FEDERAL UNIVERSITY OF MINAS GERAIS
BIOLOGICAL SCIENCES INSTITUTE
INTERUNIT PROGRAM OF POST-GRADUATION IN BIOINFORMATICS

PhD Thesis

Exploring *in silico* transcriptional and post-transcriptional Gene Regulatory Networks in the *Corynebacterium* genus

PhD student: Mariana Teixeira Dornelles Parise
SUPERVISORS: Prof. Dr. Vasco Ariston de Carvalho Azevedo
Prof. Dr. Jan Baumbach
CO-SUPERVISOR: Dr. Rodrigo Bentes Kato

BELO HORIZONTE
MAY - 2021

Mariana Teixeira Dornelles Parise

Exploring *in silico* transcriptional and post-transcriptional Gene Regulatory Networks in the *Corynebacterium* genus

Thesis presented as partial requirement for the degree of Doctor of Philosophy in Bioinformatics, to the Post-graduation Program of Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais

SUPERVISORS: Prof. Dr. Vasco Ariston de Carvalho Azevedo
Prof. Dr. Jan Baumbach
CO-SUPERVISOR: Dr. Rodrigo Bentes Kato

BELO HORIZONTE

MAY - 2021

043 Parise, Mariana Teixeira Dornelles.
Exploring in silico transcriptional and post-transcriptional gene regulatory networks in the Corynebacterium genus [manuscrito] / Mariana Teixeira Dornelles Parise. – 2021.
118 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Vasco Ariston de Carvalho Azevedo. Coorientadores: Prof. Dr. Jan Baumbach; Dr. Rodrigo Bentes Kato.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Redes Reguladoras de Genes. 2. Transcrição Genética. 3. Fatores de Transcrição. 4. Corynebacterium. 5. Expressão Gênica. I. Azevedo, Vasco Ariston de Carvalho. II. Baumbach, Jan. III. Kato, Rodrigo Bentes. IV. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. V. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DA DEFESA DE TESE	**132/2021 entrada 2º/2016**
Mariana Teixeira Dornelles Parise	**CPF: 026.576.900-09**

Às nove horas do dia **27 de maio de 2021**, reuniu-se, no aplicativo zoom, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Exploring in silico transcriptional and post-transcriptional Gene Regulatory Networks in the Corynebacterium genus"**, requisito para obtenção do grau de Doutora em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Vasco Ariston de Carvalho Azevedo**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professor /Pesquisador	Instituição	Indicação
Dr. Vasco Ariston de Carvalho Azevedo	Universidade Federal de Minas Gerais	**Aprovada**
Dr. Rodrigo Bentes Kato	Universidade Federal de Minas Gerais	**Aprovada**
Dr. Jan Baumbach	University of Hamburg	**Aprovada**
Dr. Richard Röttger	University of Southern Denmark	**Aprovada**
Dr. Josch Konstantin Pauling	Technical University of Munich	**Aprovada**
Dr. Aristóteles Góes Neto	Universidade Federal de Minas Gerais	**Aprovada**
Dr. Lucas Bleicher	Universidade Federal de Minas Gerais	**Aprovada**

Pelas indicações, a candidata foi considerada: **Aprovada**
O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.
Belo Horizonte, 27 de maio de 2021.

Dr. Vasco Ariston de Carvalho Azevedo – Orientador

Dr. Rodrigo Bentes Kato - Coorientador

Dr. Jan Baumbach – Coorientador

Dr. Richard Röttger

Dr. Josch Konstantin Pauling

Dr. Aristóteles Góes Neto

Dr. Lucas Bleicher

Dr. Siomar de Castro Soares

“The truth may be puzzling. It may take some work to grapple with. It may be counterintuitive. It may contradict deeply held prejudices. It may not be consonant with what we desperately want to be true. But our preferences do not determine what’s true.”

Carl Sagan

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisors Prof. Vasco Azevedo, Prof. Jan Baumbach, and Dr. Rodrigo Kato for their guidance, support, and opportunity. I am also thankful to Prof. Dr. Andreas Tauch, Prof. Dr. Rolf Backfen, Dr. Martin Raden, Dr. Josh Pauling, Dr. Flávia Aburjaile, Dr. Anne Gomide for their collaboration in the research and review articles presented in this thesis. I would like to thank Prof. Dr. Aristóteles Góes-Neto, Dr. Tim Kacprowski, and Dr. Markus List, Dr. Nina Wenke for helpful discussions and overall support.

I would like to thank the funding agencies which made this work possible: CAPES, CNPq, the German Research Foundation, and FAPEMIG for their financial support. I am very thankful for the UFMG and TUM for welcoming my work and providing safe and sound work environments. Moreover, I am grateful for the overall support of the secretaries Sheila, Natália, Fernanda, Martina, Leonie, and Tiago as well as for the technical support provided by Lion and Drazen.

I would like to thank my life and work partner Doglas for his support on all levels throughout this thesis, the very useful brainstorming, encouragement, and collaboration. I am very thankful for all my colleagues in LGCM, Exbio and LBMCF, their support, friendship and partnership, numerous shared meals and conversations: Marisol, Mônica, Manuela, Amit, Philipp, Tobias, Gihanna, Olga L., Olga T., Chit Tong, Lorenzo, Weilong, Philipp, Mhaned, Zakaria, Amir, Michael, Kundai, Thiago, Marcus, Rodrigo D., Alfonso, Francielly, Renata, Rodrigo P., Alessandra, Sandeep, Arun, Jamal, Gabriel Q. and Marcelo T.

I am very thankful to all my family, especially to my parents, Elvis and Daniela, and to my parents-in-law, Imgrid and Moacir. A very big thanks to all my friends, relatives and ancestors, specially to Letícia, Diogo, Carolina B., Sílvia, Elizete, Monetti, Neva, Pedro E., João M. and Cláudia.

Last but not least, I would like to thank Samantha for her invaluable English language support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

σ	Sigma
ChIP-chip	Chromatin Immunoprecipitation and DNA microarray
ChIP-seq	Chromatin Immunoprecipitation and sequencing
GRN	Gene Regulatory Network
GRIL-seq	Global small non-coding RNA target identification by ligation and sequencing
mRNA	Messenger RNA
MAPS	MS2-affinity purification coupled with RNA sequencing
RIL-seq	RNA interaction by ligation and sequencing
sRNA	Small regulatory RNA
TF	Transcription Factor
TR	Transcriptional Regulators

RESUMO

A regulação transcricional e pós-transcricional são consideradas etapas importantes de controle na expressão gênica em bactérias. Moléculas regulatórias como fatores de transcrição (FT), fatores sigma e pequenos RNAs (sRNAs) são responsáveis pela modulação da expressão gênica, permitindo a sobrevivência e adaptação bacteriana em ambientes extremos. Tanto estas moléculas quanto seus papéis na expressão gênica de bactérias modelo, como *Escherichia coli* e *Bacillus subtilis*, têm sido amplamente exploradas. Porém, a nossa compreensão do panorama regulatório do gênero *Corynebacterium* ainda é limitada apesar de sua importância médica, veterinária e biotecnológica. Nesta tese, expandimos o conhecimento transcricional e pós-transcricional em corinebactérias, e a dividimos em dois capítulos com três artigos. Em nosso primeiro artigo científico, nós exploramos a regulação da transcrição por FTs e fatores sigma na sétima versão do CoryneRegNet, o banco de dados de referência para interações regulatórias do gênero *Corynebacterium* desde 2006. Foram preditas interações regulatórias para 224 organismos deste gênero, aumentando em 20 vezes a quantidade de organismos deste gênero com Redes Regulatórias Gênicas (RRGs) e atribuindo valores de significância estatística à estas interações regulatórias. Posteriormente, nós apresentamos um artigo de revisão, no qual nós coletamos, sintetizamos e organizamos conhecimento em relação aos sRNAs e seus mecanismos de ação no gênero *Corynebacterium.* Neste artigo, apresentamos sRNAs envolvidos em captação de glicose, controle do número de cópias do plasmídeo e produção de glutamato. Em nosso terceiro artigo científico, nós expandimos o conhecimento da regulação pós-transcricional em corinebactérias através do enriquecimento de RRGs transcricionais com regulações por sRNAs em seis organismos deste gênero. Além disso, sugerimos que sRNAs, FTs e fatores sigma influenciam em conjunto a expressão gênica no gênero *Corynebacterium*. Em suma, esta tese apresenta o maior banco de dados de RRGs neste gênero, o qual contém regulações por FTs, fatores sigma e sRNAs. Isto representa um passo em direção à compreensão dos mecanismos de regulação do gênero *Corynebacterium*, tanto apresentando prováveis interações regulatórias quanto sugerindo alvos para futuras investigações experimentais.

ABSTRACT

Transcriptional and post-transcriptional regulation of gene expression are considered key regulatory steps in bacteria. Regulatory molecules such as transcription factors (TF), sigma-factors and small RNAs (sRNAs) modulate gene expression, allowing bacteria to survive and adapt to constantly changing and challenging environments. These molecules and their role in gene expression have been widely explored in model bacteria such as *Escherichia coli* and *Bacillus subtilis*. However, our understanding regarding the regulatory landscape of the *Corynebacterium* genus is still very limited despite its biotechnological, medical and veterinary importance. Here, we expand the transcriptional and post-transcriptional knowledge of the *Corynebacterium* genus. This thesis manuscript is divided into two chapters and presents three articles. In our first research article, we explore the regulation by TFs and sigma-factors in the seventh version of CoryneRegNet, the reference database for corynebacterial regulatory interactions since 2006. We predicted transcriptional regulatory interactions for 224 corynebacterial organisms, increasing by twenty times the quantity of corynebacterial organisms with known transcriptional Gene Regulatory Networks (GRNs) and assigning statistical significance values to the predicted regulatory interactions. Subsequently, we present a review article, in which, we collect, summarize and organize the knowledge regarding sRNAs and their mechanisms of action in this genus, highlighting sRNAs involved in glucose uptake, plasmid copy-number control and glutamate production. In our third research article, we expand corynebacterial post-transcriptional regulatory knowledge by enriching the GRN of six corynebacterial organisms with sRNA-driven regulatory interactions and suggest sRNAs, TF and sigma-factors jointly orchestrate gene expression regulation in this genus. In brief, this thesis presents the largest GRN database for the *Corynebacterium* genus, unraveling TF-, sigma- and sRNA-driven regulatory interactions. It represents a step towards the understanding of the regulatory mechanisms of the *Corynebacterium* genus by both presenting potential regulatory interactions and suggesting targets for further experimental investigation.

I. THESIS STRUCTURE

This thesis is organized into seven sections. Section I presents the structure of the thesis. In Section II, Background and Motivation, we introduce transcriptional and post-transcriptional regulation in bacteria as well as gene regulatory networks. We highlight the importance and influence of small RNA (sRNA) regulation in these networks and present online resources holding both transcriptional and sRNA-driven Gene Regulatory Networks (GRNs). We highlight that bacterial GRN databases are mostly focused on transcriptional regulation. Subsequently, we describe CoryneRegNet's history and present the *Corynebacterium* genus along with the motivation of this thesis.

Section III presents the objectives of this work. Section IV is divided into two chapters: "Corynebacterial transcriptional GRNs" and "sRNA-enriched corynebacterial GRNs". The first presents the seventh version of CoryneRegNet, a genome-scale transcriptional GRN transfer from four model organisms to 224 corynebacterial organisms. The second chapter presents a review of our current knowledge concerning sRNAs in the *Corynebacterium* genus and the latest version of CoryneRegNet, enriching the regulatory networks of six corynebacterial organisms with sRNA-driven regulation.

In section V we discuss how the results of this thesis impact the transcriptional and post-transcriptional landscape of the *Corynebacterium* genus. We also explore the limitations of our approach and explain how we managed them. In Section VI we conclude that the results of this thesis successfully expanded the regulatory knowledge of the *Corynebacterium* genus and contributed by suggesting potential targets for future experimental assays. Finally, this section also presents the perspectives of this thesis and the appendices introduce other published, submitted and forthcoming research and review articles.

II. BACKGROUND & MOTIVATION

A. Gene expression regulation in bacteria

Gene expression in bacteria is a highly regulated process that allows bacteria to recognize and quickly adapt to changes in the environment and, thus, survive (Chandrangsu, Rensing and Helmann, 2017; Roncarati and Scarlato, 2017; Bervoets and Charlier, 2019). Transcriptional and post-transcriptional regulation are known to play central roles in the control of gene expression. Additionally, they provide aid in avoiding the accumulation of pathway intermediates and maintaining energy and resource levels (Roncarati and Scarlato, 2017). Promoter recognition and transcription initiation by RNA polymerase are considered the main regulatory steps that modulate gene expression in most bacteria. Regulatory proteins such as Transcription Factors (TF) and sigma factors (σ-factors) are considered key players in transcriptional control (Browning and Busby, 2004, 2016; Bervoets and Charlier, 2019). Recent studies have shown post-transcriptional regulation also plays a major role in the regulation of gene expression (de Sousa Abreu *et al.*, 2009; Picard *et al.*, 2012). A variety of RNA regulators such as riboswitches, RNA thermometers and small regulatory RNAs (sRNAs) have been shown to greatly influence gene expression (Shimoni *et al.*, 2007; Storz, Vogel and Wassarman, 2011; Nitzan, Rehani and Margalit, 2017; Hör, Gorski and Vogel, 2018). Furthermore, an *in vivo* sRNA-target identification suggested that at least half of the messenger RNAs (mRNAs) are targeted by sRNAs in *Escherichia coli* and *Salmonella enterica* (Waters *et al.*, 2017); these molecules are considered key players in post-transcriptional control (Storz, Vogel and Wassarman, 2011; Nitzan, Rehani and Margalit, 2017).

During the transcription process, σ-factors are crucial for promoter site recognition in the transcription unit, directly binding to the DNA strand. These molecules generally regulate gene expression through σ-factor competition, in which alternative σ-factors are used to express specific genes responding to certain stress conditions (Krebs *et al.*, 2013; Bervoets and Charlier, 2019). For an extensive explanation concerning σ-driven regulation, see (Feklístov *et al.*, 2014; Browning and Busby, 2016; Davis *et al.*, 2017). TFs regulate gene expression by interacting with specific binding sites called Transcription Factor Binding Sites (TFBSs), allowing them to activate and repress gene expression influencing one or several genes. TFs stimulating gene expression are called activators and positively regulate gene expression whereas the ones repressing it are called repressors and negatively regulate gene expression (Krebs *et al.*, 2013; Browning and Busby, 2016; Bervoets and Charlier, 2019). The induction or repression of a certain gene is modulated by the concentration of the product encoded by the gene, a substrate or even a chemical derivative of either (Krebs, Goldstein and Kilpatrick, 2017), as presented in Figure 1. For more details on TF-driven regulation see (Balleza *et al.*, 2009; Browning and Busby, 2016).

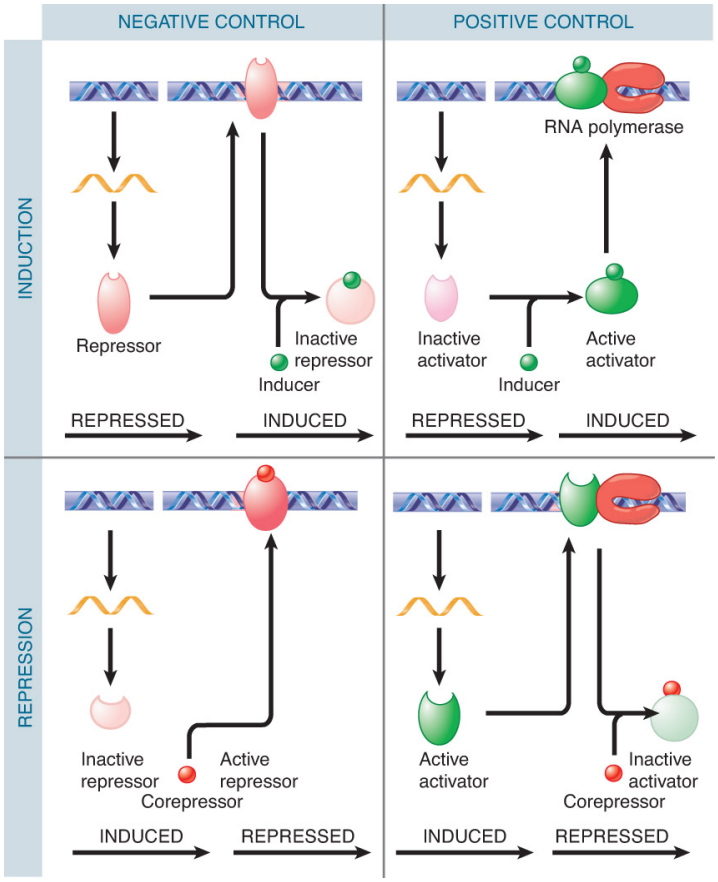


Figure 1. Regulation of gene expression by TFs, sourced from (Krebs, Goldstein and Kilpatrick, 2017). Repressor-mediated regulation being: (A) induced by an inducer molecule and (B) repressed by a co-repressor molecule. Activator-mediated regulation being: (C) induced by an inducer molecule and (D) repressed by a co-repressor molecule.

After the transcription process, mRNAs' stability and degradation are often regulated by sRNAs and RNA-binding proteins (Duval *et al.*, 2015; Bervoets and Charlier, 2019). Regulation mediated by sRNAs mostly occurs through *cis* or *trans* base-pairing with mRNAs. *Cis*-regulation occurs when *cis* anti-sense sRNAs fully bind to their target mRNAs. Likewise, *trans*-regulation occurs when *trans*-encoded sRNAs poorly bind to their target mRNAs (Waters and Storz, 2009; Storz, Vogel and Wassarman, 2011; Nitzan, Rehani and Margalit, 2017). Such molecules can bind to sRNAs to activate or repress the expression of their target genes, as presented in Figure 2. When repressing a target, the sRNA can occlude its ribosomal binding site or recruit an RNAse, inducing RNA cleavage and degradation. When activating a target, the sRNA can expose an occluded ribosomal binding site or stabilize the target by preventing its cleavage and degradation (Nitzan, Rehani and Margalit, 2017). The sRNA-mRNA base-pairing process is often mediated by RNA-binding proteins, such as Hqf and ProQ; however, such molecules can act independently, impacting on: (i) the susceptibility of the target mRNA and sRNA to degradation, (ii) ribosome binding site accessibility and (iii) transcription termination/antitermination structures (Duval *et al.*, 2015; Bervoets and Charlier, 2019). The regulation of the sRNAs' gene expression mostly happens at the transcriptional level and is

mediated by TFs and σ-factors (Brosse and Guillier, 2018). Other than the mechanisms presented here, other factors contribute to the regulation of gene expression; for an extensive description of bacterial regulation of gene expression, please see (Bervoets and Charlier, 2019).

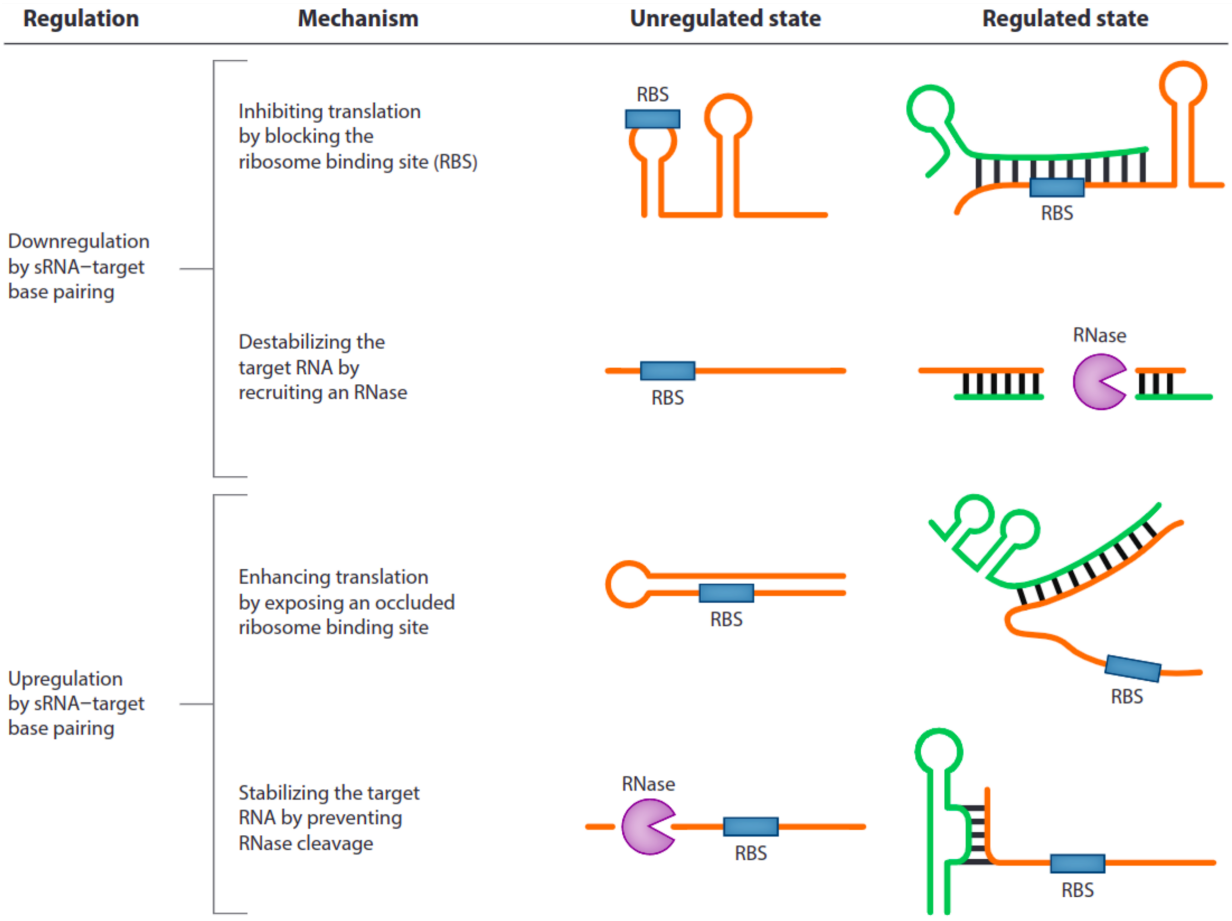


Figure 2. Overview of sRNA regulatory interactions, sourced from (Nitzan, Rehani and Margalit, 2017).

B. Gene Regulatory Networks

The set of regulatory interactions of an organism can be generally comprehended in terms of regulatory elements (TFs, σ-factors and sRNAs) that regulate specific genes or operons. This general view is used to create computational models called Gene Regulatory Networks (GRN) (Babu *et al.*, 2004; Baumbach, Tauch and Rahmann, 2009; Baumbach, 2010; Kılıç and Erill, 2016). These networks are usually represented as directed graphs in which the regulatory elements and their target genes (TGs) are the nodes. The criterion to create an edge is the regulatory interaction between the regulatory element and the TG (Babu, Lang and Aravind, 2009; Thompson, Regev and Roy, 2015). As presented in the previous section, the key regulatory elements in transcriptional and post-transcriptional regulation are TFs, σ-factors and sRNAs (Storz, Vogel and Wassarman, 2011; Nitzan, Rehani and Margalit, 2017; Bervoets and

Charlier, 2019). Such regulators jointly coordinate bacterial gene expression by, for instance, targeting the same genes, forming regulatory cascades and forming regulatory circuits. Figure 3 shows a sample GRN presenting the following regulatory circuits: (i) feed feedback: TF_1 regulates $sRNA_1$ and $sRNA_1$ regulates TF_1; (ii) feedforward loop: sRNA1 regulates σ1, consequently, directly and indirectly regulating TG3; (iii) single input module: sRNA2 regulates TF2, which, in turn, regulates TG_4, TG_5, TG_6 and TG_7; and (iv) dense overlapping regulon: TG_4, TG_5, TG_6 and TG_7 are densely regulated by TF_2, $sRNA_3$ and TF_3. For an extensive description of how sRNA, TFs and σ-factors interact in GRNs see (Beisel and Storz, 2010; Nitzan, Rehani and Margalit, 2017; Brosse and Guillier, 2018).

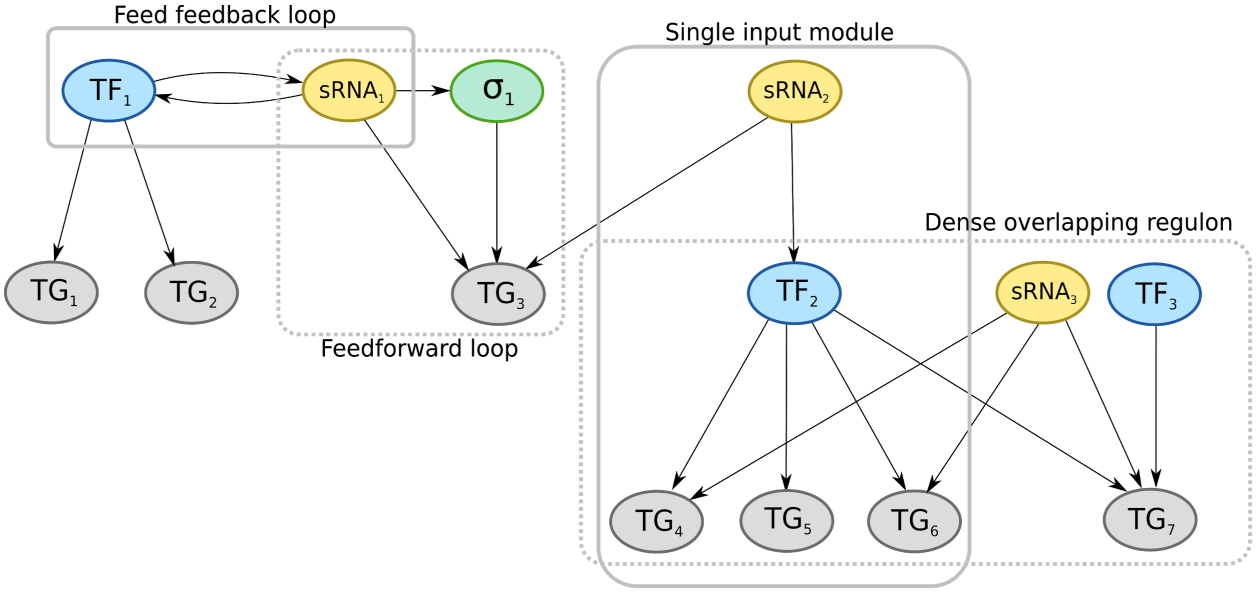


Figure 3. Sample GRN. The nodes represent regulatory elements and the colors represent their regulatory role: sRNAs in yellow, TFs in blue, σ factors in green and TGs in gray. The arrows indicate the regulatory interactions.

Transcriptional and post-transcriptional GRNs have been experimentally reconstructed by performing *in vivo* and *in vitro* experiments for the identification of (i) expressed regulatory elements and target genes and (ii) regulatory interactions. Table 1 summarizes the experimental techniques generally used to reconstruct GRNs driven by both sRNAs and transcriptional regulators (TRs, referring to TFs and σ-factors). These methods have been extensively applied to reconstruct GRNs for model organisms, such as *Escherichia coli* (Salgado *et al.*, 2006; Santos-Zavaleta *et al.*, 2019), *Bacillus subtilis* (Arrieta-Ortiz *et al.*, 2015, 2020) and *Salmonella enterica* (Smith *et al.*, 2016; Métris *et al.*, 2017). However, these methods are considered labor- and cost-intensive and, thus, infeasible to be applied to every bacterium (Baumbach, Rahmann and Tauch, 2009; Leyn *et al.*, 2016). For comprehensive descriptions of the experimental reconstruction of GRNs, see (Balleza *et al.*, 2009; Thompson, Regev and Roy, 2015; Mercatelli *et al.*, 2020) for TR-driven networks and (Altuvia, 2007; Ahmed, Hafeez and Mahmood, 2018; Diallo and Provost, 2020) for sRNA-driven networks.

Table 1. Experimental assays employed in the experimental reconstruction of GRNs. In the table, Transcriptional Regulators (TR) refer to TFs or σ-factors.

	TR-TG interaction	sRNA-mRNA
Expression	RNA-seq, microarray	Co-immunoprecipitation of RNA binding proteins, total RNA-labeling, deep-sequencing
Interaction	EMSA, ChIP-chip, ChIP-seq	Pulse-expression, MAPS, RIL-seq, GRIL-seq

The computational reconstruction of these networks reveals promising genes and regulatory elements for further experimental testing without performing exhaustive wet-lab assays (Babu, Lang and Aravind, 2009; Baumbach, Rahmann and Tauch, 2009; Wright *et al.*, 2014). Like experimental reconstruction, computational reconstruction of GRNs can also be divided into the prediction of (i) regulatory elements and (ii) regulatory interactions. For TR-driven regulations, known regulatory interactions from taxonomically related organisms are used as models for transferring the GRNs; a regulatory interaction is considered conserved when the TR, the TG and the binding site are conserved (Baumbach, Rahmann and Tauch, 2009; Pauling, Röttger, Tauch, *et al.*, 2012). For TR-driven regulations, the sRNAs are first predicted, either *de novo* or based on already known sRNAs (homology) (Zhang et al., 2017b; Backofen et al., 2018). The prediction of the regulatory interaction can be done by analyzing local interactions or full-hybrids of the potential sRNA-mRNA interaction (Pain et al., 2015). For further explanations and assessments of the computational reconstruction of TR-driven networks see (Babu, Lang and Aravind, 2009; Baumbach, Rahmann and Tauch, 2009; Kılıç and Erill, 2016). For the reconstruction of sRNA-driven networks, see (Pain *et al.*, 2015; Zhang *et al.*, 2017; Backofen *et al.*, 2018).

Both experimentally and computationally reconstructed GRNs can be found in online databases for the most-studied bacterial organisms: RegulonDB for *E. coli* (Santos-Zavaleta *et al.*, 2019), MTB Portal for *Mycobacterium tuberculosis* (Turkarslan et al., 2015), Subtiwiki for *Bacillus subtilis* (Zhu and Stülke, 2018), CoryneRegNet for the *Corynebacterium* genus (Parise *et al.*, 2020) and Abasy Atlas for several bacteria (Ibarra-Arellano *et al.*, 2016). An extensive list of online databases is presented in Table 2; note that most of the GRN databases focus on TRs despite the importance of sRNAs in regulatory networks. Recently, RegulonDB (Santos-Zavaleta *et al.*, 2019) has integrated sRNA regulatory interactions into the *E. coli* GRNs. Even more recently, Escorcia-Rodríguez and collaborators integrated sRNAs and protein-protein interaction data into *Corynebacterium glutamicum*'s GRN (Escorcia-Rodríguez, Tauch and Freyre-González, 2021). Generally, sRNA regulatory interactions are available on dedicated databases, such as sRNAMap (Huang *et al.*, 2009) for Gram-negative bacteria. BSRD (Li *et al.*, 2013), sRNATarBase 3 (Wang *et al.*, 2016) and RNAInter (Wang *et al.*, 2016; Lin *et al.*, 2020) present sRNA regulatory interactions for several bacterial species. Likewise, databases such as Rfam (Kalvari et al., 2018), RNA central (The RNAcentral Consortium, 2019) and sRNAdb (Pischimarov *et al.*, 2012) present sRNA collections for several bacterial species. Even though several model bacterial species have known GRNs, only *E. coli* and *C. glutamicum* have GRNs enriched with sRNA regulation data. This shows databases integrating sRNA-based and transcriptional regulatory networks are largely missing.

Table 2. Online resources for bacterial GRNs.

