transcript obtained by 5’ and 3’ RACE from *B. cenocepacia* J2315 chromosome 1.  
**Figure 3.5** Detection of the transcript from the non-annotated ORF3 from *B. cenocepacia* J2315 by RT-PCR.  
**Figure 3.6** Genetic organization of the IGR *locus* interrupted by the plasposon.  
**Figure 3.7** Prediction of the consensus secondary structure of *B. cenocepacia* J2315 ORF3 protein, by JPre4 web server.  
**Figure 3.8** Growth curves of *B. cenocepacia* strains J2315 and SJ2.  
**Figure 3.9** Tricine-SDS-PAGE analysis of the total soluble proteins obtained from *B. cenocepacia* J2315 or *B. cenocepacia* SJ2 strains.  
**Figure 3.10** Assessment of biofilm formation by *B. cenocepacia* J2315 or *B. cenocepacia* SJ2 strains in polystyrene microtiter plates.  
**Figure 3.11** Assessment of *B. cenocepacia* J2315 and *B. cenocepacia* SJ2 cells hydrophobicity.  
**Figure 3.12** Comparison of *B. cenocepacia* J2315 and *B. cenocepacia* SJ2 swimming and twitching motility.  
**Figure 3.13** *B. cenocepacia* SJ2 mutant strain is more susceptible to heat-shock stress than the WT strain.  
**Figure 3.14** *B. cenocepacia* SJ2 mutant strain exhibits an increased cell lysis rate upon exposure to SDS, when compared to *B. cenocepacia* J2315.  
**Figure 3.15** Comparison of *E. coli* OP50, *B. cenocepacia* J2315 and *B. cenocepacia* SJ2 ability to kill the nematode *C. elegans*.  
**Figure 3.16** Prediction of putative promoters and *Rho*-independent transcription terminators of MavA.  
**Figure 3.17** Prediction of the secondary structure of MavA sRNA.  
**Figure 3.18** Results of the predicted homologue retrieved from the analysis with sRNAMap tool.  
**Figure 3.19** Tricine-SDS-PAGE analysis of the total soluble proteins obtained from Bcc strains under study.  
**Figure 3.20** Growth curve comparison.  
**Figure 3.21** MavA overexpression in *B. cenocepacia* J2315 cells increases its resistance to heat-shock stress.  
**Figure 3.22** Biofilm formation by *B. cenocepacia* J2315, *B. cenocepacia* J2315 (pMLS7) and *B. cenocepacia* J2315 (pMya2) strains in polystyrene microtiter plates.
Figure 3.23 MavA sRNA influences swimming and twitching motility

Figure 3.24 Comparison of *E. coli* OP50 (purple bars), *B. cenocepacia* J2315, *B. cenocepacia* J2315 (pMLS7) and *B. cenocepacia* J2315 (pMya2) ability to kill the nematode *C. elegans*.

LIST OF TABLES

Table 1.1 *Burkholderia cepacia* complex (Bcc) species, their main sources and biotechnological applications according to the List of Prokaryotic Names with Standing in Nomenclature (LPSN) 8

Table 2.1 Bacterial strains, plasmids and nematode used in this study. 13

Table 3.1 Antibiotic susceptibility of *B. cenocepacia* J2315 and *B. cenocepacia* J2315 derived mutant SJ2. 32

Table 3.2 Predicted mRNA targets of MavA in *B. cenocepacia* J2315 genome with TargetRNA2 tool. 37
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>Amino acids</td>
</tr>
<tr>
<td>ANI</td>
<td>Average nucleotide identity</td>
</tr>
<tr>
<td>BATH</td>
<td>Bacterial adhesion to hydrocarbon</td>
</tr>
<tr>
<td>Bcc</td>
<td><em>Burkholderia cepacia</em> complex</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>cAMP</td>
<td>3’,5’-cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CF</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CFTR</td>
<td>Cystic Fibrosis Transmembrane Conductance Regulator</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony forming units</td>
</tr>
<tr>
<td>CGD</td>
<td>Chronic granulomatous disease</td>
</tr>
<tr>
<td>Csr</td>
<td>Carbon storage regulator protein</td>
</tr>
<tr>
<td>EPS</td>
<td>Exopolysaccharides</td>
</tr>
<tr>
<td>iBB</td>
<td>Institute for Bioengineering and Biosciences</td>
</tr>
<tr>
<td>IGR</td>
<td>Intergenic region</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>LB</td>
<td>Lennox Broth</td>
</tr>
<tr>
<td>LBB</td>
<td>Luria-Bertani Broth</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>LPSN</td>
<td>List of Prokaryotic Names with Standing in Nomenclature</td>
</tr>
<tr>
<td>Lrp</td>
<td>Leucine-responsive regulatory protein</td>
</tr>
<tr>
<td>MH</td>
<td>Mueller Hinton</td>
</tr>
<tr>
<td>MIC</td>
<td>Minimal inhibitory concentration</td>
</tr>
<tr>
<td>MLST</td>
<td>Multilocus sequence typing</td>
</tr>
<tr>
<td>MMC</td>
<td>Mytomicin C</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NCBI</td>
<td>National center for biotechnology information</td>
</tr>
<tr>
<td>NGM</td>
<td>Nematode growth medium</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotides</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>Omp</td>
<td>Outer membrane protein</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PIA</td>
<td><em>Pseudomonas</em> Isolation Agar</td>
</tr>
<tr>
<td>QS</td>
<td>Quorum-sensing</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosome-binding site</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>RNAseq</td>
<td>RNA sequencing</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RPM</td>
<td>Rotations per min</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>SD</td>
<td>Shine-Dalgaro</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulphate</td>
</tr>
<tr>
<td>sRNA</td>
<td>Small non-coding RNA</td>
</tr>
<tr>
<td>Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Streptomycin resistant</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start sites</td>
</tr>
<tr>
<td>UTRs</td>
<td>Untranslated regions</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>x g</td>
<td>Unit of time gravity</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 Cells harbour different classes of RNAs

The work developed by Hoagland et al. [1] has played a major role in helping us to understand the DNA translation mechanism to proteins. Transfer RNAs (tRNAs) were first described by this research group as “soluble RNAs” that when bound to amino acids (aa) caused their activation. However, an important piece of the puzzle was still missing and it was only in 1961 that the concept of messenger RNA (mRNA) based on the Jacob-Monod Model [2] was biochemically demonstrated by Brenner et al. [3] and Gros et al. [4].

Cells harbor several types of RNA molecules with different functions, which can be divided into two major classes i) protein coding RNAs and ii) non-coding RNAs [5]. These two classes are discriminated based on the presence or absence of an open reading frame (ORF), respectively [5–7]. mRNA contains the genetic information copied from DNA coded in triplets of “letters” and is the only RNA belonging to the class (i) [8]. Non-coding RNAs (ii) are a much more diverse group and due to their function as regulators in different steps of gene expression, they are recognized also as regulatory RNAs. Therefore, non-coding RNA class includes, among others, a) tRNAs which are responsible for the growth of the polypeptide chains due to the presence of a three base sequence that base pairs with its complementary triplet present in mRNA [1,8]; b) ribosomal RNAs (rRNAs) that are associated with proteins consisting of essential structures of the translation process - ribosomes, and c) small non-coding RNAs (sRNAs), a relatively new class of non-coding RNAs that can have housekeeping or exclusively posttranscriptional regulatory roles [9,10].

1.2 Bacterial small non-coding RNAs (sRNAs)

Bacteria are present in different ecological niches where they are often exposed to changing and hostile environments such as those encountered during host infection processes. The changes that occur in these environments can be fast and enclose fluctuations in pH, nutrient availability, temperature and presence of reactive oxygen and nitrogen species (reviewed in [11,12]).

However, bacteria can adapt to these conditions, i.e. they are able to maintain cellular homeostasis and thrive, due to modifications in gene and protein expression [7,13]. These alterations are caused by a regulation that can occur at different stages of gene expression (such as transcription and translation) and can affect mRNA stability, as well as DNA maintenance or silencing. The mechanisms involved in these gene regulatory networks are based on changes in RNA conformation, protein-DNA, RNA-RNA and RNA-DNA interactions (Fig. 1.1) (reviewed in [14]).

Technical advances in nucleic acid sequencing methodologies have significantly contributed to the amount of known microbial genomes presently available. Consequently, it is possible to produce biocomputational genome-wide searches. Allied to these techniques, approaches relying on direct detection are important to define transcription start sites (TSS) and 5’ untranslated regions (5’ UTRs), such as tilling arrays and high-throughput RNA sequencing (RNAseq). These techniques allowed the discovery of sRNAs loci not only in several pathogenic and non-pathogenic bacteria but also in the other two domains of life (Archaea and Eukarya) [9,12,15,16].

sRNAs are an heterogeneous group of molecules [14]. They are primarily encoded in chromosomes, being most of them located in intergenic regions (IGRs), between annotated protein-coding genes and in the complementary strand of ORFs [17,18], although they have also been found in plasmids [19]. There are two relevant features in sRNAs namely a) their size and b) structure. Despite the absence of a
consensus among the authors about the size of these molecules, it is in general within the range of 50-500 nucleotides (nt) [12,15,20]. Regarding its structure, sRNAs enclose an initial stable stem-loop and they are often associated with Rho-independent terminators of transcription which consist in a GC-rich stem-loop followed by a set of U’s. These stem-loops may be the main reason for the high stability observed in these molecules [21,22].

Figure 1.1 Schematic representation of the three levels of gene expression regulation involving regulatory RNAs. Transcription regulation is a process mainly mediated by protein-DNA interactions which occurs at the gene promoter region (i). Posttranscriptional regulation. The majority of the regulatory RNAs known so far act at this level of regulation causing conformational changes in the binding molecules or enzymes (ii). Posttranscriptional modification is relevant due to the fact that some proteins are produced in an inactive form and have to undergo covalent or conformational changes to display their activity (iii). Adapted from [23].

The majority of the sRNAs identified so far belong to Escherichia coli. Although their functions are not fully described, it has been shown that they mainly act as negative regulators of gene expression [21]. In fact, the first chromosomally encoded regulator identified was the E. coli MicF RNA, which binds imperfectly to the 5’ region of the ompF mRNA near the start codon. In view of this, translation of the mRNA encoding the major outer membrane porin F type (OmpF) is inhibited [24].

1.3 General regulatory RNAs classes

In a first approach, two classes of regulatory non-coding RNAs can be distinguished based on the location of their target: cis-encoded and trans-encoded molecules. The cis-encoded molecules are RNA regulators encoded in the same genetic location of the target RNA (directly on the same - sense- or on the opposite strand - antisense), whereas trans-encoded refers to regulators that modulate the expression of a mRNA transcribed from a distinct genetic locus [12,14].

1.3.1 Regulatory 5’ UTR elements

Riboswitches are a group of non-coding RNAs with a regulatory function. However, they are not a standard sRNA since they are transcribed from the 5’ UTR of the mRNA that they regulate, being cis-encoded and cis-acting elements [25]. In response to several environmental signals, both chemical and physical, these regulatory elements change their conformation leading to the modulation of expression of downstream genes [25]. The riboswitch conformation is altered upon binding of the leader sequence to small molecules, repressing transcription or translation through the disruption of transcriptional terminators/antiterminators or preventing binding to the ribosome-binding site (RBS) region [14].
RNA thermometers are a group of these molecules which specifically respond to fluctuations in temperature, being particularly important in pathogenic bacteria since the host body temperature induces expression of specific virulence genes [26]. Their regulation is achieved by an interchange between two conformations: a) “closed” structure, present at low temperatures where the Shine-Dalgarno (SD) sequence and/or AUG start codon are inaccessible for translation initiation and b) “open” structure, present at high temperatures when the secondary inhibitory structure around the RBS melts [12].

These RNA elements are composed by the “aptamer region” domain which functions as ligand-binding domain and the “expression platform” domain, being the effector domain on the downstream genes [27].

1.3.2 *Cis-encoded antisense sRNAs*

In addition to the environmental sensors aforementioned there is another class of *cis*-encoded RNA regulators, being the majority of them constitutively expressed [14]. sRNAs residing on plasmids were the first to be described as members of this class [28].

Their mechanism of action consists on base pairing with their target mRNAs. Since these sRNAs are mainly transcribed from the DNA strand opposite of their target, generally overlapping the 5’ end, they are named antisense sRNAs [13]. Consequently, they display extended regions of full complementarity with their target, around 75 or more nucleotides, and can exert a downregulation on the target genes. This downregulation can occur at the level of mRNA stability, transcription or translation (Fig. 1.2) [14].

![Figure 1.2 Regulatory mechanisms employed by *cis*-encoded antisense sRNAs are based on two configurations. Base pairing between a sRNA (encoded opposite to the 5’ UTR of its target mRNA) and its target, prevents the ribosome binding to the RBS. This event inhibits the translation process and promotes sRNA-mRNA complex degradation (A). A few sRNAs are encoded in regions complementary to the sequence separating ORF’s (B) and base pairing of these sRNAs can target RNases causing mRNA cleavage (1) or transcriptional termination through the cessation of RNA polymerase activity due to a putative loop formation (2). In the last case, a reduction in the expression levels of downstream genes is induced. The grey arrows represent the promoter. Adapted from [14].](image)

A few years ago, a new subset of these sRNAs was identified as negative regulators of the expression of mRNAs encoding potentially toxic proteins, when present in high intracellular concentrations [29]–toxin-antitoxins (reviewed in [30]). In this case, the antisense sRNA acts as an antitoxin and it is
responsible for maintaining a low basal level of toxin expression. Although this mechanism is not fully understood, it is thought that it occurs substantially by a translation block (reviewed in [19,29]).

In addition to the already mentioned role as antitoxin, the group of cis-encoded antisense sRNAs can also be involved in direct processing of mRNAs encoded in its opposite strands (Fig. 1.2. B1) [19]. The E. coli GadY is a cis-encoded RNA that it is encoded in the opposite strand of the operon gadXW, overlapping the 3’ end of the gadX mRNA [31]. Opdyke et al. [31] and Tramonti et al. [32] have shown that in cells in the stationary phase, the GadY antisense RNA levels increase, promoting the generation of base pairing with the 3’ UTR of gadX mRNA. This binding promotes the cleavage of the duplex gadX-gadW, which is involved in the response to acidic conditions, increasing the intracellular levels of gadX transcript. Therefore, GadY acts as a positive posttranscriptional regulator of gadX.  

1.3.3 Trans-encoded sRNAs

Unlike cis-encoded sRNAs, trans-encoded sRNAs are only expressed under stress conditions. These sRNAs have been encountered in specific conditions such as i) iron limitation [33], ii) oxidative stress [34], iii) low temperature [35] and iv) accumulation of non-metabolizable phosphoglucone molecules such as glucose-phosphate [36].

As already mentioned, trans-encoded sRNAs are encoded in a different genetic location relatively to their mRNA target [12,14]. In view of this, a limited complementarity is shared between the sRNA and its target, usually in discontinuous patches. In this case, the region where the sRNA overlaps the target mRNA is known as the “seed sequence” [19]. Seed regions of only 7 contiguous bases have been found for some sRNAs as being required for regulation [37,38].

The base pairing between sRNA and mRNA can result in positive or negative regulations. In both cases, mRNA translation and/or stability can be affected. However, the majority of trans-encoded sRNAs known so far, display a negative regulation through translational inhibition, mRNA degradation or both (Fig. 1.3).

Almost 50 years ago, Franze de Fernandz et al. [39] identified Hfq protein in E. coli as an important host factor to the replication of RNA bacteriophage Qβ. However, it was only in the last decade that this global metabolism regulator, acting as a RNA chaperone, was found to be required by several Gram-
positive and Gram-negative bacteria at the posttranscriptional level of regulation. This Sm-like protein appears to act as a mediator of the short and imperfect interaction established between *trans*-encoded sRNAs and their target mRNAs, due to their limited complementarity. In this case, it stimulates pairing between sRNA-mRNA by increasing the annealing rates (reviewed in [40]). The protein has the ability to bind not only sRNAs, but also mRNAs and in both cases it binds A/-U- rich sequences near hairpins [41]. Furthermore, Hfq is also involved in regulation through the modulation of the sRNA levels (Fig. 1.4).

Unlike what would be expected, the Hfq protein is not required in all cases of limited pairing but the reason behind this observation is still poorly understood.

![Figure 1.4 Major mechanisms of action of the RNA-binding protein Hfq](image)

**1.3.3.1 mRNAs 3’ regions are rich reservoirs for sRNAs**

The majority of sRNAs that have been investigated and characterized are encoded by free-standing genes located in IGRs and having a single output function associated. However, some protein-coding genes have been observed to possess dual-function since they produce from its 5’ region a mRNA template for translation and from its 3’ region a regulatory RNA, which will act as an antisense regulator on other mRNAs, through a shared promoter and/or terminator [43–46]. In the last years, a few examples of sRNAs overlapping mRNAs 3’ region were reported [47,48]. Recent studies based on co-immunoprecipitation of Hfq and associated RNAs, allied to RNA-seq techniques with higher resolution, confirmed that mRNAs 3’ regions are large reservoirs for new sRNAs [49,50].

Regarding their biogenesis, sRNAs that are originated from 3’ regions of known mRNAs can be distinguished into two different types: sRNAs transcribed from an insider promoter of the overlapping mRNA gene or 3’ UTR (type I) (Fig. 1.5 A) and sRNAs resulting from an internal processing of the initial mRNA (type II) (Fig.1.5 B). However, in both types there is a shared use of the Rho- independent transcription terminator present in the mRNA. The *Salmonella* dapB mRNA represents the first example of dual output via the 3’ UTR of a conserved gene being included in type I sRNAs, since DapZ sRNA overlaps the 3’ UTR of the *dapB* mRNA and is transcribed from an mRNA-internal promoter that exists upstream of the stop codon of the conserved *dapB* gene (Fig.1.5 A). This sRNA is completely independent from the transcription of *dapB* and has a different outcome associated, since the *dapB* gene is essential for the biosynthesis of lysine and DapZ causes a translational repression of genes that encode major ABC transporters [50]. The cpxP mRNA belonging to the Cpx regulon of *Salmonella* is a perfect
example of type II sRNAs, due to its associated dual output. CpxQ sRNA is the first processed 3’ UTR multi-target regulator to be known. It overlaps the 3’ UTR of the cpxP mRNA and both share the same promoter, which makes CpxQ sRNA dependent of cpxP transcription (Fig. 1.5. B). After transcription of the parental mRNA, CpxQ is generated through processing in 3’ region of the coding sequence by RNase E cleavage, a few nucleotides downstream of cpxP stop codon (Fig. 1.5. B) [48]. Although CpxP protein and CpxQ noncoding RNA act in the same pathway, they perform different but complementary functions since CpxP is a periplasmic protein involved in envelope stress [51] and CpxQ sRNA is a cytoplasmic effector that repress mRNAs from several envelope proteins [48].

![Diagram of sRNA processing](image)

**Figure 1.5 sRNAs from bacterial 3’ UTRs.** Transcription of sRNAs from bacterial 3’ UTRs can occur by its own promoter (type I) (A) or through internal processing of the associated mRNA by RNase cleavage (type II) (B). The sRNA resulting from processing in the mRNA 3’ region carry a 5’ monophosphate end, which is a sensor domain of RNase E. In both cases Hfq protein is involved in the stabilization of the sRNA and mediates its interaction with their targets. Furthermore, Hfq is also involved in the internal processing since it can form a complex with RNase E and direct it to the 3’ end of mRNAs. The grey arrows represent the promoters. Adapted from [48].

### 1.3.4 Regulatory sRNAs of protein activity

The majority of the already characterized sRNAs act through base pairing with their target transcripts. Nevertheless, a small number of sRNAs capable of binding to proteins and modify directly their activities have also been identified. These RNAs, which include the *E. coli* 6S RNA and the CsrB family of sRNAs, exert their function by mimicking the structures of other nucleic acids [19].

A few years ago, the 6S RNA was found to bind to the active site of a RNA polymerase holoenzyme comprising a σ70 subunit (housekeeping form of RNA polymerase). Consequently, it inhibits the expression of certain promoters and increases the association with alternative stress sigma factors when the transition between exponential and stationary phase of growth occurs. Moreover, this RNA has a remarkable accumulation during the late stationary phase which has been associated with the sequestration of the majority of σ70 holoenzyme [52]. Later studies have revealed that the structure of 6S RNA governs the σ70 holoenzyme sequestration with the double-stranded RNA hairpin containing a critical central bulge mimicking the promoter open complex [53].

The *E. coli* carbon storage regulator protein (CsrA) was identified as a RNA-binding protein, as well as a global key posttranscriptional regulator of the carbon metabolism and bacterial motility, mainly during
the entry into the stationary phase [54]. CsrA homodimers bind to GGA motifs in stem-loops of a RNA hairpin located in the 5’ UTR of their mRNA targets. Therefore, the stability and/or translation of the mRNA will be affected (reviewed in [19]). The intracellular levels of this protein are mostly negatively regulated by two sRNAs enclosing multiple CsrA binding sites, CsrB and CsrC. When the levels of these RNAs increase they interact and sequester the RNA-binding protein. Opposite to the regulation of the housekeeping form of RNA polymerase, the CsrB and CsrC antagonize CsrA protein, precluding the binding to the target mRNAs (reviewed in [54]).

1.4 Cell advantages of regulating with sRNAs

Regulatory RNAs, such as sRNAs, appear to be widely distributed in the Bacteria domain. Moreover, they are also present in several vital functions performing a regulatory role in diverse environmental conditions and overcoming other regulators, which include protein transcription factors.

As already mentioned, bacteria are often present in hostile environments with a high rate of changing conditions and often exposed to external stimuli. Thus, it is beneficial for cells to store energy to be further used in growth and/or maintenance.

sRNAs are single-stranded RNA molecules with a medium size ranging around a few hundreds of base pairs. Due to the fact that they are shorter in size than the majority of mRNAs they provide a reduced metabolic cost for the cell and faster synthesis when compared to protein transcription factors. In addition, they do not require translation minimizing the metabolic cost (reviewed in [20]). Furthermore, sRNAs act at the posttranscriptional level of regulation and with these advantages it is possible to conclude that not only the response to the unforeseen stress conditions will be faster but also the recovery will be more rapid after the removal of the external stimulus. In both cases, the faster regulation confers a selective advantage [55].

The capacity of a unique sRNA to be transcribed under different conditions together with the irreversible regulatory effects achieved by the ability of some sRNAs to direct the cleavage of their targets are also important advantages of sRNAs [14].

1.5 Burkholderia cepacia complex (Bcc) – overview

During the late 1940’s, Walter Burkholder [56] studied the phytopathogenic agent responsible for the soft onion rot bulbs, a disease known as “sour skin”. During his investigation, the bacterial species Pseudomonas cepacia was isolated and described for the first time [56].

In 1992, a molecular taxonomical analysis of the genus Pseudomonas based on the 16S rRNA sequences, DNA-DNA homology, cellular lipid and fatty acid composition and also in phenotypic characteristics was performed. This analysis resulted in the proposal of a new genus named Burkholderia and, consequently, some species previously inserted in the genus Pseudomonas, particularly Pseudomonas cepacia, were transferred to the new genus. Pseudomonas cepacia was further renamed as Burkholderia cepacia [57]. The Burkholderia genus is phylogenetically diverse. By November 2016, this genus included approximately 105 species [58] that are widely distributed in ecological niches [59]. Subsequently, Burkholderia species were shown to present several biotechnological properties such as bioremediation and biocontrol/biopesticides agents. These applications in biotechnology are due to their ability to i) degrade natural and man-made complex aromatic pollutants in water and soils, ii) produce antifungal compounds, iii) fix atmospheric nitrogen and iv) promote growth of plants and crop production (reviewed in [59]). However, Burkholderia species have emerged in the 1980s as important pathogens to both humans and animals, in particular to cystic fibrosis (CF) patients [60].
The polyphasic taxonomic study performed by Vandamme et al. [61] with *Burkholderia cepacia* isolates from CF patients demonstrated that these bacteria are grouped into five distinct species, then designated as genomovars (I-V), together forming a group of closely related species denominated *Burkholderia cepacia* complex (Bcc).

Bcc bacteria display a diverse nutrition, since they are able to use a broad range of simple and complex molecules as carbon sources, including xenobiotics (reviewed in [59]).

The genome of Bcc species has some interesting properties: i) its large size, which is correlated with their high functional versatility [62], and ii) its organization in three replicons and, in some cases, plasmids. Recent studies performed by Agnoli et al. [63] demonstrated that these microorganisms are able to tolerate the deletion of the third chromosome. These findings suggest that the third replicon could be a plasmid. Currently, this cluster is known to comprise 20 validated and close related species (Table 1; [58,64]), exhibiting a high level of 16S rRNA (over 98%) and *recA* (94%-95%) gene sequence similarity but a moderate degree of DNA-DNA hybridization (30%-60%) [65]. It has also been demonstrated that different Bcc species share whole-genome average nucleotide identity (ANI) with values ranging from 85.04% and 89.92% [66].

Table 1.1 *Burkholderia cepacia* complex (Bcc) species, their main sources and biotechnological applications according to the List of Prokaryotic Names with Standing in Nomenclature (LPSN) [58]. CF: Cystic Fibrosis.

<table>
<thead>
<tr>
<th>Bcc species</th>
<th>Main sources and biotechnological applications</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. cepacia</em></td>
<td>Main pathogen in Portuguese CF population [67]; present in the environment; used has bio remediation and biocontrol agent</td>
<td>[57]</td>
</tr>
<tr>
<td><em>B. vietnamiensis</em></td>
<td>Infections in Humans (CF and non-CF); present in the rhizosphere; used has bioremediation and biocontrol agent</td>
<td>[68]</td>
</tr>
<tr>
<td><em>B. multivorans</em></td>
<td>Infections in Humans (CF and non-CF, being one of the main CF pathogens in European CF centers [69]); present in the rhizosphere</td>
<td>[61]</td>
</tr>
<tr>
<td><em>B. pyrocinia</em></td>
<td>Infections in Humans (CF and non-CF); used has biocontrol agent</td>
<td>[70]</td>
</tr>
<tr>
<td><em>B. stabilis</em></td>
<td>Isolated from CF and non-CF patients</td>
<td>[71]</td>
</tr>
<tr>
<td><em>B. abonensis</em></td>
<td>Infections in non CF patients (isolated from nosocomial infections)</td>
<td>[72]</td>
</tr>
<tr>
<td><em>B. ambifaria</em></td>
<td>Infections in Humans (CF and non-CF); major bacterium present in the rhizosphere; used has biocontrol agent;</td>
<td>[73]</td>
</tr>
<tr>
<td><em>B. anthina</em></td>
<td>Infections in CF patients and animals; present in soils</td>
<td>[74]</td>
</tr>
<tr>
<td><em>B. cenocepacia</em></td>
<td>Infections in Humans (CF and non-CF, being one of the main CF pathogens); present in the rhizosphere; used has biocontrol agent</td>
<td>[75]</td>
</tr>
<tr>
<td><em>B. dolosa</em></td>
<td>Causes mostly infections in CF patients, but one environmental strain has been described</td>
<td>[76]</td>
</tr>
<tr>
<td><em>B. arboris</em></td>
<td>Infections in Humans (CF and non-CF); present in the environment</td>
<td>[77]</td>
</tr>
<tr>
<td><em>B. diffusa</em></td>
<td>Infections in Humans (CF and non-CF); present in the environment and soils</td>
<td>[77]</td>
</tr>
<tr>
<td><em>B. latens</em></td>
<td>Infections in CF patients</td>
<td>[77]</td>
</tr>
<tr>
<td><em>B. metallica</em></td>
<td>Infections in CF patients</td>
<td>[77]</td>
</tr>
<tr>
<td><em>B. seminalis</em></td>
<td>Infections in Humans (CF and non-CF); present in the environment</td>
<td>[77]</td>
</tr>
<tr>
<td><em>B. contaminans</em></td>
<td>Infections in Humans (main pathogen among Portuguese CF population [67] and animals</td>
<td>[66]</td>
</tr>
<tr>
<td><em>B. lata</em></td>
<td>Infections in Humans (CF and non-CF); highly present in soils</td>
<td>[66]</td>
</tr>
<tr>
<td><em>B. pseudomultivorans</em></td>
<td>Present in Humans and the environment</td>
<td>[78]</td>
</tr>
<tr>
<td><em>B. stagnalis</em></td>
<td>Present in soils and in Humans respiratory tract</td>
<td>[79]</td>
</tr>
<tr>
<td><em>B. territii</em></td>
<td>Present in the environment, especially in groundwater</td>
<td>[79]</td>
</tr>
</tbody>
</table>

In more recent years, a new approach known as multilocus sequence typing (MLST) has gained notoriety among bacterial typing methods [80]. MLST uses a set of seven genes encoding proteins with housekeeping functions and discriminates strains based on the comparison of the nucleotide sequences.
of the gene set. Nowadays, it is the standard technique used to differentiate Bcc isolates since it shows a high correlation with other methods and a high discriminatory power [66]. The primers and other information on Bcc MLST can be accessed at http://www.pubmlst.org/bcc/.

1.5.1 Bcc species as opportunistic pathogens in Humans

The first significant descriptions of Bcc infections in patients with CF were performed three decades ago by Isles et al. [60] and Tablan [81]. CF is the most common autosomal recessive monogenic disease among the Caucasian population. It has an estimated frequency of 1 to 8,000 live births in Portugal [82].

Some Bcc bacteria are highly transmissible among CF patients, particularly by social contacts. Several strains belonging to B. multivorans, B. cenocepacia (particularly lineage ET12) and B. dolosa species have been reported to cause outbreaks in CF centers worldwide [83]. During the last decade, the Bcc prevalence in CF patients had a significant decrease, representing about 3.5% of CF infected patients [84]. Although representing a small fraction, Bcc infections are the most feared and through the years have gained special attention from the patients, medical and scientific communities due to their unpredictable clinical outcome.

CF is the result of mutations occurring in a single gene present at the long arm of chromosome 7, which encodes a 1,480 aa polypeptide known as Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) [85]. The CFTR protein functions mainly as a 3',5' - cyclic adenosine monophosphate (cAMP)-dependent chloride channel of low conductance at the apical membranes of epithelial cells that are present in a variety of organs such as sweat glands, pancreas, liver and respiratory and reproductive tracts [86]. However, this protein can also regulate the function of sodium channels [87]. 1,800 mutations with different outcomes in the CFTR protein have been identified as potentially CF-causing [88]. Independently of the molecular mechanism by which the mutation disrupts the normal protein synthesis or function, an atypical transepithelial ion transport of chloride and water occurs. Although several organs are affected, the variations affecting the respiratory tract are the main cause of morbimortality in CF patients, since this defect allows the deposition of thickened mucus and prevents its removal from airway surfaces [89]. Therefore, the conditions generated inside the lungs predispose to the establishment of bacteria, leading to chronic pulmonary infections [89,90] mostly caused by Staphylococcus aureus in young individuals and by Pseudomonas aeruginosa in adults [91,92]. The majority of Bcc infected CF patients undergo a reduction of the pulmonary function caused by the chronic infection and exacerbation events. Approximately 20% of these patients develop in their lifetime a rapid evolving clinical state known as “cepacia syndrome”. This syndrome has not yet been observed for other CF pathogens and it is characterized by high fevers, necrotizing pneumonia and septicaemia, which ultimately results in the patient early death [60].

Bcc infections are also notably difficult to treat and the strategies applied rarely result in eradication of the bacteria upon colonization. One of the major contributors to this problem is their intrinsic resistance to the majority of clinically available antimicrobials, which include aminoglycosides, quinolones, polymyxins and β-lactams [93]. Moreover, these microorganisms have several efflux pumps that can withdraw antibiotics from the cell and have the ability to modify the cell envelope leading to a reduced permeability of the membrane to antibiotics (reviewed in [94]).

Bcc bacterial infections have also emerged in patients with other pathologies such as chronic granulomatous disease (CGD) [95] and immunocompromised (patients with cancer and HIV) [96,97]. Similarly, to CF, CGD is also an hereditary disease, although with a much lower incidence than CF, where bacterial infections caused by Bcc species and their consequences represent the second most common cause of death [95]. This disease is considered to be a primary immune disorder since it is
caused by mutations in the subunits of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex belonging to the phagocytes. These mutations lead to a defective production of reactive oxygen species (ROS) by the phagocytes and therefore they are unable to kill bacteria through oxidative processes [95].

In addition to spread by social contact among CF patients, several outbreaks of bacteremia have been reported among hospitalized non-CF patients such as individuals having chronic hemodialysis, diabetes mellitus, congestive heart failure and malignancy, mainly caused by *B. cenocepacia*, *B. multivorans* [98,99] and *B. contaminans* [100,101]. Therefore, Bcc species are now recognized as emergent nosocomial pathogens [69,97].

### 1.5.2 Major virulence factors of the Bcc

In order to enter in their host and reach the blood stream, as well as to proliferate inside eukaryotic cells, Bcc bacteria produce a wide variety of virulence factors, being the best characterized those produced by the *B. cenocepacia ET12* lineage representative J2315 (reviewed in [59]).