Database	Regulators	Organism	Evidence type	References
RegulonDB	TF, sigma and sRNA	*Escherichia coli*	Predicted and experimental	(Santos-Zavaleta *et al.*, 2019)
EHECRegNet	TF, sigma	Human pathogenic *Escherichia coli*	Predicted and experimental	(Pauling, Röttger, Neuner, *et al.*, 2012)
Abasy Atlas	TF, sigma, sRNA	Several bacteria	Predicted and experimental	(Ibarra-Arellano *et al.*, 2016)
SubtiWiki	TF, sigma	*Bacillus subtilis*	Experimental	(Zhu and Stülke, 2018)
PRODORIC	TF, sigma	Several bacteria	Predicted and experimental	(Eckweiler *et al.*, 2018)
CoryneRegNet	TF, sigma, sRNA	*Corynebacterium* genus	Predicted and experimental	(Parise *et al.*, 2020)
CMRegNet	TF, sigma	*Corynebacterium* and *Mycobacterium* genera	Predicted and experimental	(Abreu *et al.*, 2015)
MTB Network Portal	TF, sigma	*Mycobacterium tuberculosis*	Experimental	(Turkarslan *et al.*, 2015)
DBTBS	TF, sigma	*Bacillus subtilis*	Predicted and experimental	(Sierro *et al.*, 2008)
SalmoNet	TF, sigma	*Salmonella enterica*	Predicted and experimental	(Métris *et al.*, 2017)

C. GRNs in the *Corynebacterium* genus

The *Corynebacterium* genus currently has 125 validly published pathogenic and non-pathogenic species (*LPSN-DSMZ*, 2021). These are Gram-positive bacteria that present high GC content, 1.8~15.8Mb of genome size and diverge in shape, oxygen requirement and preferred environment depending on the species (Oliveira *et al.*, 2017). Amongst non-pathogenic corynebacteria, we can highlight the amino-acid producers *C. glutamicum* and *Corynebacterium efficiens*. The first is the corynebacterial model and is extensively used in L-glutamate and L-lysine production (Ikeda and Katsumata, 1992; Georgi, Rittmann and Wendisch, 2005). Pathogenic corynebacteria affects several hosts, such as cattle, dogs, sheep, horses, monkeys and humans (von Graevenitz and Bernard, 2006). *Corynebacterium diphtheriae* is the most prominent pathogen in this genus, causing diphtheria in humans (Murphy, 1996; Zasada and Mosiej, 2018). The species *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis* can cause diphtheria-like diseases in humans and also affect other hosts (Wagner *et al.*, 2011; Hacker *et al.*, 2016). The latter affects several hosts causing different diseases: ulcerative lymphangitis (UL) in horses, ulcerative granulomatous lesions and mastitis in cattle, oedematous skin disease (OSD) in buffaloes and caseous lymphadenitis (CA) in sheep and goats (Selim, 2001; Yeruham *et al.*, 2004; Dorella *et al.*, 2006). *Corynebacterium jeikeium* is a multidrug-resistant non-diphtherial species of this genus that can act as a causative agent of hospital infections (Bookani *et al.*, 2018; Shleeva, Savitsky and Kaprelyants, 2020). For an extensive review on this genus, see (Oliveira *et al.*, 2017).

In the corynebacterial context, CoryneRegNet has stood out as the reference database for this genus since 2006 (Baumbach, Brinkrolf, Czaja, *et al.*, 2006). Its first version was focused on the reconstruction of the *C. glutamicum* transcriptional network (Baumbach, Brinkrolf, Wittkop, *et al.*, 2006). The second version presented transcriptional data regarding *C. diphtheriae, C. efficiens* and *C. jeikeium* together with binding site analysis (Baumbach, Brinkrolf, Wittkop, *et al.*, 2006). CoryneRegNet 3.0 integrated *E. coli* K-12 regulatory data and allowed for comparative visualization of the regulons (Baumbach *et al.*, 2007), whereas CoryneRegNet 4.0 presented novel features as gene expression data analysis and homology detection (Baumbach, 2007). In the fifth version, transcriptional GRN transfer was performed using *C. glutamicum* as the model organism for *C. diphtheriae*, *C. efficiens* and *C. jeikeium* (no release manuscript). CoryneRegNet 6.0 presented transcriptional regulatory data for 12 strains distributed throughout the *Corynebacterium* and *Escherichia* genera. Transcriptional GRNs were transferred from *C. glutamicum* to all the fully sequenced and annotated corynebacterial strains at that time (Pauling, Röttger, Tauch, *et al.*, 2012). The seventh version of CoryneRegNet presents GRNs for 225 corynebacterial strains and will be presented in the first chapter of this thesis. The latest version of CoryneRegNet integrates sRNA regulatory interactions into the GRNs of six corynebacterial species of medical, veterinary and biotechnological interest and will be presented in the second chapter of this thesis.

D. Motivation

Transcriptional and post-transcriptional regulation are considered crucial steps in the regulation of gene expression, allowing bacteria to adapt and survive in continuously-challenging environments. This ability to adapt is essential for the survival of pathogens inside and outside the host. Furthermore, the understanding of bacterial adaptation and response mechanisms allows researchers to, for instance, engineer amino acid production pathways. Experimental assays aiming to unravel such mechanisms have been extensively performed for model organisms, such as *E. coli*, *B. subtilis* and *S. enterica*. However, they are considered laborious and financially infeasible to be performed at a genomic scale for every member of a bacterium genus. In this context, the computational reconstruction of GRNs reveals potential regulatory interactions, indicating promising targets for future experimental assays. The *Corynebacterium* genus presents species of medical, veterinary and biotechnological interest, including the pathogens *C. diphtheriae, C. jeikeium, C. ulcerans* and *C. pseudotuberculosis*, as well as those of industrial interest, such as *C. glutamicum* and *C. efficiens*. Despite the importance of this genus, the knowledge of how these bacteria respond to stress, survive and adapt is still limited. This thesis presents an effort to unveil the transcriptional and post-transcriptional regulation of the *Corynebacterium* genus by, first, reconstructing *in silico* the transcriptional GRNs of 224 corynebacterial organisms and, then, enriching these networks with sRNA-driven regulatory interactions for the six most prominent species of pathogenic- and biotechnological-interest corynebacteria.

III. OBJECTIVES

A. Main objective

To expand transcriptional and post-transcriptional regulatory knowledge in the *Corynebacterium* genus.

B. Specific objectives

- To perform the transcriptional GRN transfer from *E. coli* K-12, *C. glutamicum* ATCC 13032, *B. subtilis* 168 and *M. tuberculosis* H37Rv to all corynebacterial genomes with complete sequence available on NCBI;
- To assign statistical significance values to all predicted regulations;
- To re-implement CoryneRegNet's back- and front-end in order to meet the requirements of the updated database and support new features as well as future database growth;
- To collect, organize and summarize state-of-the-art knowledge on sRNAs and their roles in the regulation of gene expression within the *Corynebacterium* genus.
- To predict sRNA-driven GRNs for *C. glutamicum* ATCC 13032, *C. diphtheriae* NCTC 12129, *C. jeikeium* K411, *C. ulcerans* NCTC7910, *C. pseudotuberculosis* 1002B and *C. efficiens* YS-314;
- To integrate the sRNA-driven GRNs into the existing transcriptional regulatory networks on CoryneRegNet;
- To update CoryneRegNet's back- and front-end in order to integrate sRNA regulation into the GRNs.

IV. CHAPTERS

A. CHAPTER I - Corynebacterial transcriptional GRNs

In this chapter, we explore transcriptional GRNs in the *Corynebacterium* genus. We present the seventh version of CoryneRegNet, the reference database for corynebacterial transcriptional regulatory interactions.

1. Research Article

Title: CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks

Authors: **Mariana Teixeira Dornelles Parise**, Doglas Parise, Rodrigo Bentes Kato, Josch Konstantin Pauling, Andreas Tauch, Vasco Ariston de Carvalho Azevedo & Jan Baumbach

Status: Published in the journal Scientific Data (2-year impact factor: 5.541)

One of the main objectives of this thesis is to expand the transcriptional knowledge concerning the *Corynebacterium* genus. This research article describes the seventh version of CoryneRegNet, which holds GRNs for 225 corynebacterial organisms and increases by 20 times the number of corynebacterial strains with transcriptional GRNs. In total, 82,268 regulatory interactions, 57,747 regulated genes and 9,590 transcriptional regulators are available in the seventh version of CoryneRegNet. The information presented here could aid in elucidating transcriptional mechanisms spurring adaptation, survival and infection of this genus, as well as in guiding future experimental assays.

SCIENTIFIC DATA

OPEN

ARTICLE

CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks

Mariana Teixeira Dornelles Parise[1,2,5] ✉, Doglas Parise[1,2,5], Rodrigo Bentes Kato[1], Josch Konstantin Pauling[3], Andreas Tauch[4], Vasco Ariston de Carvalho Azevedo[1] & Jan Baumbach[2]

We present the newest version of CoryneRegNet, the reference database for corynebacterial regulatory interactions, available at www.exbio.wzw.tum.de/coryneregnet/. The exponential growth of next-generation sequencing data in recent years has allowed a better understanding of bacterial molecular mechanisms. Transcriptional regulation is one of the most important mechanisms for bacterial adaptation and survival. These mechanisms may be understood via an organism's network of regulatory interactions. Although the *Corynebacterium* genus is important in medical, veterinary and biotechnological research, little is known concerning the transcriptional regulation of these bacteria. Here, we unravel transcriptional regulatory networks (TRNs) for 224 corynebacterial strains by utilizing genome-scale transfer of TRNs from four model organisms and assigning statistical significance values to all predicted regulations. As a result, the number of corynebacterial strains with TRNs increased twenty times and the back-end and front-end were reimplemented to support new features as well as future database growth. CoryneRegNet 7 is the largest TRN database for the *Corynebacterium* genus and aids in elucidating transcriptional mechanisms enabling adaptation, survival and infection.

Introduction

Next-generation sequencing (NGS) has unraveled the genomic sequence of a multitude of bacterial genomes[1]. Despite the amount of information, these data do not fully explain how organisms orchestrate their survival on a molecular level. To understand the mechanisms that coordinate an organism's adaptation to environmental changes, it is crucial to understand how a cell maintains transcription[2,3]. The main players in the transcriptional regulation of bacterial organisms are transcription factors (TFs). These regulatory proteins recognize transcription factor binding sites (TFBSs) in the upstream region of the respective target genes (TGs), stimulating or repressing their expression[4–6]. Experimental studies such as RNA-Seq[7], microarray[8], ChIP-chip and ChIP-seq[9] have been applied in order to reveal regulatory interactions in a cell. Nevertheless, performing these experiments for all bacterial strains would be labor-intensive and, thus, financially infeasible[4,10]. As a result, these experimental data are not available for every member of a bacterial genus.

To alleviate this lack of data, genome-scale transfer of TRNs has been applied providing insights into the regulatory mechanisms of bacterial organisms[10,11]. In this context, a model organism is an organism with the most complete and experimentally validated TRN which can be utilized to predict regulatory interactions in other organisms, called target organisms, with incomplete or less validated TRNs. TRNs are constructed as directed graphs where the nodes represent TFs and their TGs while the criterion to create an edge from a TF to a TG is the regulatory interaction between them[6,12,13]. Edge labels may then indicate the corresponding TFBSs and/or the type of the regulatory interaction. A reliable method to transfer TRNs from a model organism to taxonomically

[1]Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [2]Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany. [3]LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany. [4]Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany. [5]These authors contributed equally: Mariana Teixeira Dornelles Parise, Doglas Parise. ✉e-mail: mparise@wzw.tum.de

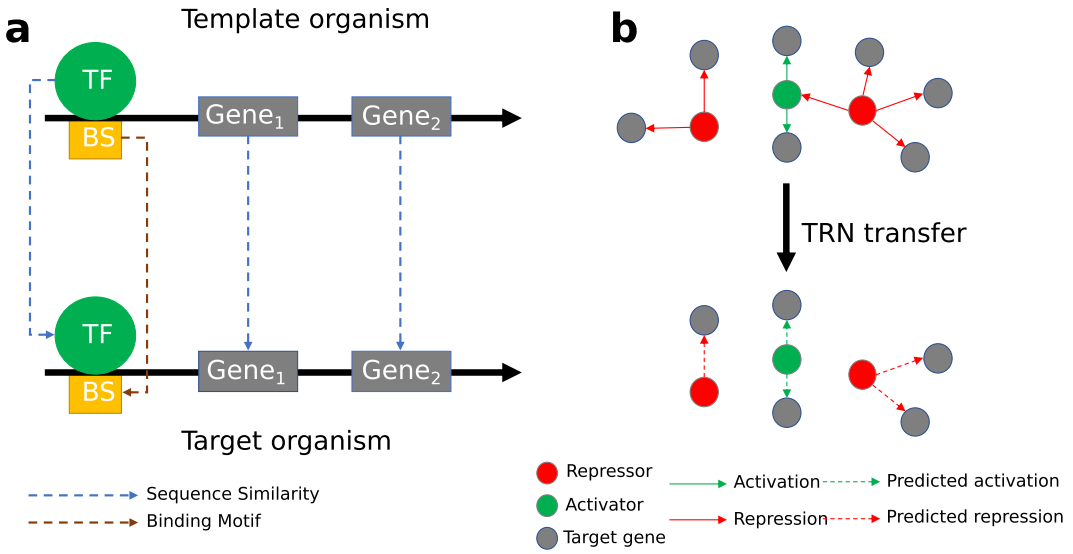


Fig. 1 Overview of the computational reconstruction of TRNs. (**a**) General concept of regulatory conservation. (**b**) TRN transfer scheme: The TRN of a model organism (top) and the predicted TRN of the target organism including all transferred regulations (bottom). In the networks, nodes represent the genes and arrows represent the regulatory interactions.

related target organisms is to consider a regulatory interaction to be conserved between two organisms when the TF, the TG and the TFBS are all conserved[4]. Figure 1a illustrates a conserved regulatory interaction between two organisms. An example of a TRN transferred from a model organism to a target organism is represented in Fig. 1b. More extensive explanations on genome-scale TRN transfer methods can be found in Baumbach *et al.*[11] and Kiliç *et al.*[14].

Both experimentally and computationally reconstructed TRNs are publicly available in databases such as RegulonDB[15] for *Escherichia coli*, EHECRegNet[16] for human pathogenic *Escherichia coli*, TB Portal[17] and MTB Network Portal[18] for *Mycobacterium tuberculosis*, DBTBS[19] and Subtiwiki[20] for *Bacillus subtilis*, and CoryneRegNet[21] for the *Corynebacterium* genus. RegulonDB[15] focuses on detailed and manually curated transcriptional regulation data retrieved from literature for *E. coli*. MTB Network Portal[18] and Subtiwiki[20] provide literature-mined transcriptional regulation data on *M. tuberculosis* and *B. subtilis*. Abasy Atlas[22] is an online collection of regulatory data covering 42 bacteria retrieved from both literature and other online databases. However, there is no resource focusing on corynebacterial gene regulatory networks, and no database that stores predicted TRNs based on evolutionary conservation across a whole collection of model and target organisms. CoryneRegNet has served as the reference database of the genus *Corynebacterium* since 2006[23]. This genus includes organisms with medical, veterinary and biotechnological relevance[24–27]. While the National Center for Biotechnology Information (NCBI) database contains more than 60 corynebacterial species with fully sequenced and annotated genomes, there are TRNs of only eight of these species available in online databases.

The previous version of CoryneRegNet[21] was released in 2012 presenting predicted (transferred) TRNs for eleven corynebacterial strains. The steady increase in corynebacterial genomic sequences contained in public databases allows us to unravel further transcriptional regulatory interactions. In the seventh version of CoryneRegNet, we now present 82,268 regulatory interactions, an increase of more than eleven times compared to the sixth version, as well as 228 TRNs, increasing the number of corynebacterial strains with known TRNs by twenty times. It contains up-to-date regulatory information about the model organisms *C. glutamicum* ATCC 13032, *E. coli* K-12, *M. tuberculosis* H37Rv and *B. subtilis* 168, and predicted TRNs of 224 target organisms of the *Corynebacterium* genus. Furthermore, we present an increase of more than seven times the number of corynebacterial species with TRNs available in public databases representing a great improvement for the bacterial gene regulatory network research community.

Results

In this section we present the results of the re-implemented back- and front-end, the updated database content and the predicted TRNs of all fully sequenced and annotated corynebacterial genomes.

Updated database content. In CoryneRegNet 7 we updated the database content by adding new model and target organisms. As in previous versions, TRNs are categorized as either experimentally validated or computationally predicted. The former contains up-to-date TRNs of *C. glutamicum* ATCC 13032, *E. coli* K-12, *M. tuberculosis* H37Rv and *B. subtilis* 168, the latter contains predicted TRNs of a total of 224 corynebacterial strains. A full list of these strains as well as more details regarding the experimental and predicted databases are presented in Supplementary Table S1. The resulting number of predicted TFs, regulated genes, regulations, binding motifs and profile Hidden Markov Models (HMMs) are presented in Table 1 together with the evolution of the database content throughout previous CoryneRegNet versions.

Version	Organisms	TFs	Reg. genes	Regulations	BMs	PWMs	profile HMMs	Publication
1.0	1	53	331	430	192	23	—	Baumbach *et al.*[23]
2.0	4	64	499	607	274	29	—	Baumbach *et al.*[41]
3.0	5	213	1632	2912	1522	130	—	Baumbach *et al.*[53]
4.0	7	213	1632	2912	1522	130	—	Baumbach *et al.*[54]
5.0e	11	245	1986	3712	1759	144	—	—
5.0p	11	350	2888	4928	2553	249	—	—
6.0e	12	245	1986	3712	1759	144	—	Pauling *et. al.*[21]
6.0p	12	482	3946	6352	3429	381	—	Pauling *et. al.*[21]
7.0e	4*	539	3921	8162	4974	—	446	This
7.0p	228	9590	57747	82268	56870	—	9497	This

Table 1. Content of experimental (e) and predicted (p) databases in this version and previous versions of CoryneRegNet. TFs, transcription factors; Reg. genes, regulated genes; BMs, binding motifs; PWMs position weight matrices; profile HMMs, profile hidden Markov models. Adapted from Pauling *et al.*[21]. *Please note that in previous versions of CoryneRegNet, all organisms were considered experimentally validated if their **genomes** had been experimentally verified through sequencing assays. However, to be more accurate, we changed this definition in this version. We now only consider an organism as experimentally validated if its experimentally verified **TRN** is available. This change of definition led to a decrease in the number of actual model organisms in the experimental view. In previous terms this number would have been 228.

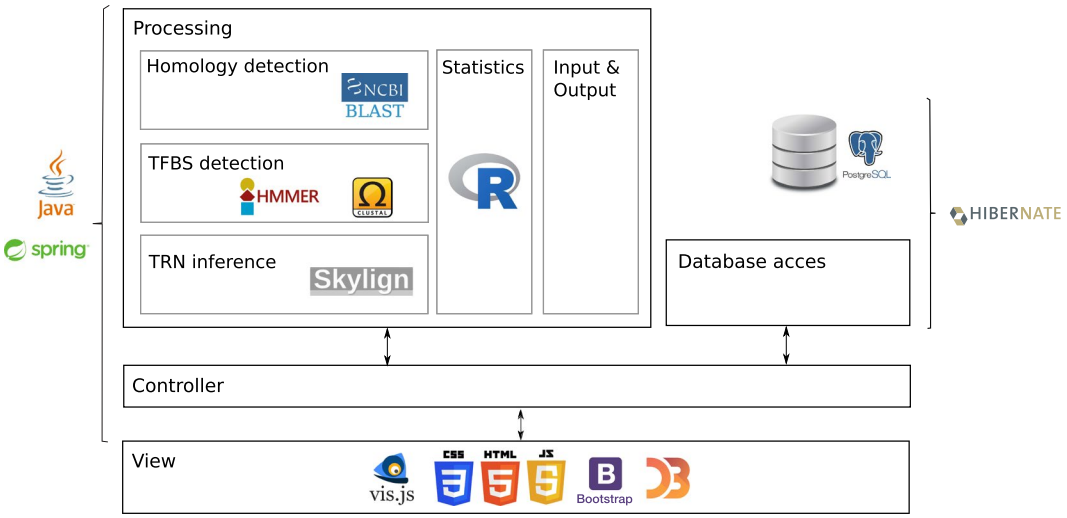


Fig. 2 Schematic of the CoryneRegNet 7 architecture. The processing layer handles the TRN transfer and data parsing. The database access layer manages any query, update, insertion or deletion in the database. The view is responsible for user interaction through the browser. The controller conducts all the communication and data handling among the other three layers.

Novel back- and front-end. In this work, we re-implemented CoryneRegNet's back- and front-end, allowing the user to browse the database via a modern and easy-to-use web-interface. The new architecture (Fig. 2) is inspired by the Model-View-Controller (MVC) architectural pattern[28,29]. This modular structure allows components to be modified or replaced, facilitating maintenance and future updates.

CoryneRegNet web-interface. The CoryneRegNet 7 website contains TRNs of 224 corynebacterial target genomes and 4 model organisms. Information about quantities of regulator types, distribution of TFs, distribution of co-regulating TFs and distribution of HMM profile lengths are shown in the statistics page. Those are shown for each database (predicted and experimental) as well as for each organism. Figure 3 represents the statistics page for the experimental database.

Through the web-interface, the user is able to browse the TRNs in both table and network format. In table format, the list of regulatory interactions (RIs) provides source, target and operon information (Fig. 4a). The network visualization comes with two different layout options: a gene-centered layout and an operon-centered layout, see Fig. 4b,c, respectively. Both visualizations give access to gene information by clicking on the gene and/or operon of interest. Each network may also be downloaded in.sif file format. Additionally, the user can visualize networks of genes or operons of interest by using the network visualization in the gene information pop-up.

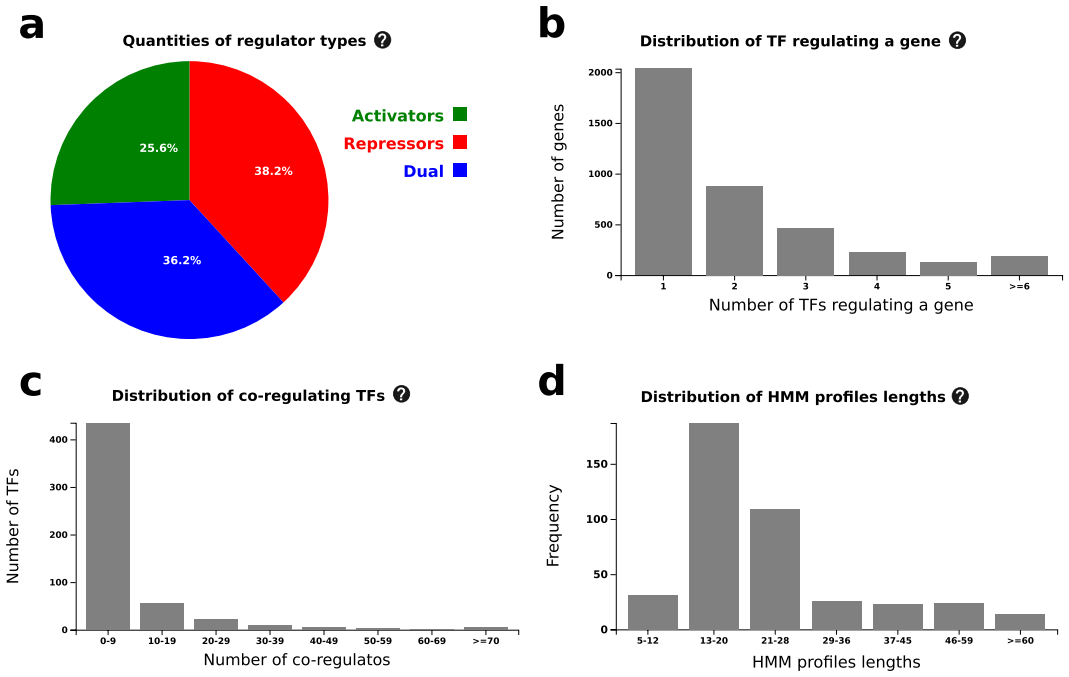


Fig. 3 Overview of main statistics present in CoryneRegNet. In (**a**) the pie chart presents the quantities of regulator types in percentages, (**b**) represents the distribution of the numbers of TFs regulating a gene, (**c**) presents the distribution of co-regulating TFs, and (**d**) demonstrates the distribution of HMM profiles lengths.

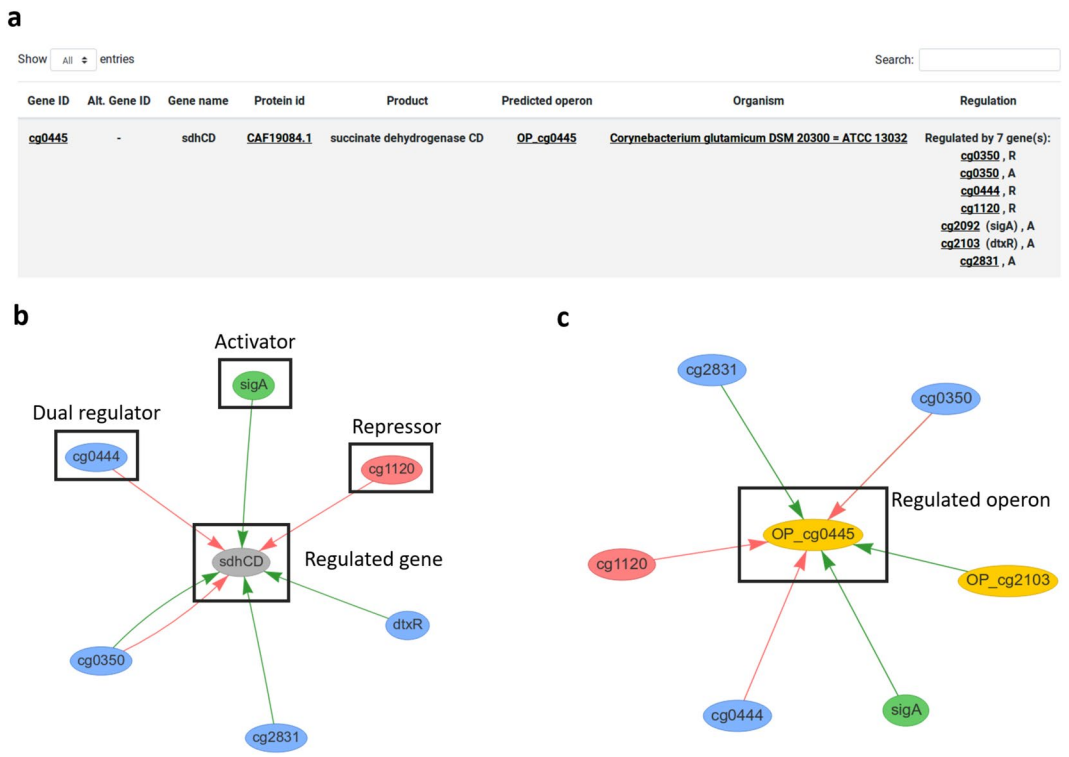


Fig. 4 Example of a RI. (**a**) TG (cg0445) and its regulators in table format. (**b**,**c**) present the same TG as in (**a**), but in a graph format with gene-centered layout (**b**) and an operon-centered layout (**c**).

Furthermore, we offer a detailed gene information page which shows gene identifiers linked to NCBI, nucleotide and protein sequences, homologous genes, and regulatory information. This page also allows the user to make additional motif searches in the database. A view of this page showing the putative homologous genes of

Gene: cg0199 (msmA) ProteinID: CAF18727.1 **Homologous proteins** Regulated by: Gene position Binding site prediction

Candidate homologous of gene cg0199 (msmA):

Show 10 entries

Homologous gene ID	Homologous gene name	Homologous protein id	Organism
ARO_RS01040	-	WP_038582077.1	Corynebacterium glutamicum ATCC 21831
SB89_RS00825	-	WP_011013433.1	Corynebacterium glutamicum B253
cgc1_RS00855	mmsA	WP_011013433.1	Corynebacterium glutamicum C1
AC079_RS00975	-	WP_011013433.1	Corynebacterium glutamicum CP
COI99_RS004555	mmsA	WP_011013433.1	Corynebacterium glutamicum HA

Fig. 5 Detailed gene information page. Here, the "Homologous proteins" tab lists homologs of the cg0199 protein in various other organisms.

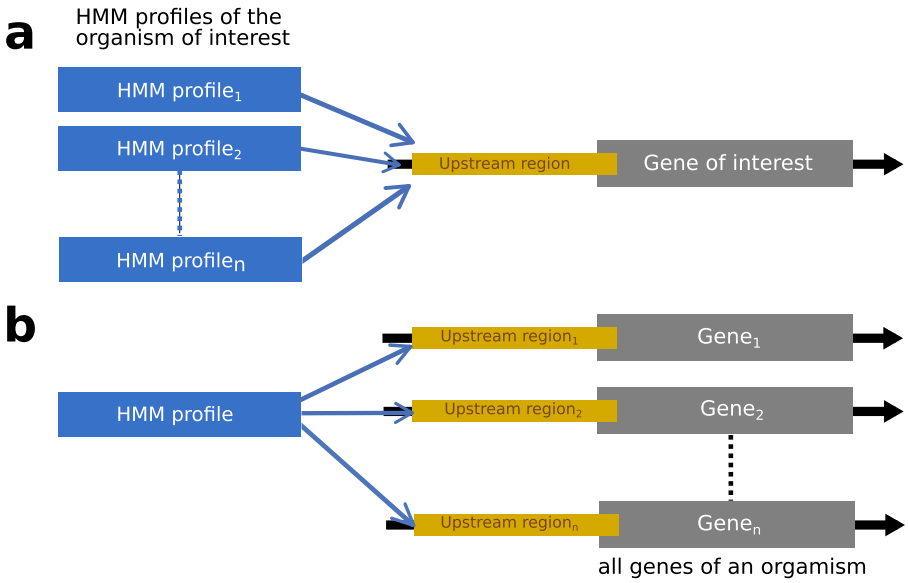


Fig. 6 Kinds of motif search provided in CoryneRegNet 7. (**a**) HMM profiles of one organism being used to identify potential binding sites in the upstream region of a gene of interest. (**b**) HMM profile of interest being used to identify potential binding sites in all genes of an organism.

cg0199 is presented in Fig. 5. The user is provided with information of all genes that are predicted to be homologous to the gene of interest that are present in CoryneRegNet 7.

At this point, CoryneRegNet 7 offers the biggest collection of profile HMMs publicly available for the *Corynebacterium* genus. See the methodology section for an explanation of how those were generated. These profiles and their logos are available for download in the gene information page of genes encoding transcription factors. In addition, the user can utilize the profile HMMs stored in the database to search the upstream region of genes present in CoryneRegNet 7. Two kinds of motif searches are provided: (i) the upstream region of the gene of interest can be searched with HMM profiles of an organism of interest (Fig. 6a) and (ii) the HMM profile of the TF of interest can then be used to identify potential binding sites in the upstream regions of all genes of an organism in the database (Fig. 6b).

Finally, the website provides a comprehensive help page with theoretical and practical explanations of the website content, methodology and navigation, including a broad collection of published literature concerning Corynebacterial transcriptional regulation. This help page can be accessed at www.exbio.wzw.tum.de/coryneregnet/docs&help.htm.

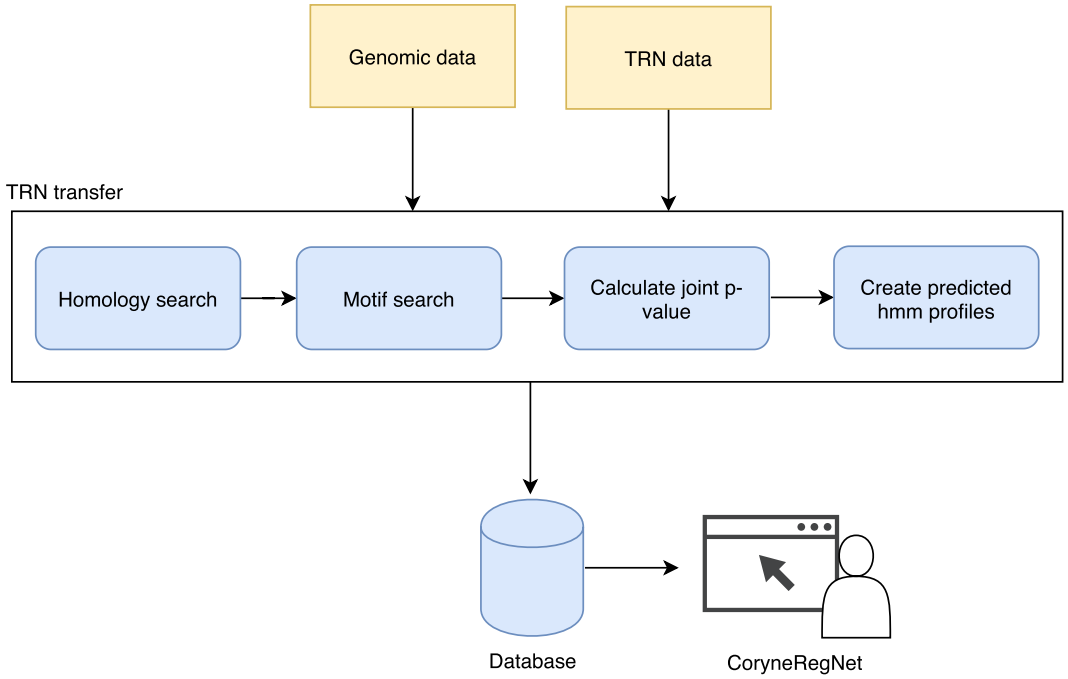


Fig. 7 Schematic overview of the TRN transfer pipeline.