The lipopolysaccharide (LPS) is one of the fundamental virulence factors in Gram-negative bacteria. The particular composition of the core oligosaccharide has been reported as implicated in the resistance of Bcc bacteria to cationic antimicrobial peptides and polymyxins [102,103]. Only a few polysaccharides synthesized by prokaryotes can be excreted, which are known as exopolysaccharides – EPS [104]. Among Bcc bacteria, production of EPS is very fluctuating, however some clinical and environmental isolates are known to produce a particular EPS – Cepacian [105]. Several studies have associated Cepacian production with the persistence of infection by Bcc bacteria, defective phagocytosis process of bacteria by neutrophils and an inhibition of neutrophil chemotaxis and ROS production [106,107]. Studies performed by Cunha et al. [108] have demonstrated that Cepacian also plays an important role in the formation and establishment of biofilms.

Bacteria have a remarkable capacity to form communities known as biofilms, in order to be protected from environmental factors and host-defence mediators so they can establish and persist in the host airways [109]. Caraher et al. [110] have demonstrated an enhanced *in vitro* antibiotic resistance (particularly to ceftazidime and ciprofloxacin) upon the formation of biofilms by Bcc bacteria, presumably due to a decrease of the bacterial surface exposed to the antibiotics. The resistance to multiple antibiotics by Bcc strains constitutes also an important virulent characteristic and has already been addressed in section 1.5.1.

In addition to the generation of communities Bcc also possess the quorum-sensing systems (QS). QS is density-dependent and supports the prompt adaptation of bacteria to the environment. The CepIR was the first QS system identified in a Bcc strain [111] and it is responsible for the regulation of toxins, lipases, proteases, iron-chelating siderophores, swarming motility and biofilm production (reviewed in [112]).

The acquisition of clusters of genes through horizontal gene transfer by Bcc strains (genomic islands) is of extreme importance since these islands encode virulence and metabolism associated genes. On the other hand, these genes promote both the survival and pathogenesis of Bcc bacteria in the CF respiratory epithelium and, consequently, the development of an epidemic lineage [113].

Moreover, these bacteria also express a particular type of fimbriae, the cable pili, with adhesins associated and flagella. Both the cable pili and the flagella help in bacterial adhesion and entry into the respiratory epithelium of host cells [114].
Specialized secretion systems such as type III and type VI also have determining roles in Bcc species, regarding to bacterial intracellular survival and replication. These complex structures inject bacterial molecules into the host cytoplasm, triggering host-signalling pathways and leading to the establishment of a bacterial intracellular niche [115,116].

1.5.3 Identification of novel Bcc virulence factors

In order to identify genes involved in Bcc virulence, researchers from iBB (Institute for Bioengineering and Biosciences) have prepared two mutant libraries by plasposon mutagenesis from strains *B. contaminans* IST408 and *B. cenocepacia* J2315.

The strategy used to identify genes involved in virulence was based on the identification of mutants with attenuated virulence towards the nematode *Caenorhabditis elegans*. Several mutants were identified and the sequencing of the interrupted genes allowed the identification in *B. contaminans* of a gene encoding a protein involved in RNA-RNA interactions – the *hfq* gene. Subsequent research work led to the identification of a second gene, named *hfq2*, within the genome of all Bcc strains sequenced. This is extremely rare among prokaryotes, which usually harbour one *hfq* gene in their genome.

Bcc Hfq proteins have different sizes, since the Hfq has 79 aa residues and the Hfq2 has 188 aa residues [117]. Experiments with the clinical isolate *B. cenocepacia* J2315 and respective mutants in the Hfq encoding gene showed that the *hfq* mutant had an increased susceptibility to conditions mimicking the lung environment of a CF patient. Furthermore, infection experiments with *hfq* and *hfq2* mutants demonstrated that both genes are required for full pathogenesis and virulence in the nematode *C. elegans* [117,118].

The iBB research team also identified another attenuated mutant found to have a single plasposon insertion in a 555 bp intergenic region of *B. cenocepacia* J2315 chromosome 1, upstream of the start codon of BCAL2958 gene, which encodes an outer membrane protein of the OmpA family, and downstream of the 5’ UTR of BCAL2957 gene, which in turn encodes a DNA gyrase subunit A enzyme. Bioinformatics analysis and RT-PCR experiments revealed the presence of a putative sRNA and non-annotated ORF in the region where the plasposon was inserted. In this context, the aim of this work is the functional characterization of the identified putative sRNA and ORF, later found to be widely conserved among the *Burkholderia* genus. To achieve this goal, molecular techniques were adopted in order to characterize the sRNA and the hypothetical ORF. In addition, stress, mobility and virulence assays in the nematode *C. elegans*, as well as an extensive bioinformatics analysis were performed.

This study presents a contribute to better understanding of the regulatory mechanisms played by sRNAs in the expression of virulence determinants by Bcc bacteria. Despite the results obtained are somehow limited they contribute to a better understanding of Bcc biology, which is of central importance for the future development of novel and more accurate therapeutic strategies against the severe infections caused by Bcc bacteria.
2. MATERIALS AND METHODS

2.1 Bacterial strains, plasmids, nematode and culture conditions

All bacterial strains, plasmids and the nematode used in this work are listed in Table 2.1. *B. cenocepacia* and *E. coli* strains were maintained in *Pseudomonas* Isolation Agar (PIA; Benton Dickinson) and Lennox Broth (LB; Pronadisa), respectively, with 2% (w/v) agar (Iberagar). Unless otherwise stated, *B. cenocepacia* and *E. coli* strains were cultured on solid media and incubated at 37 °C for 24 h. In order to maintain selective pressure, when required, solid growth media was supplemented with adequate antibiotics. Antibiotics were used as follows: 300 µg/mL chloramphenicol or 600 µg/mL trimethoprim for *Burkholderia* strains and 150 µg/mL ampicillin, 50 µg/mL kanamycin or 100 µg/mL trimethoprim for *E. coli* strains.

Table 2.1 Bacterial strains, plasmids and nematode used in this study.

<table>
<thead>
<tr>
<th>Strains/ plasmids/ nematode</th>
<th>Genotype/ Description</th>
<th>Reference/ Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> α-DH5</td>
<td>F− · Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (r−, mcr+) phoA supE44 λ− thi-1 gyrA96 relA1</td>
<td>Invitrogen</td>
</tr>
<tr>
<td><em>E. coli</em> HB101</td>
<td>F−, thi-1, hsdS20 (r−, mcr−), supE44, recA13, ara-14, leuB6, proA2, lacY1, galK2, rpsL20 (str′), xyl-5, mtl-1</td>
<td>Promega</td>
</tr>
<tr>
<td><em>E. coli</em> OP50</td>
<td>Uracil auxotroph strain for <em>C. elegans</em> feeding</td>
<td>[119]</td>
</tr>
<tr>
<td><em>Burkholderia</em> strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. cenocepacia</em> J2315</td>
<td>Clinical isolate from a CF patient, Edinburgh (UK). Reference strain from ET12 lineage</td>
<td>[98]</td>
</tr>
<tr>
<td><em>B. cenocepacia</em> J2315 (pMLS7)</td>
<td><em>B. cenocepacia</em> J2315 transformed with pMLS7</td>
<td>This study</td>
</tr>
<tr>
<td><em>B. cenocepacia</em> J2315 (pMya2)</td>
<td><em>B. cenocepacia</em> J2315 transformed with pMya2</td>
<td>This study</td>
</tr>
<tr>
<td><em>B. cenocepacia</em> SJ2</td>
<td>Mutant derived from <em>B. cenocepacia</em> J2315 by random plasposon mutagenesis with pTnMod-OCm</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC19</td>
<td>Cloning vector, lacZ α-peptide, Amp′</td>
<td>[120]</td>
</tr>
<tr>
<td>pRK2013</td>
<td>Mobilizing vector, ColEI tra(RK2)r; Km′</td>
<td>[121]</td>
</tr>
<tr>
<td>pMLS7</td>
<td>Cloning vector, pBBR1 ori, mob, Tp′</td>
<td>[122]</td>
</tr>
<tr>
<td>pMya2</td>
<td>pMLS7 containing the 502 bp fragment with the mavA sequence and promoter region from <em>B. cenocepacia</em> J2315</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Nematode</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. elegans</em> mutant strain DH26</td>
<td><em>C. elegans</em> with a temperature-sensitive mutation, rendering worms sterile at 25 ºC</td>
<td>[123]</td>
</tr>
</tbody>
</table>
Liquid cultures were performed at 37 °C overnight with orbital agitation (250 rpm) in 4 mL of LB liquid medium supplemented with the appropriate antibiotics. Bacterial cultivation for growth kinetics characterization was carried out in 25 mL of LB or M9 [1X M9 salts (374 mM NH₄Cl, 441 mM KH₂PO₄ and 448 mM Na₂HPO₄·7H₂O), 22.2 mM Glucose, 1 mM MgSO₄ and 8.6 mM NaCl] medium, inoculated at an initial optical density of 0.1 measured at 640 nm (OD₆₄₀ nm), and incubated at 37 °C with orbital agitation. Bacterial growth was followed for 24 h in LB medium or for 72 h in M9 medium through measurements of the culture OD₆₄₀ nm. The specific growth rate was calculated for LB medium, using the exponential equation from GraphPad Prism version 6.01 for Windows (GraphPad Software, La Jolla, California USA). The results herein presented are mean values from two independent experiments for the growth curve in LB and from three independent experiments for the growth curve in M9 medium. When not in use, bacterial strains were maintained at -70 °C in 40% (v/v) glycerol.

2.2 Molecular biology techniques

2.2.1 Extraction and purification of genomic and plasmid DNA

Plasmid DNA was extracted from E. coli cultures grown overnight in LB medium supplemented with the adequate antibiotics, using the NZYMiniprep kit (NZYTech) in accordance to the manufacturer’s instructions. Total genomic DNA from overnight liquid cultures of B. cenocepacia J2315 was extracted using the High Pure PCR Template Preparation kit (Roche), according to the manufacturer’s instructions. The concentrations of both plasmid and genomic DNA were estimated spectrophotometrically by measuring samples absorbance at 260 nm with a ND-1000 spectrophotometer (NanoDrop). When necessary, DNA fragments were purified from agarose gels with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel), after gel electrophoresis with 1X TAE buffer (40 mM Tris-ba, 20 mM Acetic acid and 1 mM EDTA) at 4 V/cm [124] and subsequent gel staining with GelRed™ Nucleic Acid Gel Stain (Biotium), according to the manufacturer’s instructions. DNA was visualized with a UV transilluminator (Bio-Rad).

2.2.2 Construction of a plasmid expressing the MavA sRNA

In order to analyse the MavA sRNA function in B. cenocepacia J2315, the B. cenocepacia J2315 MavA putative encoding sequence was amplified with the primers Hip_UP (5’ TTGGATCCATGAACGCGAGCT 3’) and P8D4_3’ (5’ TTGTGACGAATTCGCCTTCC 3’), retrieving a 502 bp fragment. This fragment was first ligated to the BamHI/ SalI restriction sites of pUC19 and then to the BamHI/HindIII restriction sites of the constitutive expression vector pMLS7 (B. cepacia promoter Pₛ₇) [122], with T4 DNA Ligase (Fermentas). The resulting construction, named pMya2, was introduced into E. coli α-DH5 cells by heat-shock transformation and into B. cenocepacia J2315 by triparental mating.

2.2.3 Insertion of foreign DNA in bacterial cells

2.2.3.1 Heat-shock transformation of E. coli cells

E. coli α-DH5 cells were prepared according to Sambrook and Russell [124]. Briefly, 100 mL of LB liquid medium were inoculated to a final OD₆₄₀ nm of 0.1 from overnight liquid cultures (at 37 °C) and maintained under the same conditions until the culture reached the mid-log phase (OD₆₄₀ nm ≈ 0.6–0.8). The liquid culture was then centrifuged (2,700 x g, 5 min, 4 °C) and the cell pellet was carefully resuspended in 100 mL of 0.1 M MgCl₂. After a second centrifugation (1000 x g, 5 min, 4 °C), the cell
pellet was carefully resuspended in 100 mL of 0.1 M CaCl₂, incubated on ice for 30 min and subjected to another centrifugation step under the same conditions. The resulting cell pellets were resuspended in 22 mL of 0.1 M CaCl₂. 3.5 mL of 86% (v/v) glycerol was added and aliquots were frozen at -70 °C or immediately used for transformation.

The competent E. coli cells were transformed by heat-shock transformation. Briefly, 150 μL of competent cells and 50 μL of TCM salts (10 mM CaCl₂, 10 mM MgCl₂ and 10 mM Tris, to a final pH of 7.5) were added to the 1.5 mL eppendorf containing the plasmid DNA, followed by incubation on ice for 30 min. The cells were then heat-pulsed at 42 °C for 3 min and transferred to ice for 5 min. Finally, cells were incubated in liquid LB medium for 1 h and then plated in selective medium [124,125].

2.2.3.2 Triparental mating of B. cenocepacia cells

Triparental conjugation experiment was adapted from Engledow et al. [126] and Dubarry et al. [127]. The E. coli donors [E. coli α-DH5 (pMLS7) and E. coli α-DH5 (pMya2)] and helper strain (E. coli HB101 carrying pRK2013), as well as the recipient strain (B. cenocepacia J2315) were grown overnight, separately, in liquid LB medium supplemented with the appropriate antibiotics. The overnight cultures were then diluted to a final OD₆₅₀nm of 0.1 in 25 mL of fresh LB medium, supplemented with antibiotics, and were grown under the same conditions to an OD₆₅₀nm ≈ 0.6. Cells from 0.4 mL of the recipient strain cultures and of 0.2 mL of the helper and donor strains cultures were separately harvested by centrifugation, washed and resuspended with sterile 0.9% (w/v) NaCl. Then, the three bacterial cultures (recipient, helper and donor strain) were mixed together and harvested at low speed centrifugation for 5 min. The resulting pellet was resuspended in 100 μL of 0.9% (w/v) NaCl and spot-inoculated on the surface of a sterile 47 mm diameter filter disk (Supor®, 0.2 µm pore size, GelmanSciences), previously placed onto the surface of a LB agar plate. After 16 h of incubation at 30 °C, the bacterial layer formed in the surface of the filter was washed and resuspended in 1 mL of sterile 0.9% (w/v) NaCl. Serial dilutions were prepared and 100 μL of each dilution were plated on appropriate selective medium (PIA agar plates supplemented with 600 µg/mL of trimethoprim) for positive selection of transconjugants.

2.2.4 DNA sequence analysis

Nucleotide sequences were retrieved from National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/) and the Burkholderia Genome Database [128]. DNA sequencing was performed by Operon MWG Eurofins (Germany) as a paid service. The nucleotide sequences retrieved from Operon MWG Eurofins were analysed with VecScreen (http://www.ncbi.nlm.nih.gov/VetScreen/VetScreen.htm) in order to identify and eliminate vector contamination. BLASTN 2.5.1 algorithm [129] was also used to compare the obtained sequences with those available at NCBI. Additionally, multiple alignments were performed using the MUSCLE tool (http://www.ebi.ac.uk/Tools/msa/muscle/) [130]. The intergenic region studied in this work was analysed for the potential presence of encoded proteins with the ORF Finder tool (https://www.ncbi.nlm.nih.gov/orffinder/).

2.2.5 Extraction and purification of total RNA

Cells from liquid cultures of B. cenocepacia J2315 strain and its derivative mutant SJ2 were harvested by centrifugation (5,500 x g, 3 min, 4 °C) at mid-log phase (6 h of incubation), mixed with RNA Later Solution (Ambion) and frozen at -20 °C. Total RNA was extracted from these samples using the Quick-RNA™ MiniPrep (Zymo Research), according to the manufacturer’s instructions. DNase I treatment was performed in order to remove trace amounts of contaminating DNA. RNA concentration was
estimated by a ND 1000 spectrophotometer (Nanodrop). Samples with ratios $A_{260\text{ nm}}/A_{280\text{ nm}}$ and $A_{260}\text{ nm}/A_{230}\text{ nm}$ below 1.9 were not considered. RNA quality was evaluated by electrophoresis of 1 µg of RNA samples, mixed with Gel Loading Buffer II (Ambion) and incubated at 95 ºC for 3-5 min, in denaturing gels (7 M urea - 10% (v/v) polyacrylamide) with 1X TBE buffer (89 mM Tris.base, 89 mM Boric acid and 2 mM EDTA) at 200 V.

After electrophoresis, the acrylamide gel was stained for 15 min with SYBR® Gold Nucleic Acid Gel Stain (Thermo Fischer Scientific), according to the manufacturer’s instructions, and exposed to short wave UV light in a transilluminator (Bio-Rad).