Discussion

In CoryneRegNet version 7, we entirely redesigned the back- and front-end to support the updated database content, new functional features as well as future database growth. As we did in the other versions, we present TRNs for all fully sequenced and annotated corynebacterial genomes available in NCBI (June 2019). Consequently, CoryneRegNet 7 currently offers the biggest knowledge base available regarding TRNs of corynebacterial organisms. Along with our newly-designed web interface, we include an operon network layout and the option to download network views from each organisms network visualization as a file (.sif). This allows users to modify and enrich the network locally and create personalized visualizations based on their own research using third-party software.

We also significantly improved our transfer pipeline in two ways. First, we replaced the use of Position Weight Matrices (PWMs) with profile HMMs in our motif conservation analysis. Profile HMMs enable the modelling of insertions and deletions, greatly improving the detection of remote homologous and model nucleotide dependency as well as length variations in the model's binding sites[30–32]. TFBSs are considered to have low evolutionary conservation between species[11] and this strategy provides more robustness and flexibility when predicting them[33]. It is an advantage considering that mutations in these sites are expected to occur from one species to another. Second, we added the calculation of p-values, i.e. the likelihood of observing this conservation by chance, for each regulatory interaction which provides important information in interpreting the results.

Even though great progress has been made in the TRN field, there are still a few limitations concerning bacterial TRNs. The transfer of TRNs has been hindered by the limited availability of experimentally validated data on bacterial TRNs which are available for only a few model organisms such as *E. coli*, *B. subitilis* and *C. glutamicum*[4,14]. TRN transfer from one organism to another largely depends on known regulatory interactions in the model organism as well as genome similarity between model and target organism[4,11,13,14]. Thus, the more experimental TRN data is available for a greater diversity of bacterial species, the higher the quality of any predicted TRNs. Furthermore, the ability of detecting regulatory interactions acquired by horizontal gene transfer (HGT) becomes relevant[4,11,14] considering regulatory interactions related to life-style may not be identified by using only one model organism. Using several model organisms allows us to identify these regulations. Previous studies point out that (i) known virulence determinants in Enterohemorrhagic *E. coli* are located on mobile genetic elements, which are generally acquired in HGT events[16] and (ii) in *E. coli* neighboring regulators were co-transferred with their TGs in HGT[34]. Methodologies that support the use of more model organisms together with more high quality experimentally validated TRNs will result in more complete TRNs that will consider lifestyle-related regulatory interactions (e.g. pathogenic or non-pathogenic, free-living or host-associated). A first step in this direction was presented in this study, since we transferred regulatory interactions between organisms of different phyla and presented the joint p-value allowing the researcher to evaluate the degree of conservation of each predicted regulatory interaction.

Methods

Database content update. Genomic data of the 228 organisms used in this work were retrieved from the NCBI database[35] in June 2019 (for more details see Supplementary Table S1). TRN data of the model organisms were retrieved from RegulonDB[15] for *E. coli* K-12, Minch *et al.*[36] for *M. tuberculosis* H37Rv, DBTBS[19] for *B. subtilis*. *C. glutamicum* ATCC 13032 data from CoryneRegNet 6[21] were updated with new data from Freyre-González and Tauch (2017)[37].

In order to predict TRNs for the 224 corynebacterial strains, we extended the transfer methodology described by Baumbach and collaborators[4]. First, TF binding profiles were generated for every TF of the model organisms. For this purpose, binding sites of each TF were collected, aligned with Clustal Omega[38] and binding profiles were generated using HMM-build from the HMMER package[39]. Second, we performed all-vs-all protein BLAST[40] search and selected the best bidirectional BLAST hits (BBHs) using a cutoff of 10^{-10} to predict homologous proteins. The upstream regions (−560, +20) of all genes and operons in the analysis were identified. Third, the upstream regions of all homologous TGs in the target organisms were scanned using HMMER[30] to predict conserved TFBSs. Figure 1a illustrates this approach. The HMM profiles of the conserved TFs were applied to the upstream regions of the potentially regulated TGs by using HMMER's default parameters, which corresponds to a p-value of ~10^{-5} [41]. Genes with an intergenic distance of less than 50 base pairs were considered to be part of the same operon and predicted regulatory interactions to the first gene were extended to the operon[4]. The role of a predicted regulatory interaction is inherited from the model regulatory interaction used in the transfer. Finally, profile HMMs were generated for predicted TFs as described above for model TFs. The interaction p-value was obtained by applying Tippet's method[42]. The R package Metap[43] was used to calculate the joint p-value of the p-values obtained in the homology and motif searches. These steps are summarized in Fig. 7.

Implementation of CoryneRegNet 7. The CoryneRegNet7 back-end was developed using Java and the database was implemented using the PostgreSQL[44] management system for relational databases. Hibernate[45] was used for object-relational mapping between PostgreSQL and Java. The Spring framework[46] was used to implement a Model-View-Controller (MVC) architectural pattern[28,29]. The new front-end was developed using HTML5[47], CSS3[48], Bootstrap[49] and JavaScript. We used vis.js[50] to implement the network visualization and graphs in the statistics page were drawn using D3[51]. The website was tested and runs on Google Chrome and Mozilla Firefox. Furthermore, the screen resolution was optimized for tablet (1024 × 768), laptop (1920 × 1080) and desktop (2560 × 1440) resolutions.

Data availability

All data generated in this work is provided to the research community free of charge as comma separated values (.csv format) via the figshare repository[52] and in CoryneRegNet's download section (http://www.exbio.wzw.tum.de/coryneregnet/processToDownalod.htm).

Code availability

CoryneRegNet7 code is available on GitHub: https://github.com/baumbachlab/CoryneRegNet7.

Received: 6 November 2019; Accepted: 15 April 2020;
Published online: 11 May 2020

References

1. Haft, D. H. *et al*. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* **46**, D851–D860 (2018).
2. Röttger, R., Rückert, U., Taubert, J. & Baumbach, J. How little do we actually know? On the size of gene regulatory networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **9**, 1293–1300 (2012).
3. Park, J. & Wang, H. H. Systematic and synthetic approaches to rewire regulatory networks. *Curr Opin Syst Biol* **8**, 90–96 (2018).
4. Baumbach, J., Rahmann, S. & Tauch, A. Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms. *BMC Syst. Biol.* **3**, 8 (2009).
5. Voordeckers, K., Pougach, K. & Verstrepen, K. J. How do regulatory networks evolve and expand throughout evolution? *Curr. Opin. Biotechnol.* **34**, 180–188 (2015).
6. Hao, T. *et al*. The Genome-Scale Integrated Networks in Microorganisms. *Front. Microbiol.* **9**, 296 (2018).
7. Nagalakshmi, U. *et al*. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
8. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
9. Barski, A. *et al*. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
10. Leyn, S. A. *et al*. Comparative genomics and evolution of transcriptional regulons in Proteobacteria. *Microb Genom* **2**, e000061 (2016).
11. Baumbach, J. On the power and limits of evolutionary conservation—unraveling bacterial gene regulatory networks. *Nucleic Acids Res.* **38**, 7877–7884 (2010).
12. Babu, M. M., Lang, B. & Aravind, L. Methods to Reconstruct and Compare Transcriptional Regulatory Networks. *Methods in Molecular Biology* **541**, 163–180 (2009).
13. Thompson, D., Regev, A. & Roy, S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu. Rev. Cell Dev. Biol.* **31**, 399–428 (2015).
14. Kılıç, S. & Erill, I. Assessment of transfer methods for comparative genomics of regulatory networks in bacteria. *BMC Bioinformatics* **17**(Suppl 8), 277 (2016).
15. Santos-Zavaleta, A. *et al*. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).
16. Pauling, J. *et al*. On the trail of EHEC/EAEC–unraveling the gene regulatory networks of human pathogenic Escherichia coli bacteria. *Integr. Biol.* **4**, 728–733 (2012).
17. Galagan, J. E. *et al*. TB database 2010: overview and update. *Tuberculosis* **90**, 225–235 (2010).
18. Turkarslan, S. *et al*. A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis. *Sci Data* **2**, 150010 (2015).
19. Sierro, N., Makita, Y., de Hoon, M. & Nakai, K. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res.* **36**, D93–6 (2008).
20. Zhu, B. & Stülke, J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res.* **46**, D743–D748 (2018).
21. Pauling, J., Röttger, R., Tauch, A., Azevedo, V. & Baumbach, J. CoryneRegNet 6.0–Updated database content, new analysis methods and novel features focusing on community demands. *Nucleic Acids Res.* **40**, D610–4 (2012).
22. Ibarra-Arellano, M. A., Campos-González, A. I., Treviño-Quintanilla, L. G., Tauch, A. & Freyre-González, J. A. Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria. *Database* **2016**, 1–16 (2016).

23. Baumbach, J., Brinkrolf, K., Czaja, L. F., Rahmann, S. & Tauch, A. CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. *BMC Genomics* **7**, 24 (2006).
24. Jamal, S. B., Tiwari, S., Silva, A. & Azevedo, V. Pathogenesis of Corynebacterium diphtheriae and available vaccines; an overview. *Glob. J. Infect. Dis. Clin. Res* **3**, 20–24 (2017).
25. Viana, M. V. C. *et al*. Comparative genomic analysis between Corynebacterium pseudotuberculosis strains isolated from buffalo. *PLoS One* **12**, e0176347 (2017).
26. Parise, D. *et al*. First genome sequencing and comparative analyses of Corynebacterium pseudotuberculosis strains from Mexico. *Stand. Genomic Sci.* **13**, 21 (2018).
27. Baltz, R. H., Demain, A. L. & Davies, J. E. *Manual of Industrial Microbiology and Biotechnology*. (American Society for Microbiology Press, 2010).
28. Pope, K. G. & Krasner, S. A cookbook for using the model-view-controller user interface paradigm in Smalltalk-80. *Journal of Object-Oriented Programming* **1** (1988).
29. Gupta, P. & Govil, M. C. MVC Design Pattern for the multi framework distributed applications using XML, spring and struts framework. *International Journal on Computer Science and Engineering* **2**, 1047–1051 (2010).
30. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
31. Madera, M. & Gough, J. A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic Acids Res.* **30**, 4321–4328 (2002).
32. Delorenzi, M. & Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625 (2002).
33. Riva, A. The MAPPER2 Database: a multi-genome catalog of putative transcription factor binding sites. *Nucleic Acids Res.* **40**, D155–61 (2012).
34. Price, M. N., Dehal, P. S. & Arkin, A. P. Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli. *Genome Biol.* **9**, R4 (2008).
35. Sayers, E. W. *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **47**, D23–D28 (2019).
36. Minch, K. J. *et al*. The DNA-binding network of Mycobacterium tuberculosis. *Nat. Commun.* **6**, 5829 (2015).
37. Freyre-González, J. A. & Tauch, A. Functional architecture and global properties of the Corynebacterium glutamicum regulatory network: Novel insights from a dataset with a high genomic coverage. *J. Biotechnol.* **257**, 199–210 (2017).
38. Sievers, F. *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539 (2011).
39. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**, W29–W37 (2011).
40. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
41. Baumbach, J., Brinkrolf, K., Wittkop, T., Tauch, A. & Rahmann, S. CoryneRegNet 2: An Integrative Bioinformatics Approach for Reconstruction and Comparison of Transcriptional Regulatory Networks in Prokaryotes. *J. Integr. Bioinform.* **3**, 1–13 (2006).
42. Heard, N. A. & Rubin-Delanchy, P. Choosing between methods of combining p-values. *Biometrika* **105**, 239–246 (2018).
43. Dewey, M. *metap: meta-analysis of significance values. R package version 1.1* (2019).
44. PostgreSQL Team. PostgreSQL: The World's Most Advanced Open Source Relational Database, version 9.3, https://www.postgresql.org/ (2019).
45. Hibernate Team. Hibernate Object/Relational Mapping, version 4.3, https://hibernate.org/orm/ (2019).
46. Spring Framework Team. Spring Framework, version 4.1, https://spring.io/ (2019).
47. HTML Team. HTML Standard, version 5, https://html.spec.whatwg.org/multipage/ (2019).
48. CSS Team. CSS Snapshot 2018, version3, https://www.w3.org/TR/css3-roadmap/ (2019).
49. Otto, M., Thornton, J. & Bootstrap contributors. Bootstrap Introduction, version 4.0, https://getbootstrap.com/docs/4.0/getting-started/introduction/ (2019).
50. Vis.js Team. Vis.js, community edition, https://visjs.org/ (2019).
51. Bostock, M., Ogievetsky, V. & Heer, J. D^3: Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.* **17**, 2301–2309 (2011).
52. Tauch, A. *et al*. CoryneRegNet 7 - The reference database and online analysis platform for corynebacterial gene regulatory networks. *figshare* https://doi.org/10.6084/m9.figshare.c.4720991 (2020).
53. Baumbach, J. *et al*. CoryneRegNet 3.0—An interactive systems biology platform for the analysis of gene regulatory networks in corynebacteria and Escherichia coli. *Journal of Biotechnology* **129**, 279–289 (2007).
54. Baumbach, J. CoryneRegNet 4.0–A reference database for corynebacterial gene regulatory networks. *BMC Bioinformatics* **8**, 429 (2007).

Acknowledgements

JB is grateful for support from H2020 grant RepoTrial (nr. 777111) and his VILLUM Young Investigator grant (nr. 13154). MP received support from CNPq (nr. 201336/2018-9), and DP from CAPES (nr. 88887.364607/2019-00) for their work at TUM in Germany. MP's work was also supported by the German Research Foundation (under SFB924). Contributions by J.P. were funded by the Bavarian State Ministry of Science and the Arts in the framework of the Center Digitisation.Bavaria (ZD.B, Zentrum Digitalisierung.Bayern). V.A. is grateful for support from his CNPq Research Productivity grant (nr. 305093/2015-0), CNPq Universal grant (nr. 405233/2016-7) and FAPEMIG grant (nr. APQ 02600-17). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Author contributions

M.P. and D.P. developed the new version of CoryneRegNet and wrote the manuscript. J.B. and V.A. designed the project. J.B., J.P., V.A., A.T. and R.K. reviewed the manuscript. M.P., D.P., J.P. and J.B. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41597-020-0484-9.

Correspondence and requests for materials should be addressed to M.T.D.P.

Reprints and permissions information is available at www.nature.com/reprints.

B. CHAPTER II - sRNA-enriched corynebacterial GRNs

In this chapter, we explore sRNA-driven GRNs in the *Corynebacterium* genus. First, we present a review article regarding sRNAs and their role in the regulation of gene expression within the Corynebacterium genus. In this manuscript, we state that its sRNA knowledge has been poorly explored and highlight the necessity of revealing this genus' regulation by sRNAs. In the second manuscript of this chapter, we explore sRNA regulatory interactions in *C. efficiens*, *C. glutamicum*, *C. ulcerans*, *C. diphtheriae*, *C. pseudotuberculosis* and *C. jeikeium* and integrate their sRNA-driven regulatory interactions into the transcriptional GRNs of CoryneRegNet 7.

1. Review Article

Title: Exploring sRNAs in the Corynebacterium genus

Authors: **Mariana T. D. Parise**, Doglas Parise, Olga Zolotareva, Flavia Alburjaile, Rodrigo Kato, Vasco Azevedo, Jan Baumbach

Status: To be submitted in the journal Microbiological Research (2-year impact factor: 3.970).

One of the main objectives of this thesis is to expand the post-transcriptional regulatory knowledge of the *Corynebacterium* genus. This review article describes both predicted and experimentally verified sRNAs and their roles in this genus. Moreover, we present sRNAs databases and the distribution of sRNA families across the *Corynebacterium* genus. The information presented here contributes by collecting, organizing and summarizing the entire set of sRNAs and their regulatory interactions in this genus.

Title: Exploring small RNAs in the *Corynebacterium* genus

Authors: Mariana T. D. Parise [1,2*], Doglas Parise [1,2*], Olga Zolotareva [1], Flavia Alburjaile [3], Rodrigo Kato [2], Vasco Azevedo [2], Jan Baumbach [1,4]

[1] Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany.
[2] Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
[3] Oswaldo Cruz Foundation (Fiocruz), Brazil.
[4] Computational Biomedicine lab, Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Germany.

* These authors contributed equally

Abstract:

Gene expression regulation allows bacteria to quickly adapt to changes in the environment. Together with transcription factors, bacterial small regulatory RNAs (sRNAs) regulate the expression of genes participating in cell division, stress response, virulence, iron homeostasis and many other crucial processes. However, despite the medical, veterinary and economic importance of the *Corynebacterium* genus, the post-transcriptional layer of regulation in this genus remains poorly investigated. To date, only a few studies of corynebacterial sRNAs and their mechanisms of action have been conducted. Here, we present the first review article on sRNAs in the *Corynebacterium* genus, covering both experimentally verified and predicted sRNAs. This review focuses on sRNAs involved in cell division control, plasmid copy-number control, glucose uptake and glutamate production as well as three sRNA families conserved across several corynebacterial species. We conclude that sRNA knowledge regarding the *Corynebacterium* genus is still hindered by the lack of experimental and predicted investigations; therefore, future studies are necessary to elucidate how these molecules impact corynebacterial gene expression.

Introduction

The *Corynebacterium* genus comprises species of biotechnological, medical, and veterinary interest. *Corynebacterium glutamicum,* the model organism of this genus, has been extensively used in L-glutamate and L-Lysine production (Georgi et al., 2005; Wang et al., 2018). The human pathogen *Corynebacterium diphtheriae* causes an acute, rapidly-developing, and feverish infection named diphtheria and is considered the

most prominent pathogen in this genus (Mandlik et al., 2007; Oliveira et al., 2017). The zoonotic pathogens *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis* cause a diphtheria-like disease in humans and mastitis in cattle. Furthermore, the latter also causes caseous lymphadenitis (CLA) in sheep and goats, oedematous skin disease (OSD) in buffaloes, and ulcerative granulomatous lesions in cattle.

Understanding how corynebacterial regulation is necessary for the development of efficient vaccines, treatments and diagnostic approaches as well as for the optimization of amino acid production pathways. In bacteria, regulation of gene expression consists of two key steps: transcriptional, governed by transcription factors (TFs) and post-transcriptional, mediated by small regulatory RNAs (sRNAs) (Bervoets and Charlier, 2019; Hör et al., 2018). While transcriptional regulation in *Corynebacteria* has been mostly explored in the model organism *C. glutamicum* (Oliveira et al., 2017), the post-transcriptional level remains poorly investigated. Our recent study in *C. glutamicum* demonstrated that regulation by TFs and sigma-factors (σ-factors) do not fully explain its gene expression (Parise et al., 2021). We suggested that more regulatory layers, such as sRNA-driven regulation, should be integrated into the regulatory networks in order to yield more realistic gene regulatory networks (GRNs).

Bacterial sRNAs are known to play critical roles in the regulation of multiple processes such as virulence, biofilm formation, iron homeostasis, and adaptation to stress conditions (Ahmed et al., 2018; Bak et al., 2015; Chakravarty and Massé, 2019; Chareyre and Mandin, 2018). These molecules are 50-500 nucleotides long (Carrier et al., 2018) and generally perform the post-transcriptional regulation of gene expression via base-pairing with messenger RNAs (mRNAs) (Carrier et al., 2018; Hör et al., 2018). *Cis*-acting sRNAs bind perfectly to their targets (Fig 1A), whereas *trans*-acting sRNAs present short, imperfect base-pairings (Fig 1B) (Hör et al., 2018). *Cis*-acting sRNAs that are enconded in the opposite strand of a known coding region are named antisense RNAs (asRNAs) (Storz et al., 2011) and often referred to as counter-transcribed RNAs (ctRNAs) when plasmid-encoded (Kong et al., 2018; Yip et al., 2015). The binding of sRNAs can activate or repress the expression of their target mRNAs (Fig. 1C). Activating sRNAs can up-regulate expression from their target mRNA by (i) preventing the occlusion of the ribosome binding site (RBS) or (ii) protecting mRNA from degradation by RNases (Ahmed et al., 2018; Nitzan et al., 2017). Inhibitory sRNAs repress the expression of their targets via (i) blocking the RBS or (ii) inducing mRNA cleavage by recruiting an RNase. For comprehensive explanations of these mechanisms, see references (Brosse and Guillier, 2018; Carrier et al., 2018; Storz et al., 2011).

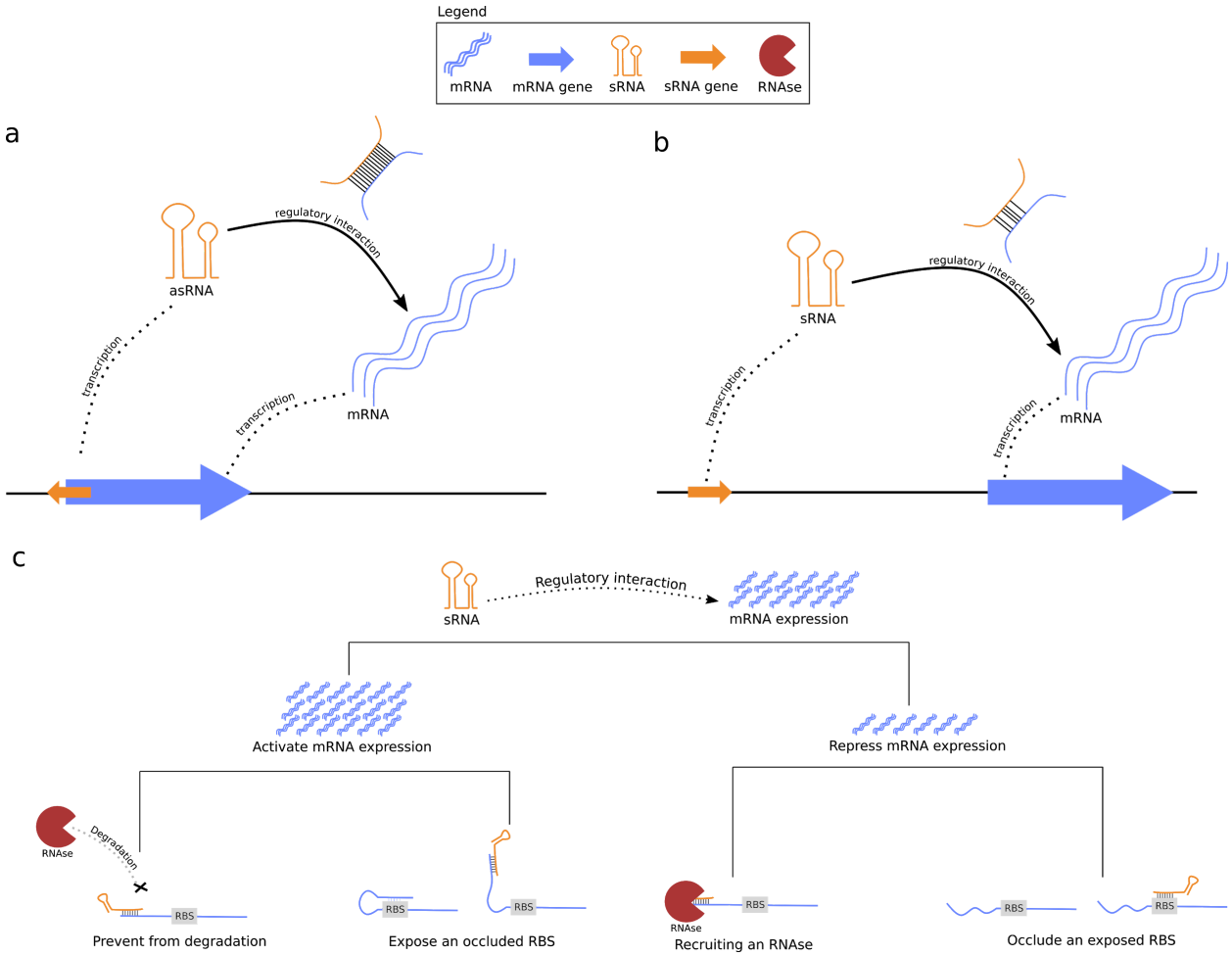


Fig. 1. Cis-acting (A) and trans-acting (B) sRNAs base pairing with mRNA. In (C) the mechanisms for activating and repressing targets are presented.

Despite their important role in the regulation of expression, sRNAs have been extensively studied only in model organisms, such as *Escherichia coli* (Santos-Zavaleta et al., 2019; Thompson et al., 2007), *Bacillus subtilis* (American Society for Microbiology. General Meeting, 2008; Dambach et al., 2013), and *Salmonella enterica* (Hébrard et al., 2012; Hör et al., 2020). In the *Corynebacterium* genus, only one genome-scale study conducted on the amino-acid producer *C. glutamicum* revealed 805 sRNAs expressed in response to various stress conditions: thermic (heat and cold shock), oxidative, diamide, and salt (Mentz et al., 2013). Currently, a total of 887 sRNAs are known in 39 out of the 125 publicly validated corynebacterial species, including cis-encoded, trans-encoded, and plasmid-encoded sRNAs. Further studies regarding the post-transcriptional regulation in corynebacteria are expected to uncover sRNAs and their role in the regulatory mechanisms of this genus in the foreseeable future.

In this review, we collect and summarize corynebacterial sRNA knowledge and elicit the need to fill the gap in the current understanding of corynebacterial regulation. It compiles both experimentally verified and computationally predicted sRNA knowledge in the *Corynebacterium* genus available in literature and public databases. This review begins with the molecular and computational methods for the discovery of sRNAs. Next, it presents bacterial sRNA databases and highlights the ones with corynebacterial data. Then, we discuss both experimental and predicted sRNAs in this genus and show the distribution of the predicted sRNA families across the corynebacterial species. Finally, we conclude that sRNA investigation in the *Corynebacterium* genus is still in its infancy and both experimental and computational studies are largely missing for these organisms of medical-veterinary and biotechnological importance.

sRNA discovery tools and techniques

The investigation of sRNAs roles in bacterial gene expression usually begins from solving two key problems: (i) the discovery of sRNAs and determination of their sequences and (ii) the identification of sRNA targets. To solve both problems, many molecular techniques and computational tools have been developed. Such techniques and methods are summarized in Figure 2 and explained in the next two subsections.

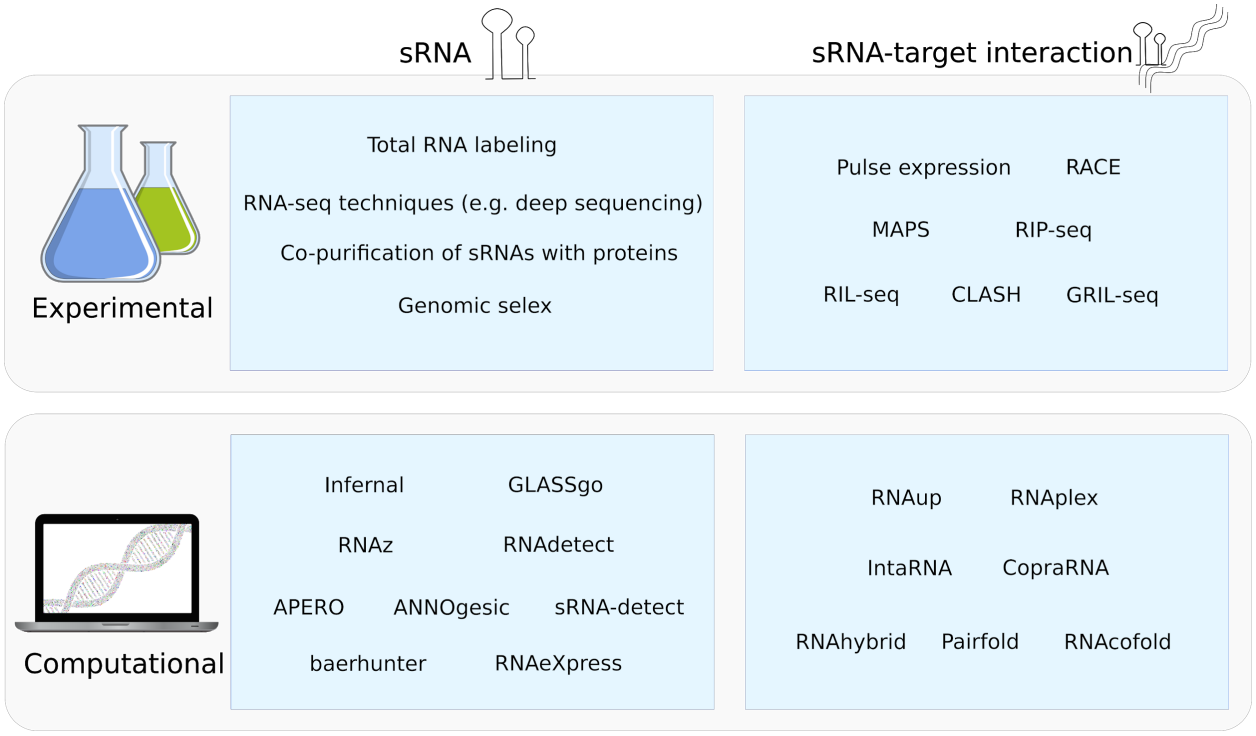


Fig. 2. Overview of techniques and tools used to identify sRNAs and their targets.

Molecular techniques

Due to sRNAs' importance and influence in gene expression regulation, several techniques have been developed to identify these molecules and their targets, as shown in Fig. 2. Total RNA labeling (Wu et al., 1996) as well as RNA-seq techniques, such as deep sequencing (Sharma and Vogel, 2009; Sittka et al., 2008), have been used to identify sRNAs in a transcriptome-wide fashion. Likewise, co-purification of sRNAs with proteins (Wassarman et al., 2001) and genomic SELEX (Lorenz et al., 2006) have been used for RNA-binding protein dependent sRNA identification. To identify sRNA-target interactions, 3-rapid amplification of cDNA ends (RACE) is designed to search for sRNAs regulating a gene of interest. When searching for sRNA targets, pulse expression (Massé et al., 2005) and MAPS (Lalaouna et al., 2015) can be used to investigate the targets of an sRNA of interest. GRIL-seq (Han et al., 2016) identifies RNA-binding proteins dependent and independent sRNA targets of an sRNA of interest *in vivo*. RIP-seq (Vakulskas et al., 2016) identifies RNA targets for RNA-binding proteins of interest. Finally, RIL-seq (Melamed et al., 2016) and CLASH (Waters et al., 2017) identify sRNA interactions associated with RNA-binding proteins. For extensive explanations and comparisons of such techniques, see (Ahmed et al., 2018; Altuvia, 2007; Georg et al., 2020; Hör et al., 2018).

Computational techniques

The computational prediction of sRNAs and their targets provides valuable insights and targets for further experimental assays (Zhang et al., 2017), as shown in Fig. 2. Tools such as Infernal (Barquist et al., 2016; Nawrocki and Eddy, 2013) and GLASSgo (Lott et al., 2018) are able to predict homologous sRNAs, whereas RNAz (Gruber et al., 2010) and RNAdetect (Chen et al., 2019) can be used for predicting functional sRNAs. Moreover, tools such as APERO (Leonard et al., 2019), ANNOgesic (Yu et al., 2018), sRNA-detect (Peña-Castillo et al., 2016), RNA-eXpress (Forster et al., 2013) and baerhunter (Ozuna et al., 2020) can be used to annotate sRNAs from bacterial RNA-seq data. For sRNA interaction prediction, RNAup (Lorenz et al., 2011), IntaRNA (Mann et al., 2017) and RNAplex (Tafer and Hofacker, 2008) can be used to predict the best local RNA-RNA interaction. CopraRNA (Wright et al., 2014, 2013) predicts evolutionarily conserved local RNA-RNA interactions. Finally, full hybrid sRNA-RNA interaction can be predicted with RNAhybrid (Krüger and Rehmsmeier, 2006), RNAcofold (Bernhart et al., 2006) and Pairfold (Tulpan et al., 2005). Several other computational tools are available for exploring the bacterial sRNA landscape; for an extensive listing, explanation and assessment of such tools, see (Backofen et al., 2018; Georg et al., 2020; Li et al., 2012; Naskulwar et al., 2021; Pain et al., 2015; Zhang et al., 2017).

Experimentally verified sRNAs

The first sRNA in *C. glutamicum* was detected by Zemanová et al. in 2008 using the RACE technique. This heat shock-induced asRNA was shown to up-regulate *gntR2* (cg1935) (Brinkrolf et al., 2010; Zemanová et al., 2008), the regulator of gluconate catabolism and glucose uptake in *C. glutamicum* (Frunzke et al., 2008). Later, Mentz and collaborators presented the first genome-wide study of sRNAs in *C. glutamicum* ATCC 13032 (Mentz et al., 2013). The authors used deep sequencing to identify sRNA transcripts under five stress conditions: salt, oxidative, diamide and thermic (heat and cold shock). The total amount of transcripts were further classified into cis-antisense sRNAs (543) and trans-encoded sRNAs (262). The cis-antisense sRNAs' genes were functionally classified, revealing the cell envelope biogenesis and secretion process to have the higher proportion of asRNAs within the genes associated with such functions. In this study, three *trans*-encoded sRNAs were selected for validation with northern blot and further investigation: cgb_03605, cgb_00105 and cgb_20715. The first encodes the highly conserved 6C RNA, while the other two are highly expressed sRNAs: cgb_00105 in the upstream region of cg0010 and cgb_20715 in the upstream region of cg2071. Neither 6C RNA nor cgb_20715 were differentially expressed under any stress condition. Cgb_0105 was expressed in all conditions but heat shock (Mentz et al., 2013). Pahlke et al. demonstrated that 6C RNA is connected to the LexA-dependent SOS response and suggested that it may also be involved in cell division control (Pahlke et al., 2016). In the human pathogen *C. diphtheriae*, two antisense sRNAs were identified in an RNA-seq study investigating its DtxR-mediated transcriptional regulation (Wittchen et al., 2018). Asthese two sRNAs are antisense to the gene encoding diphtheria toxin (*tox*), the detailed understanding of their functions may be of medical importance. The biotechnological potential of corynebacterial sRNAs has been demonstrated by Sun et al., who applied a synthetic sRNA system to enhance glutamate production in *C. glutamicum* through gene knockdown (Sun et al., 2019). The knockdown of *pyk*, *ldhA*, and *odh* genes resulted in 1.5 to 2.8-fold increased extracellular glutamate concentrations. The achieved gain in glutamate production via sRNA metabolic engineering was comparable to the outcome of CRISPRi glutamate enhancement system (Cleto et al., 2016; Sun et al., 2019).

Besides chromosomal sRNAs, plasmid-encoded sRNAs were also experimentally verified for *Corynebacterial* plasmids. The antisense sRNA pGA1 was experimentally identified in the pGA1 plasmid from *C. glutamicum* LP-6 (Venkova-Canova et al., 2003). The authors demonstrated that this sRNA has a role in the negative control of pGA1 copy-number and hypothesized that it might also negatively control the plasmidial replication initiator *rep (Venkova-Canova et al., 2003)*. A more recent study suggested that copy-number control mediated by *repA* antisense

sRNA is a mechanism shared amongst most plasmids of the pCG1 family (Okibe et al., 2010). Hashiro and collaborators also identified a plasmid-encoded antisense sRNA, named sRNA1, in the upstream region of *repA* in the *E. coli-C. glutamicum* shuttle vector pVC7N. This antisense sRNA was considered a potential regulator of pAM330-like plasmids by occluding the ribosome binding-site of the *repA* mRNA (Hashiro et al., 2019).

Databases

The data on corynebacterial sRNA can be found on three types of databases: (i) large comprehensive RNA databases, aggregating the molecular knowledge on various kinds of RNA from a wide range of species, (ii) databases of regulatory interactions and (iii) specialized databases focused on bacterial sRNAs. While specialized databases may be expected to provide the most comprehensive and detailed information of corynebacterial sRNAs and their targets, the practical use of the specialized sRNA databases for corynebacterial research seems to be very limited, as presented in Table 1. The first specialized bacterial sRNA database sRNAMap (Huang et al., 2009) provided the data on sRNAs, TFs and their targets only for Gram-negative bacteria. Later in 2012, Pischimarov et al. presented the database and a comparative analysis platform for sRNA in Gram-positive bacteria called sRNAdb (Pischimarov et al., 2012). BSRD, created by (L. Li et al., 2013), integrated bacterial sRNAs from sRNAMap, Rfam and RegulonDB with expression profiles collected from manual literature curation. Unfortunately, at the time of this review submission, three of the four specialized bacterial sRNA databases (sRNAdb, sRNAMap and BSRD) were unavailable. The only exception was sRNATarBase 3 (http://ccb1.bmi.ac.cn:81/srnatarbase/), which gathers experimentally verified and predicted sRNAs-target interactions from the literature and provides them in a form of sRNA-driven regulatory networks (Wang et al., 2016). Although the latest version of sRNATarBase comprises 201 sRNAs from 53 strains, none of them belongs to the *Corynebacterium* genus.

Being that current sRNA databases provided no data relevant for corynebacterial organisms, we searched for it in databases of regulatory interactions. CoryneReNet 7 (Parise et al., 2020) is the reference database for corynebacterial regulatory interactions, which holds GRNs for 225 corynebacterial strains, focused on regulation by TFs and σ-factors. Abasy Atlas 2.4 (https://abasy.ccg.unam.mx/) holds TF- and σ-driven regulatory networks for 9 corynebacterial organisms. Escorcia-Rodríguez et. al. have recently integrated the 543 asRNAs from (Mentz et al., 2013) in the *C. glutamicum*'s regulatory network of Abasy Atlas, assuming the asRNAs regulate their own genes (Escorcia-Rodríguez et al., 2021). Since this data has been published as a preprint manuscript, it will not be considered as consolidated corynebacterial literature.

To find corynebacterial sRNA from species other than *C. glutamicum,* we queried two RNA databases of broader purpose, RNAcentral (RNAcentral Consortium, 2021) and Rfam (Kalvari et al., 2021). These databases aggregate knowledge on various kinds of RNA molecules derived from a wide range of organisms, including multiple corynebacterial species. Rfam (https://rfam.xfam.org/) is an RNA family database providing multiple sequence alignments, secondary structures, and covariance models for each family. RNAcentral (https://rnacentral.org/) is an online portal that integrates the data on more than 18 million unique non-coding RNAs from 44 databases and provides versatile tools for search, retrieval and analysis of sequence, structure and functional data. Our search identified 47 and 29 corynebacterial sRNAs in Rfam (v14.5) and RNAcentral (v17), respectively. Interestingly, although RNACentral integrated the data from Rfam, these sets of corynebacterial sRNA did not overlap. Moreover, sRNAs found in Rfam and RNAcental had almost no overlap with the set of 805 sRNA from Abasy Atlas and (Mentz et al., 2013). A sole sRNA, cgb_0105, was found in both Abasy Atlas and Rfam. This weak overlap between different data sources indicates that only a small fraction of all the existing sRNA in the *Corynebacterium* genus is discovered and highlights the need for further research.

Table 1. Overview of sRNA databases.

Database	Organism	sRNAs		sRNA structures	sRNA-target regulation		Year	Reference
		All organisms	*Corynebacterium*		All organisms	*Corynebacterium*		
sRNAdb	Gram-positive bacteria	10,561	0	no	N/A	N/A	2012	(Pischimarov *et al.*, 2012)
sRNATarBase 3	Several bacteria	201	0	yes	771	0	2016	(Wang *et al.*, 2016)
BSRD	Several bacteria	9,719	116	yes	194	0	2013	(L. Li *et al.*, 2013)
sRNAMap	Gram-negative bacteria	397	0	yes	60	0	2009	(Huang *et al.*, 2009)
Rfam 14.3	Mouse, human, bacteria and others	62,033	47	yes	N/A	N/A	2021	(Kalvari *et al.*, 2021)
RNACentral	Mouse, human, bacteria and others	57,757	29	yes	N/A	N/A	2021	(RNAcentral Consortium, 2021)
RegulonDB	*Escherichia coli*	132	0	no	235	0	2019	(Santos-Zavaleta *et al.*, 2019)
SRD	*Staphylococcus aureus*	575	0	yes	N/A	N/A	2015	(Sassi *et al.*, 2015)

Predicted sRNAs

Rfam (Kalvari et al., 2021) provides 47 sRNAs distributed into 37 corynebacterial organisms; this database predicts sRNAs based on known sRNA families using the software Infernal (Barquist et al., 2016; Nawrocki and Eddy, 2013). The 47 sRNAs belong to three sRNA families: 5' ureB, Ms_IGR-7 and Cgb105. The 5' ureB sRNA family was first reported to assist in the downregulation of urease acting in response to changes in environmental pH levels in *Helicobacter pylori* (Wen et al., 2011). Rfam presents 27 corynebacterial species with predicted sRNAs belonging to the 5' ureB family. Ms_IGR-7 was initially discovered in *Mycobacterium smegmatis*; it is thought to participate in the regulation of growth and adaptation responding to changes in the nutrient pool (S.-K. Li et al., 2013). Rfam predicted 17 corynebacterial species with predicted sRNAs belonging to this family. *Corynebacterium* sRNA Cgb105 family was first identified in *C. glutamicum* (cgb_0105) in response to environmental stress conditions (Mentz et al., 2013). Rfam presents three corynebacterial species with predicted sRNAs belonging to the Cgb105 family. These sRNA families were initially validated on northern blots, Ms_IGR-7 and Cgb105 were also identified by deep sequencing (S.-K. Li et al., 2013; Mentz et al., 2013; Wen et al., 2011). Likewise, RNAcentral presents 29 ctRNAs for the genus *Corynebacterium* distributed into 15 species, belonging to the ctRNA_pGA1 family. It was initially discovered in the pGA1 plasmid from *C. glutamicum* LP-6 (Venkova-Canova et al., 2003). None of the pathogenic *Corynebacterium* species is predicted to have the sRNAs Ms_IGR-7 and sRNA 105. The distribution of the sRNA families presented in this section is shown in Fig. 3 and the distribution of the sRNAs per organism is shown in Supplementary Table 1.

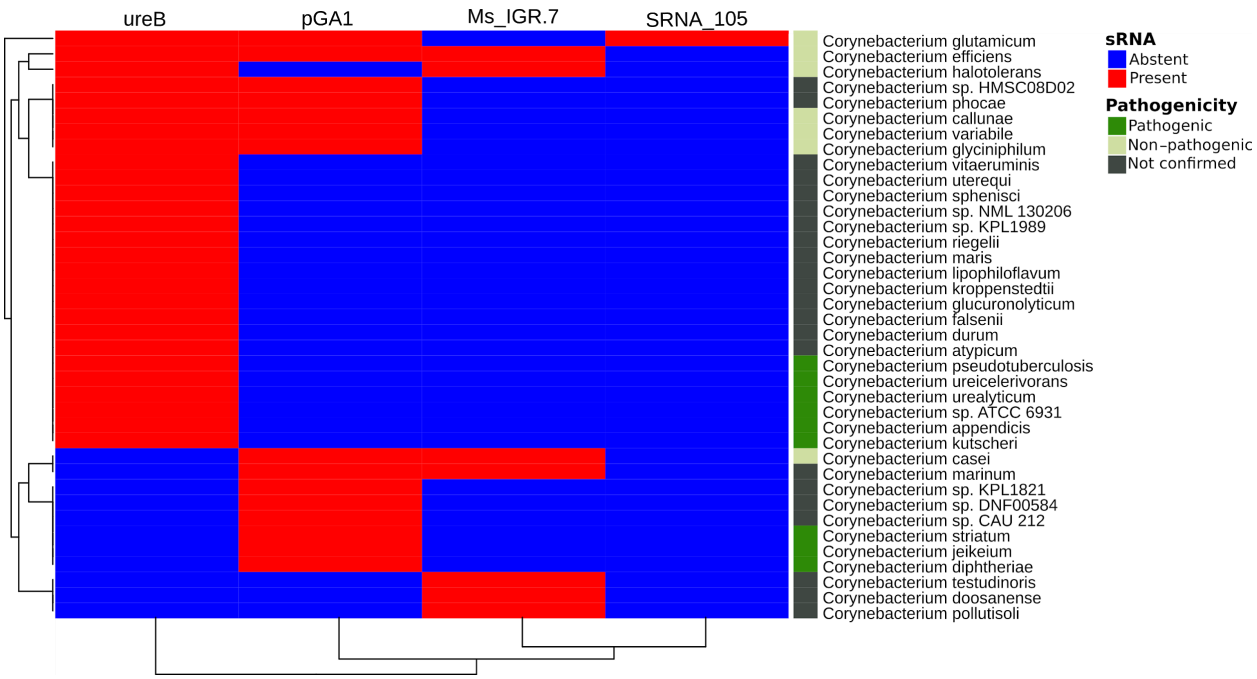


Fig. 3. Distribution of the sRNA families in the corynebacterial species. Blue cells represent the absence of the sRNA and red cells represent its presence.

Discussion and conclusion

Currently, our knowledge concerning sRNAs and their mechanisms of action in the *Corynebacterium* genus is limited by the lack of experimental and *in silico* investigations. To date, corynebacterial sRNAs may be associated with the following: (i) regulation of gluconate metabolism and glucose uptake as well as control of cell division in *C. glutamicum*, (ii) control of the *tox* gene in *C. diphtheriae*, and (iii) control of plasmid copy-number in pGA1 family and pAM330-like plasmids. Despite these molecules' importance in bacterial survival, adaptation and pathogenicity in other genera, the study of sRNAs is still in its infancy in the *Corynebacterium* genus.

The data on corynebacterial sRNAs remains incomplete and scattered among the relevant publications and seems to be underrepresented in sRNA databases. There are currently only 887 sRNAs described in 39 of 125 species in the *Corynebacterium* genus. Most of the known corynebacterial sRNA are only *in silico*-predicted and lack experimental validation. *C. diphtheriae and C. glutamicum* are the only two species with very few sRNAs and sRNA-mediated regulatory interactions experimentally confirmed, suggesting that an important part of the corynebacterial regulatory network remains unexplored. Although the *Corynebacterium* genus includes several pathogenic species, their RNAs are poorly studied despite the key role of these molecules in pathogenicity and virulence. Amongst 51 pathogenic corynebacteria (Zasada and Mosiej, 2018), only nine species have predicted or experimentally verified sRNAs. Likewise, amongst 10

species of biotechnological interest (Oliveira et al., 2017), only *C. glutamicum*'s sRNAs have been used for biotechnological purposes (Sun et al., 2019). This indicates that there is still much to be revealed concerning sRNAs in biotechnological and medical-veterinary corynebacterial organisms.

Further investigation of sRNAs roles in regulatory mechanisms may aid to elucidate *C. pseudotuberculosis*' pathogenesis, virulence and biofilm formation. Interestingly, the master regulator of iron DtxR was not differentially expressed in a recent *C. pseudotuberculosis* transcriptomics study under iron starvation (Ibraim et al., 2019), indicating other regulatory mechanisms may be influencing DtxR expression. Since sRNAs play an important role in iron homeostasis, investigating the role of sRNAs in the iron homeostasis of this zoonotic pathogen may be of general interest. Likewise, investigating the role of sRNAs in the human-affecting diphtheria toxin producers *C. diphtheriae* and *C. ulcerans* may also be interesting. Moreover, *C. diphtheriae*'s biofilm formation has been associated with its pathogenesis and its *tox* asRNAs remains to be studied. Other relevant biofilm-forming pathogens in this genus are *Corynebacterium striatum*, *Corynebacterium pseudodiphtheriticum*. Furthermore, an sRNA-mediated system designed to influence gene expression has been successfully applied in *C. glutamicum* to enhance glutamate production (Sun et al., 2019). This system may also be applied to enhance the production of other biomolecules, such as glutamate in *C. efficiens* and L-Lysine in *C. glutamicum*.

Deciphering sRNA-mediated regulatory interactions may aid in the understanding of the regulatory mechanisms and elucidate how transcriptional and post-transcriptional regulation jointly influence the gene expression in this genus. In this review, we collected, summarized and organized the available corynebacterial sRNA data, presenting an overview of the sRNA knowledge in *Corynebacterium* and indicating promising future investigations on this genus. Our major contribution is highlighting the lack of knowledge concerning sRNAs of a genus of medical, veterinary and industrial relevance.

References

Ahmed, W., Hafeez, M.A., Mahmood, S., 2018. Identification and functional characterization of bacterial small non-coding RNAs and their target: A review. Gene Reports 10, 167–176. https://doi.org/10.1016/j.genrep.2018.01.001

Altuvia, S., 2007. Identification of bacterial small non-coding RNAs: experimental approaches. Curr. Opin. Microbiol. 10, 257–261. https://doi.org/10.1016/j.mib.2007.05.003

American Society for Microbiology. General Meeting, 2008. Abstracts of the ... General Meeting of the American Society for Microbiology.

Backofen, R., Gorodkin, J., Hofacker, I.L., Stadler, P.F., 2018. Comparative RNA Genomics. Comparative Genomics. https://doi.org/10.1007/978-1-4939-7463-4_14

Bak, G., Lee, J., Suk, S., Kim, D., Young Lee, J., Kim, K.-S., Choi, B.-S., Lee, Y., 2015. Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in Escherichia coli. Sci. Rep. 5, 15287. https://doi.org/10.1038/srep15287

Barquist, L., Burge, S.W., Gardner, P.P., 2016. Studying RNA Homology and Conservation with Infernal: From Single Sequences to RNA Families. Curr. Protoc. Bioinformatics 54, 12.13.1–12.13.25. https://doi.org/10.1002/cpbi.4

Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F., Hofacker, I.L., 2006. Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol. Biol. 1, 3. https://doi.org/10.1186/1748-7188-1-3

Bervoets, I., Charlier, D., 2019. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. FEMS Microbiol. Rev. 43, 304–339. https://doi.org/10.1093/femsre/fuz001

Brinkrolf, K., Schröder, J., Pühler, A., Tauch, A., 2010. The transcriptional regulatory repertoire of Corynebacterium glutamicum: reconstruction of the network controlling pathways involved in lysine and glutamate production. J. Biotechnol. 149, 173–182. https://doi.org/10.1016/j.jbiotec.2009.12.004

Brosse, A., Guillier, M., 2018. Bacterial Small RNAs in Mixed Regulatory Networks. Microbiol Spectr 6. https://doi.org/10.1128/microbiolspec.RWR-0014-2017

Carrier, M.-C., Lalaouna, D., Massé, E., 2018. Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action. Annu. Rev. Microbiol. 72, 141–161. https://doi.org/10.1146/annurev-micro-090817-062607

Chakravarty, S., Massé, E., 2019. RNA-Dependent Regulation of Virulence in Pathogenic Bacteria. Front. Cell. Infect. Microbiol. 9, 337. https://doi.org/10.3389/fcimb.2019.00337

Chareyre, S., Mandin, P., 2018. Bacterial Iron Homeostasis Regulation by sRNAs. Microbiol Spectr 6. https://doi.org/10.1128/microbiolspec.RWR-0010-2017

Chen, C.-C., Qian, X., Yoon, B.-J., 2019. RNAdetect: efficient computational detection of novel non-coding RNAs. Bioinformatics 35, 1133–1141. https://doi.org/10.1093/bioinformatics/bty765

Cleto, S., Jensen, J.V., Wendisch, V.F., Lu, T.K., 2016. Corynebacterium glutamicum Metabolic Engineering with CRISPR Interference (CRISPRi). ACS Synth. Biol. 5, 375–385. https://doi.org/10.1021/acssynbio.5b00216

Dambach, M., Irnov, I., Winkler, W.C., 2013. Association of RNAs with Bacillus subtilis Hfq. PLoS One 8, e55156. https://doi.org/10.1371/journal.pone.0055156

Escorcia-Rodríguez, J.M., Tauch, A., Freyre-González, J.A., 2021. Corynebacterium glutamicum regulation beyond transcription: Organizing principles and reconstruction of an extended regulatory network incorporating regulations mediated by small RNA and protein-protein interactions. arXiv preprint arXiv:2101.02865.

Forster, S.C., Finkel, A.M., Gould, J.A., Hertzog, P.J., 2013. RNA-eXpress annotates novel transcript features in RNA-seq data. Bioinformatics 29, 810–812. https://doi.org/10.1093/bioinformatics/btt034

Frunzke, J., Engels, V., Hasenbein, S., Gätgens, C., Bott, M., 2008. Co-ordinated regulation of gluconate catabolism and glucose uptake in Corynebacterium glutamicum by two functionally equivalent transcriptional regulators, GntR1 and GntR2. Mol. Microbiol. 67, 305–322. https://doi.org/10.1111/j.1365-2958.2007.06020.x

Georgi, T., Rittmann, D., Wendisch, V.F., 2005. Lysine and glutamate production by Corynebacterium glutamicum on glucose, fructose and sucrose: roles of malic enzyme and fructose-1,6-bisphosphatase. Metab. Eng. 7, 291–301. https://doi.org/10.1016/j.ymben.2005.05.001

Georg, J., Lalaouna, D., Hou, S., Lott, S.C., Caldelari, I., Marzi, S., Hess, W.R., Romby, P., 2020. The power of cooperation: Experimental and computational approaches in the functional characterization of bacterial sRNAs. Mol. Microbiol. 113, 603–612. https://doi.org/10.1111/mmi.14420

Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L., Stadler, P.F., 2010. RNAz 2.0: improved noncoding RNA detection. Pac. Symp. Biocomput. 69–79.

Han, K., Tjaden, B., Lory, S., 2016. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. Nat Microbiol 2, 16239. https://doi.org/10.1038/nmicrobiol.2016.239

Hashiro, S., Mitsuhashi, M., Yasueda, H., 2019. High copy number mutants derived from Corynebacterium glutamicum cryptic plasmid pAM330 and copy number control. J. Biosci. Bioeng. 127, 529–538. https://doi.org/10.1016/j.jbiosc.2018.10.012

Hébrard, M., Kröger, C., Srikumar, S., Colgan, A., Händler, K., Hinton, J.C.D., 2012. sRNAs and the virulence of Salmonella enterica serovar Typhimurium. RNA Biol. 9, 437–445. https://doi.org/10.4161/rna.20480

Hör, J., Gorski, S.A., Vogel, J., 2018. Bacterial RNA Biology on a Genome Scale. Mol. Cell 70, 785–799. https://doi.org/10.1016/j.molcel.2017.12.023

Hör, J., Matera, G., Vogel, J., Gottesman, S., Storz, G., 2020. Trans-Acting Small RNAs and Their Effects on Gene Expression in Escherichia coli and Salmonella enterica.

EcoSal Plus 9. https://doi.org/10.1128/ecosalplus.ESP-0030-2019

Huang, H.-Y., Chang, H.-Y., Chou, C.-H., Tseng, C.-P., Ho, S.-Y., Yang, C.-D., Ju, Y.-W., Huang, H.-D., 2009. sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. Nucleic Acids Res. 37, D150–4. https://doi.org/10.1093/nar/gkn852

Ibraim, I.C., Parise, M.T.D., Parise, D., Sfeir, M.Z.T., de Paula Castro, T.L., Wattam, A.R., Ghosh, P., Barh, D., Souza, E.M., Góes-Neto, A., Gomide, A.C.P., Azevedo, V., 2019. Transcriptome profile of Corynebacterium pseudotuberculosis in response to iron limitation. BMC Genomics 20, 663. https://doi.org/10.1186/s12864-019-6018-1

Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., Rivas, E., Eddy, S.R., Finn, R.D., Bateman, A., Petrov, A.I., 2021. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. 49, D192–D200. https://doi.org/10.1093/nar/gkaa1047

Kong, H.-K., Liu, X., Lo, W.U., Pan, Q., Law, C.O.K., Chan, T.F., Ho, P.L., Lau, T.C.K., 2018. Identification of Plasmid-Encoded sRNAs in a blaNDM-1-Harboring Multidrug-Resistance Plasmid pNDM-HK in Enterobacteriaceae. Front. Microbiol. 9, 532. https://doi.org/10.3389/fmicb.2018.00532

Krüger, J., Rehmsmeier, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34, W451–4. https://doi.org/10.1093/nar/gkl243

Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J.T., Massé, E., 2015. A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. Mol. Cell 58, 393–405. https://doi.org/10.1016/j.molcel.2015.03.013

Leonard, S., Meyer, S., Lacour, S., Nasser, W., Hommais, F., Reverchon, S., 2019. APERO: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data. Nucleic Acids Res. 47, e88. https://doi.org/10.1093/nar/gkz485

Li, L., Huang, D., Cheung, M.K., Nong, W., Huang, Q., Kwan, H.S., 2013. BSRD: a repository for bacterial small regulatory RNA. Nucleic Acids Res. 41, D233–8. https://doi.org/10.1093/nar/gks1264

Li, S.-K., Ng, P.K.-S., Qin, H., Lau, J.K.-Y., Lau, J.P.-Y., Tsui, S.K.-W., Chan, T.-F., Lau, T.C.-K., 2013. Identification of small RNAs in Mycobacterium smegmatis using heterologous Hfq. RNA 19, 74–84. https://doi.org/10.1261/rna.034116.112

Li, W., Ying, X., Lu, Q., Chen, L., 2012. Predicting sRNAs and their targets in bacteria. Genomics Proteomics Bioinformatics 10, 276–284. https://doi.org/10.1016/j.gpb.2012.09.004

Lorenz, C., von Pelchrzim, F., Schroeder, R., 2006. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. Nat. Protoc. 1, 2204–2212. https://doi.org/10.1038/nprot.2006.372

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler,

P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26. https://doi.org/10.1186/1748-7188-6-26

Lott, S.C., Schäfer, R.A., Mann, M., Backofen, R., Hess, W.R., Voß, B., Georg, J., 2018. GLASSgo - Automated and Reliable Detection of sRNA Homologs From a Single Input Sequence. Front. Genet. 9, 124. https://doi.org/10.3389/fgene.2018.00124

Mandlik, A., Swierczynski, A., Das, A., Ton-That, H., 2007. Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. Mol. Microbiol. 64, 111–124. https://doi.org/10.1111/j.1365-2958.2007.05630.x

Mann, M., Wright, P.R., Backofen, R., 2017. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. Nucleic Acids Res. 45, W435–W439. https://doi.org/10.1093/nar/gkx279

Massé, E., Vanderpool, C.K., Gottesman, S., 2005. Effect of RyhB small RNA on global iron use in Escherichia coli. J. Bacteriol. 187, 6962–6971. https://doi.org/10.1128/JB.187.20.6962-6971.2005

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y.E., Reiss, N., Bar, A., Altuvia, Y., Argaman, L., Margalit, H., 2016. Global Mapping of Small RNA-Target Interactions in Bacteria. Mol. Cell 63, 884–897. https://doi.org/10.1016/j.molcel.2016.07.026

Mentz, A., Neshat, A., Pfeifer-Sancar, K., Pühler, A., Rückert, C., Kalinowski, J., 2013. Comprehensive discovery and characterization of small RNAs in Corynebacterium glutamicum ATCC 13032. BMC Genomics 14, 714. https://doi.org/10.1186/1471-2164-14-714

Naskulwar, K., Chevez-Guardado, R., Peña-Castillo, L., 2021. Computational Methods for Elucidating Gene Expression Regulation in Bacteria. Methods Mol. Biol. 2190, 95–114. https://doi.org/10.1007/978-1-0716-0826-5_4

Nawrocki, E.P., Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Nitzan, M., Rehani, R., Margalit, H., 2017. Integration of Bacterial Small RNAs in Regulatory Networks. Annu. Rev. Biophys. 46, 131–148. https://doi.org/10.1146/annurev-biophys-070816-034058

Okibe, N., Suzuki, N., Inui, M., Yukawa, H., 2010. Antisense-RNA-mediated plasmid copy number control in pCG1-family plasmids, pCGR2 and pCG1, in Corynebacterium glutamicum. Microbiology 156, 3609–3623. https://doi.org/10.1099/mic.0.043745-0

Oliveira, A., Oliveira, L.C., Aburjaile, F., Benevides, L., Tiwari, S., Jamal, S.B., Silva, A., Figueiredo, H.C.P., Ghosh, P., Portela, R.W., De Carvalho Azevedo, V.A., Wattam, A.R., 2017. Insight of Genus Corynebacterium: Ascertaining the Role of Pathogenic and Non-pathogenic Species. Front. Microbiol. 8, 1937. https://doi.org/10.3389/fmicb.2017.01937

Ozuna, A., Liberto, D., Joyce, R.M., Arnvig, K.B., Nobeli, I., 2020. baerhunter: an R package for the discovery and analysis of expressed non-coding regions in bacterial RNA-seq data. Bioinformatics 36, 966–969. https://doi.org/10.1093/bioinformatics/btz643

Pahlke, J., Dostálová, H., Holátko, J., Degner, U., Bott, M., Pátek, M., Polen, T., 2016. The small 6C RNA of Corynebacterium glutamicum is involved in the SOS response. RNA Biol. 13, 848–860. https://doi.org/10.1080/15476286.2016.1205776

Pain, A., Ott, A., Amine, H., Rochat, T., Bouloc, P., Gautheret, D., 2015. An assessment of bacterial small RNA target prediction programs. RNA Biol. 12, 509–513. https://doi.org/10.1080/15476286.2015.1020269

Parise, D., Parise, M.T.D., Kataka, E., Kato, R.B., List, M., Tauch, A., de Carvalho Azevedo, V.A., Baumbach, J., 2021. On the Consistency between Gene Expression and the Gene Regulatory Network of Corynebacterium glutamicum. Network and Systems Medicine. https://doi.org/10.1089/nsm.2020.0014

Parise, M.T.D., Parise, D., Kato, R.B., Pauling, J.K., Tauch, A., Azevedo, V.A. de C., Baumbach, J., 2020. CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. Sci Data 7, 142. https://doi.org/10.1038/s41597-020-0484-9

Peña-Castillo, L., Grüell, M., Mulligan, M.E., Lang, A.S., 2016. DETECTION OF BACTERIAL SMALL TRANSCRIPTS FROM RNA-SEQ DATA: A COMPARATIVE ASSESSMENT. Pac. Symp. Biocomput. 21, 456–467. https://doi.org/10.1142/9789814749411_0042

Pischimarov, J., Kuenne, C., Billion, A., Hemberger, J., Cemič, F., Chakraborty, T., Hain, T., 2012. sRNAdb: a small non-coding RNA database for gram-positive bacteria. BMC Genomics 13, 384. https://doi.org/10.1186/1471-2164-13-384

RNAcentral Consortium, 2021. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. Nucleic Acids Res. 49, D212–D220. https://doi.org/10.1093/nar/gkaa921

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J.S., Alquicira-Hernández, K., Muñiz-Rascado, L.J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D.A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C.-F., Galagan, J., Collado-Vides, J., 2019. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. Nucleic Acids Res. 47, D212–D220. https://doi.org/10.1093/nar/gky1077

Sassi, M., Augagneur, Y., Mauro, T., Ivain, L., Chabelskaya, S., Hallier, M., Sallou, O., Felden, B., 2015. SRD: a Staphylococcus regulatory RNA database. RNA 21, 1005–1017. https://doi.org/10.1261/rna.049346.114

Sharma, C.M., Vogel, J., 2009. Experimental approaches for the discovery and characterization of regulatory small RNA. Curr. Opin. Microbiol. 12, 536–546. https://doi.org/10.1016/j.mib.2009.07.006

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C.D., Vogel, J., 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. 4, e1000163. https://doi.org/10.1371/journal.pgen.1000163

Storz, G., Vogel, J., Wassarman, K.M., 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell 43, 880–891. https://doi.org/10.1016/j.molcel.2011.08.022

Sun, D., Chen, J., Wang, Y., Li, M., Rao, D., Guo, Y., Chen, N., Zheng, P., Sun, J., Ma, Y., 2019. Metabolic engineering of Corynebacterium glutamicum by synthetic small regulatory RNAs. J. Ind. Microbiol. Biotechnol. 46, 203–208. https://doi.org/10.1007/s10295-018-02128-4

Tafer, H., Hofacker, I.L., 2008. RNAplex: a fast tool for RNA-RNA interaction search. Bioinformatics 24, 2657–2663. https://doi.org/10.1093/bioinformatics/btn193

Thompson, K.M., Rhodius, V.A., Gottesman, S., 2007. SigmaE regulates and is regulated by a small RNA in Escherichia coli. J. Bacteriol. 189, 4243–4256. https://doi.org/10.1128/JB.00020-07

Tulpan, D., Andronescu, M., Chang, S.B., Shortreed, M.R., Condon, A., Hoos, H.H., Smith, L.M., 2005. Thermodynamically based DNA strand design. Nucleic Acids Res. 33, 4951–4964. https://doi.org/10.1093/nar/gki773

Vakulskas, C.A., Leng, Y., Abe, H., Amaki, T., Okayama, A., Babitzke, P., Suzuki, K., Romeo, T., 2016. Antagonistic control of the turnover pathway for the global regulatory sRNA CsrB by the CsrA and CsrD proteins. Nucleic Acids Res. 44, 7896–7910. https://doi.org/10.1093/nar/gkw484

Venkova-Canova, T., Pátek, M., Nesvera, J., 2003. Control of rep gene expression in plasmid pGA1 from Corynebacterium glutamicum. J. Bacteriol. 185, 2402–2409. https://doi.org/10.1128/jb.185.8.2402-2409.2003

Wang, J., Liu, T., Zhao, B., Lu, Q., Wang, Z., Cao, Y., Li, W., 2016. sRNATarBase 3.0: an updated database for sRNA-target interactions in bacteria. Nucleic Acids Res. 44, D248–53. https://doi.org/10.1093/nar/gkv1127

Wang, Y., Cao, G., Xu, D., Fan, L., Wu, X., Ni, X., Zhao, S., Zheng, P., Sun, J., Ma, Y., 2018. A Novel Corynebacterium glutamicum l-Glutamate Exporter. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/AEM.02691-17

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., Gottesman, S., 2001. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. 15, 1637–1651. https://doi.org/10.1101/gad.901001

Waters, S.A., McAteer, S.P., Kudla, G., Pang, I., Deshpande, N.P., Amos, T.G., Leong, K.W., Wilkins, M.R., Strugnell, R., Gally, D.L., Tollervey, D., Tree, J.J., 2017. Small RNA interactome of pathogenic E. coli revealed through crosslinking of RNase E. EMBO J. 36, 374–387. https://doi.org/10.15252/embj.201694639

Wen, Y., Feng, J., Scott, D.R., Marcus, E.A., Sachs, G., 2011. A cis-encoded antisense small RNA regulated by the HP0165-HP0166 two-component system controls expression of ureB in Helicobacter pylori. J. Bacteriol. 193, 40–51. https://doi.org/10.1128/JB.00800-10

Wittchen, M., Busche, T., Gaspar, A.H., Lee, J.H., Ton-That, H., Kalinowski, J., Tauch, A., 2018. Transcriptome sequencing of the human pathogen Corynebacterium diphtheriae NCTC 13129 provides detailed insights into its transcriptional

landscape and into DtxR-mediated transcriptional regulation. BMC Genomics 19, 82.