2.2.6 Reverse transcription-polymerase chain reaction (RT-PCR)

The total RNA obtained from *B. cenocepacia* J2315 and its derivative mutant SJ2 cells harvested at mid-log phase (6 h of incubation), was extracted as described in section 2.2.5. In order to completely remove any trace amounts of contaminating DNA, total RNA was subjected to two DNase (RiboPure Bacteria kit, Life Technologies) treatments of 60 min each. Treated RNA was precipitated overnight at -20 ºC with 2.5 volumes of cold 100% (v/v) ethanol, 0.5 M ammonium acetate and 100 µg/mL glycogen. Next, the mixture was centrifuged (16,000 x g, 30 min, 4 ºC) and the RNA was washed with 70% (v/v) ethanol, air dried and re-dissolved in water.

Complementary DNA (cDNA) was generated with the NZY First-Strand cDNA Synthesis kit (NZYtech), using 250 ng of purified RNA from each strain. All the amplifications performed were done with Phusion DNA polymerase (Thermo Scientific). In order to confirm the transcription of MavA sRNA in both strains, the cDNA previously obtained was amplified with the primers LW_MavA (5’ TTCTAGAGTGTTCACGCACACGAG 3’) and RACE 3 (5’ AAAATCCTCCCCACCCCAAG 3’). The transcription of the OmpA protein was confirmed with the primers BCAL2958-UP (5’ TTGGATCCATGAATAAACTTT 3’) and BCAL2958-LW (5’ AAAAGCTTGTGGTGCCGGAAC 3’), whereas the transcription of the putative ORF was evaluated by amplification with the primers RACE 1 (5’ AGTTGAAGCGATCTTGGA 3’) and Hip_UP (described in section 2.2.2). The initial samples of total RNA were used as negative controls and a sample of genomic DNA from *B. cenocepacia* J2315 was used as a positive control. The samples obtained in the Polymerase Chain Reactions (PCR) were separated in a 2% (w/v) agarose gel electrophoresis.

2.2.7 Rapid amplification of cDNA ends (RACE)

RNA purified from *B. cenocepacia* J2315 at mid-log phase (6 h of incubation) was used for amplification of the cDNA ends corresponding to MavA sRNA, which was performed using the FirstChoice® RLM-RACE kit (Ambion), according to the manufacturer’s instructions. For 5’ RACE, total RNA was treated in order to preserve a 5’-monophosphate, enabling the ligation of the 45 bp 5’ RACE Adapter (5’ GCUGAUGCGGAUGAAPAACACUCUGCUUUUGCUUGAUGGAA 3’), and then, reverse transcribed. The tailed cDNA was nest amplified first with 5’ RACE Outer primer (5’ CCTGATGGCCGCTGAATGACACTG 3’) and a 5’ RACE gene specific outer primer (Int3_UP: 5’ CTCTAAAGTTACGCAGGC 3’), followed by amplification with 5’ RACE Inner primer (5’ CGCCGATCCGAACACTGCGTTTGCTGGCTTTGATG 3’) and a 5’ RACE gene specific inner primer (RACE 1, described in the above section). Previously to 3’ RACE, total RNA was 3’ end poly(A) tailed using a Poly(A) polymerase I (Ambion) and then reverse transcribed with 3’ RACE Adapter primer (5’ GCAGGACAAGATTAAATAGCAGCTACTAGTACGT 3’). After these initial steps, tailed cDNA was amplified with 3’ RACE Outer primer (5’ GCAGGACAAGATTAAATAGCAGCTACTAGTACGT 3’) and a 3’ RACE gene specific outer primer (RACE 2: 5’ GCCATATCGTCCCATCTCTC 3’). The resulting
products were independently purified through 2% (w/v) agarose gel electrophoresis and blunt-end cloned into the HincII site of pUC19, followed by *E. coli* α-DH5 heat-shock transformation, for sequencing.

2.2.8 Analysis of total soluble proteins by Tricine-SDS-PAGE

All *B. cenocepacia* strains under study were inoculated in 4 mL of LB and M9 media with the appropriated antibiotics at an initial OD$_{640}$ of 0.1 and incubated at 37 ºC with orbital agitation. After 8 h of incubation in LB medium and 24 or 48 h in M9 medium, 1 mL of the cultures at OD$_{640}$ 0.6 were collected and the resulting pellet was resuspended in 40 μL of sample buffer solution [10% (v/v) Glycerol, 4% (w/v) Sodium Dodecyl Sulphate (SDS), 0.2% (w/v) Bromophenol blue, 2% (w/v) Mercaptoethanol, 1.2% (w/v) Tris (pH 6.8)]. Samples were incubated at 37 ºC for 1 h and loaded in a discontinuous tricine gel system, allowing proteins to be concentrated [stacking gel, 4% (v/v)] and separated according to their relative molecular mass [running gel, 15% (v/v)] [131].

Protein electrophoresis was performed at room temperature using 1X anode (0.1 M Tris, 22.5 mM HCl, pH ≈ 8.9) and cathode (0.1 M Tris, 0.1 M Tricine, 34.7 mM SDS, pH ≈ 8.25) buffers. Voltage was kept at 30 V until samples enter the running gel, then the voltage was increased to 150 V, 80 mA and 10 W, as described by Schägger [131]. Proteins were visualized after gel staining with previously boiled Coomassie blue staining solution [0.2% (w/v) Coomassie blue R-250, 10% (v/v) Acetic acid, 47.5% (v/v) Ethanol] for 30 min. After washing with deionized water, the gel was placed into a previously boiled destaining solution [10% (v/v) Acetic acid, 26.3% (v/v) Ethanol], for 15-60 min.

2.2.9 Western blot experiments

Western blot analysis against OmpA-like protein was carried out using *B. cenocepacia* J2315 and *B. cenocepacia* SJ2 cells extracts harvested by centrifugation (5,500 x g, 3 min, room temperature) at mid-log phase (6 h of incubation). The resulting cell pellet was resuspended in 40 μL of the sample buffer solution [20% (v/v) Glycerol, 4% (w/v) SDS, 0.2% (w/v) Bromophenol blue, 3% (w/v) 1,4-Dithiothreitol (DTT), 1.2% (w/v) Tris (pH 6.8)]. Samples were incubated at 100 ºC for 5 min and loaded in a 12.5% (v/v) SDS-PAGE [132]. Total protein electrophoresis was carried out at room temperature using 1X running buffer (0.25 M Tris.base, 1.92 M Glycine, 34.7 mM SDS) at a constant voltage of 150 V, 80 mA and 10 W.

After electrophoresis, gel, the nitrocellulose (NC) membrane (Biotrace™ NT, Pall Corporation) and 3MM Whatmann® filter papers were immersed in Bjerrum and Schafer-Nielsen buffer [0.58% (w/v) Tris, 0.29% (v/v) Glycine, 20% (v/v) Methanol, 0.04% (w/v) SDS] and the system was assembled. Total proteins were electrotransfered to the membrane at 15 V, 120 mA for 50 min in a semi-dry electrophoretic transfer unit (Trans-Blot® SD, Bio-Rad). The membrane was blocked by incubation with 5 % (w/v) skimmed milk (DIvCO) in 1X Phosphate Buffered Saline solution (PBS; 140 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$ and 1.8 mM KH$_2$PO$_4$, pH 7.4), overnight at 4 ºC with agitation. Afterwards, the membrane was washed three times with 1X PBS containing 0.05% Tween® 20 and probed with the primary Goat antibody anti-BCAL2958 (1:3,000 dilution, SICGEN), developed by Sousa *et al.* [133], for 2 h at room temperature. The membrane was then incubated with the secondary antibody HRP—conjugated Mouse anti-Goat IgG (1:10,000 dilution, Santa Cruz Biotechnology), for 1 h at room temperature. Lastly, the membrane was briefly incubated with the peroxidase substrate ECL (Sigma) and the chemiluminescent signals were detected with the device FUSION Solo (Vilber Lourmat).
2.3 Nematode killing assays

2.3.1 Maintenance of \textit{C. elegans} and egg preparation

\textit{C. elegans} worms were cultured at 20 \textdegree C on nematode growth medium I [NGM I agar plates contained 0.3\% (w/v) NaCl, 0.25\% (w/v) Tryptone and 1.7\% (w/v) Agar. This medium was supplemented with 0.05 mM Nystatin and 25 mM KPO\textsubscript{4} buffer (796 mM KH\textsubscript{2}PO\textsubscript{4}, 204 mM K\textsubscript{2}HPO\textsubscript{4}, pH 6.0), 1 mM CaCl\textsubscript{2}, 1 mM MgSO\textsubscript{4}, 0.02 mM Uracil and 0.01 mM Cholesterol] [134], previously inoculated with 100 \(\mu\)L of a fresh overnight culture of \textit{E. coli} OP50 and incubated at 37 \textdegree C overnight. Every week, nematodes were maintained in culture by cutting a piece of agar covered with \textit{E. coli} and \textit{C. elegans} for a fresh new \textit{E. coli} plate.

\textit{C. elegans} eggs were collected by rinsing four times the plate containing both worms and eggs with 1 mL of sterile deionized water and the suspension was divided in three tubes. 500 \(\mu\)L of hypochlorite solution (4\% (v/v) NaClO and 6\% (w/v) NaOH) were added to each tube and vortexed for 8-10 min. Afterwards, this suspension was centrifuged (700 x g, 1 min, room temperature) and the supernatant was carefully discarded. The pellet was washed with 1 mL of sterile bidestilled water, centrifuged in the same conditions above mentioned and then, resuspended in 100-200 \(\mu\)L of M9 buffer (22 mM KH\textsubscript{2}PO\textsubscript{4}, 28 mM Na\textsubscript{2}HPO\textsubscript{4}, 85 mM NaCl, 1 mM MgSO\textsubscript{4}). Finally, \textit{C. elegans} eggs were inoculated in a NGM I plate already prepared with \textit{E. coli} OP50, which was kept at 20 \textdegree C.

2.3.2 \textit{C. elegans} killing assays

Nematode killing assays were performed essentially as described by Sousa \textit{et al.} [135]. Liquid cultures of the \textit{Burkholderia} strains used in this work were grown overnight and 50 \(\mu\)L of each suspension were inoculated on 35 mm diameter plates of nematode growth medium II [NGM II agar plates contained 0.3\% (w/v) NaCl, 0.35\% (w/v) Bactopeptone and 1.7\% (w/v) Agar. This medium was supplemented with 0.05 mM Nystatin, 25 mM KPO\textsubscript{4} buffer (796 mM KH\textsubscript{2}PO\textsubscript{4}, 204 mM K\textsubscript{2}HPO\textsubscript{4}, pH 6.0), 1 mM CaCl\textsubscript{2}, 1 mM MgSO\textsubscript{4}, 0.02 mM Uracil and 0.01 mM Cholesterol] [136]. NGM II plates were incubated at 37 \textdegree C for 24 h, allowing the formation of a bacterial lawn, in which 18 to 30 hypochlorite-synchronized L4 larvae of \textit{C. elegans} DH26 (obtained from the \textit{Caenorhabditis} Genetics Centre, University of Minnesota, Minneapolis, USA) were inoculated. For worm counting, a Stemi 2000-C stereomicroscope (Zeiss) at 50\times of magnification was used. In order to keep \textit{C. elegans} DH26 strain sterile, plates were incubated during 5 days at 25 \textdegree C. Morphological appearance and viability of the worms were assessed daily. Nematodes were considered dead when failed to respond to touch. The results presented are mean values of two independent experiments, each comprising three replicates. NGM II plates inoculated with \textit{E. coli} OP50 were used as a control.

2.4 Biofilm formation and cell surface hydrophobicity assays

2.4.1 Biofilm formation ability

Biofilm formation assays were performed based on the methodology described by Sass \textit{et al.} [16] and O’Toole and Kolter [137]. Overnight liquid cultures of the \textit{Burkholderia} strains under study were performed in 4 mL of Luria-Bertani Broth (LBB; 1\% (w/v) Triptone, 1\% (w/v) NaCl, 0.5\% (w/v) Yeast extract) at 37 \textdegree C with orbital agitation. After 16 h of incubation, 25 mL of LBB medium were inoculated at an initial OD\textsubscript{640 nm} of 0.1 and incubated under the same conditions, until mid-log phase was reached. These cultures were diluted to a standardized OD\textsubscript{640 nm} of 0.5 and 20 \(\mu\)L were used to inoculate each well.
of a 96-well polystyrene microtiter plate (Greiner Bio-One) containing 180 μL of LBB. Plates were incubated at 37 °C for 24 or 48 h without agitation. Sterile growth medium was used as negative control.

In order to quantify the biofilm formation, culture medium and the unattached bacterial cells were removed by rinsing the wells with 200 μL of distilled water, at least three times. Adherent bacteria were stained with 200 μL of 1% (w/v) crystal violet during 15 min at room temperature. After three gentle rinses with 200 μL of distilled water, the dye associated to the attached cells was solubilized in 200 μL of 95% (v/v) ethanol. Biofilm was quantified by measuring the absorbance of the solution at 590 nm in the SPECTROstar Nano Microplate Reader (BMG LABTECH). Three independent experiments comprising four dependent replicates were performed.

2.4.2 Bacterial adhesion to hexadecane

The bacterial adhesion to hydrocarbon (BATH) [138] was used to determine the cell surface hydrophobicity of the B. cenocepacia J2315 and B. cenocepacia SJ2 strains. Cultures were grown in LB liquid medium for 24 h and harvested by centrifugation (16,000 x g, 5 min, room temperature). Cell pellet was washed twice and diluted in 1X PBS buffer pH 7.4 (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ and 1.8 mM KH₂PO₄) to an OD₆40 nm of 0.6. These cell suspensions were divided in 1.5 mL aliquots and then mixed with volumes of hexadecane (Merck) ranging from 0-800 μL. Glass test tubes were vortexed for 20 s and the phases (aqueous phase – hexadecane phase) were allowed to separate for 30 min. After this time, the OD₆40 nm was measured in the Hitachi U-2000 spectrophotometer. Hydrophobicity (%) was calculated using the formula [1-(OD₆40 nm/OD₆40)]×100, where OD₆40 represents the initial cell suspension OD₆40 nm and OD₆40 is the aqueous phase OD₆40 nm. Results are mean values of two independent experiments.

2.5 Motility assays

2.5.1 Swimming assays

Swimming assays were performed as described by Deziel et al. [139]. Tryptone swim agar plates containing 1 % (w/v) Triptone, 0.5 % (w/v) NaCl, and 0.3 % (w/v) Agar were spot-inoculated with 2 μL of exponentially growing Bcc cultures, previously diluted to a standardized OD₆40 nm of 1. The swimming diameters were measured after incubation for 48 h at 37 °C. Results are the mean values of two independent experiments.

2.5.2 Twitching motility assays

Twitching assays were performed accordingly to the method described by Deziel et al. [139]. With the help of a toothpick, B. cenocepacia colonies from fresh plates were stab-inoculated through a thin LB agar layer [1% (w/v) Agar] until reach the polystyrene surface of the plate. The blurred zone of growth at the agar-plate interface was measured after 72 h of incubation at 37 °C. Results are the mean values of two independent experiments.

2.6 Stress susceptibility experiments

2.6.1 Heat stress susceptibility

The survival of the B. cenocepacia strains to heat stress was tested by incubating 4 mL of LB medium until the cultures reached late-exponential phase. After this time, aliquots of 100 μL were incubated at
50 °C and after 0, 5 and 15 min of incubation, aliquots were serially diluted and inoculated on LB plates, incubated at 37 °C for 24 h and the number of CFUs/mL was counted.

The relative number of bacterial survivors was calculated by dividing the number of CFUs/mL after the heat stress by the initial number of CFUs/mL. Results are mean values of two independent experiments.

2.6.2 Anionic detergent shock experiments

In order to evaluate the susceptibility of B. cenocepacia J2315 and SJ2 strains to SDS, 25 mL of LB medium were inoculated and incubated (OD\textsubscript{640 nm} 0.1) until mid-log phase. The cells were then collected, washed and diluted to an OD\textsubscript{640 nm} of 0.7 with 20 mM Tris (pH 8). To evaluate the mortality kinetics of each strain, pulses of SDS to a final concentration of 0.1%, 0.5%, 1% and 5% (v/v) were added and the bacterial mixture OD\textsubscript{640 nm} was measured at room temperature during 10 min, using a double-beam spectrophotometer (Hitachi U-2000). Positive controls, without SDS, were also performed.

The cell lysis rate was calculated based on the decay rate of each curve. Results are mean values of two independent experiments.

2.7 Antibiotic susceptibility testing

B. cenocepacia strains J2315 and SJ2 susceptibility to the antibiotics imipenem, amikacin, tetracycline, streptomycin, ciprofloxacin, aztreonam, ceftazidime, gentamicin and trimethoprim, as well as B. cenocepacia J2315 (pMLS7) and B. cenocepacia J2315 (pMya2) susceptibility to tetracycline and streptomycin were assessed by determining the minimal inhibitory concentration (MIC) by broth micro dilution technique [140]. Stock solutions of each antibiotic were prepared in Mueller Hinton (MH) liquid medium (Fluka) at a final concentration of 1,024 μg/mL. A 96-well polystyrene microtiter plate (Greiner Bio-One) was prepared containing 75 μL of MH liquid medium in each well. Serial 1:2 dilutions of each antibiotic were made in order to obtain final test concentrations ranging from 1,024 μg/mL to 0.125 μg/mL. The wells were inoculated with 75 μL of each tested organism with approximately 1×10^6 CFU/mL (OD\textsubscript{640 nm} of 0.0025). The plates were incubated at 37 °C for 24 h. After this time, the wells were examined for turbidity (growth) and their OD\textsubscript{640 nm} was measured in a SPECTROstar Nano Microplate Reader (BMG LABTECH). The lowest concentration of antimicrobial agent capable of preventing growth of the tested organism was considered the MIC value. Wells without antibiotic or without organism inoculum were considered positive and negative controls, respectively. These experiments were performed four times.

2.8 Statistical analysis

The results here presented for biofilm formation assays were tested for statistical significance with SPSS statistical package (version 21, SPSS Inc., Chicago, IL, USA). Results are presented as mean values and standard deviations (SD). The Kolmogorov-Smirnov test was used to assess the normality of continuous variables. For the variables with a normal distribution the homogeneity of variances was evaluated using the Levene test. The Student’s paired t-test was used to evaluate the differences in mean values of the results observed in cultures with different treatments. For non-normal variables, the nonparametric Mann-Whitney U test was used. A p-value < 0.05 was considered statistically significant. The figures presented in this work correspond to representative results of the assays performed.
2.9 Bioinformatics analysis

2.9.1 MavA sRNA bioinformatics analysis

In order to confirm MavA sRNA conservation among bacterial genus, BLAST searches were performed against publicly available genomes using the BLASTN algorithm present at NCBI, as mentioned in section 2.2.4. Conserved domains were searched using NCBI’s conserved domain database [141]. The region of the putative sequence encoding MavA sRNA was extensively analysed for the presence of putative promoters using the Virtual Footprint and PRODORIC promoter analysis server [142] and the promoter recognition program BPROM [143], followed by a prediction of Rho-independent transcription terminators with the FindTerm [143] and ARNold [22] web tools. The search for functional homologues of MavA was performed using the sRNAMap database [144]. The putative consensus secondary structure of MavA was retrieved with the RNAalifold web tool [145], whereas a previously multiple alignment of RNA sequences belonging to Bcc bacteria was performed with the LocARNa [146] tool, both tools available at the Freiburg RNA Tools web server [147]. Putative mRNA targets of MavA sRNA were searched using TargetRNA2 (version 2.01) [148] and RNApredator (version 1.55) [149] web tools. A minimum seed of 7 nt and a hybridization target region window size of -80 to +20 around the translation start site were used.

2.9.2 Hypothetic ORF bioinformatics analysis

The nucleotide and amino acid sequence obtained to the hypothetical protein (section 2.2.4) were used for a BLAST search against the genome and annotated ORFs from several bacterial species, using the BLASTN and BLASTP algorithms, respectively, present at NCBI, in order to analyse its conservation. The presence of putative conserved domains was performed in NCBI’s conserved domain database [141], while the presence of putative promoters was analysed using the Virtual Footprint and PRODORIC promoter analysis server [142] and the promoter recognition program BPROM [143]. Clustal Omega tool was used to perform multiple alignments for the protein sequences retrieved from BLASTP (https://www.ebi.ac.uk/Tools/msa/clustalo/) [150] and the consensus secondary structure was predicted by the JNet algorithm from JPred4 web server [151]. The various physicochemical properties of the hypothetical protein were computed with ProtParam tool [152], available at ExPASY bioinformatics resource portal (http://web.expasy.org/protparam/). With the ultimate goal to identify a potential sub-cellular localization for the hypothetic protein, SignalP 4.1 [153] and TatP 1.0 [154] web servers were used to predict the presence and location of signal peptide cleavage sites, whereas SecretomeP 2.0 [155] and LocTree3 [156] web servers were used to predict the presence of a classical or non-classical protein secretion pathway. Moreover, TMHMM Server v. 2.0 was used to predict the presence of transmembrane helices in the hypothetical protein [157].
3. RESULTS AND DISCUSSION

3.1 Characterization of the B. cenocepacia J2315 chromosome 1 intergenic region flanked by genes gyrA and ompA

The B. cenocepacia J2315 derived mutant SJ2 was previously obtained by the iBB research team by random plasposon mutagenesis with pTnMod-OCm [135]. This mutant was identified after the qualitative screening for mutants exhibiting reduced virulence towards the nematode C. elegans. The region interrupted by the plasposon insertion is located in B. cenocepacia J2315 chromosome 1, between the gyrA and ompA putative encoding genes. This 555 bp intergenic region (IGR) was examined in this study.

The plasposon was found to be inserted at 133 nt from the start codon of the ORF putatively encoding an outer membrane protein A (OmpA), and 413 nt from the start codon near the ORF putatively encoding a GyrA enzyme. Recently, the coding sequence of the B. cenocepacia J2315 OmpA-like protein was shown to display an extensive 5’ UTR region, comprising approximately 151 nt [16].

In order to gain further insights on the possible function of this IGR, we have initiated this work by analysing the possible effects of the plasposon insertion on ompA gene integrity and functionality, since the transcript from the putative OmpA family protein could be affected in its 5’ UTR. Total RNA was extracted from exponentially growing B. cenocepacia J2315 and B. cenocepacia SJ2 cells, cultured in rich medium and the corresponding cDNA was synthesized. The nucleotide sequence of the ompA gene was retrieved from the Burkholderia Genome Database and specific primers were designed for the amplification by PCR of the 669 bp fragment putatively encoding this protein. The results obtained (Fig. 3.1 A) demonstrated that the transcription of the gene encoding the OmpA-like protein is not affected by the plasposon insertion, since in both cDNA samples the band obtained presented the predicted size.

A Western blot assay was performed to evaluate if the ompA mRNA of the mutant strain was functional. The detection of a band with a molecular mass inferior to 25 kDa is consistent with the molecular mass of OmpA-like protein (23.97 kDa), confirming the ompA mRNA translation in the mutant SJ2 (Fig. 3.1 B). In E. coli ompA mRNA stability is acquired due to the presence of two stem-loops in the 5’ UTR. The Hfq protein binds to the single-stranded regions adjacent to the stem-loops, modifying the mRNA stability (reviewed in [158]). Since the plasposon is inserted 8 nt after the predicted 5’ UTR beginning of the B. cenocepacia J2315 OmpA-like protein and assuming the mRNA presents the same structure as described for E. coli ompA mRNA, then the plasposon could not interfere with the single-stranded regions and affect ompA mRNA functionality.

Figure 3.1 Detection by RT-PCR of the transcript from the ORF putatively encoding an OmpA-like protein, located in B. cenocepacia J2315 chromosome 1 (A) and assessment of the ompA mRNA functionality by Western blot (B). The cDNA was generated from total RNA collected from exponentially grown cells. Lanes: L1 - GeneRuler™ Low Range DNA Ladder; 1 – cDNA from B. cenocepacia J2315; 2 – cDNA from B. cenocepacia SJ2; C – negative control (without template); 3 and 4 - negative controls, using as template the initial total RNA samples of B. cenocepacia J2315 and B. cenocepacia SJ2, respectively; L2 – Molecular mass ladder (Unstained Protein Molecular Weight Marker, Fermentas); 5 – OmpA from B. cenocepacia J2315; 6 – OmpA from B. cenocepacia SJ2.
Since the OmpA-like protein was not affected by the plasposon insertion, the previously observed change in the virulence of the SJ2 mutant towards *C. elegans* was hypothesized to be due to a transcript interrupted in the IGR between *ompA* and *gyrA* genes. As a first approach, an ORF search was performed for the IGR nucleotide sequence. Five non-annotated ORFs were predicted. However, due to their predicted locations, the hypothetical ORFs were not interrupted by the plasposon (data not shown). Therefore, a new ORF search was performed adding 38 nt of the sequence encoding the OmpA-like protein to the query sequence, since an initiation codon was located in the 5’ region of the *ompA* encoding gene. Six non-annotated ORFs (ORF1 to ORF6) were predicted (Fig. 3.2). The putative ORF3 was found to be the ORF more likely affected by the plasposon, since its insertion could result in the interruption of the putative ORF3 sequence.

<table>
<thead>
<tr>
<th>Label</th>
<th>Strand</th>
<th>Frame</th>
<th>Start</th>
<th>Stop</th>
<th>Length (nt</th>
<th>aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>+</td>
<td>1</td>
<td>55</td>
<td>411</td>
<td>357</td>
<td>118</td>
</tr>
<tr>
<td>ORF3</td>
<td>-</td>
<td>1</td>
<td>584</td>
<td>303</td>
<td>282</td>
<td>93</td>
</tr>
<tr>
<td>ORF6</td>
<td>-</td>
<td>3</td>
<td>222</td>
<td>115</td>
<td>108</td>
<td>36</td>
</tr>
<tr>
<td>ORF2</td>
<td>+</td>
<td>2</td>
<td>470</td>
<td>559</td>
<td>90</td>
<td>29</td>
</tr>
<tr>
<td>ORF5</td>
<td>-</td>
<td>2</td>
<td>166</td>
<td>77</td>
<td>90</td>
<td>29</td>
</tr>
<tr>
<td>ORF4</td>
<td>-</td>
<td>1</td>
<td>83</td>
<td>&gt;3</td>
<td>81</td>
<td>26</td>
</tr>
</tbody>
</table>

Northern blot assays were performed by the iBB research team to identify the transcription products originating from the IGR, using DNA probes against different regions of both DNA strands. Preliminary results with Int3_UP primer (section 2.2.7) showed the presence of a transcript in the reverse strand, with approximately 100 bp (data not shown).

To further characterize the transcript obtained by the Northern blot assay, 5’ and 3’ RACE analysis were carried out in the present work against the reverse strand and using pure total RNA samples to determine the 5’ and 3’ ends of this transcript. The 5’ RACE analysis of the intergenic region was performed using as 5’ RACE gene specific outer primer Int3_UP (section 2.2.7) and a band of approximately 100 bp, without the 45 bp belonging to the 5’ RACE adapter, was obtained (Fig. 3.3 lane 2). Then, an analysis using as 5’ RACE gene specific inner primer RACE 1 (section 2.2.6) was performed and a band of approximately 40 bp was obtained, without the 45 bp of the 5’ RACE adapter (Fig. 3.3 lane 1). These results demonstrated the presence of a transcript with a shorter 5’ end than the expected for the putative ORF3. In this case, the bands obtained should comprise a size higher than 335 and 294 bp, respectively. Since the transcript obtained by 5’ RACE appears to start only 100 nt upstream the putative ORF3 stop codon and 188 nt downstream its start codon, a specific primer to assess its 3’ end was designed for the initial sequence of the transcript. 3’ RACE analysis were performed using the 3’ RACE gene specific outer primer RACE 2 (section 2.2.7). Two bands were obtained with approximately 30 bp and 200 bp (Fig. 3.3 lane 3), without the 45 bp belonging to the 3’ RACE adapter. These fragments were blunt-end cloned into the HincII site of pUC19 plasmid and the resulting plasmids were used to transform *E. coli* α-DH5 competent cells.

At the moment, it was not possible to sequence the 5’ RACE fragments. Nevertheless, the 200 bp fragment obtained by 3’ RACE allowed the identification of a transcript with a size of 167 nt in the reverse strand of *B. cenocepacia* J2315 chromosome 1, spanning from nucleotides 3,238,624 to 3,238,791. It should be noted that this transcript could belong to the hypothetical ORF3 since it is present in the same region as the predicted ORF3.
In this case, if the transcript comprised only the 167 nt obtained with the 3’ RACE, the plasposon insertion would be situated at a distance of 20 nt upstream of the first nucleotide of the transcript. However, it is possible that the transcript has a larger size, since the amplification with 5’ RACE gene specific primers Int3_UP and RACE 1 primers points out to the presence of, at least, 12 additional nt at its 5’ end. Therefore, the transcript likely comprises 179 nt and the plasposon is potentially situated only 8 nt upstream of its encoding sequence.

Specific primers were designed for the amplification of the putative sequence encoding the transcript and RT-PCR experiments were carried out using the same cDNA samples employed in the RT-PCR of OmpA-like protein. The transcript under study was confirmed to be present under the conditions used in B. cenocepacia J2315 strain (Fig. 3.4, lane 1). However, its expression was also confirmed for SJ2 mutant (Fig. 3.4, lane 2) suggesting that it is not interrupted by the plasposon.

The abovementioned genetic elements appear to have no involvement in the virulence attenuation of the B. cenocepacia SJ2 mutant strain towards C. elegans.

Since the transcript obtained by 5’ RACE analysis started approximately 176 nt downstream the predicted ORF3 start codon and 100 nt upstream its stop codon and it is not interrupted in SJ2 mutant strain, we hypothesized that this transcript could belong to another element present in the IGR. Therefore, a specific forward primer was designed for the nucleotide sequence retrieved as belonging to the hypothetical ORF and RT-PCR experiments were carried out with the cDNA templates used in the abovementioned experiments. The transcription of the putative ORF3 was confirmed in B. cenocepacia J2315 (Fig. 3.5 lane 1). Furthermore, it was also observed that the plasposon is interrupting the ORF3, since its transcript was not detected in the SJ2 strain (Fig. 3.5 lane 2).
These results indicate the presence of at least two distinct genetic elements in the IGR, namely a transcript and a putative ORF. Since the sequence encoding the transcript obtained by RACE analysis does not contain start or stop codons nearby and due to its small size, we hypothesized that this sequence constitutes a sRNA.

Although the data collected from the ORF Finder tool did not allow an accurate chromosome mapping of the non-annotated ORF3, it allowed the prediction of its presence in the reverse strand of \textit{B. cenocepacia} J2315 chromosome 1, spanning from nucleotides 3,238,700 to 3,238,982 (Fig. 3.6). It should be mentioned that the sequence corresponding to ORF3 overlaps the nucleotide sequence encoding the sRNA at least 100 nt and the putative sequence encoding the OmpA-like protein in 29 nt (Fig. 3.6). In \textit{B. cenocepacia} SJ2, the plasposon is inserted 160 nt downstream the ATG codon of the putative ORF3 and 108 nt upstream its predicted stop codon (TAA). After the genome analysis of \textit{B. cenocepacia} J2315, the 5' RACE analysis above described was found to have been performed using primers for the overlapping sequence. For this reason, the 5' RACE analysis is expected to reveal two different bands corresponding to ORF3 and sRNA.

These elements will be characterized in the further sections.

3.2 Biocomputational analysis and phenotypic characterization of the hypothetical protein encoded by ORF3

3.2.1 The hypothetical protein ORF3 is conserved among \textit{Burkholderia} and has high homology with annotated ORFs

The \textit{B. cenocepacia} J2315 derived mutant SJ2 was previously identified by the iBB research team through the qualitative screening of mutants with reduced virulence towards \textit{C. elegans} (unpublished results). The results from this work revealed that the plasposon insertion interrupted a non-annotated ORF (Fig. 3.2 and 3.6), encoding a hypothetical protein composed of 93 aa residues. A putative TSS in
the 3,239,456 nt region of the reverse strand of *B. cenocepacia* J2315 chromosome 1, *i.e.*, 150 nt from the start codon of ORF3, was recently identified by Sass *et al.* [16]. To investigate the expression of the putative protein, the 5' hypothetical regulatory region upstream the nucleotide sequence predicted to encode ORF3 was analysed with the bacterial promoter prediction tools BPROM (Softberry) [143] and Virtual Footprint (Prodoric) [142]. Although no results were obtained with the BPROM tool, the Virtual Footprint tool predicted, with a high score (9.76 out of 10), the existence of a region where the *P. aeruginosa* transcriptional regulator RhIR could bind, 190 nt upstream the putative start codon of ORF3 encoding sequence. RhIR transcription regulator plays a central role in QS response [159].

Analysis carried out with the ProtParam tool [152] showed that the hypothetical protein has a predicted molecular mass of approximately 10.53 kDa and a pI of 10.20. In addition, the protein was predicted to be unstable [instability index (II) of 41.75] and hydrophilic (grand average of hydropathicity index, GRAVY, -0.141) [160].