Wright, P.R., Georg, J., Mann, M., Sorescu, D.A., Richter, A.S., Lott, S., Kleinkauf, R., Hess, W.R., Backofen, R., 2014. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res. 42, W119–23. https://doi.org/10.1093/nar/gku359

Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., Vogel, J., Hess, W.R., Backofen, R., Georg, J., 2013. Comparative genomics boosts target prediction for bacterial small RNAs. Proc. Natl. Acad. Sci. U. S. A. 110, E3487–96. https://doi.org/10.1073/pnas.1303248110

Wu, T.P., Ruan, K.C., Liu, W.Y., 1996. A fluorescence-labeling method for sequencing small RNA on polyacrylamide gel. Nucleic Acids Res. 24, 3472–3473. https://doi.org/10.1093/nar/24.17.3472

Yip, C.B., Ding, H., Hynes, M.F., 2015. Counter-transcribed RNAs of Rhizobium leguminosarum repABC plasmids exert incompatibility effects only when highly expressed. Plasmid 78, 37–47. https://doi.org/10.1016/j.plasmid.2014.12.003

Yu, S.-H., Vogel, J., Förstner, K.U., 2018. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. Gigascience 7. https://doi.org/10.1093/gigascience/giy096

Zasada, A.A., Mosiej, E., 2018. Contemporary microbiology and identification of Corynebacteria spp. causing infections in human. Lett. Appl. Microbiol. 66, 472–483. https://doi.org/10.1111/lam.12883

Zemanová, M., Kaderábková, P., Pátek, M., Knoppová, M., Silar, R., Nesvera, J., 2008. Chromosomally encoded small antisense RNA in Corynebacterium glutamicum. FEMS Microbiol. Lett. 279, 195–201. https://doi.org/10.1111/j.1574-6968.2007.01024.x

Zhang, Y., Huang, H., Zhang, D., Qiu, J., Yang, J., Wang, K., Zhu, L., Fan, J., Yang, J., 2017. A Review on Recent Computational Methods for Predicting Noncoding RNAs. Biomed Res. Int. 2017, 9139504. https://doi.org/10.1155/2017/9139504

Supplementary Table 1. sRNA distribution in the *Corynebacterium* genus

Organism	Nr. of sRNAs	Source
Corynebacterium glutamicum	817	(Mentz, 2013), Rfam, RNA central, (Venkova-Canova, Pátek and Nesvera, 2003) (Hashiro, Mitsuhashi and Yasueda, 2019),(Sun et al., 2019)
Corynebacterium appendicis	1	Rfam
Corynebacterium atypicum	1	Rfam
Corynebacterium callunae	2	Rfam, RNA central
Corynebacterium casei	3	Rfam, RNA central
Corynebacterium doosanense	1	Rfam
Corynebacterium durum	1	Rfam
Corynebacterium efficiens	11	Rfam, RNA central
Corynebacterium falsenii	1	Rfam
Corynebacterium glucuronolyticum	1	Rfam
Corynebacterium glyciniphilum	2	Rfam, RNA central
Corynebacterium halotolerans	2	Rfam
Corynebacterium kroppenstedtii	1	Rfam
Corynebacterium kutscheri	1	Rfam
Corynebacterium lipophiloflavum	1	Rfam
Corynebacterium marinum	5	Rfam, RNA central
Corynebacterium maris	1	Rfam
Corynebacterium phocae	2	Rfam, RNA central
Corynebacterium pollutisoli	1	Rfam
Corynebacterium pseudotuberculosis	1	Rfam
Corynebacterium riegelii	1	Rfam
Corynebacterium sp. ATCC 6931	1	Rfam
Corynebacterium sp. HMSC08D02	2	Rfam, RNA central
Corynebacterium sp. KPL1989	1	Rfam
Corynebacterium sp. NML 130206	1	Rfam
Corynebacterium sphenisci	1	Rfam
Corynebacterium testudinoris	2	Rfam
Corynebacterium urealyticum	1	Rfam
Corynebacterium ureicelerivorans	1	Rfam
Corynebacterium uterequi	1	Rfam

Corynebacterium variabile	2	Rfam, RNA central
Corynebacterium vitaeruminis	1	Rfam
Corynebacterium jeikeium	7	RNA central
Corynebacterium sp. DNF00584	1	RNA central
Corynebacterium sp. KPL1821	1	RNA central
Corynebacterium sp. CAU 212	1	RNA central
Corynebacterium diphtheriae	4	RNA central, (Wittchen et al. 2018)
Corynebacterium striatum	2	RNA central
Total	887	

2. Research Article

Title: An integrated database of small RNAs and their interplay with transcriptional gene regulatory networks in corynebacteria.

Authors: **Mariana Teixeira Dornelles Parise**, Doglas Parise, Flavia Figueira Aburjaile, Anne Cybelle Pinto Gomide, Rodrigo Bentes Kato, Martin Raden, Rolf Backofen, Vasco Ariston de Carvalho Azevedo, Jan Baumbach

Status: Submitted to the journal Frontiers in Microbiology (2-year impact factor: 4.235).

One of the main objectives of this thesis is to expand the post-transcriptional regulatory knowledge of the *Corynebacterium* genus. This methodological article describes (i) the prediction of sRNAs and their targets for six corynebacterial species of medical, veterinary and biotechnological interest and (ii) the subsequent integration of their sRNA-driven regulation to the CoryneRegNet's GRNs. The information presented here contributes by suggesting that sRNAs play an important role in the regulatory networks of corynebacterial species. Furthermore, we highlight sRNA regulatory interactions predicted for a virulence factor in *C. pseudotuberculosis*, a candidate virulence factor in *C. ulcerans* and a potential drug target in *C. diphtheriae* for further investigation.

Supplementary material:
https://www.dropbox.com/s/ut8srp4sp1txky0/656435_SupMaterial.zip?dl=0

frontiers

An integrated database of small RNAs and their interplay with transcriptional gene regulatory networks in corynebacteria

Mariana Teixeira Dornelles Parise[1, 2*], Doglas Parise[1, 2], Flavia F. Aburjaile[3], Anne C. Pinto Gomide[1], Rodrigo B. Kato[1], Martin Raden[4], Rolf Backofen[4], Vasco A. Azevedo[1], Jan Baumbach[2, 5]

[1]Federal University of Minas Gerais, Brazil, [2]Technical University of Munich, Germany, [3]Oswaldo Cruz Foundation (Fiocruz), Brazil, [4]University of Freiburg, Germany, [5]University of Southern Denmark, Denmark

Submitted to Journal:
Frontiers in Microbiology

Specialty Section:
Systems Microbiology

Article type:
Methods Article

Manuscript ID:
656435

Received on:
20 Jan 2021

Revised on:
04 May 2021

Journal website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

MP, MR, RB, VA and JB conceptualized this work. MP and DP developed the software and wrote the manuscript. MP performed the analysis. VA, RK and JB supervised the work. MR, RB, RK, FA, AG VA and JB reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Keywords

small RNAs, sRNA targets, Corynebacterium, CoryneRegNet, Gene Regulatory Networks

Abstract

Word count: 168

Small RNAs (sRNAs) are one of the key players in the post-transcriptional regulation of bacterial gene expression. These molecules, together with transcription factors, form regulatory networks and greatly influence the bacterial regulatory landscape. Little is known concerning sRNAs and their influence on the regulatory machinery in the genus Corynebacterium, despite its medical, veterinary and biotechnological importance. Here, we expand corynebacterial regulatory knowledge by integrating sRNAs and their regulatory interactions into the transcriptional regulatory networks of six corynebacterial species, covering four human and animal pathogens, and integrate this data into the CoryneRegNet database. To this end, we predicted sRNAs to regulate 754 genes, including 206 transcription factors, in corynebacterial gene regulatory networks. Amongst them, the sRNA Cd-NCTC13129-sRNA-2 is predicted to directly regulate ydfH, which indirectly regulates 66 genes, including the global regulator glxR in C. diphtheriae. All of the sRNA-enriched regulatory networks of the genus Corynebacterium have been made publicly available in the newest release of CoryneRegNet (www.exbio.wzw.tum.de/coryneregnet/) to aid in providing valuable insights and to guide future experiments.

Contribution to the field

Small RNAs (sRNAs) are one of the key players in the post-transcriptional regulation of bacterial gene expression. These molecules form complex regulatory circuits by (i) co-regulating genes alongside transcription factors and sigma factors, and (ii) regulating such regulatory proteins. Despite their influence and importance on gene expression, bacterial databases integrating sRNA-based and transcriptional Gene Regulatory Networks (GRNs) are largely missing. Here, expand corynebacterial GRNs knowledge by presenting the first attempt of integrating sRNA regulation into GRNs in the genus Corynebacterium. We applied a conservative bioinformatics prediction approach resulting in sRNA regulatory networks, which we have integrated with the GRNs from CoryneRegNet, yielding more than 6,000 sRNA regulatory interactions and revealing 754 genes in the GRN to be regulated by both sRNAs and transcription factors. In a case study, we used the sRNA-enriched database of CoryneRegNet to identify sRNAs regulating virulence factors, potential drug targets and global regulators. To integrate such regulatory interactions, we updated CoryneRegNet's front- and back-end, available at https://www.exbio.wzw.tum.de/coryneregnet/. To sum up, we believe this integrated network data to provide valuable candidates for future experimental assays and suggest sRNAs may have a great influence in the corynebacterial regulation of gene expression.

Funding statement

JB is grateful for support from H2020 grant RepoTrial (nr. 777111) and his VILLUM Young Investigator grant (nr. 13154). DP received support from CAPES (nr. 88887.364607/2019-00) and MP from CNPq (nr. 201336/2018-9), for their work at TUM in Germany. MP's work was also supported by the German Research Foundation (under SFB924). VA is grateful for support from his CNPq Research Productivity grant (nr. 305093/2015-0), CNPq Universal grant (nr. 405233/2016-7) and FAPEMIG grant (nr. APQ 02600-17). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: No human studies are presented in this manuscript.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.exbio.wzw.tum.de/coryneregnet/processToDownalod.htm.

An integrated database of small RNAs and their interplay with transcriptional gene regulatory networks in corynebacteria

Mariana Teixeira Dornelles Parise [1,2], Doglas Parise [1,2], Flavia Figueira Aburjaile [3], Anne Cybelle Pinto Gomide [2], Rodrigo Bentes Kato [2], Martin Raden [4], Rolf Backofen [4], Vasco Ariston de Carvalho Azevedo [2], Jan Baumbach [1,5,6]

[1] Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany.
[2] Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
[3] Oswaldo Cruz Foundation (Fiocruz), Brazil.
[4] Bioinformatics, Department of Computer Science, University of Freiburg, Freiburg, Germany
[5] Computational Biomedicine lab, Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Germany
[6] Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany

Abstract

Small RNAs (sRNAs) are one of the key players in the post-transcriptional regulation of bacterial gene expression. These molecules, together with transcription factors, form regulatory networks and greatly influence the bacterial regulatory landscape. Little is known concerning sRNAs and their influence on the regulatory machinery in the genus *Corynebacterium,* despite its medical, veterinary and biotechnological importance. Here, we expand corynebacterial regulatory knowledge by integrating sRNAs and their regulatory interactions into the transcriptional regulatory networks of six corynebacterial species, covering four human and animal pathogens, and integrate this data into the CoryneRegNet database. To this end, we predicted sRNAs to regulate 754 genes, including 206 transcription factors, in corynebacterial gene regulatory networks. Amongst them, the sRNA Cd-NCTC13129-sRNA-2 is predicted to directly regulate *ydfH*, which indirectly regulates 66 genes, including the global regulator *glxR* in *C. diphtheriae*. All of the sRNA-enriched regulatory networks of the genus *Corynebacterium* have been made publicly available in the newest release of CoryneRegNet (www.exbio.wzw.tum.de/coryneregnet/) to aid in providing valuable insights and to guide future experiments.

Keywords: Small RNAs, sRNA targets, *Corynebacterium*, CoryneRegNet, Gene Regulatory Networks

1. Introduction

Small RNAs (sRNAs) have been proven to be important players in the regulatory mechanisms of bacteria (Waters and Storz, 2009; Gripenland et al., 2010; Waters et al., 2017). These molecules interact with messenger RNAs (mRNAs) to induce or repress gene expression post-transcriptionally (De Lay et al., 2013; Papenfort and Vanderpool, 2015). Regulatory sRNAs can both co-regulate genes alongside transcription factors (TFs) and sigma factors, as well as regulate these regulatory proteins, forming regulatory circuits (Lee and Gottesman, 2016; Mandin et al., 2016; Nitzan et al., 2017). Consequently, sRNA regulations have been recently integrated into gene regulatory networks (GRNs), granting these networks a more comprehensive view of gene expression regulation (Beisel and Storz, 2010; Nitzan et al., 2017; Brosse and Guillier, 2018; Arrieta-Ortiz et al., 2020).

Due to its importance, both computational and experimental techniques have been developed for identifying sRNAs and their interactions. Experimental methods such as total RNA labeling (Wu et al., 1996), deep sequencing (Sittka et al., 2008; Sharma and Vogel, 2009; Barquist and Vogel, 2015) and co-immunoprecipitation of RNA-binding proteins (Faner and Feig, 2013) have been used to discover novel sRNAs. Other techniques, such as pulse-expression (Massé et al., 2005), MAPS (Lalaouna and Massé, 2015), RIL-seq (Melamed et al., 2016) and GRIL-seq (Han et al., 2016) have been applied to identify sRNA-mRNA interactions. For a comprehensive description see (Altuvia, 2007; Ahmed et al., 2018; Diallo and Provost, 2020). Computational methods stand out by revealing promising sRNA candidates for further experimental testing without exhaustive wet-lab assays (Wright and Georg, 2018). In general, sRNA prediction software can be grouped into three types of methods: *de novo*, homology-based and experimental-data dependent (Zhang et al., 2017b; Backofen et al., 2018). sRNA target prediction software can be divided into two types of methods: local-interaction based and full-hybrid based (Pain et al., 2015). For further explanations and comparisons of these methods see (Pain et al., 2015; Zhang et al., 2017b; Backofen et al., 2018).

Both predicted and experimental bacterial sRNAs have been made publicly available in databases such as Rfam (Kalvari et al., 2018) and RNA central (The RNAcentral Consortium, 2019) for several organisms, including bacteria. Likewise, sRNA data for Gram-positive bacteria is available on sRNAdb (Pischimarov et al., 2012). BSRD (Li et al., 2013), sRNATarBase (Wang et al., 2016), sRNAMap (Huang et al., 2009) and RNAInter (Lin et al., 2020) provide sRNA regulatory information for several bacterial species. Despite the influence and importance of these molecules on gene expression, databases integrating sRNA-based and transcriptional regulatory networks are largely missing. To the best of our knowledge, RegulonDB (Santos-Zavaleta et al., 2019), the reference database for *Escherichia coli* GRNs, is the only one to have done this integration though exclusively for *E. coli* K12.

In the context of the *Corynebacterium* genus, CoryneRegNet (Parise et al., 2020) is the reference database for Corynebacterial transcriptional regulatory networks, containing more than 80,000 regulatory interactions but lacking sRNA data. A few Corynebacterial sRNAs can be found in BSRD (Li et al., 2013), Rfam (Bujnicki and Ghosh, 2018) and RNA central (The RNAcentral Consortium, 2019). For *Corynebacterium glutamicium,* the model organism for this genus, 805 sRNAs were experimentally identified using deep sequencing and were reported in (Mentz et al., 2013). However, there are no experimental or predicted sRNA regulations for the *Corynebacterium* genus.

Here, we present the first study about the integration of sRNA regulations with transcriptional regulation in corynebacteria. We predicted sRNAs and their targets for six *Corynebacterium* species of either medical, veterinary or industrial interest, yielding 922 sRNAs and 6,389 sRNA regulatory interactions. This data was integrated into CoryneRegNet 7.5, revealing 754 genes in the GRN to be regulated by both sRNAs and transcription factors and 206 regulatory proteins to be regulated by sRNAs. In a case study of human pathogenic corynebacteria using the CoryneRegNet 7.5 sRNA-enriched database content, we predict the sRNAS *Cd-NCTC13129-sRNA-2* and *scjk1464.1* to form regulatory cascades with TFs. *Cd-NCTC13129-sRNA-2* is predicted to regulate the *ydfH* homologue, indirectly regulating 66 genes in *C. diphtheriae* and *scjk1464.1* is predicted to regulate *mcbR* and *dtxR*, indirectly regulating 35 genes in *C. jeikeium*. In the animal pathogen *C. pseudotuberculosis*, the virulence factor *fagC* is also predicted to be regulated by the sRNA *Cp-1002B-sRNA-1*. To sum up, the integration of sRNAs

and their interactions into the transcriptional regulatory networks in CoryneRegNet provides a more comprehensive view on corynebacterial regulatory mechanisms.

2. Material and Methods

The CoryneRegNet sRNA integration pipeline consists of seven steps: sRNA collection and prediction, homology detection, alignment, sRNA classification, filter, structure prediction and target prediction. An overview of these steps is shown in Figure 1. We started with compiling a dataset of 805 experimentally verified sRNAs from (Mentz et al., 2013) and 70 predicted sRNAs from BSRD (Li et al., 2013). In order to predict novel sRNAs, we used cmsearch (Nawrocki and Eddy, 2013) on the target genomes with no experimental sRNAs publicly available. Details about the sRNA datasets and the genomes used in this analysis are given in Table 1.

Table 1 - The sRNA datasets and target species.

Strain	Accession number	sRNA dataset		
		(Menz et. al, 2013)	BSRD	This study
Corynebacterium diphtheriae NCTC 13129	NC_002935.2		x	x
Corynebacterium efficiens YS-314	NC_004369.1		x	x
Corynebacterium glutamicum ATCC 13032	BX927147.1	x	x	
Corynebacterium jeikeium K411	NC_007164.1		x	x
Corynebacterium ulcerans NCTC7910	NZ_LS483400.1			x
Corynebacterium pseudotuberculosis 1002B	NZ_CP012837.1			x

Afterwards, we identified homologues for every sRNA in the analysis by using GLASSgo (Lott et al., 2018). Homologous sRNAs belonging to the genomes of interest were incorporated into the analysis. For each sRNA in the analysis, we selected its most distant homologues from the same species and from the same genus with ≥80% of similarity. Thus, these sequences were aligned by using clustalo (Sievers et al., 2011). The sRNAs were classified as either functional or non-functional by running RNAz (Gruber et al., 2010) and RNAdetect (Chen et al., 2019) based on the stability and the conservation of the predicted RNA structures as well as on sequence homology. Predicted sRNAs that were classified as non-functional were removed from the analysis. The secondary structure was predicted using RNAalifold (Bernhart et al., 2008) for every sRNA in the analysis. Furthermore, sRNA targets were predicted by running CopraRNA (Wright et al., 2013) with default settings. Adjusted p-values were

calculated using the Beijamini-Hochberg correction from the R package stats, method p.adjust (p.adjust). Then, we selected the fifteen best-ranked interactions predicted with a p-value <0.01, as suggested in (Wright and Georg, 2018). The sRNAs and their targets were integrated into CoryneRegNet (Parise et al., 2020) by updating the front-end and back-end, as well as the database. Finally, we predicted gene ontologies for every gene regulated by sRNAs by running Go Feat (Araujo et al., 2018). A detailed explanation of these methods as well as an example can be seen in the Supplementary Material, section II.

3. Results

3.1. Database Content

We presented CoryneRegNet 7.5, an updated release of the corynebacterial reference database and analysis platform, now including sRNA networks integrated with the transcriptional regulatory networks of the genus *Corynebacterium*. A total of 922 sRNAs and 6,389 regulatory interactions for six corynebacterial strains were integrated into our database, as shown in Table 2. In total, CoryneRegNet release 7.5 now holds 88,657 regulatory interactions, 10,077 regulators and 59,848 regulated genes. The updated database content is publicly available on CoryneRegNet's download page: https://www.exbio.wzw.tum.de/coryneregnet/processToDownload.htm.

Table 2 - New, sRNA-related database content of CoryneRegNet 7.5.

Strain	sRNA		sRNA regulatory interaction
	Experimental	Predicted	Predicted
Corynebacterium diphtheriae NCTC 13129	-	19	176
Corynebacterium efficiens YS-314	-	44	439
Corynebacterium glutamicum ATCC 13032	805	17	5,324
Corynebacterium jeikeium K411	-	27	343
Corynebacterium ulcerans NCTC7910	-	6	65
Corynebacterium pseudotuberculosis 1002B	-	4	42
Total	805	117	6,399

3.2. Website

We updated CoryneRegNet's user interface to present information concerning sRNAs and their targets. Both the regulatory interaction table view and the network view were updated and enriched with corresponding sRNA-related features. The search page now allows the user to (i) search for gene identifiers (Figure 2B) when querying the database for mRNA or sRNA genes (Figure 2A) and (ii) search for a list of genes.

Depending on the search choice (Figure 2A), the user will be directed to the gene-centered or sRNA-centered network view, as presented in Figures 2C and 2D, respectively. sRNAs and their regulatory interactions have been integrated into the network visualization as orange nodes and directed edges. Considering there is no annotation of activation/repression prediction for the sRNA-mRNA interactions, we represent every sRNA regulatory interaction as an orange, directed edge. The complete sRNA-mRNA interactions set of a genome can also be visualized in case no specific gene or sRNA is selected.

In addition, users can now find genes and sRNAs of interest by using the new filtering and sorting features in the table-oriented view, as presented in Supplementary Figures 2A and 2B, respectively. In the sRNA view, we included filters for: (i) sRNAs regulating transcription factors, (ii) sRNAs regulating genes in the TRN and (iii) functional sRNAs. Likewise, in the gene view we included filters for: (i) genes encoding regulatory proteins, (ii) genes regulated by regulatory proteins, (iii) genes regulated by sRNAs and (iv) genes regulated by sRNAs and/or regulatory proteins.

A sample sRNA page is displayed in Figure 3A. It presents essential information of the sRNA of interest such as: type of evidence, position and orientation in the genome, whether or not the sRNA was classified as functional, and the sRNAs' nucleotide sequence. The predicted structure of the selected sRNA is also presented along with its dot plot and alignment graph. The former illustrates the interaction between the nucleotides (Supplementary Figure 2A) and the latter the conservation between the sRNA of interest and its homologous sRNAs (Supplementary Figure 2B). Additionally, the user can visualize the sRNA regulatory interactions in the "Regulates" tab (Figure 3B). This tab shows information regarding each regulatory interaction predicted by CopraRNA (Wright et al., 2013) of the selected sRNA such as its position, minimum energy, hybridization energy and p-value.

Furthermore, we integrated the sRNA interaction network into the statistics section with three new analyses: (i) quantities of sRNA types (Supplementary Figure 2A), (ii) distribution of sRNAs regulating a gene (Supplementary Figure 2B) and (iii) distribution of co-regulating sRNAs (Supplementary Figure 2B). Finally, we updated the documentation and workflow sections at the website accordingly.

3.3. Case Study

We illustrate the utility of the sRNA-enriched CoryneRegNet 7.5 by utilizing the updated filtering features to identify 206 regulatory proteins regulated by sRNAs and 754 genes regulated by both sRNAs and TFs in our six genomes. We selected the genes regulated by both sRNAs and TFs in the following four pathogenic bacteria: *C. diphtheriae* NCTC 13129, *C. jeikeium* K411, *C. pseudotuberculosis* 1002B and *C. ulcerans* NCTC7910. In addition, we selected gene circuits in these pathogenic bacteria and in the model organism *C. glutamicum* and presented whether these observations are conserved in *C. efficiens*. We visualized the regulatory networks of these genes using the list-based network feature in CoryneRegNet 7.5, where we also collected their homologous genes.

In *C. glutamicum*, we predicted 662 genes to be co-regulated by sRNAs and TFs. Amongst them, we can highlight cg0350, *sdhCD*, *acn*, *cgtR3*, *pstA* and the sigma factor *sigA*, as presented in Figure 4A. The sRNA cgb_1195 potentially co-regulates cg0350 (*glxR* homologue) together with four transcriptional regulators: cg2544 (*ydfH* homologue), cg0146 (*sucR* homologue), *sigA* and cg0444 (*ramB* homologue). Additionally, cg0350 has been reported to regulate itself in this organism. The sRNA is predicted to directly and indirectly regulate the highly regulated genes *sdhCD* and *acn*, forming feed forward loop Cg-FF-1 (Figure 4A). These two genes are also part of the dense overlapping regulon Cg-DOR-1, in which three other sRNAs potentially co-regulate them together with five TFs and *sigA*. The membrane anchor subunit *sdhCD* jointly encodes with *sdhA* and *sdhB* the succinate dehydrogenase enzyme, a component of the TCA cycle (Polen et al., 2007; Bussmann et al., 2009). The *acn* gene is also a component of the TCA cycle; it encodes an aconitase enzyme and its inactivation is detrimental to cell growth (Yoon and Woo, 2018). Both the *sdhCD* and *acn* genes were found differentially expressed in acetate medium when compared with glucose medium (Bott, 2007). Figure 4B presents the highly regulated *pstA* as being potentially co-regulated by six sRNAs, two transcription factors and *sigA*. The sRNA cgb_04174 is predicted to directly and indirectly regulate *pstA*, forming the feed forward loop Cg-FF-2. In total, *pstA* is predicted to be directly regulated by six sRNAs and indirectly regulated by eigth sRNAs. This gene is part of the Pst system, which is part of

the inorganic orthophosphate (P_i) starvation stimulon in *C. glutamicum* (Ishige et al., 2003). The transcriptional regulators *sigA*, *cgtR3* and cg0350 are also predicted to be regulated by sRNAs. *SigA* is the primary sigma factor in *C. glutamicum* and is potentially regulated by five sRNAs; this regulator is considered responsible for the transcription of the majority of the housekeeping genes in this organism (Oguiza et al., 1996; Schröder and Tauch, 2010). The global regulator cg0350 (*glxR* homologue) has been reported to be involved in the regulation of 195 genes in *C. glutamicum* (Freyre-González and Tauch, 2017; Parise et al., 2020) and is potentially regulated by one sRNA. The regulator *cgtR3* (*phoR*) is the master regulator of phosphate metabolism in *C. glutamicum* and is potentially regulated by two sRNAs (Schröder and Tauch, 2010). None of the observations mentioned so far is conserved in the other organisms analyzed in this study. Furthermore, *mraZ* is predicted to be regulated by 22 sRNAs, as presented in Figure 4C. This gene is highly conserved in bacteria and is part of the division cell cluster (*dcw*) (Eraso et al., 2014). The cleavage of the coding region of its mRNA is required for efficient cell division in *C. glutamicum* (Maeda et al., 2016). The other genes from the *mraZ* operon, *mraW* and cg2376 (*ftsL* homologue), are potentially regulated by sRNAs. *MraW* is potentially regulated by six sRNAs; amongst them, cgb_03605 is also predicted to regulate *mraZ*. Cg2376 is predicted to be regulated by one sRNA. *MraZ* homologue genes in *C. efficiens*, *C. jeikeium* and *C. pseudotuberculosis* are also potentially regulated by 10 sRNAs, two sRNAs and one sRNA, respectively. In *C. ulcerans,* the *mraW* homologue is potentially regulated by one sRNA, whereas none of the cg2376 homologs are predicted to be regulated by sRNAs in this study.

In *C. diphtheriae* NCTC 13129, we predicted 16 genes to be co-regulated by sRNAs and TFs; the regulatory network of these genes can be seen in Figure 5. Amongst them, the sRNA Cd-NCTC13129-sRNA-2 potentially regulates the transcription factor DIP_RS19435 (*ydfH* homologue), forming a single-input module inside the dense overlapping regulon Cd-DOR-1 (Figure 5). The *ydfH* homologue is predicted to auto-regulate itself and to regulate DIP_RS12895 (*glxR* homologue). It forms a regulatory cascade where the complete set of genes regulated by *glx*R may be indirectly regulated by this sRNA, accounting for 66 genes. The complete regulon of *ydfH* and *glxR* is presented in Supplementary Figure 4. As presented in the dense overlapping regulon Cd-DOR-1 (Figure 5), the GlxR homologue TF potentially

co-regulates four genes with sRNAs: DIP_RS15610 (*ispE* homologue), *gap*, *odhA* and DIP_RS12055. The sRNA Cd-NCTC13129-sRNA-4 potentially regulates both the *ispE* homologue and DIP_RS14355, a methionine ABC transporter substrate-binding. The latter is also regulated by the TetR/AcrR-family regulator DIP_RS23775 (*mcbR* homologue). In C. efficiens, the homologous methionine ABC transporter substrate-binding (CE_RS03295) is also potentially co-regulated by one sRNA (Ce-YS314-sRNA-28) and a TetR/AcrR family TF (CE_RS13790). Also in Cd-DOR-1 (Figure 5), *gap* and *odhA* are predicted to be regulated by the same sRNA, scdi510.1, which also co-regulates *mdh* along with the LuxR family regulator DIP_RS20635 (*ramA* homologue). Likewise, *gap* (cg1791) is also predicted to be co-regulated by cg0350 (*glxR* homologue) and the sRNAs scgl2151.1, cgb_23426 and cgb_10355 in *C. glutamicum*. In general, the genes in Cd-DOR-1 are involved in the TCA cycle and in carbohydrate metabolism.

Also in *C. diphtheriae*, five other genes are potentially co-regulated by both sRNAs and TFs. The hemin-binding protein *hmuT (Draganova et al., 2015)* is potentially co-regulated by scdi175.1 and *dtxR*. The sRNA scdi28.1 is predicted to co-regulate the heat-shock protein GroEL2 along with the transcription factor *hrcA*. In *C. efficiens*, the GroEL2 homologue (CE_RS12690) is also predicted to be regulated by a sRNA (Ce-YS314-sRNA-3) and a *hrcA* homologue (CE_RS10870). In *C. diphtheriae*, Cd-NCTC13129-sRNA1 potentially regulates DIP_RS12535 (*pdxS* homologue) and *pyk*, which are also regulated by DIP_RS18315 (*gatR* homologue) and DIP_RS12530 (*pdxR* homologue), respectively. We also observed the DIP_RS18360 gene (*hflX* homologue) being potentially co-regulated by an XRE family transcriptional regulator and the sRNA scdi1478.1.