The ORF3 nucleotide sequence was found to be conserved among 16 Bcc members and other *Burkholderia* species with minimal identity values of 94%. Moreover, the deduced amino acid sequence of ORF3 was found to be 98% identical to that of a recently annotated ORF belonging to the Bcc strain *B. multivorans* ATCC 17616 and 96% identical to the amino acid sequences of 5 annotated ORFs from *B. pseudomallei* strains (Fig. 3.7). No functions were described so far for these hypothetical proteins. Homologues with a sequence identity of 69% and 29% were found with annotated ORFs of unknown function from a *Paraburkholderia rhizoxinica* strain and from a Bacteroidetes bacterium, respectively. However, no conserved motifs could be predicted. The *in silico* consensus secondary structure of ORF3 was predicted with JPred4 web server [151] after the alignment of *Burkholderia* amino acid sequences (Fig. 3.7). Two α-helices and four β-sheets were predicted for the protein secondary structure. Moreover, extended regions of the amino acid sequence seem to be conserved among all the identical proteins from the *Burkholderia* genus.

![Figure 3.7 Prediction of the consensus secondary structure of B. cenocepacia J2315 ORF3 protein, by JPred4 web server.](image-url)
The predicted cellular localization of the *B. cenocepacia* J2315 ORF3 was analysed using different web tools. While SignalP 4.1 [153] and TaP 1.0 [154] servers failed to predict the presence of a N-terminal signal peptide and, consequently its localization, LocTree3 [156] webserver predicted with 94% accuracy the occurrence of secretion, suggesting that this protein is extracellular. The ClyA cytotoxin is an example of a bacterial protein secreted via the non-classical secretion pathway, being released from *E. coli* cells in vesicles detached from the outer membrane [161]. Indeed, SecretomeP 2.0 Server [155] predicted the secretion of the hypothetical protein by the non-classical pathway. Based on these bioinformatic results, it is possible that ORF3 protein is an extracellular protein.

### 3.2.2 *B. cenocepacia* J2315 ORF3 does not impair growth ability

The growth ability of the wild-type (WT) strain and SJ2 mutant was compared in rich medium (LB) and minimal medium (M9) at 37 °C, with orbital agitation (Fig. 3.8). The strains showed similar specific growth rates when growing in LB medium, since the calculated specific growth rate was approximately 0.38 h⁻¹ ± 0.01 for the WT strain and 0.37 h⁻¹ ± 0.01 for the SJ2 mutant (Fig 3.8 A). Although it was not possible to determine the specific growth rate in M9 medium due to the lack of time points during the exponential and stationary growth phases, a reduction in the OD₆₄₀ nm after 72 h of incubation was registered for the SJ2 (Fig. 3.8 B). This result suggests that the ORF3 is relevant for the bacterial growth in stationary phase in minimal media.

![Figure 3.8 Growth curves of B. cenocepacia strains J2315 (■) and SJ2 (●) in LB medium (n=2) (A) and M9 medium (n=3) (B) at 37 °C. Values are presented as mean ± SD.](image)

We also analysed the soluble proteins profile of *B. cenocepacia* J2315 and *B. cenocepacia* J2315 derived mutant SJ2. For this purpose, *B. cenocepacia* J2315 and *B. cenocepacia* SJ2 strains were incubated for 8 h in rich medium (LB) and for 24 or 48 h in minimal medium (M9). Samples of crude extracts prepared from cells harvested at each time point were analysed by Tricine-SDS-PAGE to study the effect of the interrupted ORF on the protein pattern of the total soluble proteins. As shown in Fig. 3.9, no significant differences on protein patterns could be detected.
Figure 3.9 Tricine-SDS-PAGE analysis of the total soluble proteins obtained from *B. cenocepacia* J2315 or *B. cenocepacia* SJ2 strains. Total soluble proteins were obtained from cells of *B. cenocepacia* J2315 and *B. cenocepacia* SJ2 after incubation at 37 °C for 8 h in LB rich medium (A) and 24 or 48 h in M9 minimal medium (B) and diluted to a final OD$_{640}$ nm of 0.4. Lanes: L – Molecular mass ladder (Unstained Protein Molecular Weight Marker, Fermentas); 1- total soluble proteins from *B. cenocepacia* J2315; 2 - total soluble proteins from *B. cenocepacia* SJ2; 3 and 5- total soluble proteins from *B. cenocepacia* J2315 after an incubation period of 24 or 48 h in M9, respectively; 4 and 6- total soluble proteins from *B. cenocepacia* SJ2 after an incubation period of 24 or 48 h in M9, respectively.

3.2.3 *B. cenocepacia* SJ2 mutant strain forms thicker biofilms and has an increased cell surface hydrophobicity

To evaluate the role of ORF3 on biofilm formation the *B. cenocepacia* J2315 and SJ2 strains were cultivated in microtiter plates and the amount of biofilms formed was assessed after 24 and 48 h in LBB medium [1% (w/v) NaCl], at 37 °C. The results presented in Fig. 3.10 show that the mutant strain forms significantly thicker biofilms than the WT strain after 24 h (*p < 0.001) and 48 h (*p < 0.05).

The hydrophobic properties of microbial cells are correlated with cell surface composition and has been ascribed an important role in the adherence of microorganisms to substrata, contributing to biofilms formation (reviewed in [162]). A hypothesis for the ability of SJ2 strain to form thicker biofilms is the increased hydrophobicity of its cells surface. Therefore, experiments of bacterial adhesion to the solvent hexadecane (BATH) [138] were performed (Fig. 3.11). The mutant SJ2 cells exhibited an increased cell surface hydrophobicity when compared to the WT strain. An alteration in the outer membrane of the SJ2 mutant cells might be responsible for the increased cell surface hydrophobicity observed.


3.2.4 The twitching motility of \textit{B. cenocepacia} SJ2 strain is apparently affected

Since SJ2 mutant strain forms thicker biofilms than the WT strain, experiments regarding flagellar-mediated (swimming) [163] and type IV pilus-mediated (twitching) motility [164] were performed. While no alterations were observed for the swimming motility, the twitching motility was decreased in the mutant (Fig. 3.12). These results correlate with those obtained in section 3.2.3, indicating that the SJ2 strain outer membrane topology could be affected.

![Figure 3.12 Comparison of \textit{B. cenocepacia} J2315 and \textit{B. cenocepacia} SJ2 swimming (A) and twitching (B) motility. Bars indicate the colony diameters obtained for \textit{B. cenocepacia} J2315 (black) or \textit{B. cenocepacia} SJ2 (grey), after 48 h of incubation in adequate solid media at 37 °C (A) or 72 h at 37 °C (B). Results are expressed as mean ± SD. Statistical analysis performed with the unpaired \( t \)-test (n=2).]

3.2.5 \textit{B. cenocepacia} SJ2 strain shows increased mortality at 50 °C and cell lysis rate in the presence of SDS

To further investigate the role of the hypothetical protein in the bacterial membrane stability, stress susceptibility experiments involving exposure to heat or SDS stress were performed.

Membrane fatty acid composition and fluidity are two components that have been shown to affect bacterial thermal resistance [165,166]. Thus, \textit{B. cenocepacia} J2315 and SJ2 strains resistance to heat-shock stress was evaluated. Although only two replicates have been performed and consequently, a statistical analysis could not be done, an accentuated decrease in SJ2 survival was observed when compared with the WT strain (Fig. 3.13). After 5 min of exposure to 50 °C, the percentage of \textit{B. cenocepacia} J2315 survivors was approximately 5-fold higher than the percentage of \textit{B. cenocepacia} SJ2 survivors. Increasing the exposure time to the thermal stress to 15 min led to a general decay in the

![Figure 3.11 Assessment of \textit{B. cenocepacia} J2315 (■) and \textit{B. cenocepacia} SJ2 (○) cells hydrophobicity. Bacterial adhesion to hexadecane was evaluated by the addition of 0-800 µL of hexadecane to 1.5 mL bacterial suspensions with an OD\(_{640\text{nm}}\) of 0.6, followed by OD\(_{640\text{nm}}\) measurement of the aqueous phase. Values represent mean ± SD and are expressed as percentages relative to the OD\(_{640\text{nm}}\) of the initial cell suspension. Statistical analysis performed with the unpaired \( t \)-test (n=2).]

Figure 3.11 Assessment of \textit{B. cenocepacia} J2315 (■) and \textit{B. cenocepacia} SJ2 (○) cells hydrophobicity. Bacterial adhesion to hexadecane was evaluated by the addition of 0-800 µL of hexadecane to 1.5 mL bacterial suspensions with an OD\(_{640\text{nm}}\) of 0.6, followed by OD\(_{640\text{nm}}\) measurement of the aqueous phase. Values represent mean ± SD and are expressed as percentages relative to the OD\(_{640\text{nm}}\) of the initial cell suspension. Statistical analysis performed with the unpaired \( t \)-test (n=2).
percentage of bacterial survivors. However, the percentage of survivors for the WT strain was approximately 2%, while SJ2 survivors were almost non-detectable (0.2%).

Annous et al. [165] and Sampathkumar et al. [166] have shown for Pediococcus spp. and Salmonella enteritidis, respectively, that the presence of unsaturated fatty acids in the membranes leads to higher membrane fluidity and, consequently, reduced thermal resistance to high temperatures. Our results suggest that the hypothetical protein could be related to the maintenance of the membrane fluidity, since an interruption in the respective ORF drastically decreased the percentage of survivors upon exposure to 50 °C. The involvement of other mechanisms, including heat-shock proteins responses cannot be excluded.

Denaturing detergents, such as the anionic detergent SDS, can cause membrane disruption by interfering with the inter- and intra-molecular protein-protein interactions [167].

Since B. cenocepacia SJ2 strain possibly presents structural changes at the membrane level, the mortality kinetics of B. cenocepacia J2315 and SJ2 strains to pulses of different SDS concentrations were performed by measuring the OD$_{660}$ nm at room temperature during 10 min. The cell lysis rate of SJ2 mutant strain was higher than that of the WT for all the concentrations of SDS tested (Fig. 3.14). For a final concentration of 0.1% (v/v), the cell lysis rate of SJ2 strain presented approximately a 3-fold increase when compared to the WT cell lysis rate. For the final SDS concentrations of 0.5% and 1% (v/v) it was observed an increase of 2.5-fold and 1.5-fold increase for 5% (v/v) of SDS.

Indeed, a higher cell lysis rate was expected for the mutant strain since the outer membrane is likely altered and consequently, protein-protein interactions can be weakened. This would cause an effortless disruption of the intramembrane interactions by SDS.

![Figure 3.13 B. cenocepacia SJ2 mutant strain is more susceptible to heat-shock stress than the WT strain.](image)

![Figure 3.14 B. cenocepacia SJ2 mutant strain (●) exhibits an increased cell lysis rate upon exposure to SDS, when compared to B. cenocepacia J2315 (○).](image)
3.2.6 The *B. cenocepacia* SJ2 strain is more susceptible to ceftazidime, imipenem and tetracycline

The effects of the interrupted ORF on antibiotics susceptibility were evaluated by the broth microdilution technique [140]. A total of nine antibiotics from different classes and currently used in the clinical setting were used in this study. These antibiotics included β-lactams, more specifically the cephalosporin ceftazidime, a carbapenem (imipenem) and a monobactam (aztreonam); aminoglycosides, such as amikacin, streptomycin and gentamicin; tetracycline; fluoroquinolones, in particular ciprofloxacin, and the antimetabolite trimethoprim. The MIC values obtained are presented in Table 3.1, showing an increased susceptibility of the SJ2 mutant to two β-lactams (imipenem and ceftazidime) with a 2 and 4-fold decrease in the MIC value, respectively, and to the antibiotic tetracycline by presenting a 2-fold decrease in the MIC value.

Table 3.1 Antibiotic susceptibility of *B. cenocepacia* J2315 and *B. cenocepacia* J2315 derived mutant SJ2. MIC values were assessed by the microdilution method (n=4) and are presented in µg/mL. Abbreviations: IPM - Imipenem; AMK - Amikacin; TET - Tetracycline; CIP - Ciprofloxacin; ATM - Aztreonam; SM - Streptomycin; CAZ - Ceftazidime; GEN - Gentamicin; TMP – Trimethoprim.

<table>
<thead>
<tr>
<th>Strains</th>
<th>IPM</th>
<th>AMK</th>
<th>TET</th>
<th>SM</th>
<th>CIP</th>
<th>ATM</th>
<th>CAZ</th>
<th>GEN</th>
<th>TMP</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. cenocepacia</em> J2315</td>
<td>128</td>
<td>256</td>
<td>64</td>
<td>&gt;1,024</td>
<td>8</td>
<td>&gt;1,024</td>
<td>128</td>
<td>1,024</td>
<td>1,024</td>
</tr>
<tr>
<td><em>B. cenocepacia</em> SJ2</td>
<td>64</td>
<td>256</td>
<td>32</td>
<td>&gt;1,024</td>
<td>8</td>
<td>&gt;1,024</td>
<td>32</td>
<td>1,024</td>
<td>1,024</td>
</tr>
</tbody>
</table>

β-lactams are small hydrophilic drugs that use pore-forming porins to enter the cells [168] and act by interfering with penicillin-binding proteins involved in the synthesis of the bacterial peptidoglycan layer, leading to an inhibition of the cell wall synthesis (reviewed in [169]). Tetracyclines are hydrophobic drugs that can enter the cells by lipid-mediated or porin-mediated pathways [170] and act by blocking protein synthesis [171]. These two classes of antibiotics are very distinct in their mode of action, however they can gain access to the cell interior through similar porins [170].

Interestingly, a few charged residues of the OmpF porin constriction zone have been demonstrated to affect β-lactams flux and sensitivity, when mutated [172]. However, the bioinformatic tool TMHMM Server v. 2.0 [157] does not predict the occurrence of any transmembrane domain in ORF3 protein.

Further work must be performed to elucidate the role of ORF3 in antibiotic resistance.

3.2.7 *B. cenocepacia* SJ2 mutant strain exhibits decreased ability to kill the nematode *C. elegans*

To investigate the role of ORF3 protein in *B. cenocepacia* J2315 virulence, slow-killing experiments using the infection model *C. elegans* were performed as mentioned in section 2.3.

The *B. cenocepacia* SJ2 mutant showed a time-dependent decrease on its ability to kill the nematodes (Fig. 3.15). While *B. cenocepacia* J2315 strain was able to reduce to 50% the initial nematode population within approximately 1.5 days, the *B. cenocepacia* SJ2 strain achieved the same level of nematode population after, approximately, 2.5 days. The difference between the WT and SJ2 strain in terms of nematodes killing efficiency was clearer during days 3 and 4. The WT strain reduced the initial population to 18% and 10% by days 3 and 4, respectively, whereas SJ2 mutant caused a more attenuated reduction to, approximately, 43% and 30%, respectively, during the same period. Moreover, in the last day of this experiment almost no surviving worms could be detected for the WT (approximately 2%), while for the SJ2 mutant 15% of the initial worms were still alive.
The reduced ability of SJ2 to kill the nematodes population cannot be attributed to growth defects, since growth curve assays performed in rich medium and under starvation conditions (LB and M9, respectively) showed no significant differences (Fig. 3.8). Although no statistical significance was obtained, the data herein presented suggests that the ORF3 plays a role in the virulence of *B. cenocepacia* J2315 and might be particularly important in the progression of the infection. It should be noted that to confirm these data more experiments should be performed and the SJ2 mutant ability to colonize the nematode intestinal tract should be investigated.

3.3 Biocomputational analysis and phenotypic characterization of MavA sRNA

3.3.1 MavA sRNA is specific of the *Burkholderia* genus and has one functional homologue

The sRNA identified through the characterization of the *B. cenocepacia* J2315 derived mutant SJ2 was named MavA and comprises an estimated size of 179 nt. To investigate the expression of MavA, the 5’ putative regulatory region upstream the nucleotide sequence obtained by 3’ RACE was analysed with the bacterial promoter prediction tools BPROM (Softberry) [143] and Virtual Footprint (Prodoric) [142]. Both programs retrieved a typical bacterial sigma 70 (σ70) promoter harbouring the canonical -10/-35 elements [16,173] (Fig. 3.16 A). However, the functionality of this promoter is still not proven. A putative TSS in the region 3,238,600 nt of the reverse strand in *B. cenocepacia* J2315 chromosome 1 was recently reported [16]. Thus, the predicted TSS is located at a distance of 7 nt relatively to the consensus -10 box, according with previous reports [174]. The presence of this promoter and TSS support the hypothesis that MavA is larger than 167 bp. In this case, the size obtained by RACE analysis would be 236 nt and, opposite to our results, MavA transcription should be affected by the plasposon.

The 3’ putative regulatory region downstream the mavA nucleotide sequence and upstream the sequence encoding GyrA enzyme was analysed using the Rho-independent transcription terminators web tools, FindTerm [143] and ARNold [22]. Although only one putative terminator was obtained from each tool, both programs retrieved the same sequence, approximately at 55 nt downstream the mavA encoding sequence (Fig. 3.16 B).

Figure 3.15 Comparison of *E. coli* OP50 (purple bars), *B. cenocepacia* J2315 (black bars) and *B. cenocepacia* SJ2 (grey) ability to kill the nematode *C. elegans*. Values represent mean ± SD and are expressed as percentages relative to the initial number of hypochlorite-synchronized L4 larvae of *C. elegans* DH26 inoculated (between 18-30). Statistical analysis performed with the unpaired t-test (n=2).
Regarding the currently available genomes, *B. cenocepacia* J2315 mavA nucleotide sequence was found to be conserved among 17 Bcc members and other *Burkholderia* species with minimal identity values of 88% and 80%, respectively. However, no conserved domains could be detected and it was not possible to retrieve mavA homologues within non-*Burkholderia* genomes, suggesting that MavA sRNA is specific to members of the *Burkholderia* genus.

Figure 3.16 Prediction of putative promoters and Rho-independent transcription terminators of MavA. Relevant nucleotide sequence of the 5’ putative regulatory region upstream MavA encoding sequence. The predicted -10 and -35 elements of the promoter are underlined. The predicted TSS [16] is represented in blue by an arrow at the position +1 (A). Relevant nucleotide sequence of the 3’ putative regulatory region upstream MavA encoding sequence. The predicted region of the Rho-independent transcription terminator is represented in bold, while the palindromic sequence and the stretch of T’s, that demonstrate the possible presence of a GC-rich hairpin [22] are highlighted in purple with arrows and with a box, respectively (B).

The *in silico* secondary structure of MavA was predicted with LocARNA [146] and RNAalifold [145] web tools after the alignment of MavA Bcc sequences (Fig. 3.17 A). The secondary structure obtained is presented in Fig. 3.17 B. MavA likely folds into a highly paired and thermostable structure due to the absence of unpaired regions and presence of stem-loops in the 5’ and 3’ ends with a high GC content, respectively.
Figure 3.17 Prediction of the secondary structure of MavA sRNA. Multiple alignment of MavA RNA sequences from eight members of the Bcc and computed with LocARNA tool [146]. The black rectangle delineates the 12 nt that are likely to constitute MavA sequence, as predicted by MavA amplification with different primers (A). MavA sRNA consensus secondary structure with a free energy of -59.46 kcal/mol was predicted by RNAalifold web tool [145] (B). Both figures are coloured following a structural conservation of the base pair: red – all sequences used to generate the consensus structure have the base pair; yellow – two types of base pairs can occur; green – three types of base pairs can occur. The saturation decreases with the number of incompatible base pairs.

A search for MavA functional homologues within all the bacterial genomes with available sRNA annotations was performed using the sRNAMap web tool [144]. This tool searches for homologs base pairs within families of functional sRNAs. However, the use of these bioinformatic tools are challenging since the majority of the information present in the databases was obtained from studies performed with E. coli, Salmonella and Yersinia species. Therefore, there is a lack of information about functional sRNAs in β-proteobacteria and, particularly, in organisms from the Burkholderiales order. For MavA sRNA four functional homologues were predicted, however the identity values obtained were very low. The homologue with the lowest E-value was considered for further analysis (Fig. 3.18 A). Despite MavA and its predicted homologue (IstR-2) have only 12 nt shared in their sequences (Fig. 3.18 B), they present some similar features. IstR-2 is a 140 nt Hfq-dependent sRNA encoded in an intergenic region present in the reverse strand of E. coli K-12 MG1655 genome. Similar to MavA (Fig.3.6) IstR-2 is another sRNA encoded in the same region of IstR-1 and although both sRNAs are transcribed from different promoters, there is an overlap of their 3’ ends [175]. Besides encoding two sRNAs, this locus also encodes a toxic peptide (TisB) whose translation is inactivated by IstR-1 RNA and tisAB mRNA base pairing [175]. The transcription of IstR-2 is induced in response to DNA damage via the SOS system, being particularly expressed in the presence of a SOS response triggering factor - Mytomicin C (MMC) [175]. Unlike IstR-1, the role of IstR-2 sRNA in tisAB regulation is still unknown. Ortega et al. [176] have recently demonstrated that IstR-2 sRNA is upregulated in non-growing intracellular Salmonella enterica serovar Typhimurium and it is the least abundant sRNA when bacteria are extracellular. These findings suggest that this sRNA plays an important role in the adaptation of S. enterica Typhimurium to an intracellular lifestyle, although its target remains unknown.
3.3.2 MavA putative targets in *B. cenocepacia J2315* genome

In order to functionally characterize MavA sRNA a search for putative mRNA targets was conducted. Each of the three chromosomes from *B. cenocepacia J2315* was inspected for putative MavA targets. 30 putative MavA targets with the lowest energy were predicted with TargetRNA2 [148] (Table 3.2) and RNAPredator [149] (Annex) web tools. From all the data obtained only a total of twelve targets were common to both web tools. This result was not unexpected since each tool applies a different algorithm.
The putative targets retrieved are involved in various cell functions, such as cell metabolism and inorganic ion transport; energy production and conversion; cell motility and secretion; signal transduction; lipids metabolism; cell envelope biogenesis and outer membrane; amino acid transport; DNA replication, recombination and repair; transcription and translation (ribosomal structure). Based on the analysis performed MavA appears to exert mainly a downregulation on its targets, since it was predicted a base pairing between the sRNA and the SD sequence/AUG codon (Fig. 1.3 C) [12]. Moreover, due to the distant genomic localization of its targets MavA is likely a trans-encoded sRNA.

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>mRNA information</th>
<th>Energy (kJ/mol)</th>
<th>p-value</th>
<th>Locus Tag</th>
<th>mRNA information</th>
<th>Energy (kJ/mol)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCAL3476</td>
<td>putative type-b cytochrome</td>
<td>- 14.72</td>
<td>0.001</td>
<td>BCAL1862</td>
<td>(pbbA gene) acetyl-CoA acetyltransferase binding-protein-dependent transport system protein</td>
<td>- 10.04</td>
<td>0.023</td>
</tr>
<tr>
<td>BCAL1572</td>
<td>hypothetical protein</td>
<td>- 13.30</td>
<td>0.003</td>
<td>BCAM0955</td>
<td></td>
<td>- 9.95</td>
<td>0.024</td>
</tr>
<tr>
<td>BCAL0749</td>
<td>putative cytochrome c oxidase</td>
<td>- 12.44</td>
<td>0.006</td>
<td>BCAL0634</td>
<td>putative lipoprotein</td>
<td>- 9.83</td>
<td>0.025</td>
</tr>
<tr>
<td>BCAL0135</td>
<td>chemotaxis protein CheY</td>
<td>- 11.89</td>
<td>0.008</td>
<td>BCAM0178</td>
<td>putative periplasmic solute-binding protein</td>
<td>- 9.78</td>
<td>0.026</td>
</tr>
<tr>
<td>BCAS0071</td>
<td>hypothetical protein</td>
<td>- 11.86</td>
<td>0.008</td>
<td>BCAM2225</td>
<td>ABC transporter ATP-binding membrane protein</td>
<td>- 9.78</td>
<td>0.026</td>
</tr>
<tr>
<td>BCAM0055</td>
<td>GTP cyclohydrolase I</td>
<td>- 11.75</td>
<td>0.009</td>
<td>BCAL3269</td>
<td>(dnaJ gene) chaperone protein DnaJ</td>
<td>- 9.69</td>
<td>0.027</td>
</tr>
<tr>
<td>BCAL0037</td>
<td>FoF1 ATP synthase subunit epsilon</td>
<td>- 11.74</td>
<td>0.009</td>
<td>BCAL3265</td>
<td></td>
<td>- 9.61</td>
<td>0.028</td>
</tr>
<tr>
<td>BCAL1472</td>
<td>succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A</td>
<td>- 11.50</td>
<td>0.010</td>
<td>BCAM1815</td>
<td>short chain dehydrogenase</td>
<td>- 9.54</td>
<td>0.029</td>
</tr>
<tr>
<td>BCAL2353</td>
<td>putative sulfate transporter</td>
<td>- 11.14</td>
<td>0.013</td>
<td>BCAL1128</td>
<td>putative DNA-binding protein</td>
<td>- 9.51</td>
<td>0.030</td>
</tr>
<tr>
<td>BCAS0331</td>
<td>hypothetical protein</td>
<td>- 10.89</td>
<td>0.015</td>
<td>BCAL1915</td>
<td>putative decarboxylase</td>
<td>- 9.38</td>
<td>0.031</td>
</tr>
<tr>
<td>BCAL0728</td>
<td>hypothetical protein</td>
<td>- 10.82</td>
<td>0.015</td>
<td>BCAL0229</td>
<td>30S ribosomal protein S12</td>
<td>- 9.31</td>
<td>0.032</td>
</tr>
<tr>
<td>BCAM1199</td>
<td>two-component regulatory system response regulator protein</td>
<td>- 10.63</td>
<td>0.017</td>
<td>BCAL3099</td>
<td>putative branched-chain amino acid transporter permease</td>
<td>- 9.28</td>
<td>0.033</td>
</tr>
<tr>
<td>BCAL3061</td>
<td>hypothetical protein</td>
<td>- 10.39</td>
<td>0.019</td>
<td>BCAL1691</td>
<td>(orbC gene) putative iron transport-related ATP-binding protein</td>
<td>- 9.16</td>
<td>0.035</td>
</tr>
<tr>
<td>BCAL1493</td>
<td>hypothetical protein</td>
<td>- 10.30</td>
<td>0.020</td>
<td>BCAS0104</td>
<td>(fliD2 gene) A-type flagellar hook-associated protein 2 (HAP2)</td>
<td>- 9.02</td>
<td>0.037</td>
</tr>
<tr>
<td>BCAL2598</td>
<td>putative DNA-binding protein</td>
<td>- 10.14</td>
<td>0.022</td>
<td>BCAM1545</td>
<td>LuxR superfamily regulatory protein</td>
<td>- 9.00</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Table 3.2 Predicted mRNA targets of MavA in *B. cenocepacia* J2315 genome with TargetRNA2 tool [148].
although both programs predicted a sRNA-mRNA interaction involving 10 to 20 nt. The first 13 to 36 nt of MavA sequence were the most implicated in the predicted interactions.

Phenotypes studies were performed to confirm if any of the predicted targets could be in fact a real target of MavA.

3.3.3 MavA overexpression leads to S12 protein overexpression

To confirm that MavA sRNA targets mRNAs, the sRNA was cloned on the constitutive expression vector pMLS7 containing the Ps; B. cepacia specific promoter [122]. The resulting plasmid, pMya2, was used to transform B. cenocepacia J2315 cells by triparental mating.

The B. cenocepacia J2315 cells carrying the plasmid overexpressing MavA sRNA (pMya2) were incubated for 8 h in rich medium (LB) and for 24 and 48 h in minimal medium (M9). Samples of cell extracts prepared for each time point were analysed by Tricine-SDS-PAGE in order to study the influence of MavA overexpression in the total soluble proteins. A protein with an apparent molecular mass of 14 kDa was found as overexpressed by B. cenocepacia J2315 (pMya2) cells (Fig. 3.19) after incubation in LB medium and M9 medium. The levels of this protein increased only in the total soluble proteins samples of B. cenocepacia overexpressing MavA sRNA. This result suggests a possible role for MavA sRNA.

Identification of the 14 kDa protein overexpressed was performed by mass spectrometry (MS) in the group of Professor Pedro Santos (Centre of Molecular and Environmental Biology, at the University of Minho). The sequencing results showed that the separated bands correspond to the S12 ribosomal protein, a component of the 30S ribosomal subunit encoded by rpsL gene [177].

With the parameters used, the locus BCAL0229, where the rpsL gene is located, was only predicted with the TargetRNA2 tool. This tool computed rpsL gene as a possible MavA target (Table 3.2), yielding −9.31 kJ/mol of energy and a p-value of 0.032. Interestingly, the sRNA-mRNA interaction was predicted as occurring within the first 11 nt of MavA sequence, which are conserved among Burkholderia species. These results demonstrate that bioinformatic predictions are not 100% reliable and the best strategy is the use of both bioinformatics and experimental techniques.
Additionally, MavA could exert a downregulation of *rpsL* gene, since the sRNA was predicted by TargetRNA2 to bind to the start codon of *rpsL* mRNA (Fig. 1.3 C). According to the experimental data obtained (Fig. 3.19) MavA sRNA does not exert a negative regulation, at least directly, on *rpsL* gene. Therefore, we hypothesized that MavA can be involved in an anti-antisense mechanism, negatively regulating a negative regulator of *rpsL* gene. *E. coli* MicF sRNA is a recent example of such mechanism, since it negatively represses the translation of leucine-responsive regulatory protein (Lrp), which in turn negatively regulates operons involved in amino acid catabolism and peptide transport [178].

High-resolution structural and biochemical studies of ribosomes and ribosomal complexes have characterized S12 protein as an important player in the process of translation, particularly to discriminate cognate from near-cognate tRNA species. Furthermore, S12 protein also promotes global changes in the small ribosomal subunit favouring its transition to a domain “closure” during the translation cycle progression [179,180].

The best characterized phenotype involving S12 protein and *rpsL* gene was first described in *E. coli* [177]. Point mutations in *rpsL* gene and, consequently, in S12 protein resulted in a reduced sensitivity to the antibiotic streptomycin, a broad-spectrum aminoglycoside that interacts with the 30S ribosomal subunit, leading to a misreading of the mRNA [181] and affecting ribosomal proofreading [182]. More recently, the spontaneous mutations responsible for the increase of bacterial resistance to this aminoglycoside were shown to be often accompanied by a decrease of the bacterial resistance to tetracyclines [183,184]. Tetracycline functions differently from streptomycin, blocking the attachment of aminoacyl-tRNA to the ribosomal acceptor (A) site [185].

To investigate if MavA sRNA overexpression could influence S12 protein and consequently streptomycin and tetracycline resistance, the susceptibility to these antibiotics was determined by broth micro dilution technique based on standard techniques described by Andrews [140]. *B. cenocepacia J2315* (pMLS7) and *B. cenocepacia J2315* (pMya2) strains exhibited a MIC higher than 1,024 µg/mL for streptomycin. It should be noted that *B. cenocepacia J2315* strain is highly resistant to this antibiotic. The highest concentration tested in this experiment was 1,024 µg/mL and the WT strain had a MIC value higher than this concentration. Thus, if MavA sRNA affects S12 protein levels, causing increased resistance to antibiotics we were not able to detect them. As abovementioned, certain types of mutations in this protein have been described as frequently associated to a higher susceptibility to other antibiotics, such as tetracycline. The MIC values obtained for *B. cenocepacia J2315* (pMya2), *B. cenocepacia J2315* (pMLS7) and *B. cenocepacia J2315* strains were identical, 64 µg/mL. Moreover, the expression vector used in this study did not influence the bacterial resistance to the tested antibiotics. Together, these results suggest that the overexpression of MavA sRNA does not affect the resistance to the tested antibiotics by means of S12 protein overproduction.
3.3.4 MavA overexpression does not affect *B. cenocepacia* J2315 ability to grow in minimal medium

The growth ability of the WT and *B. cenocepacia* J2315 overexpressing MavA strain was compared in cultures growing in minimal medium (M9) at 37 ºC with orbital agitation. No differences were observed, with both strains showing similar growth curves and indicating that MavA overexpression does not affect the ability of the strain to grow on minimal medium (Fig. 3.20).

![Growth curve comparison](image)

**Figure 3.20 Growth curve comparison.** Comparison of *B. cenocepacia* strains J2315 (■) and *B. cenocepacia* J2315 (pMya2) (▲) growth in M9 medium with orbital agitation at 37 ºC (n=3). Values represent mean ± SD.

3.3.5 MavA overexpression affects *B. cenocepacia* J2315 resistance to heat-shock

Some trans-acting sRNAs have been discovered to be induced under heat-shock conditions (reviewed in [186]). MicA sRNA was first described in *E. coli* and is an example of a sRNA whose expression is highly induced under this type of stresses [187], acting through the repression of the translation of *omp* mRNAs [188,189].