In *C. jeikeium* K411, we predicted twenty genes to be jointly regulated by sRNAs and TFs; the regulatory network of these genes is presented in Figure 6. Amongst these genes we identified two dense overlapping regulons, highlighted as Cj-DOR-1 and Cj-DOR-2. In Cj-DOR-1, the sRNAs scjk260.2, scjk885.1, scjk557.1, scjk1019.1 are predicted to co-regulate five genes (*rhtC*, *fadH*, *rpfB*, *cat1* and JK_RS05010) with the global regulator *glxR*. The gene JK_RS05010 (*rpfI* homologue) was predicted to have hydrolase activity and is potentially co-regulated by *glxR*, *mtrA* and scjk577.1. The *rpfI* gene, which encodes a resuscitation-promoting factor interacting protein, is a virulence factor in *C. ulcerans* (Trost et al., 2011). The deletion of this gene impaired the growth of

long-stored cells in *C. glutamicum* (Hartmann et al., 2004). The other resuscitation-promoting factor, *rpfB*, is also potentially regulated by *mtrA*. In *C. efficiens*, the *rpfB* homologue is also potentially co-regulated by the sRNA Ce-YS314-sRNA-12, the *glxR* homologue (CE_RS01675) and the *mtrA* homologue (CE_RS03955). Also in Cj-DOR-1, *metB* and *metX* are potentially co-regulated by *metR* and one sRNA, these genes are involved in the metabolism of methionine in *C. glutamicum* (Rückert et al., 2003). In the single-input module Cj-SIM-1, the sRNA Cj-K411-sRNA2 potentially regulates the transcription factor JK_RS05100 (*sufR* homologue), indirectly regulating the *sufBDCS* gene cluster and the *nif* operon (*nifU*-JK_RS05070). The genes in this circuit are involved in the formation of iron-sulfur clusters in bacteria (Frazzon, 2003; Outten and Wayne Outten, 2015). In *C. efficiens*, the *sufR* homologue (CE_RS08375) is also potentially regulated by two sRNAs (scef1290.1 and scef1536.1) and regulates the *nif* operon (*nifU*-CE_RS08405) as well as the *sufBDCS* gene cluster (CE_RS08400, CE_RS08395, CE_RS08390, CE_RS08385).

Cj-DOR-2 (Figure 6) contains a cluster of 10 sRNAs potentially co-regulating two genes along with the transcription factors TcsR4 and ClgR. When analyzing these sRNAs, we noticed sRNAs scjk2061.1, scjk118.1, scjk463.1, scjk1484.1, scjk1444.1, scjk2091.1, scjk1857.1, scjk1861.1, scjk620.1 and scjk833.1 are identical copies of the same sRNA located in different regions of the genome. The genomic coordinates of these sRNAs are presented in Supplementary Table III. the following regions of the genome: 117083-117197, 462452-462566, 619808-619922, 832580-832694, 1443235-1443349, 1483232-1483346, 1856182-1856296, 1860886-1861000, 2060398-2060.512, 2090313-2090427. The genes potentially regulated by these sRNAs, *clpC* and JK_RS07360, encode a Clp ATPase subunit and a hypothetical protein, respectively. In addition to regulating *clpC*, ClgR is also predicted to co-regulate two other genes with sRNAs, *clpP2* and *clpX.* Both *clpC* and *clpP2* are part of a protein quality control system of the cell along with the other proteolytic subunit *clpP1* (Schröder and Tauch, 2010). *ClpX* is also an ATPase subunit that belongs to the Clp/Hsp100 superfamily, which is involved in stress response, energy metabolism, NADPH synthesis and glucose consumption (Huang et al., 2020). This observation is not conserved amongst the Corynebacterial species analyzed in this manuscript. In Cj-DOR-2, the sRNA scjk1464.1 and *tscR4* potentially co-regulate the sensor histidine kinase *tcsS4,* which belongs to a two-component signal transduction system. These systems are

important to bacteria due to their capacity to detect and adapt to changes in the environment (Pao and Saier, 1995). *TscR4* is also predicted to regulate the copper chaperone JK_RS07345 alongside the sRNA Cj-K411-sRNA-3. Likewise, this sRNA potentially co-regulates the heat shock protein *groES* and the flavin-dependent oxidoreductase JK_RS00955, which are also regulated by the *hrcA* and JK_RS10540 (*maR1* homologue), respectively. GroES is involved in the transport of proteins and in the post-translational folding, along with the heat shock protein GroEL (Rinke et al., 1992). In general, genes in Cj-DOR-2 are potentially involved in growth and cell proliferation.

In *C. jeikeium*, the diphtheria toxin repressor DtxR, regulates many genes associated with iron metabolism and forms the feed forward loop Cj-FFL-1 with the sRNA scjk1464.1 by directly and indirectly regulating *rpsH* (Figure 6). This sRNA is also predicted to directly regulate the transcription factor *mcbR* (Supplementary Figure 5). By potentially regulating *mcbR* and *dtxR*, scjk1464.1 is predicted to indirectly regulate thirty-five genes. Additionally, two other sRNAs (scjk830.1 and scjk1448.1) are predicted to regulate *rpsH*. This gene encodes a 30S ribosomal protein that is associated with the small ribosomal subunit and has been considered as a potential drug target in *C. diphtheriae* (Jamal et al., 2017; Hassan et al., 2018). By analyzing these sRNAs in Rfam, we observed that they do not belong to the same sRNA family. Furthermore, the sRNA scjk1019 is predicted to co-regulate *rhtC* with *glxR* and JK_04405 (*argR* homologue). This gene was used to increase the production of L-threonine in *C. glutamicum (Diesveld et al., 2009)*.

In *C. pseudotuberculosis* 1002B, four genes were predicted to be co-regulated by sRNAs and TFs; the regulatory network of these genes is presented in Figure 7A. The *fagC* (Cp1002B_RS00130) gene is potentially regulated by sRNA Cp-1002B-sRNA-1, as well by the diphtheria toxin repressor (*dtxR*), and is part of the operon *fagABC*. This operon is an active part of the iron acquisition system and is a known virulence factor in *C. pseudotuberculosis* (Billington et al., 2002). Likewise, *fagC* is also potentially regulated by one sRNA (Cu-NCTC7910-sRNA-6) and *dtxR* (CKV68_RS01925) in *C. ulcerans*, as shown in Figure 7B. In *C. pseudotuberculosis*, Cp-1002B-sRNA-1 potentially co-regulates the *azoR* gene along with *marR1*; this gene encodes a flavin mononucleotide (FMN)-dependent homodimeric azobenzene reductase and is involved in the response of oxidative stress. In *C. efficiens*, the a*zoR* homologue (CE_RS08755)

is also potentially regulated by one sRNA (scef1673.1) and the *marR1* homologue (CE_RS06390), whereas in *C. glutamicum,* the a*zoR* homologue (cg1850) is potentially regulated by three sRNAs (cgb_31975, cgb_30915 and scgl2371.1) and the *marR1* homologue (cg1324). In *C. pseudotuberculosis* (Figure 7A), the gene *pfkA* (phosphofructokinase) is predicted to be regulated by Cp-1002B-sRNA-2, *glxR* and Cp1002B_RS04515 (*ramA* homologue). This gene is involved in the reduction of the amount of fructose-6-phosphate during the L-serine fermentation process with sucrose as a carbon resource in *C. glutamicum (Zhang et al., 2017a)*. The *PfkA* homologue in *C. glutamicum* is also potentially regulated by sRNAs and TFs, as presented in Figure 4A. Also in *C. pseudotuberculosis*, Cp-1002B-sRNA-2 also regulates the *recX* gene along with LexA; both *lexA* and *recX* are involved in the bacterial SOS response, acting in DNA damage repair (Resende et al. 2011; Pogson et al. 1996; Jochmann et al. 2009). In *C. glutamicum*, the *recX* homologue (cg2140) is also potentially regulated by two sRNAs (cgb_10545 and cgb_17865) and the *lexA* homologue (cg2114).

In *C. ulcerans* NCTC7910*,* we also predicted other 2 genes to be regulated by sRNAs and TFs; the regulatory network of these genes is presented in Figure 7B. The *pckG* gene, which encodes a phosphoenolpyruvate carboxykinase, was predicted to be regulated by one sRNA and three transcription factors (*glxR*, *ramA* and *ramB*). The transcription factor *DnaK* is regulated by one sRNA and the transcription factor *glnR*. Additionally, it regulates the expression of both genes involved in bacterial adhesion and virulence factors in other bacteria (Hanawa et al., 2002; Gomide et al., 2018). These observations are not conserved in the other genomes analyzed in this study.

4. Discussion

Although several databases on sRNAs and GRNs exist, the integration of these regulatory networks is still a missing point in deciphering gene expression. Several studies have shown the interplay between TFs and sRNAs when regulating gene expression by forming regulatory circuits, as reviewed by (Beisel and Storz, 2010; Nitzan et al., 2017; Brosse and Guillier, 2018). Furthermore, consistency assessments in *E. coli* (Larsen et al., 2019) and *C. glutamicum* (Parise et al., 2021) showed that regulation driven by transcription factors is not able to satisfactorily explain gene

expression and suggested other layers of regulation to be integrated into the networks in order to model the complexity of gene expression. Our work contributes to expanding the regulatory landscape of two biotechnological and four pathogenic *Corynebacterium* species by predicting their sRNA regulatory networks and by integrating them into the corresponding GRNs.

Regarding sRNA prediction, we searched for (i) sRNA homologous of the experimentally validated ones from (Mentz et al., 2013) using GLASSgo (Lott et al., 2018) and (ii) novel sRNAs belonging to known sRNA families from Rfam (Kalvari et al., 2021) using cmsearch (Nawrocki and Eddy, 2013). The former uses iterative blast search, pairwise identity filtering and graph-based clustering based on secondary structures to find sRNA homologous (Lott et al., 2018). It allows us to search for homologous sRNAs not belonging to a specific sRNA family. Meanwhile, cmsearch allows us to use covariance models to search for novel members of curated sRNA families from Rfam. Cmsearch has been considered the most specific and sensitive sRNA homology tool (Freyhult et al., 2007; Lott et al., 2018) and GLASSgo presented results comparable to cmsearch in a recent benchmark (Lott et al., 2018). RNAz and RNAdetect identify functional sRNA candidates amongst the ones predicted by GLASSgo and cmsearch, yielding strong candidates for further investigation as well as target prediction (Gruber et al., 2010; Backofen et al., 2018; Chen et al., 2019). Regarding the sRNA target prediction, CopraRNA is currently considered the best bacterial sRNA-mRNA interaction prediction software (Pain et al., 2015; Georg et al., 2020). It constructs a combined prediction based on the conservation of sRNA interactions across a given set of organisms, which significantly decreases the false positive rate (Wright et al., 2013; Backofen et al., 2018). In order to maximize the reliability of our regulatory interactions, we selected the most dissimilar sRNA homologues from the same genus and from the same species predicted by GLASSgo with more than 80% of similarity for the sRNA interaction prediction with CopraRNA (Wright et al., 2013). This procedure increases the chances of our regulatory interactions to be true because they will be conserved on a genus- or species-level.

This, along with the filtering of the fifteen best-ranked CopraRNA predictions with p-value < 0.01 makes our conservative predictions yielding strong candidates for hypothesis generation and future experimental assay design. Even though these predicted regulatory interactions can either activate or repress the mRNA expression, we provide no functional annotation for them.

By applying our GRN sRNA-enrichment pipeline, we identified TFs, sRNAs and sigma factors jointly forming regulatory circuits in the regulatory networks. We were able to identify feed forward loops, single input modules and dense overlapping regulons. With no information on TFs regulating sRNAs, feedback loops were not possible to be identified for these networks. Furthermore, we presented the occurrences in which the co-regulation by sRNAs and TFs were also observed in other studied organisms. We highlighted genes in regulatory circuits involved in the following pathways: methionine biosynthesis and metabolism of cofactors and vitamins in *C. jeikeium;* TCA cycle and carbohydrate metabolism in *C. diphtheriae*; and TCA cycle, phosphate metabolism and cell division in *C. glutamicum*.

In our gene ontology analysis, ATP-binding is the molecular process with the most amount of genes potentially regulated by sRNAs in all studied organisms. This is not surprising, given the immense importance of ATP for the survival, growth and replication of all living organisms. In bacteria, ATP is associated with virulence factors and can even regulate virulence genes, e.g. the *mgtC* gene in *Salmonella* (Klein and Lewinson, 2011; Lee and Groisman, 2012; Mempin et al., 2013). Besides that, the other molecular processes with which most genes are associated are DNA binding and Metal ion binding, showing a probable strong influence of sRNA in these molecular functions. In *C. diphtheria* NCTC 13129, the sRNA Cd-NCTC13129-sRNA-2 potentially regulates the transcription factor *ydfH*, which regulates the global regulator *glxR*. Additionally, it is the regulator with the largest amount of regulations known in the *Corynebacterium* species. Likewise, in *C. jeikeium,* the sRNA scjk1464.1 regulates the transcription factors *dtxR* and *mcbR*. *DtxR* is the master regulator of iron metabolism in *C. glutamicum* (Wennerhold and Bott, 2006; Schröder and Tauch, 2010) and the TetR

family regulator *mcbR* is involved in biofilm formation in *E. coli* (Zhang et al., 2008). Note that in *C. glutamicum* cg0350 (*glxR* homologue) is potentially regulated by the sRNA cgb_1195 and forms a feed forward loop together with this sRNA, *sdhCD* and *acn*.

Amongst the genes potentially regulated by sRNAs, note the virulence factor *fagC* in *C. pseudotuberculosis*, the candidate virulence factor *rpfI* in *C. ulcerans* and the potential drug target *rpsH* in *C. diphtheriae*. We also observed the heat shock protein GroEL and the histidine kinase TcsS4 being regulated by sRNAs in *C. jeikeium*. While heat shock proteins are essential for bacterial survival and were recently associated with virulence and drug resistance (Neckers and Tatu, 2008), two-component systems are known as regulators of virulence factors and genes related to adhesion, pilus formation and drug resistance (López-Goñi et al., 2002; Matsushita and Janda, 2002; Tiwari et al., 2014). Moreover, the genes related to survival and adaptation in the *nif* operon and in the *suf* gene cluster (Stock et al., 1989; Huet et al., 2005) are regulated by the same sRNA and transcription factor in *C. jeikeium*. Genes of biotechnological interest, such as *pfkA in C. pseudotuberculosis, rhtC* in *C. jeikeium*, and *pyk* in *C. diphtheriae,* were also pointed out as sRNA targets. These genes are associated with L-threonine production, L-serine fermentation and lactic acid production in *C. glutamicum*, respectively. These molecules are largely used in the food industry (Diesveld et al., 2009; Chai et al., 2016; Zhang et al., 2017a). The presented regulations show the potential of sRNAs to regulate genes of medical, veterinary and biotechnological interest in corynebacterial species.

5. Conclusion

We introduce the sRNA regulatory networks integrated with the transcriptional gene regulatory networks of *C. glutamicum*, *C. pseudotuberculosis*, *C. ulcerans*, *C. diphtheriae*, *C. jeikeium* and *C. efficiens*. This integration allowed us to identify sRNAs and TFs forming generalizable patterns, such as feed forward loops, dense overlapping regulons and single-input modules. It indicates sRNAs and TFs jointly orchestrating the regulation of corynebacterial gene expression, suggesting that sRNAs may have a great impact in modeling the gene expression of important biological processes in corynebacteria. Our results suggest several genes for further experimental investigation

in the studied organisms. Amongst them, note the potential regulation of *mraZ,* which is conserved in four organisms of this study, and of the virulence factor *fagC*, which is potentially regulated by *dtxR* and one sRNA in both *C. pseudotuberculosis* and *C. ulcerans*. We believe that with CoryneRegNet 7.5, in which we implemented the integrated networks with extended visualization and querying functionality, we move an additional step towards understanding the corynebacterial regulatory mechanisms and provide new starting points to guide future experimental assays to comprehend the regulatory mechanisms underlying pathogenicity, survival, adaptation and amino acid production in the *Corynebacterium* genus.

Funding

JB is grateful for support from H2020 grant RepoTrial (nr. 777111) and his VILLUM Young Investigator grant (nr. 13154). DP received support from CAPES (nr. 88887.364607/2019-00) and MP from CNPq (nr. 201336/2018-9), for their work at TUM in Germany. MP's work was also supported by the German Research Foundation (under SFB924). VA is grateful for support from his CNPq Research Productivity grant (nr. 305093/2015-0), CNPq Universal grant (nr. 405233/2016-7) and FAPEMIG grant (nr. APQ 02600-17). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Author Contributions

MP, MR, RB, VA and JB conceptualized this work. MP and DP developed the software and wrote the manuscript. MP performed the analysis. VA, RK and JB supervised the work. MR, RB, RK, FA, AG VA and JB reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Ahmed, W., Hafeez, M. A., and Mahmood, S. (2018). Identification and functional characterization of bacterial small non-coding RNAs and their target: A review. *Gene Reports* 10, 167–176. doi:10.1016/j.genrep.2018.01.001.

Altuvia, S. (2007). Identification of bacterial small non-coding RNAs: experimental approaches. *Curr. Opin. Microbiol.* 10, 257–261. doi:10.1016/j.mib.2007.05.003.

Araujo, F. A., Barh, D., Silva, A., Guimarães, L., and Ramos, R. T. J. (2018). GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Sci. Rep.* 8, 1794. doi:10.1038/s41598-018-20211-9.

Arrieta-Ortiz, M. L., Hafemeister, C., Shuster, B., Baliga, N. S., Bonneau, R., and Eichenberger, P. (2020). Inference of Bacterial Small RNA Regulatory Networks and Integration with Transcription Factor-Driven Regulatory Networks. *mSystems* 5. doi:10.1128/mSystems.00057-20.

Backofen, R., Gorodkin, J., Hofacker, I. L., and Stadler, P. F. (2018). Comparative RNA Genomics. *Comparative Genomics*, 363–400. doi:10.1007/978-1-4939-7463-4_14.

Barquist, L., and Vogel, J. (2015). Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annu. Rev. Genet.* 49, 367–394. doi:10.1146/annurev-genet-112414-054804.

Beisel, C. L., and Storz, G. (2010). Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.* 34, 866–882. doi:10.1111/j.1574-6976.2010.00241.x.

Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., and Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9, 474. doi:10.1186/1471-2105-9-474.

Billington, S. J., Esmay, P. A., Songer, J. G., and Jost, B. H. (2002). Identification and role in virulence of putative iron acquisition genes from Corynebacterium pseudotuberculosis. *FEMS Microbiol. Lett.* 208, 41–45. doi:10.1111/j.1574-6968.2002.tb11058.x.

Bott, M. (2007). Offering surprises: TCA cycle regulation in Corynebacterium glutamicum. *Trends Microbiol.* 15, 417–425. doi:10.1016/j.tim.2007.08.004.

Brosse, A., and Guillier, M. (2018). Bacterial Small RNAs in Mixed Regulatory Networks.

Microbiol Spectr 6. doi:10.1128/microbiolspec.RWR-0014-2017.

Bujnicki, J., and Ghosh, P. (2018). Faculty Opinions recommendation of Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. doi:10.3410/f.732083510.793543310.

Bussmann, M., Emer, D., Hasenbein, S., Degraf, S., Eikmanns, B. J., and Bott, M. (2009). Transcriptional control of the succinate dehydrogenase operon sdhCAB of Corynebacterium glutamicum by the cAMP-dependent regulator GlxR and the LuxR-type regulator RamA. *J. Biotechnol.* 143, 173–182. doi:10.1016/j.jbiotec.2009.06.025.

Chai, X., Shang, X., Zhang, Y., Liu, S., Liang, Y., Zhang, Y., et al. (2016). A novel pyruvate kinase and its application in lactic acid production under oxygen deprivation in Corynebacterium glutamicum. *BMC Biotechnol.* 16, 79. doi:10.1186/s12896-016-0313-6.

Chen, C.-C., Qian, X., and Yoon, B.-J. (2019). RNAdetect: efficient computational detection of novel non-coding RNAs. *Bioinformatics* 35, 1133–1141. doi:10.1093/bioinformatics/bty765.

De Lay, N., Schu, D. J., and Gottesman, S. (2013). Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J. Biol. Chem.* 288, 7996–8003. doi:10.1074/jbc.R112.441386.

Diallo, I., and Provost, P. (2020). RNA-Sequencing Analyses of Small Bacterial RNAs and their Emergence as Virulence Factors in Host-Pathogen Interactions. *Int. J. Mol. Sci.* 21. doi:10.3390/ijms21051627.

Diesveld, R., Tietze, N., Fürst, O., Reth, A., Bathe, B., Sahm, H., et al. (2009). Activity of exporters of Escherichia coli in Corynebacterium glutamicum, and their use to increase L-threonine production. *J. Mol. Microbiol. Biotechnol.* 16, 198–207. doi:10.1159/000142530.

Draganova, E. B., Akbas, N., Adrian, S. A., Lukat-Rodgers, G. S., Collins, D. P., Dawson, J. H., et al. (2015). Heme Binding by Corynebacterium diphtheriae HmuT: Function and Heme Environment. *Biochemistry* 54, 6598–6609. doi:10.1021/acs.biochem.5b00666.

Eraso, J. M., Markillie, L. M., Mitchell, H. D., Taylor, R. C., Orr, G., and Margolin, W. (2014). The highly conserved MraZ protein is a transcriptional regulator in Escherichia coli. *J. Bacteriol.* 196, 2053–2066. doi:10.1128/JB.01370-13.

Faner, M. A., and Feig, A. L. (2013). Identifying and characterizing Hfq-RNA interactions. *Methods* 63, 144–159. doi:10.1016/j.ymeth.2013.04.023.

Frazzon, J. (2003). Formation of iron–sulfur clusters in bacteria: an emerging field in bioinorganic chemistry. *Current Opinion in Chemical Biology* 7, 166–173. doi:10.1016/s1367-5931(03)00021-8.

Freyhult, E. K., Bollback, J. P., and Gardner, P. P. (2007). Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA. *Genome Res.* 17, 117–125. doi:10.1101/gr.5890907.

Freyre-González, J. A., and Tauch, A. (2017). Functional architecture and global properties of the Corynebacterium glutamicum regulatory network: Novel insights from a dataset with a high genomic coverage. *J. Biotechnol.* 257, 199–210. doi:10.1016/j.jbiotec.2016.10.025.

Georg, J., Lalaouna, D., Hou, S., Lott, S. C., Caldelari, I., Marzi, S., et al. (2020). The power of cooperation: Experimental and computational approaches in the functional characterization of bacterial sRNAs. *Mol. Microbiol.* 113, 603–612. doi:10.1111/mmi.14420.

Gomide, A. C. P., de Sá, P. G., Cavalcante, A. L. Q., de Jesus Sousa, T., Gomes, L. G. R., Ramos, R. T. J., et al. (2018). Heat shock stress: Profile of differential expression in Corynebacterium pseudotuberculosis biovar Equi. *Gene* 645, 124–130. doi:10.1016/j.gene.2017.12.015.

Gripenland, J., Netterling, S., Loh, E., Tiensuu, T., Toledo-Arana, A., and Johansson, J. (2010). RNAs: regulators of bacterial virulence. *Nat. Rev. Microbiol.* 8, 857–866. doi:10.1038/nrmicro2457.

Gruber, A. R., Findeiß, S., Washietl, S., Hofacker, I. L., and Stadler, P. F. (2010). RNAz 2.0: improved noncoding RNA detection. *Pac. Symp. Biocomput.*, 69–79. Available at: https://www.ncbi.nlm.nih.gov/pubmed/19908359.

Hanawa, T., Yamanishi, S., Murayama, S., Yamamoto, T., and Kamiya, S. (2002). Participation of DnaK in expression of genes involved in virulence of Listeria monocytogenes. *FEMS Microbiol. Lett.* 214, 69–75. doi:10.1111/j.1574-6968.2002.tb11326.x.

Han, K., Tjaden, B., and Lory, S. (2016). GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. *Nat Microbiol* 2, 16239. doi:10.1038/nmicrobiol.2016.239.

Hartmann, M., Barsch, A., Niehaus, K., Pühler, A., Tauch, A., and Kalinowski, J. (2004). The glycosylated cell surface protein Rpf2, containing a resuscitation-promoting factor motif, is involved in intercellular communication of Corynebacterium glutamicum. *Arch. Microbiol.* 182, 299–312. doi:10.1007/s00203-004-0713-1.

Hassan, S. S., Jamal, S. B., Radusky, L. G., Tiwari, S., Ullah, A., Ali, J., et al. (2018). The Druggable Pocketome of Corynebacterium diphtheriae: A New Approach for in

silico Putative Druggable Targets. *Frontiers in Genetics* 9. doi:10.3389/fgene.2018.00044.

Huang, H.-Y., Chang, H.-Y., Chou, C.-H., Tseng, C.-P., Ho, S.-Y., Yang, C.-D., et al. (2009). sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic Acids Res.* 37, D150–4. doi:10.1093/nar/gkn852.

Huang, M., Zhao, Y., Feng, L., Zhu, L., Zhan, L., and Chen, X. (2020). Role of the ClpX from Corynebacterium crenatum involved in stress responses and energy metabolism. *Appl. Microbiol. Biotechnol.* 104, 5505–5517. doi:10.1007/s00253-020-10597-w.

Huet, G., Daffé, M., and Saves, I. (2005). Identification of the Mycobacterium tuberculosis SUF Machinery as the Exclusive Mycobacterial System of [Fe-S] Cluster Assembly: Evidence for Its Implication in the Pathogen's Survival. *Journal of Bacteriology* 187, 6137–6146. doi:10.1128/jb.187.17.6137-6146.2005.

Ishige, T., Krause, M., Bott, M., Wendisch, V. F., and Sahm, H. (2003). The phosphate starvation stimulon of Corynebacterium glutamicum determined by DNA microarray analyses. *J. Bacteriol.* 185, 4519–4529. doi:10.1128/jb.185.15.4519-4529.2003.

Jamal, S. B., Hassan, S. S., Tiwari, S., Viana, M. V., de Jesus Benevides, L., Ullah, A., et al. (2017). An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae. *PLOS ONE* 12, e0186401. doi:10.1371/journal.pone.0186401.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., et al. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46, D335–D342. doi:10.1093/nar/gkx1038.

Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* 49, D192–D200. doi:10.1093/nar/gkaa1047.

Klein, J. S., and Lewinson, O. (2011). Bacterial ATP-driven transporters of transition metals: physiological roles, mechanisms of action, and roles in bacterial virulence. *Metallomics* 3, 1098–1108. doi:10.1039/c1mt00073j.

Lalaouna, D., and Massé, E. (2015). Identification of sRNA interacting with a transcript of interest using MS2-affinity purification coupled with RNA sequencing (MAPS) technology. *Genom Data* 5, 136–138. doi:10.1016/j.gdata.2015.05.033.

Larsen, S. J., Röttger, R., Schmidt, H. H. H. W., and Baumbach, J. (2019). E. coli gene regulatory networks are inconsistent with gene expression data. *Nucleic Acids Res.* 47, 85–92. doi:10.1093/nar/gky1176.

Lee, E.-J., and Groisman, E. A. (2012). Control of a Salmonella virulence locus by an ATP-sensing leader messenger RNA. *Nature* 486, 271–275. doi:10.1038/nature11090.

Lee, H.-J., and Gottesman, S. (2016). sRNA roles in regulating transcriptional regulators: Lrp and SoxS regulation by sRNAs. *Nucleic Acids Res.* 44, 6907–6923. doi:10.1093/nar/gkw358.

Li, L., Huang, D., Cheung, M. K., Nong, W., Huang, Q., and Kwan, H. S. (2013). BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res.* 41, D233–8. doi:10.1093/nar/gks1264.

Lin, Y., Liu, T., Cui, T., Wang, Z., Zhang, Y., Tan, P., et al. (2020). RNAInter in 2020: RNA interactome repository with increased coverage and annotation. *Nucleic Acids Res.* 48, D189–D197. doi:10.1093/nar/gkz804.

López-Goñi, I., Guzmán-Verri, C., Manterola, L., Sola-Landa, A., Moriyón, I., and Moreno, E. (2002). Regulation of Brucella virulence by the two-component system BvrR/BvrS. *Vet. Microbiol.* 90, 329–339. doi:10.1016/S0378-1135(02)00218-3.

Lott, S. C., Schäfer, R. A., Mann, M., Backofen, R., Hess, W. R., Voß, B., et al. (2018). GLASSgo - Automated and Reliable Detection of sRNA Homologs From a Single Input Sequence. *Front. Genet.* 9, 124. doi:10.3389/fgene.2018.00124.

Maeda, T., Tanaka, Y., Takemoto, N., Hamamoto, N., and Inui, M. (2016). RNase III mediated cleavage of the coding region of mraZ mRNA is required for efficient cell division in Corynebacterium glutamicum. *Mol. Microbiol.* 99, 1149–1166. doi:10.1111/mmi.13295.

Mandin, P., Chareyre, S., and Barras, F. (2016). A Regulatory Circuit Composed of a Transcription Factor, IscR, and a Regulatory RNA, RyhB, Controls Fe-S Cluster Delivery. *mBio* 7. doi:10.1128/mbio.00966-16.

Massé, E., Vanderpool, C. K., and Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in Escherichia coli. *J. Bacteriol.* 187, 6962–6971. doi:10.1128/JB.187.20.6962-6971.2005.

Matsushita, M., and Janda, K. D. (2002). Histidine kinases as targets for new antimicrobial agents. *Bioorg. Med. Chem.* 10, 855–867. doi:10.1016/s0968-0896(01)00355-8.

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y. E., Reiss, N., Bar, A., et al. (2016). Global Mapping of Small RNA-Target Interactions in Bacteria. *Molecular Cell* 63, 884–897. doi:10.1016/j.molcel.2016.07.026.

Mempin, R., Tran, H., Chen, C., Gong, H., Kim Ho, K., and Lu, S. (2013). Release of extracellular ATP by bacteria during growth. *BMC Microbiol.* 13, 301.

doi:10.1186/1471-2180-13-301.

Mentz, A., Neshat, A., Pfeifer-Sancar, K., Pühler, A., Rückert, C., and Kalinowski, J. (2013). Comprehensive discovery and characterization of small RNAs in Corynebacterium glutamicum ATCC 13032. *BMC Genomics* 14, 714. doi:10.1186/1471-2164-14-714.

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi:10.1093/bioinformatics/btt509.

Neckers, L., and Tatu, U. (2008). Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell Host Microbe* 4, 519–527. doi:10.1016/j.chom.2008.10.011.

Nitzan, M., Rehani, R., and Margalit, H. (2017). Integration of Bacterial Small RNAs in Regulatory Networks. *Annu. Rev. Biophys.* 46, 131–148. doi:10.1146/annurev-biophys-070816-034058.

Oguiza, J. A., Marcos, A. T., Malumbres, M., and Martín, J. F. (1996). Multiple sigma factor genes in Brevibacterium lactofermentum: characterization of sigA and sigB. *J. Bacteriol.* 178, 550–553. doi:10.1128/jb.178.2.550-553.1996.

Outten, F. W., and Wayne Outten, F. (2015). Recent advances in the Suf Fe–S cluster biogenesis pathway: Beyond the Proteobacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1853, 1464–1469. doi:10.1016/j.bbamcr.2014.11.001.

p.adjust Available at: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/p.adjust [Accessed August 16, 2020].

Pain, A., Ott, A., Amine, H., Rochat, T., Bouloc, P., and Gautheret, D. (2015). An assessment of bacterial small RNA target prediction programs. *RNA Biol.* 12, 509–513. doi:10.1080/15476286.2015.1020269.

Pao, G. M., and Saier, M. H., Jr (1995). Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. *J. Mol. Evol.* 40, 136–154. doi:10.1007/BF00167109.

Papenfort, K., and Vanderpool, C. K. (2015). Target activation by regulatory RNAs in bacteria. *FEMS Microbiol. Rev.* 39, 362–378. doi:10.1093/femsre/fuv016.

Parise, D., Parise, M. T. D., Kataka, E., Kato, R. B., List, M., Tauch, A., et al. (2021). On the Consistency between Gene Expression and the Gene Regulatory Network of Corynebacterium glutamicum. *Network and Systems Medicine* 4, 51–59. doi:10.1089/nsm.2020.0014.

Parise, M. T. D., Parise, D., Kato, R. B., Pauling, J. K., Tauch, A., Azevedo, V. A. de C., et al. (2020). CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. *Sci Data* 7, 142. doi:10.1038/s41597-020-0484-9.

Pischimarov, J., Kuenne, C., Billion, A., Hemberger, J., Cemič, F., Chakraborty, T., et al. (2012). sRNAdb: a small non-coding RNA database for gram-positive bacteria. *BMC Genomics* 13, 384. doi:10.1186/1471-2164-13-384.

Polen, T., Schluesener, D., Poetsch, A., Bott, M., and Wendisch, V. F. (2007). Characterization of citrate utilization in Corynebacterium glutamicum by transcriptome and proteome analysis. *FEMS Microbiol. Lett.* 273, 109–119. doi:10.1111/j.1574-6968.2007.00793.x.

Rinke, T. F., Bekelie, S., Osland, A., Miko, T. L., Hermans, P. W. M., Soolingen, D., et al. (1992). Mycobacteria contain two groEL genes: the second Mycobacterium leprae groEL gene is arranged in an operon with groES. *Molecular Microbiology* 6, 1995–2007. doi:10.1111/j.1365-2958.1992.tb01372.x.

Rückert, C., Pühler, A., and Kalinowski, J. (2003). Genome-wide analysis of the L-methionine biosynthetic pathway in Corynebacterium glutamicum by targeted gene deletion and homologous complementation. *J. Biotechnol.* 104, 213–228. doi:10.1016/s0168-1656(03)00158-5.

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., et al. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.* 47, D212–D220. doi:10.1093/nar/gky1077.

Schröder, J., and Tauch, A. (2010). Transcriptional regulation of gene expression inCorynebacterium glutamicum: the role of global, master and local regulators in the modular and hierarchical gene regulatory network. *FEMS Microbiology Reviews* 34, 685–737. doi:10.1111/j.1574-6976.2010.00228.x.

Sharma, C. M., and Vogel, J. (2009). Experimental approaches for the discovery and characterization of regulatory small RNA. *Curr. Opin. Microbiol.* 12, 536–546. doi:10.1016/j.mib.2009.07.006.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., et al. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* 4, e1000163. doi:10.1371/journal.pgen.1000163.

Stock, J. B., Ninfa, A. J., and Stock, A. M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiological Reviews* 53, 450–490. doi:10.1128/mmbr.53.4.450-490.1989.

The RNAcentral Consortium (2019). RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.* 47, D1250–D1251. doi:10.1093/nar/gky1206.

Tiwari, S., da Costa, M. P., Almeida, S., Hassan, S. S., Jamal, S. B., Oliveira, A., et al. (2014). C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds. *Integr. Biol.* 6, 1088–1099. doi:10.1039/c4ib00140k.

Trost, E., Al-Dilaimi, A., Papavasiliou, P., Schneider, J., Viehoever, P., Burkovski, A., et al. (2011). Comparative analysis of two complete Corynebacterium ulcerans genomes and detection of candidate virulence factors. *BMC Genomics* 12, 383. doi:10.1186/1471-2164-12-383.

Wang, J., Liu, T., Zhao, B., Lu, Q., Wang, Z., Cao, Y., et al. (2016). sRNATarBase 3.0: an updated database for sRNA-target interactions in bacteria. *Nucleic Acids Res.* 44, D248–53. doi:10.1093/nar/gkv1127.

Waters, L. S., and Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136, 615–628. doi:10.1016/j.cell.2009.01.043.

Waters, S. A., McAteer, S. P., Kudla, G., Pang, I., Deshpande, N. P., Amos, T. G., et al. (2017). Small RNA interactome of pathogenic E. coli revealed through crosslinking of RN ase E. *The EMBO Journal* 36, 374–387. doi:10.15252/embj.201694639.

Wennerhold, J., and Bott, M. (2006). The DtxR regulon of Corynebacterium glutamicum. *J. Bacteriol.* 188, 2907–2918. doi:10.1128/JB.188.8.2907-2918.2006.

Wright, P. R., and Georg, J. (2018). Workflow for a Computational Analysis of an sRNA Candidate in Bacteria. *Methods Mol. Biol.* 1737, 3–30. doi:10.1007/978-1-4939-7634-8_1.

Wright, P. R., Richter, A. S., Papenfort, K., Mann, M., Vogel, J., Hess, W. R., et al. (2013). Comparative genomics boosts target prediction for bacterial small RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 110, E3487–96. doi:10.1073/pnas.1303248110.

Wu, T. P., Ruan, K. C., and Liu, W. Y. (1996). A fluorescence-labeling method for sequencing small RNA on polyacrylamide gel. *Nucleic Acids Res.* 24, 3472–3473. doi:10.1093/nar/24.17.3472.

Yoon, J., and Woo, H. M. (2018). CRISPR interference-mediated metabolic engineering of Corynebacterium glutamicum for homo-butyrate production. *Biotechnol. Bioeng.* 115, 2067–2074. doi:10.1002/bit.26720.

Zhang, X.-S., García-Contreras, R., and Wood, T. K. (2008). Escherichia coli

transcription factor YncC (McbR) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein YbiM (McbA). *ISME J.* 2, 615–631. doi:10.1038/ismej.2008.24.

Zhang, X., Yao, L., Xu, G., Zhu, J., Zhang, X., Shi, J., et al. (2017a). Enhancement of fructose utilization from sucrose in the cell for improved l-serine production in engineered Corynebacterium glutamicum. *Biochem. Eng. J.* 118, 113–122. doi:10.1016/j.bej.2016.11.021.

Zhang, Y., Huang, H., Zhang, D., Qiu, J., Yang, J., Wang, K., et al. (2017b). A Review on Recent Computational Methods for Predicting Noncoding RNAs. *BioMed Research International* 2017, 1–14. doi:10.1155/2017/9139504.

Figure Captions:

Figure 1 - Overview of the sRNA data integration workflow.

Figure 2 - CoryneRegNet's front-end updates in **(A+B)** search page and **(C+D)** in the network visualization. **(A)** The search page of CoryneRegNet's database allows for choosing between searching for mRNA genes or sRNA genes while **(B)** guiding the search with gene or sRNA identifiers. **(C)** Direct regulations of cg0012 and **(D)** genes regulated by cgb_07555. In the network, green nodes represent activator proteins, red nodes represent repressor proteins, blue nodes represent dual regulators (i.e. that can activate and repress gene expression), orange nodes represent sRNAs and grey nodes represent target genes. The arrows represent the regulatory interactions and their colors represent the same roles as in the nodes.

Figure 3 - CoryneRegNet 7.5's sRNA details page with **(A)** essential information of the sRNA cgb_07555 and **(B)** its regulations.

Figure 4 - *C. glutamicum*'s predicted sRNA-enriched regulons. (A) *sdhCC* and *acn* co-regulated by TFs and sRNAs and forming two regulatory circuits, Cg-DOR-1 and Cg-FF-1. (B) pstA being directly and indirectly regulated by TFs and sRNAs, forming the regulatory circuit Cg-FF-2. (C) marZ being regulated by 22 sRNAs. In the networks, green nodes represent activator proteins, red nodes represent repressor proteins, blue nodes represent dual regulators (i.e. that can activate and repress gene expression), orange nodes represent sRNAs and grey nodes represent target genes. The arrows

represent the regulatory interactions and their colors represent the same roles as the ones in the nodes.

Figure 5 - Genes regulated by sRNAs and regulatory proteins in *C. diphtheriae* NCTC 13129. In the network, green nodes represent activator proteins, red nodes represent repressor proteins, blue nodes represent dual regulators (i.e. that can activate and repress gene expression), orange nodes represent sRNAs and grey nodes represent target genes. The arrows represent the regulatory interactions and their colors represent the same roles as the ones in the nodes.

Figure 6 - Genes regulated by sRNAs and regulatory proteins in *C. jeikeium* K411. In the network, green nodes represent activator proteins, red nodes represent repressor proteins, blue nodes represent dual regulators (i.e. that can activate and repress gene expression), orange nodes represent sRNAs and grey nodes represent target genes. The arrows represent the regulatory interactions and their colors represent the same roles as the ones in the nodes.

Figure 7 - Genes regulated by sRNAs and regulatory proteins in *C. pseudotuberculosis* 1002B (a) and in *C. ulcerans* (b). In the network, green nodes represent activator proteins, red nodes represent repressor proteins, blue nodes represent dual regulators (i.e. that can activate and repress gene expression), orange nodes represent sRNAs and grey nodes represent target genes. The arrows represent the regulatory interactions and its colors represent the same roles as the ones in the nodes.

Figure 1.TIF

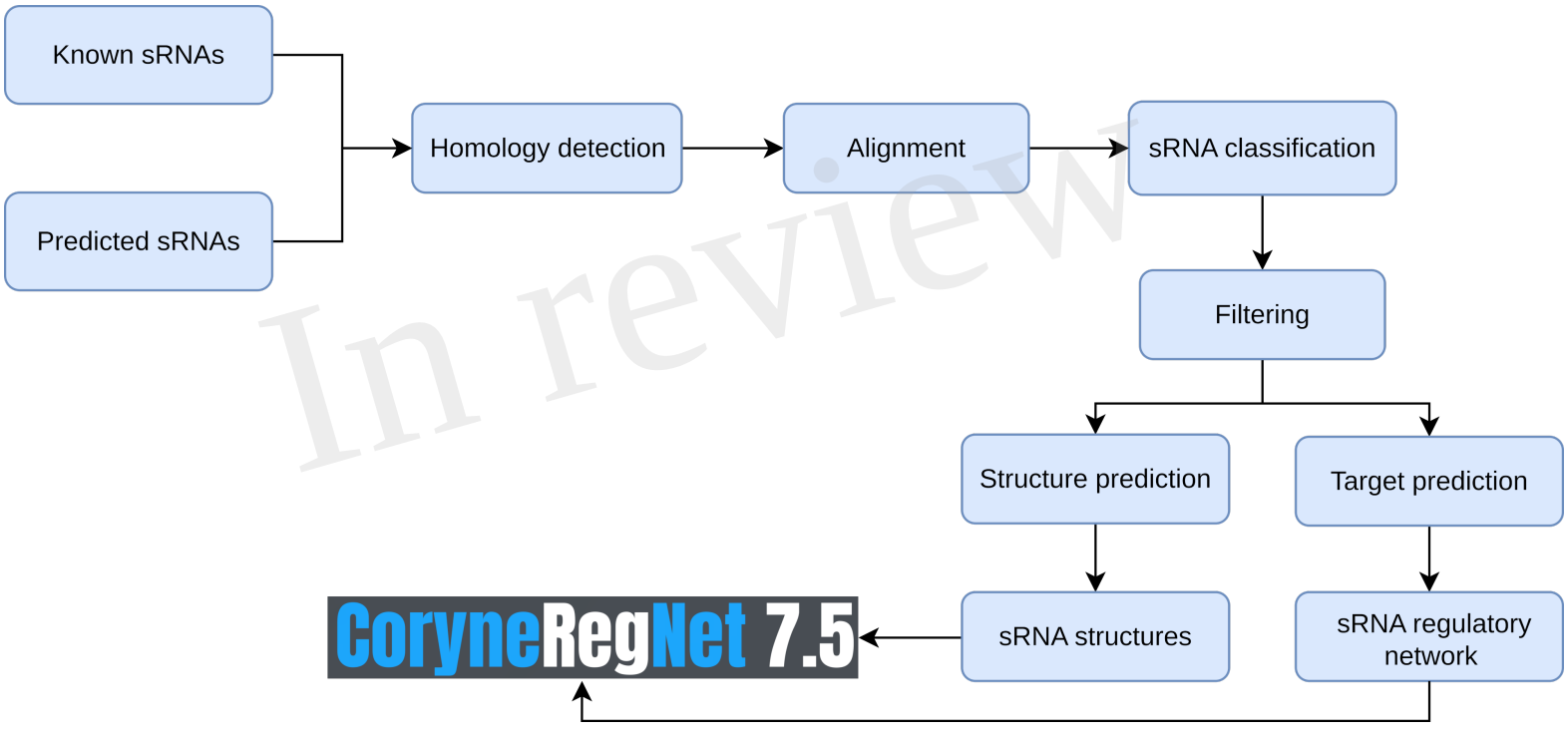


Figure 2.JPEG

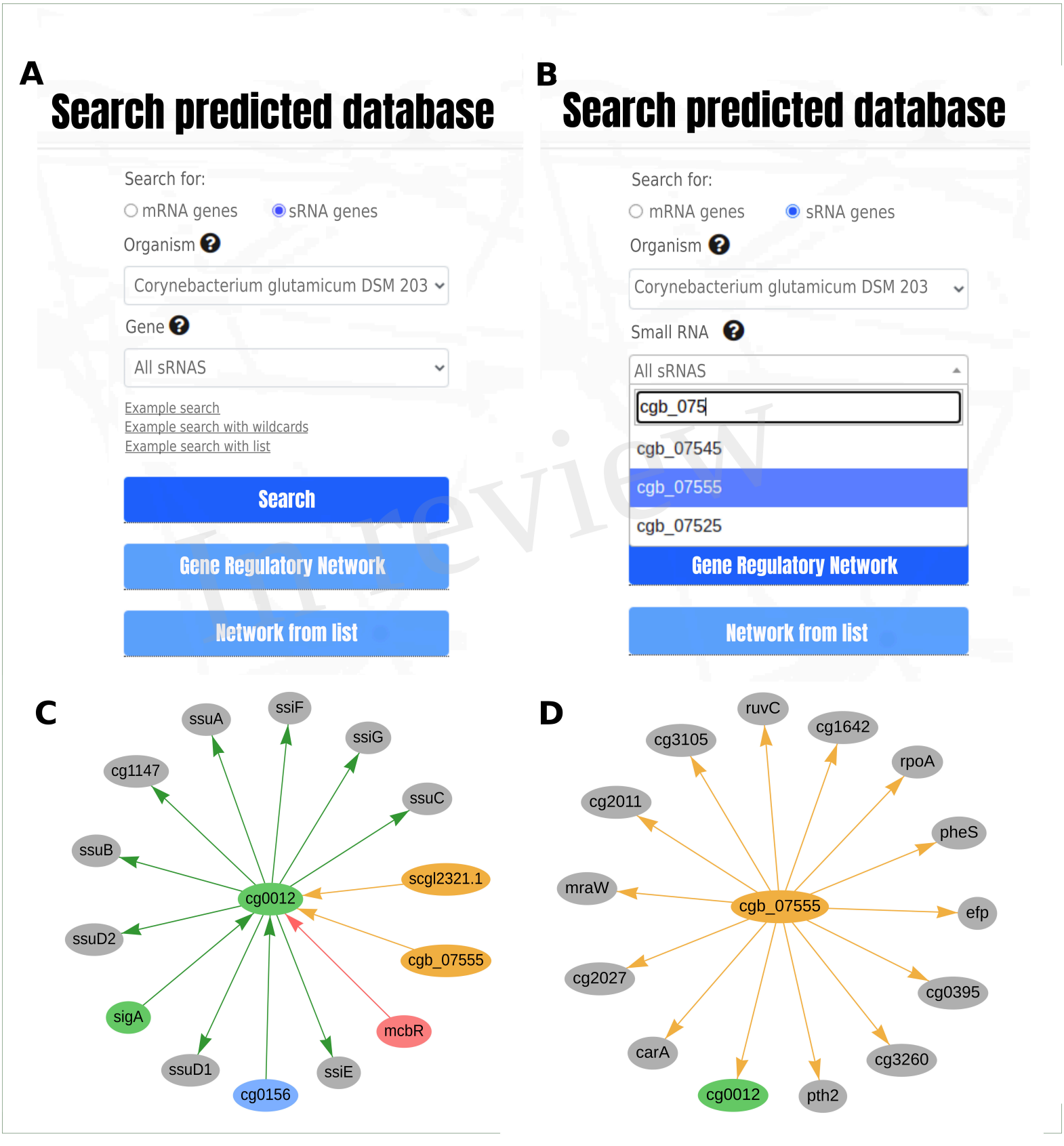


Figure 3.JPEG

A

RNA: cgb_07555 | Regulates:

Zoom structure

Locus tag: cgb_07555

Genome: Corynebacterium glutamicum DSM 20300 = ATCC 13032

Evidence: experimental

Source RNA:

Position: 669437 - 669498

Orientation: forward

Functional? true

Functional evidence: RNAdetect

Nucleotide sequence:

TCATCTCTCAATGTTACTTTTCCATTGTTAAGAATTAACAACTCTCGGTGATTTGTCGCAT

See dot plot graph | See alignment graph

B

RNA: cgb_07555 | **Regulates:**

sRNA cgb_07555 regulates:

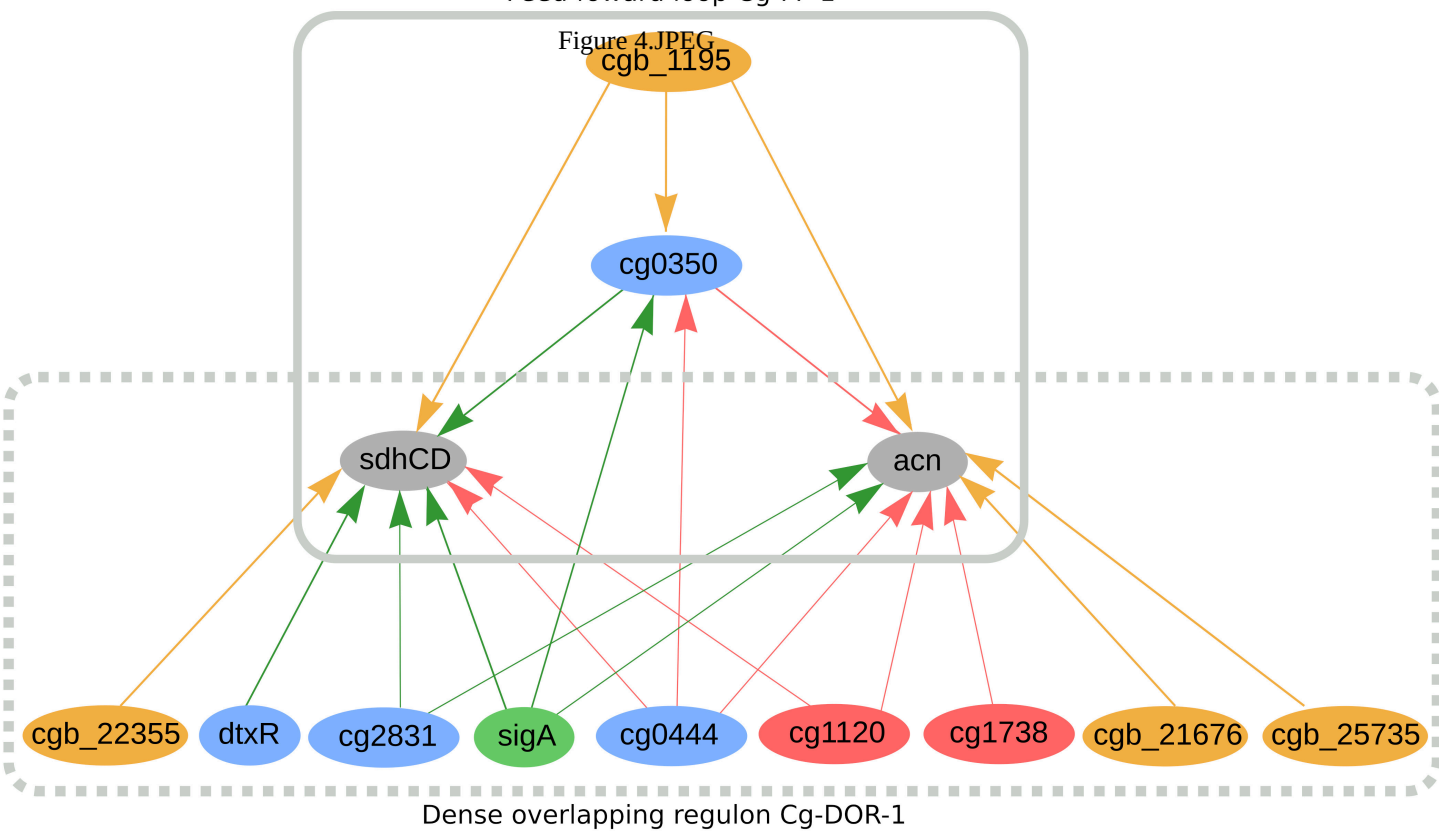
Show 10 entries Search:

mRNA	Source	P-value	Adjusted p-value	Position mRNA	Position sRNA	Minimum energy	Hybridization energy
cg2027	CopraRNA	1.281e-6	8.141e-5	53 -- 107	1 -- 59	-20.14	-38.6582
cg2377 (mraW)	CopraRNA	1.322e-5	6.143e-4	261 -- 298	4 -- 59	-18.94	-40.129
cg1871 (ruvC)	CopraRNA	1.205e-4	2.714e-3	152 -- 168	43 -- 59	-18.11	-25.4361
cg0012	CopraRNA	2.694e-4	4.097e-3	71 -- 90	41 -- 59	-16.88	-25.7237
cg2011	CopraRNA	4.002e-4	4.851e-3	136 -- 153	44 -- 61	-16.55	-25.5989

A

Feed foward loop Cg-FF-1

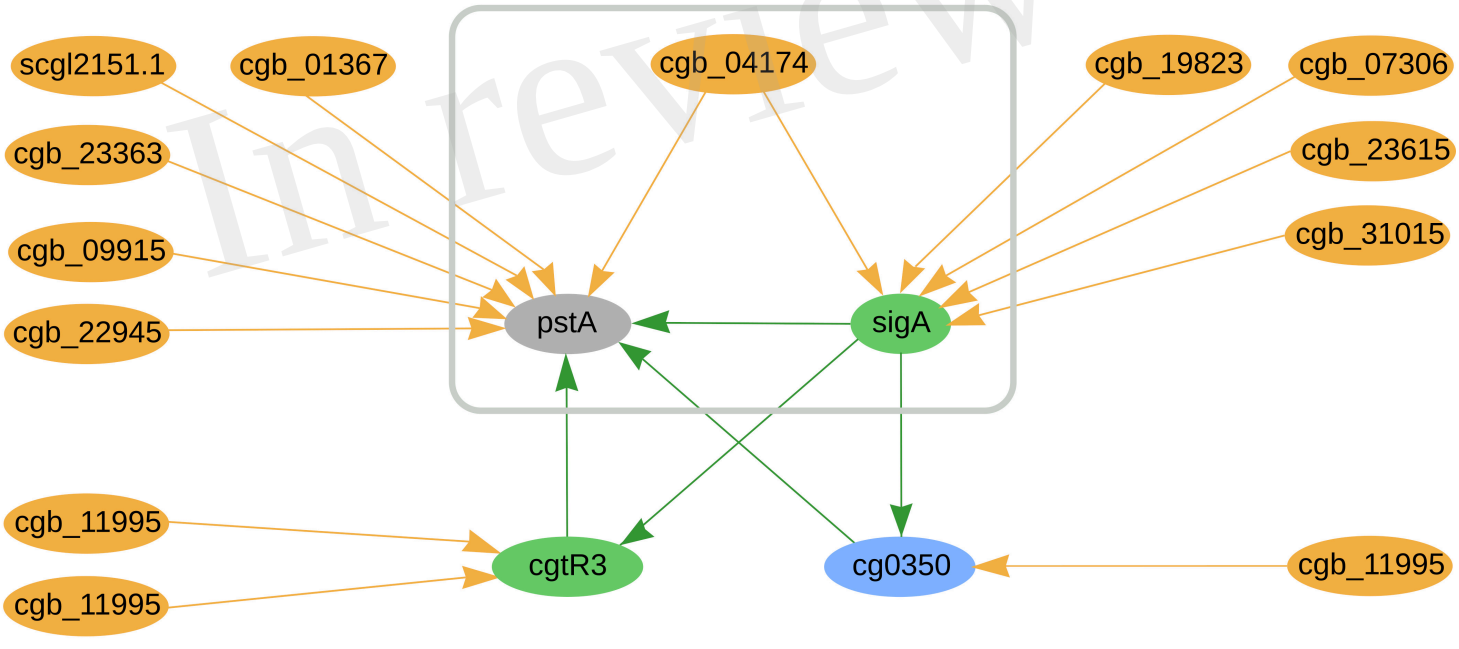
Figure 4.JPEG



Dense overlapping regulon Cg-DOR-1

B

Feed foward loop Cg-FF-2



C

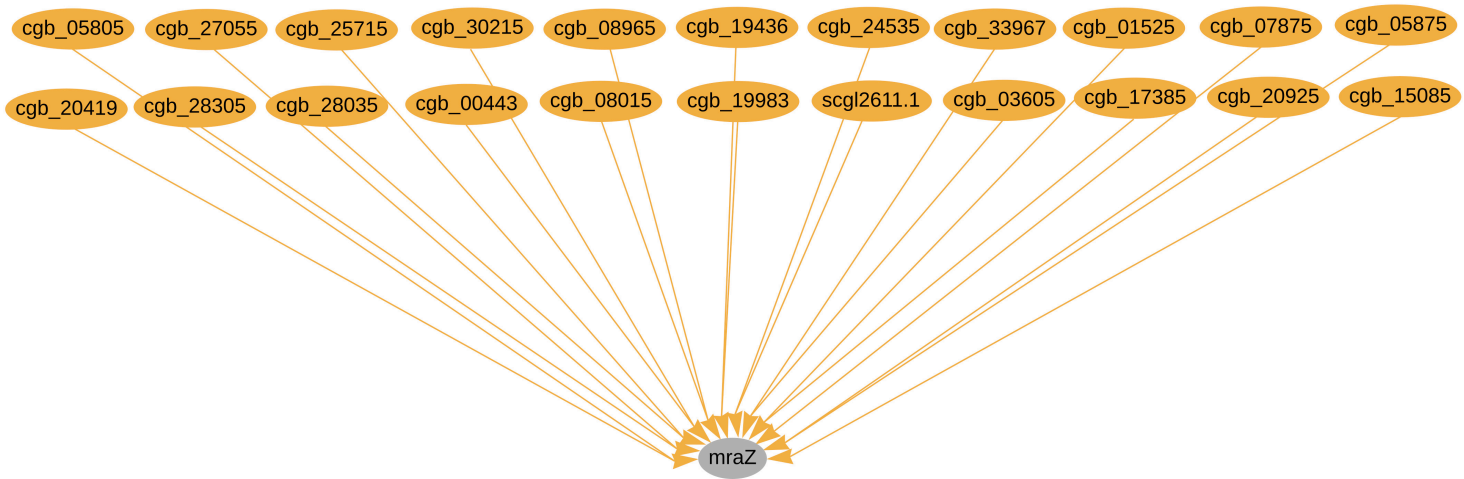


Figure 5.JPEG

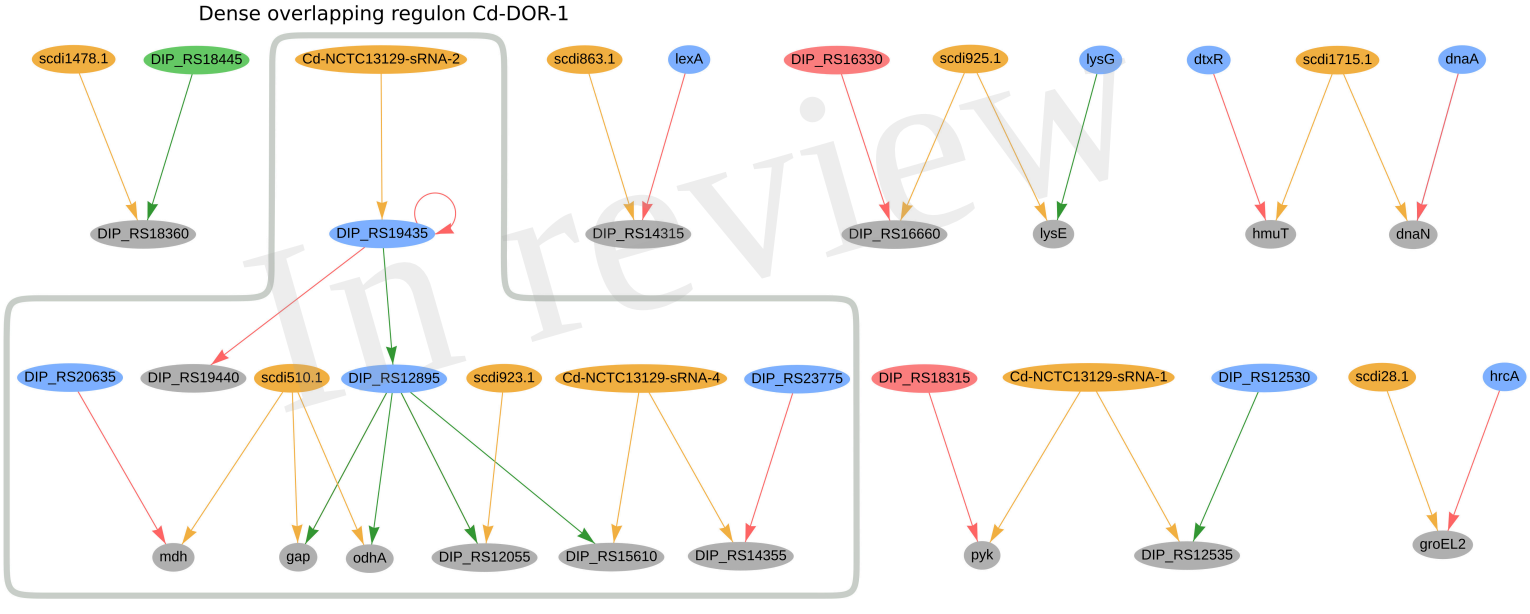


Figure 6.JPEG

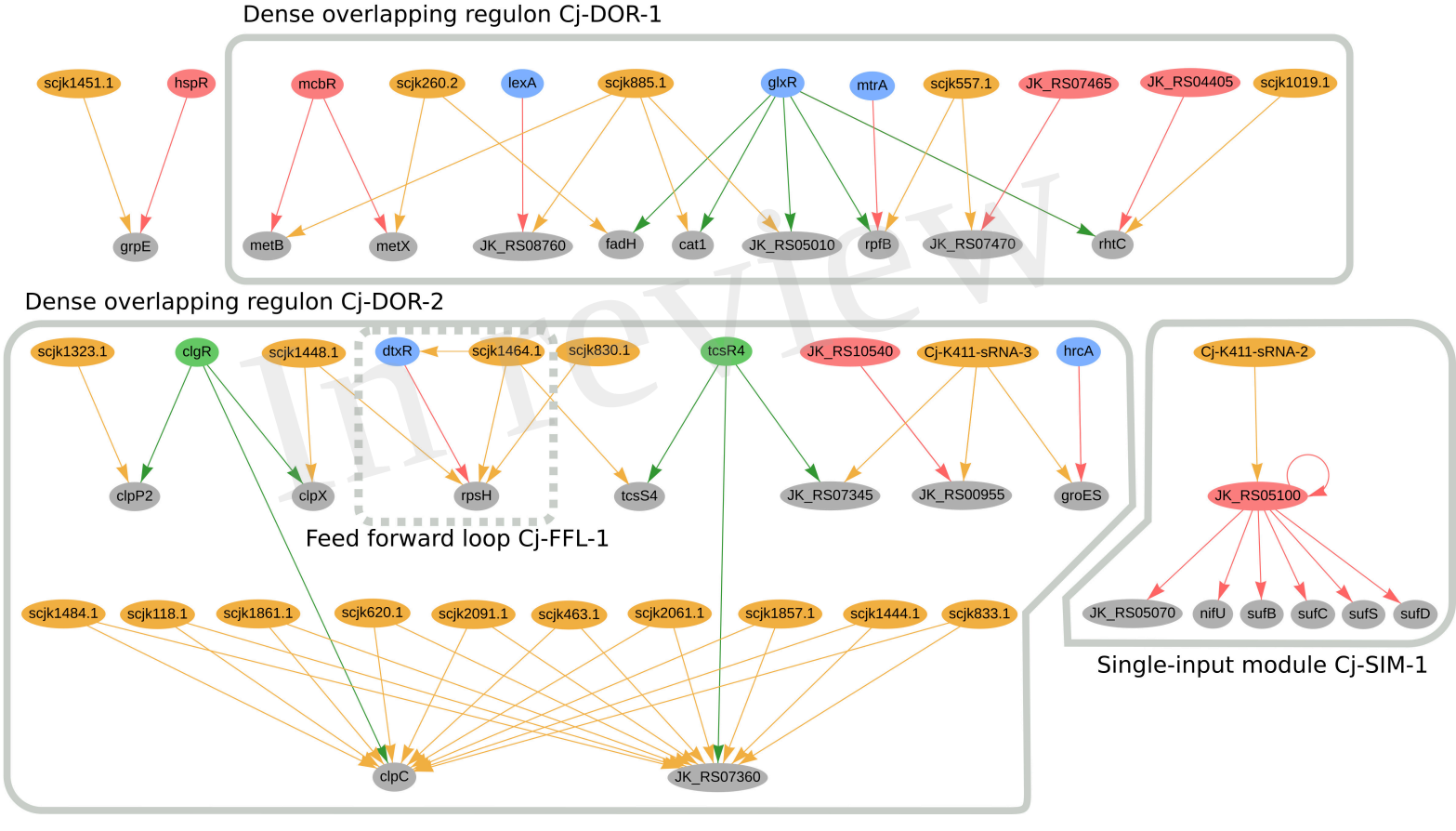
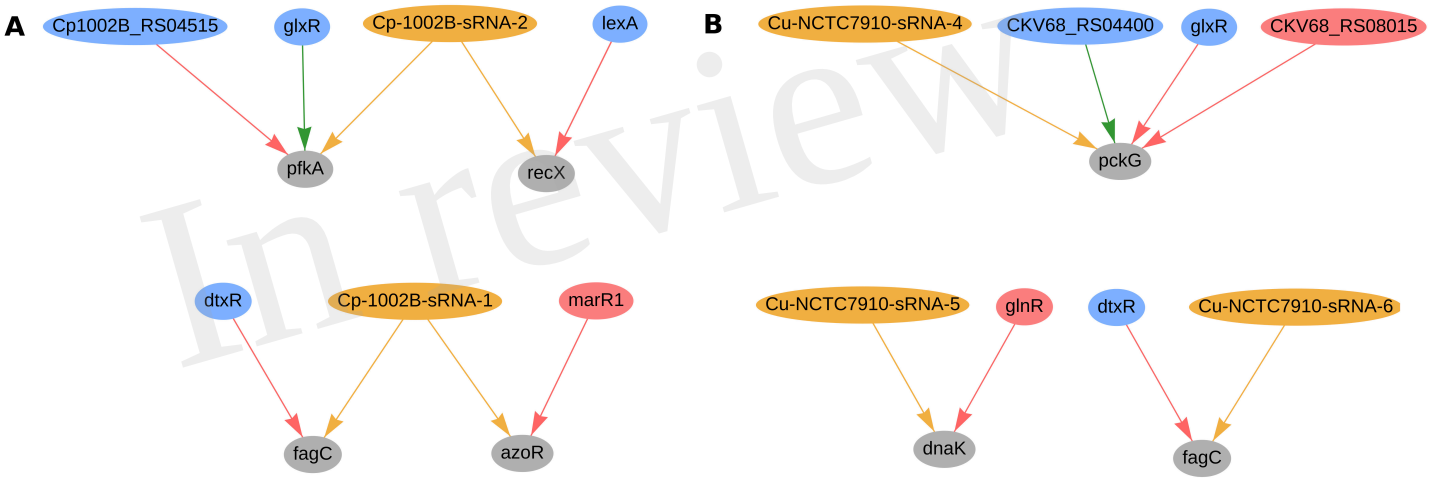


Figure 7.JPEG



V. DISCUSSION

We increased by more than seven times the number of corynebacterial species with known transcriptional GRNs available in public databases and enriched these networks with sRNA-regulation data for six corynebacterial species of either biotechnological or medical-veterinary interest. We explored both the transcriptional and post-transcriptional landscape of the *Corynebacterium* genus and suggested that TFs, sigma-factors and sRNAs jointly regulate gene expression in six species of this genus. Further investigation is necessary to both confirm our hypothesis and determine whether other corynebacterial species are also jointly regulated by TFs, sigma-factors and sRNAs. Overall, these results represent a step towards understanding the regulatory mechanisms of corynebacterial species of medical-veterinary interest and may lead us to a better understanding of how these organisms invade, colonize and infect their hosts. Thus, it may result in the development of more effective treatments, diagnosis and vaccines. Regarding biotechnological-interest species, these results may aid in the understanding of biomolecule's metabolic pathways and, consequently, open new perspectives for their manipulation. Moreover, recent studies have demonstrated that regulation by TRs alone is not able to fully explain the gene expression values of *E. coli (Larsen et al., 2019)* and *C. glutamicum* (Parise *et al.*, 2021) and suggested other regulatory layers should be explored in order to improve the quality of the regulatory networks. Our results represent a step toward this direction, considering our integration of sRNA-driven regulation into corynebacterial transcriptional GRNs.