To test the hypothesis that MavA sRNA influences *B. cenocepacia* J2315 resistance to heat-shock, the resistance of *B. cenocepacia* J2315, *B. cenocepacia* J2315 (pMLS7) and (pMya2) strains to thermal stress was evaluated. Aliquots of cultures of each strain were incubated at 50 ºC for 5 and 15 min. After these periods of exposure to 50 ºC, each culture was serially diluted and plated and bacterial survivors were counted. Although it was not possible to perform a statistical analysis with the results obtained, since only two replicates were performed, it is possible to visualize an increase in the number of bacterial survivors in the strain overexpressing MavA sRNA (Fig. 3.21). After a short-incubation period of 5 min under this stress condition, *B. cenocepacia* J2315 (pMya2) strain shows approximately a 1.4-fold increase in the percentage of bacterial survival when compared to *B. cenocepacia* J2315 (pMLS7), while *B. cenocepacia* J2315 (pMLS7) shows approximately a 1.2-fold increase in the percentage of bacterial survival when compared to the WT strain. A 3-fold increase in the exposure time at 50 ºC results in a general decay in the percentage of bacterial survivors, with the WT and *B. cenocepacia* J2315 (pMLS7) strains showing no differences between the percentage of survivors. Remarkably, *B. cenocepacia* J2315 (pMya2) strain shows approximately a 10-fold increase in the percentage of bacterial survivors when compared to the result obtained with *B. cenocepacia* J2315 (pMLS7) strain (Fig. 3.21).

According to these results, MavA sRNA increases bacterial resistance to heat-shock stress (50 ºC). However, the mechanism underlying this effect remains unknown.
3.3.6 MavA overexpression does not affect biofilm formation

The capacity of a bacteria to form biofilms indicates a higher virulent behaviour and ability to cause persistent infections [109]. Moreover, biofilms have also been associated to an enhanced in vitro antibiotic resistance [110]. The ability of Bcc strains under study to form biofilms was compared after 24 and 48 h in LBB medium at 37 ºC. Although it was possible to observe a slightly increase in the amount of biofilms formed by *B. cenocepacia* J2315 (pMya2) when compared to *B. cenocepacia* J2315 (pMLS7), no differences with statistical significance were observed (Fig. 3.22). This result suggests that MavA does not play a significant role in the persistence of infections caused by Bcc bacteria.

Despite MavA does not influence biofilm formation, the effect of this sRNA in two distinct types of motility were tested, namely in swimming and twitching motilities (Fig. 3.23). Swimming motility is an individual type of motility characterized by rotating flagella [163], whereas twitching motility occurs by extension and retraction of type IV pili [164]. Although none of the components necessary for these types of motility were predicted as possible MavA targets (Table 3.2), the results obtained for flagellar-mediated swimming motility suggest that overexpression of this sRNA can slightly increase this type of motility (Fig. 3.23 A). However, more experiments must be performed to confirm this hypothesis.
3.3.7 MavA sRNA overexpression does not affect the ability of *B. cenocepacia* J2315 to kill the nematode *C. elegans* significantly

To investigate the role of MavA sRNA in *B. cenocepacia* J2315 virulence, slow-killing experiments using the infection model *C. elegans* were performed according to Sousa *et al.* [135]. The *C. elegans* mutant strain DH26 was used in this study since it is unable to reproduce at 25 °C, being a suitable strain to count nematodes without progeny interference. The *E. coli* OP50 strain was used as a control in these experiments due to its attenuated virulence towards this nematode when compared to *B. cenocepacia* J2315 [119].

Over the five-day period of this experiment, the population of worms feeding on *B. cenocepacia* J2315 strain and their derivatives harbouring pMLS7 or pMya2 showed a gradual decrease in the number of surviving worms (Fig. 3.24). Remarkably, in the last day only 2% of the nematodes were detected as being alive. *B. cenocepacia* J2315 strain carrying the expression vector pMLS7 induced a striking reduction in the initial population during the first two days. In fact, at day 1 the percentage of surviving worms was 50% and in day 2 only 30% of the worms had survived. A lower decrease in the initial population was observed for the WT strain that revealed 60% and 40% of surviving worms at day 1 and day 2, respectively. Noteworthy, after day 3 this trend was reversed and the WT strain showed a higher decrease when compared to those feeding on *B. cenocepacia* J2315 (pMLS7). However, there was no significant differences between the *B. cenocepacia* J2315 strain carrying the expression vector pMLS7 and the WT strain.

The *B. cenocepacia* J2315 strain overexpressing MavA sRNA (pMya2) exhibited during the five-day period a slightly reduced ability to kill the nematode (Fig. 3.24). After day 1 and day 2, the percentage of surviving worms feeding on *B. cenocepacia* J2315 (pMya2) was approximately 8% higher than the percentage of surviving worms feeding on *B. cenocepacia* J2315 (pMLS7). At days 3, 4 and 5 the difference between the percentage of surviving worms obtained for the strain overexpressing MavA and for *B. cenocepacia* J2315 (pMLS7) decreased to approximately 5%. A higher percentage of surviving worms was observed for populations feeding on *B. cenocepacia* J2315 (pMya2).

Although no statistical significance was obtained, these results suggest a less virulent behaviour of *B. cenocepacia* J2315 (pMya2) strain. This behaviour could be explained by the presence of growth defects in the strain *B. cenocepacia* J2315 (pMya2), however at least under starvation no significant differences were observed (Fig. 3.20).

Overall, these results suggest that MavA sRNA appears to play no significant role on *B. cenocepacia* J2315 virulence. Nevertheless, MavA can be important for the adaptation of *B. cenocepacia* J2315 to...
an intracellular lifestyle, similarly to its homologue, IstR-2 sRNA [176]. To confirm these hypotheses further experiments have to be performed.

Figure 3.24 Comparison of *E. coli* OP50 (purple bars), *B. cenocepacia* J2315 (black bars), *B. cenocepacia* J2315 (pMLS7) (dark-grey bars) and *B. cenocepacia* J2315 (pMya2) (light-grey bars) ability to kill the nematode *C. elegans*. Values represent mean ± SD and are expressed as percentages relative to the initial number of hypochlorite-synchronized L4 larvae of *C. elegans* DH26 inoculated (between 18-30). Statistical analysis performed with the unpaired *t*-test (n=2).
4. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Several Bcc strains present beneficial effects due to their diverse abilities that include promoting plant growth, production of antifungals and degradation of xenobiotics (reviewed in [59]). However, they have also arisen as highly problematic opportunistic pathogens in vulnerable patients, mainly CF patients [60,81,95–97]. Actually, in these patients Bcc infections are often associated with a fatal outcome - “cepacia syndrome” [60]. Moreover, the resistance of several Bcc species to the majority of clinically available antimicrobials [93] and the recognition of some Bcc strains as emergent nosocomial pathogens in oncology patients [69,97] highlight the need of understanding the mechanisms underlying the Bcc pathogenesis and virulence, in order to identify novel therapeutic targets to combat life-threatening infections.

sRNAs have been widely discovered among the three domains of life (Bacteria, Eukarya and Archaea) and they are known to play important roles in bacterial virulence, acting mainly at the posttranscriptional level of gene expression regulation [9] (reviewed in [10]).

In this study, we report the presence of a non-annotated ORF and a sRNA, named ORF3 and MavA, respectively, overlapping at least 100 nt, in the 555 bp IGR of *B. cenocepacia* J2315 chromosome 1. These genetic elements were identified in the *B. cenocepacia* J2315 mutant SJ2, obtained by plasposon insertion. Stress resistance, mobility and virulence assays in the nematode *C. elegans*, as well as extensive bioinformatics analyses were performed in order to characterize the sRNA and the hypothetical ORF.

ORF3 protein is predicted to be composed of 93 aa residues and its encoding ORF was found to be conserved among the *Burkholderia* genus. Moreover, ORF3 protein sequence was identical to that of an annotated ORF of *B. multivorans* and to five annotated ORFs of *B. pseudomallei* species. Bioinformatics analyses and RT-PCR experiments showed that this hypothetical protein is affected by the plasposon insertion in *B. cenocepacia* SJ2 and it appears to play an important role in *B. cenocepacia* J2315 stress response. This conclusion was built due to the fact that its mutation leads to an increased sensitivity of *B. cenocepacia* J2315 to thermal stress and enhanced cell lysis rate when exposed SDS. In addition, *B. cenocepacia* SJ2 mutant strain also showed increased cell hydrophobicity and, consequently, a higher ability to form biofilms. Regarding the antimicrobials effects, the *B. cenocepacia* SJ2 mutant strain revealed an enhanced sensitivity to two β-lactams (ceftazidime and imipenem) and to the antibiotic tetracycline. Overall these results suggest that this hypothetical protein could be implicated in the maintenance of outer membrane integrity. Moreover, ORF3 protein was also shown to be involved in the virulence of *B. cenocepacia* J2315 strain.

The sRNA identified in this study (MavA) comprises approximately 179 bp and was found to be widely conserved among the *Burkholderia* genus. IstR-2 was identified has being the MavA functional homologue, however its biological role remains unknown. Experiments regarding the effect of MavA overexpression in total soluble proteins were performed and this work demonstrates that the levels of the ribosomal protein S12, a component of the 30S ribosomal subunit, are increased when MavA is overexpressed. However, no differences were observed in the resistance and susceptibility to the antibiotics streptomycin and tetracycline, when *B. cenocepacia* J2315 overexpressed MavA. These findings indicate that MavA does not affect S12 protein-mediated resistance to these antimicrobials. Our preliminary results suggest that the MavA sRNA might play a direct or indirect role in swimming motility, while no significant differences were observed in the amount of biofilms formed. However, MavA sRNA appears to be important for the survival of *B. cenocepacia* J2315 when exposed to heat-shock (50 °C). Our results also demonstrate that MavA sRNA is not involved in the virulence of this strain.
We hypothesize that MavA might be involved in an anti-antisense mechanism, where it negatively regulates a negative regulator of rpsL gene. However, if the observed phenotypes result from the direct or indirect influence of MavA remains unknown.

A few sRNAs have been found to be generated from 3’ regions of known mRNAs, using an internal promoter, or resulting from an internal processing of the initial mRNA (reviewed in [50]). The inhibition of the transcription of ORF3 but not MavA sRNA, caused by the plasposon insertion, together with the presence of a putative promoter for MavA sRNA in the ORF3 region support the hypothesis that these genetic elements are totally independent from each other. Thus, as described for these 3’ UTR sRNAs, MavA is likely benefiting from the ORF3 terminator, since the predicted Rho-independent terminator is located 55 bp downstream mavA encoding sequence.

The possibility that these elements represent a novel toxin-antitoxin system was considered. In the systems already characterized, the RNA acts as an antitoxin and the protein functions as a toxin responsible for membrane changes are described for proteins causing membrane lysis (reviewed in [30]). However, our results suggest that ORF3 protein is implicated in the maintenance of membrane integrity. In the recent years, a new class of proteins with a standard size of 50 aa, named small proteins (sproteins), have been widely identified through the detection of mutations in putative IGRs that interrupt non-annotated ORFs encoding these proteins or through the disclosure that some small and regulatory RNAs can also encode these small proteins (reviewed in [190]). These proteins can be involved in diverse functions, such as the modulation of cell membrane features (thickness or fluidity), the modulation of different pathways, and can also act as chaperones of both metals and nucleic acids (reviewed in [190,191]). Interestingly, some of the sproteins features coincide with the properties of ORF3 protein, since in the stress assays performed we observed that this protein is likely involved in the maintenance of the outer membrane integrity. Moreover, sproteins functioning as chaperones that promote sRNA-mRNA interactions have high values of pI, which means that the neutral protein form is present under more basic conditions (reviewed in [191]). In fact, similarly to these chaperones the zwitterionic form of the hypothetical protein encoded by ORF3 also seems to be dominant in a basic environment (high pH values) since the bioinformatically predicted pI is 10.20. Thus, it is possible that ORF3 protein belongs to the class above described.

Future work will comprise stress experiments involving MavA sRNA silencing and electrophoretic mobility shift assays (EMSA) for the assessment of MavA binding affinity to putative targets, such as rpsL mRNA and eventually to Hfq protein. In fact, the Hfq protein is often associated with sRNAs originated from the 3’ end of known mRNAs. Furthermore, the stability of MavA sRNA in the B. cenocepacia SJ2 strain should be evaluated in RNA decay experiments. The functional characterization of ORF3 protein will be also be continued, in order to understand its biological role in Bcc bacteria.
REFERENCES


57. Burkholder WH. Sour skin, a bacterial rot of onion bulbs. Phytopathology. 1950;40.
58. Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, et al. Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981)


73. Coene T, Mahenthiralingam E, Henry D, LiPuma JJ, Laevens S, Gillis M, et al. *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol


106. Conway B-AD, Chu KK, Bylund J, Altman E, Speert DP. Production of exopolysaccharide by


126. Engledow AS, Medrano EG, Lipuma JJ, Gonzalez CF, Mahenthiralingam E. Involvement of a plasmid-encoded type IV secretion system in the plant tissue watersoaking phenotype of *Burkholderia cenocepacia* involvement of a plasmid-encoded type IV secretion system in the plant tissue watersoaking phenotype of *Burkholderia cenocepacia*. 2004;186:6015–6024.


158. Smith SGJ, Mahon V, Lambert MA, Fagan RP. A molecular Swiss army knife: OmpA structure,


Predicted mRNA targets of MavA in *B. cenocepacia* J2315 genome with RNAPredator tool [149].

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>mRNA information</th>
<th>Energy (kJ/mol)</th>
<th>z-Score</th>
<th>Locus Tag</th>
<th>mRNA information</th>
<th>Energy (kJ/mol)</th>
<th>z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCAL0023</td>
<td>putative branched-chain amino acid ABC transporter ATP-binding protein</td>
<td>-22.67</td>
<td>-5.03</td>
<td>BCAL1537</td>
<td>hypothetical protein</td>
<td>-15.80</td>
<td>-2.42</td>
</tr>
<tr>
<td>BCAL1528</td>
<td>flp type pilus assembly protein</td>
<td>-17.72</td>
<td>-3.15</td>
<td>BCAL0176</td>
<td>hypothetical protein</td>
<td>-15.73</td>
<td>-2.39</td>
</tr>
<tr>
<td>BCAM0699</td>
<td>putative exported dehydrogenase</td>
<td>-17.62</td>
<td>-3.11</td>
<td>BCAM1878A</td>
<td>hypothetical protein</td>
<td>-15.69</td>
<td>-2.36</td>
</tr>
<tr>
<td>BCAL0015</td>
<td>putative branched-chain amino acid ABC transporter ATP-binding protein</td>
<td>-17.15</td>
<td>-2.93</td>
<td>BCAL0788</td>
<td>hypothetical protein</td>
<td>-15.59</td>
<td>-2.34</td>
</tr>
<tr>
<td>BCAL0636 (yagS gene)</td>
<td>putative xanthine dehydrogenase FAD binding subunit</td>
<td>-16.28</td>
<td>-2.6</td>
<td>BCAL1526</td>
<td>putative flp type pilus assembly protein</td>
<td>-15.03</td>
<td>-2.13</td>
</tr>
<tr>
<td>BCAL1644</td>
<td>permease protein</td>
<td>-16.25</td>
<td>-2.59</td>
<td>BCAM0935</td>
<td>TetR family regulatory protein</td>
<td>-15.01</td>
<td>-2.10</td>
</tr>
<tr>
<td>BCAM2441</td>
<td>putative solute-binding protein precursor</td>
<td>-16.15</td>
<td>-2.54</td>
<td>BCAM0382</td>
<td>putative lipoprotein</td>
<td>-14.94</td>
<td>-2.07</td>
</tr>
<tr>
<td>BCAM1818</td>
<td>hypothetical protein</td>
<td>-16.08</td>
<td>-2.51</td>
<td>BCAS0629</td>
<td>putative lipoprotein</td>
<td>-14.89</td>
<td>-2.15</td>
</tr>
<tr>
<td>BCAM2481</td>
<td>LysR family regulatory protein</td>
<td>-16.06</td>
<td>-2.51</td>
<td>BCAM1168</td>
<td>hypothetical protein</td>
<td>-14.88</td>
<td>-2.05</td>
</tr>
<tr>
<td>BCAL0418</td>
<td>type I restriction enzyme specificity protein</td>
<td>-15.89</td>
<td>-2.45</td>
<td>BCAL3099</td>
<td>putative branched-chain amino acid transporter permease</td>
<td>-14.86</td>
<td>-2.06</td>
</tr>
<tr>
<td>BCAS0071</td>
<td>hypothetical protein</td>
<td>-15.87</td>
<td>-2.52</td>
<td>BCAM0286</td>
<td>putative alcohol dehydrogenase</td>
<td>-14.77</td>
<td>-2.00</td>
</tr>
</tbody>
</table>