When put together, our results suggest potential regulatory interactions for over 200 corynebacterial organisms and several targets for further experimental investigation. Amongst them, note the dense overlapping regulon Cg-DOR-1 in *C. glutamicum*, in which the genes *sdhCD* and *acn* are jointly regulated by four sRNAs and five TRs. Such regulons indicate response to multiple environmental stimuli, with several regulators coordinating the expression of genes involved in multiple biological responses. This generates different amounts of relative expression of the involved genes depending on the combination of detected signals (Beisel and Storz, 2010). Likewise, the genes in both Cd-DOR-1 in *C. diphtheriae* and Cj-DOR-2 in *C. jeikeium* are also densely regulated by several TRs and sRNAs. Genes in the first regulon are also involved in the TCA cycle and in carbohydrate metabolism, while the ones in the second are involved in growth and cell proliferation. These genes may also be responding to multiple environmental stimuli, responding to different regulatory signals. Another notable result is the highly-regulated *ptsG* gene as being directly regulated by *glxR* and three other TFs in *C. diphtheriae*. Studies in other organisms show that *ptsG* is responsible for kinase activity and phosphorylation in *C. pseudotuberculosis* and is involved in signal transduction in both *C. glutamicum* and *E. coli* (Lee, 2000; Petrov et al., 2019). Finally, we observed the virulence factor *fagC* being regulated by *dtxR* and one sRNA in both *C. ulcerans* and *C. pseudotuberculosis*. Besides the regulatory networks obtained in this work, we present methodologies for predicting TR-driven and sRNA-driven regulatory networks that may be applied in other organisms and yield further regulatory interactions.

One of the limitations of this study remains in the lack of known experimental transcriptional and post-transcriptional regulatory interactions both for this genus and

other bacteria. In regard to the transfer of transcriptional GRNs, the diversity of experimental data could allow us to account for regulatory interactions acquired by horizontal gene transfer or conserved by lifestyle. The lack of corynebacterial experimental regulatory interactions affects GRN transfer, considering conserved transcriptional regulators and target genes are identified through homology detection. Consequently, only known regulators and regulatory interactions can be identified using the GRN transfer. Given that sRNAs are also identified by homology in our sRNA-enrichment methodology, the lack of experimental data is also a limiting factor for the identification of homologous sRNAs.

Moreover, the quality of the predicted GRNs is directly related to the quality of the experimental ones. Other limitations to take into account are the low conservation of TFBSs in the transcriptional GRN transfer and the high false-positive rate in the sRNA-mRNA interaction prediction. For both transcriptional GRN transfer and sRNA regulatory interaction prediction, we applied conservative methodologies in order to maximize the true positive interactions and provided p-values presenting the likelihood of the results to be true. Assigning statistical values to TF-driven regulatory interactions allows the user to interpret the results.

VI. CONCLUSIONS AND OUTLOOK

A. Conclusions

With this thesis, we expand the knowledge of *Corynebacterium* transcriptional and post-transcriptional regulation. We increased the number of organisms with known TRNs by twenty times by releasing CoryneRegNet 7.0, which holds TRNs for 225 corynebacterial organisms and over 80,000 regulatory interactions. We collected, organized and summarized sRNA information in the *Corynebacterium* genus by presenting the first review article on corynebacterial sRNAs, which explores 887 sRNAs across 39 species. In CoryneRegNet 7.5, we present sRNA-enriched GRNs for six corynebacteria of medical, veterinary and biotechnological interest. With these results, we suggest several target genes for further experimental investigation, including genes related to virulence, survival and adaptation, amino-acid production, cell division, among others.

Overall, the results presented in this thesis contribute to the research community by (i) extending transcriptional and post-transcriptional regulatory knowledge in the *Corynebacterium* genus; (ii) making this knowledge available in an easy-to-use online database; (iii) suggesting that sRNAs, TFs and sigma-factors jointly orchestrate the regulatory landscape of six corynebacterial organisms; and (iv) providing predicted TF-, sigma- and sRNA-driven regulatory interactions to be used as guidance for future experimental assays.

B. Outlook

As perspectives of this work we aim to:

- Update CoryneRegNet by predicting TRNs for every fully sequenced and annotated corynebacterial genome in NCBI;
- Extend the GRN sRNA-enrichment methodology for every genome in CoryneRegNet.
- Enrich CoryneRegNet networks with gene expression data, allowing the users to visually browse conditions.
- Develop a stand-alone version of CoryneRegNet for GRN prediction.
- Explore sRNA regulatory landscape of other bacteria, such as the biofilm-forming *Leptospira biflexa* and the antibiotic-resistant *Brucella abortus*.

VII. BIBLIOGRAPHY

Abreu, V. A. C. *et al.* (2015) ‘CMRegNet-An interspecies reference database for corynebacterial and mycobacterial regulatory networks’, *BMC genomics*, 16, p. 452.

Ahmed, W., Hafeez, M. A. and Mahmood, S. (2018) ‘Identification and functional characterization of bacterial small non-coding RNAs and their target: A review’, *Gene Reports*, 10, pp. 167–176.

Altuvia, S. (2007) ‘Identification of bacterial small non-coding RNAs: experimental approaches’, *Current opinion in microbiology*, 10(3), pp. 257–261.

Arrieta-Ortiz, M. L. *et al.* (2015) ‘An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network’, *Molecular systems biology*, 11(11), p. 839.

Arrieta-Ortiz, M. L. *et al.* (2020) ‘Inference of Bacterial Small RNA Regulatory Networks and Integration with Transcription Factor-Driven Regulatory Networks’, *mSystems*, 5(3). doi: 10.1128/mSystems.00057-20.

Babu, M. M. *et al.* (2004) ‘Structure and evolution of transcriptional regulatory networks’, *Current opinion in structural biology*, 14(3), pp. 283–291.

Babu, M. M., Lang, B. and Aravind, L. (2009) ‘Methods to reconstruct and compare transcriptional regulatory networks’, *Methods in molecular biology* , 541, pp. 163–180.

Backofen, R. *et al.* (2018) ‘Comparative RNA Genomics’, *Comparative Genomics*, pp. 363–400. doi: 10.1007/978-1-4939-7463-4_14.

Balleza, E. *et al.* (2009) ‘Regulation by transcription factors in bacteria: beyond description’, *FEMS microbiology reviews*, 33(1), pp. 133–151.

Baumbach, J., Brinkrolf, K., Wittkop, T., *et al.* (2006) ‘CoryneRegNet 2: An Integrative Bioinformatics Approach for Reconstruction and Comparison of Transcriptional Regulatory Networks in Prokaryotes’, *Journal of integrative bioinformatics*, 3(2), pp. 1–13.

Baumbach, J., Brinkrolf, K., Czaja, L. F., *et al.* (2006) ‘CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks’, *BMC genomics*, 7, p. 24.

Baumbach, J. *et al.* (2007) ‘CoryneRegNet 3.0--an interactive systems biology platform for the analysis of gene regulatory networks in corynebacteria and Escherichia coli’, *Journal of biotechnology*, 129(2), pp. 279–289.

Baumbach, J. (2007) ‘CoryneRegNet 4.0--A reference database for corynebacterial gene regulatory networks’, *BMC bioinformatics*, 8(1), p. 429.

Baumbach, J. (2010) ‘On the power and limits of evolutionary conservation—unraveling bacterial gene regulatory networks’, *Nucleic acids research*, 38(22), pp. 7877–7884.

Baumbach, J., Rahmann, S. and Tauch, A. (2009) ‘Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms’, *BMC systems biology*, 3, p. 8.

Baumbach, J., Tauch, A. and Rahmann, S. (2009) ‘Towards the integrated analysis, visualization and reconstruction of microbial gene regulatory networks’, *Briefings in bioinformatics*, 10(1), pp. 75–83.

Beisel, C. L. and Storz, G. (2010) ‘Base pairing small RNAs and their roles in global regulatory networks’, *FEMS microbiology reviews*, 34(5), pp. 866–882.

Bervoets, I. and Charlier, D. (2019) ‘Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology’, *FEMS microbiology reviews*, 43(3), pp. 304–339.

Bookani, K. R. *et al.* (2018) ‘Corynebacterium jeikeium endocarditis: A case report and comprehensive review of an underestimated infection’, *IDCases*, 11, p. 26.

Brosse, A. and Guillier, M. (2018) ‘Bacterial Small RNAs in Mixed Regulatory Networks’, *Microbiology spectrum*, 6(3). doi: 10.1128/microbiolspec.RWR-0014-2017.

Browning, D. F. and Busby, S. J. W. (2016) ‘Local and global regulation of transcription initiation in bacteria’, *Nature reviews. Microbiology*, 14(10), pp. 638–650.

Chandrangsu, P., Rensing, C. and Helmann, J. D. (2017) ‘Metal homeostasis and resistance in bacteria’, *Nature reviews. Microbiology*, 15(6), pp. 338–350.

Davis, M. C. *et al.* (2017) ‘The essential activities of the bacterial sigma factor’, *Canadian journal of microbiology*, 63(2), pp. 89–99.

Diallo, I. and Provost, P. (2020) ‘RNA-Sequencing Analyses of Small Bacterial RNAs and their Emergence as Virulence Factors in Host-Pathogen Interactions’, *International journal of molecular sciences*, 21(5). doi: 10.3390/ijms21051627.

Dorella, F. A. *et al.* (2006) ‘Corynebacterium pseudotuberculosis: microbiology, biochemical properties, pathogenesis and molecular studies of virulence’, *Veterinary research*, 37(2), pp. 201–218.

Duval, M. *et al.* (2015) ‘Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics’, *Biochimie*, 114, pp. 18–29.

Eckweiler, D. *et al.* (2018) ‘PRODORIC2: the bacterial gene regulation database in 2018’, *Nucleic acids research*, 46(D1), pp. D320–D326.

Escorcia-Rodríguez, J. M., Tauch, A. and Freyre-González, J. A. (2021) ‘Corynebacterium glutamicum regulation beyond transcription: Organizing principles and reconstruction of an extended regulatory network incorporating regulations mediated by small RNA and protein-protein interactions’, *Cold Spring Harbor Laboratory*. doi: 10.1101/2021.01.07.423633.

Feklístov, A. *et al.* (2014) ‘Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective’. doi: 10.1146/annurev-micro-092412-155737.

Georgi, T., Rittmann, D. and Wendisch, V. F. (2005) ‘Lysine and glutamate production by Corynebacterium glutamicum on glucose, fructose and sucrose: roles of malic enzyme and fructose-1,6-bisphosphatase’, *Metabolic engineering*, 7(4), pp. 291–301.

von Graevenitz, A. and Bernard, K. (2006) ‘The Genus Corynebacterium--Medical’, in Dworkin, M. et al. (eds) *The Prokaryotes: Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes*. New York, NY: Springer New York, pp. 819–842.

Hacker, E. *et al.* (2016) ‘Corynebacterium ulcerans, an emerging human pathogen’, *Future*

microbiology, 11, pp. 1191–1208.

Hör, J., Gorski, S. A. and Vogel, J. (2018) 'Bacterial RNA Biology on a Genome Scale', *Molecular cell*, 70(5), pp. 785–799.

Huang, H.-Y. *et al.* (2009) 'sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes', *Nucleic acids research*, 37(Database issue), pp. D150–4.

Ibarra-Arellano, M. A. *et al.* (2016) 'Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria', *Database: the journal of biological databases and curation*, 2016. doi: 10.1093/database/baw089.

Ikeda, M. and Katsumata, R. (1992) 'Metabolic Engineering To Produce Tyrosine or Phenylalanine in a Tryptophan-Producing Corynebacterium glutamicum Strain', *Applied and environmental microbiology*, 58(3), pp. 781–785.

Kılıç, S. and Erill, I. (2016) 'Assessment of transfer methods for comparative genomics of regulatory networks in bacteria', *BMC bioinformatics*, 17 Suppl 8, p. 277.

Krebs, J. E. *et al.* (2013) *Lewin's GENES XI*. Jones & Bartlett Publishers.

Krebs, J. E., Goldstein, E. S. and Kilpatrick, S. T. (2017) *Lewin's GENES XII*. Jones & Bartlett Learning.

Larsen, S. J. *et al.* (2019) 'E. coli gene regulatory networks are inconsistent with gene expression data', *Nucleic acids research*, 47(1), pp. 85–92.

Leyn, S. A. *et al.* (2016) 'Comparative genomics and evolution of transcriptional regulons in', *Microbial genomics*, 2(7), p. e000061.

Li, L. *et al.* (2013) 'BSRD: a repository for bacterial small regulatory RNA', *Nucleic acids research*, 41(Database issue), pp. D233–8.

Lin, Y. *et al.* (2020) 'RNAInter in 2020: RNA interactome repository with increased coverage and annotation', *Nucleic acids research*, 48(D1), pp. D189–D197.

LPSN-DSMZ (2021). Available at: https://lpsn.dsmz.de/genus/corynebacterium (Accessed: 18 February 2021).

Mercatelli, D. *et al.* (2020) 'Gene regulatory network inference resources: A practical overview', *Biochimica et Biophysica Acta, Gene Regulatory Mechanisms*, 1863(6), p. 194430.

Métris, A. *et al.* (2017) 'SalmoNet, an integrated network of ten strains reveals common and distinct pathways to host adaptation', *NPJ systems biology and applications*, 3, p. 31.

Murphy, J. R. (1996) 'Corynebacterium Diphtheriae', in Baron, S. (ed.) *Medical Microbiology*. Galveston (TX): University of Texas Medical Branch at Galveston.

Nitzan, M., Rehani, R. and Margalit, H. (2017) 'Integration of Bacterial Small RNAs in Regulatory Networks', *Annual review of biophysics*, 46, pp. 131–148.

Oliveira, A. *et al.* (2017) 'Insight of Genus Corynebacterium: Ascertaining the Role of

Pathogenic and Non-pathogenic Species', *Frontiers in microbiology*, 8, p. 1937.

Pain, A. *et al.* (2015) 'An assessment of bacterial small RNA target prediction programs', *RNA biology*, 12(5), pp. 509–513.

Parise, D. *et al.* (2021) 'On the Consistency between Gene Expression and the Gene Regulatory Network of Corynebacterium glutamicum', *Network and Systems Medicine*, 4(1), pp. 51–59.

Parise, M. T. D. *et al.* (2020) 'CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks', *Scientific data*, 7(1), p. 142.

Pauling, J., Röttger, R., Tauch, A., *et al.* (2012) 'CoryneRegNet 6.0—Updated database content, new analysis methods and novel features focusing on community demands', *Nucleic acids research*, 40(D1), pp. D610–D614.

Pauling, J., Röttger, R., Neuner, A., *et al.* (2012) 'On the trail of EHEC/EAEC--unraveling the gene regulatory networks of human pathogenic Escherichia coli bacteria', *Integrative biology: quantitative biosciences from nano to macro*, 4(7), pp. 728–733.

Picard, F. *et al.* (2012) 'Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression', *BMC genomics*, 13, p. 528.

Pischimarov, J. *et al.* (2012) 'sRNAdb: a small non-coding RNA database for gram-positive bacteria', *BMC genomics*, 13, p. 384.

Roncarati, D. and Scarlato, V. (2017) 'Regulation of heat-shock genes in bacteria: from signal sensing to gene expression output', *FEMS microbiology reviews*, 41(4), pp. 549–574.

Salgado, H. *et al.* (2006) 'RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions', *Nucleic acids research*, 34(Database issue), pp. D394–7.

Santos-Zavaleta, A. *et al.* (2019) 'RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12', *Nucleic acids research*, 47(D1), pp. D212–D220.

Selim, S. A. (2001) 'Oedematous skin disease of buffalo in Egypt', *Journal of veterinary medicine. B, Infectious diseases and veterinary public health*, 48(4), pp. 241–258.

Shimoni, Y. *et al.* (2007) 'Regulation of gene expression by small non-coding RNAs: a quantitative view', *Molecular systems biology*, 3, p. 138.

Shleeva, M., Savitsky, A. and Kaprelyants, A. (2020) 'Corynebacterium jeikeium Dormant Cell Formation and Photodynamic Inactivation', *Frontiers in microbiology*, 11. doi: 10.3389/fmicb.2020.605899.

Sierro, N. *et al.* (2008) 'DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information', *Nucleic acids research*, 36(Database issue), pp. D93–6.

Smith, C. *et al.* (2016) 'Mapping the Regulatory Network for Salmonella enterica Serovar Typhimurium Invasion', *mBio*, 7(5). doi: 10.1128/mBio.01024-16.

de Sousa Abreu, R. *et al.* (2009) 'Global signatures of protein and mRNA expression levels', *Molecular bioSystems*, 5(12), pp. 1512–1526.

Storz, G., Vogel, J. and Wassarman, K. M. (2011) 'Regulation by small RNAs in bacteria: expanding frontiers', *Molecular cell*, 43(6), pp. 880–891.

Thompson, D., Regev, A. and Roy, S. (2015) 'Comparative analysis of gene regulatory networks: from network reconstruction to evolution', *Annual review of cell and developmental biology*, 31, pp. 399–428.

Turkarslan, S. *et al.* (2015) 'A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis', *Scientific data*, 2, p. 150010.

Wagner, K. S. *et al.* (2011) 'Screening for Corynebacterium diphtheriae and Corynebacterium ulcerans in patients with upper respiratory tract infections 2007-2008: a multicentre European study', *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 17(4), pp. 519–525.

Wang, J. *et al.* (2016) 'sRNATarBase 3.0: an updated database for sRNA-target interactions in bacteria', *Nucleic acids research*, 44(D1), pp. D248–53.

Waters, L. S. and Storz, G. (2009) 'Regulatory RNAs in bacteria', *Cell*, 136(4), pp. 615–628.

Waters, S. A. *et al.* (2017) 'Small RNA interactome of pathogenic E. coli revealed through crosslinking of RNase E', *The EMBO journal*, 36(3), pp. 374–387.

Wright, P. R. *et al.* (2014) 'CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains', *Nucleic acids research*, 42(Web Server issue), pp. W119–23.

Yeruham, I. *et al.* (2004) 'A herd level analysis of a Corynebacterium pseudotuberculosis outbreak in a dairy cattle herd', *Veterinary dermatology*, 15(5), pp. 315–320.

Zasada, A. A. and Mosiej, E. (2018) 'Contemporary microbiology and identification of Corynebacteria spp. causing infections in human', *Letters in applied microbiology*, 66(6), pp. 472–483.

Zhang, Y. *et al.* (2017) 'A Review on Recent Computational Methods for Predicting Noncoding RNAs', *BioMed research international*, 2017, p. 9139504.

Zhu, B. and Stülke, J. (2018) 'SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis', *Nucleic acids research*, 46(D1), pp. D743–D748.

VIII. APPENDIX

First genome sequencing and comparative analyses of Corynebacterium pseudotuberculosis strains from Mexico

Doglas Parise*, **Mariana T D Parise***, Marcus V C Viana, Adrian V Muñoz-Bucio, Yazmin A Cortés-Pérez, Beatriz Arellano-Reynoso, Efrén Díaz-Aparicio, Fernanda A Dorella, Felipe L Pereira, Alex F Carvalho, Henrique C P Figueiredo, Preetam Ghosh, Debmalya Barh, Anne C P Gomide and Vasco A C Azevedo

*Shared first authorship.

Contribution: In this research article, I contributed to the preparation of the original manuscript genome analysis and subsequent revision of the manuscript.

Status: Published

Parise et al. Standards in Genomic Sciences (2018) 13:21
https://doi.org/10.1186/s40793-018-0325-z

Standards in Genomic Sciences

EXTENDED GENOME REPORT Open Access

First genome sequencing and comparative analyses of *Corynebacterium pseudotuberculosis* strains from Mexico

Doglas Parise[1†], Mariana T D Parise[1†], Marcus V C Viana[1], Adrian V Muñoz-Bucio[2], Yazmin A Cortés-Pérez[2], Beatriz Arellano-Reynoso[2], Efrén Díaz-Aparicio[2], Fernanda A Dorella[3], Felipe L Pereira[3], Alex F Carvalho[3], Henrique C P Figueiredo[3], Preetam Ghosh[4], Debmalya Barh[1,5,6], Anne C P Gomide[1] and Vasco A C Azevedo[1*]

Abstract

Corynebacterium pseudotuberculosis is a pathogenic bacterium which has been rapidly spreading all over the world, causing economic losses in the agricultural sector and sporadically infecting humans. Six *C. pseudotuberculosis* strains were isolated from goats, sheep, and horses with distinct abscess locations. For the first time, Mexican genomes of this bacterium were sequenced and studied in silico. All strains were sequenced using Ion Personal Genome Machine sequencer, assembled using Newbler and SPAdes software. The automatic genome annotation was done using the software RAST and in-house scripts for transference, followed by manual curation using Artemis software and BLAST against NCBI and UniProt databases. The six genomes are publicly available in NCBI database. The analysis of nucleotide sequence similarity and the generated phylogenetic tree led to the observation that the Mexican strains are more similar between strains from the same host, but the genetic structure is probably more influenced by transportation of animals between farms than host preference. Also, a putative drug target was predicted and in silico analysis of 46 strains showed two gene clusters capable of differentiating the biovars *equi* and *ovis*: Restriction Modification system and CRISPR-Cas cluster.

Keywords: Phylogenetics, Genomic sequencing, Drug target, CRISPR-Cas, Restriction-modification systems

Introduction

Corynebacterium pseudotuberculosis is a Gram-positive bacterium that infects several different species of mammals. Strains of the biovar *ovis* infect sheep and goats, and strains of the biovar *equi* infect larger mammals such as horses, camels, and buffaloes. The manifestation of the infection depends on the host [1–4]. This bacterium causes significant economic loss to animal production all over the world due to reduced production of wool, milk and meat, carcass condemnation, as well as the death of infected animals [4–6]. *C. pseudotuberculosis* can also affect humans, causing distinct kinds of lymphadenitis. Contamination occurs through contact with infected animals and consumption of infected food [4, 5, 7].

This organism affects several countries such as Australia, Brazil, Canada, Egypt, Israel, New Zealand, South Africa, United Kingdom and United States [4, 8–17]. Cases in other countries such as Portugal [18], Mexico [19] and Equatorial Guinea [20] have been reported in the recent years. In the United States, *C. pseudotuberculosis* infections are reemerging and considered endemic [19], and the state with the highest number of cases of this bacterium was Texas, which borders Mexico [21]. The spread of *C. pseudotuberculosis* to other countries brings out the importance of improving the understanding of this bacterium. In the present study, six Mexican *C. pseudotuberculosis* strains were investigated, two from the biovar *equi* and four from the biovar *ovis*. This is the first time that strains of this bacterium, isolated in Mexico, have been completely sequenced. Among those strains, these are the first isolates of the biovar *equi* coming from this country [19]. The characterization of these strains is important for achieving a better understanding of this species,

* Correspondence: vasco@icb.ufmg.br
†Doglas Parise and Mariana T D Parise contributed equally to this work.
[1]Laboratory of Cellular and Molecular Genetics, Institute of Biologic Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
Full list of author information is available at the end of the article

The Transcriptional Regulatory Network of *Corynebacterium pseudotuberculosis*

Doglas Parise*, **Mariana Teixeira Dornelles Parise***, Anne Cybelle Pinto Gomide, Flávia Figueira Aburjaile, Rodrigo Bentes Kato, Marisol Salgado-Albarrán, Andreas Tauch, Vasco Ariston de Carvalho Azevedo and Jan Baumbach

*These authors contributed equally to this study.

Contribution: I contributed to writing the preparation of the original manuscript, working on the figures and revising the manuscript.

Status: Published

microorganisms

MDPI

Review

The Transcriptional Regulatory Network of *Corynebacterium pseudotuberculosis*

Doglas Parise [1,2,*,†], Mariana Teixeira Dornelles Parise [1,2,†], Anne Cybelle Pinto Gomide [2], Flávia Figueira Aburjaile [3], Rodrigo Bentes Kato [2], Marisol Salgado-Albarrán [1,4], Andreas Tauch [5], Vasco Ariston de Carvalho Azevedo [2,‡] and Jan Baumbach [1,6,7,‡]

1 Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, 85354 Freising-Weihenstephan, Germany; mparise@wzw.tum.de (M.T.D.P.); marisol.salgado@tum.de (M.S.-A.); jan.baumbach@uni-hamburg.de (J.B.)
2 Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil; acybelle@gmail.com (A.C.P.G.); rbkato@gmail.com (R.B.K.); vasco@icb.ufmg.br (V.A.d.C.A.)
3 Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Rio de Janeiro 21040-360, Brazil; faburjaile@gmail.com
4 Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana Cuajimalpa, Mexico City 05348, Mexico
5 Center for Biotechnology (CeBiTec), Bielefeld University, 33615 Bielefeld, Germany; tauch@cebitec.uni-bielefeld.de
6 Computational BioMedicine lab, Institute of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense, Denmark
7 Chair of Computational Systems Biology, University of Hamburg, 22607 Hamburg, Germany
* Correspondence: dparise@wzw.tum.de or doglas@ufmg.br
† These authors contributed equally to this study.
‡ These authors share senior authorship.

Abstract: *Corynebacterium pseudotuberculosis* is a Gram-positive, facultative intracellular, pathogenic bacterium that infects several different hosts, yielding serious economic losses in livestock farming. It causes several diseases including oedematous skin disease (OSD) in buffaloes, ulcerative lymphangitis (UL) in horses, and caseous lymphadenitis (CLA) in sheep, goats and humans. Despite its economic and medical-veterinary importance, our understanding concerning this organism's transcriptional regulatory mechanisms is still limited. Here, we review the state of the art knowledge on transcriptional regulatory mechanisms of this pathogenic species, covering regulatory interactions mediated by two-component systems, transcription factors and sigma factors. Key transcriptional regulatory players involved in virulence and pathogenicity of *C. pseudotuberculosis*, such as the PhoPR system and DtxR, are in the focus of this review, as these regulators are promising targets for future vaccine design and drug development. We conclude that more experimental studies are needed to further understand the regulatory repertoire of this important zoonotic pathogen, and that regulators are promising targets for future vaccine design and drug development.

Keywords: *Corynebacterium pseudotuberculosis*; transcriptional regulatory mechanisms; transcription factors; two-component systems; sigma factors; virulence; pathogenicity

Citation: Parise, D.; Teixeira Dornelles Parise, M.; Pinto Gomide, A.C.; Figueira Aburjaile, F.; Bentes Kato, R.; Salgado-Albarrán, M.; Tauch, A.; Ariston de Carvalho Azevedo, V.; Baumbach, J. The Transcriptional Regulatory Network of *Corynebacterium pseudotuberculosis*. *Microorganisms* **2021**, *9*, 415. https://doi.org/10.3390/microorganisms9020415

Academic Editor: Anthony J. Sinskey

Received: 5 February 2021
Accepted: 14 February 2021
Published: 17 February 2021

1. Introduction

Transcriptional regulation is one of the most important mechanisms of bacterial adaptation to changes in the environment; in particular, pathogenic bacteria use this mechanism to contend the conditions they are exposed to when infecting the host. These conditions are associated with immune barriers imposed by the host to fight the infection; for instance, pH, oxidative and osmotic stresses, temperature and changes in nutrient availability. To quickly adapt, survive and establish the infection, transcriptional regulation controls key biological processes, such as biofilm formation, quorum sensing, temperature

On the consistency between gene expression and the gene regulatory network of *Corynebacterium glutamicum*

Doglas Parise, **Mariana Teixeira Dornelles Parise**, Evans Kataka, Rodrigo Bentes Kato, Markus List, Andreas Tauch, Vasco Ariston de Carvalho Azevedo and Jan Baumbach

Contribution: In this study, I contributed to data curation, the preparation of the original manuscript along with its revision.

Status: Published

Network and Systems Medicine
Volume 4.1, 2021
DOI: 10.1089/nsm.2020.0014
Accepted February 5, 2021

NETWORK and SYSTEMS MEDICINE
Journal of Medical Systems Biology and Network Science
Mary Ann Liebert, Inc. publishers

ORIGINAL RESEARCH Open Access

On the Consistency between Gene Expression and the Gene Regulatory Network of *Corynebacterium glutamicum*

Doglas Parise,[1,2,*] Mariana Teixeira Dornelles Parise,[1,2] Evans Kataka,[1] Rodrigo Bentes Kato,[2] Markus List,[1] Andreas Tauch,[3] Vasco Ariston de Carvalho Azevedo,[2] and Jan Baumbach[1]

Abstract

Background: Transcriptional regulation of gene expression is crucial for the adaptation and survival of bacteria. Regulatory interactions are commonly modeled as Gene Regulatory Networks (GRNs) derived from experiments such as RNA-seq, microarray and ChIP-seq. While the reconstruction of GRNs is fundamental to decipher cellular function, even GRNs of economically important bacteria such as *Corynebacterium glutamicum* are incomplete.

Materials and Methods: Here, we analyzed the predictive power of GRNs if used as in silico models for gene expression and investigated the consistency of the *C. glutamicum* GRN with gene expression data from the GEO database.

Results: We assessed the consistency of the *C. glutamicum* GRN using real, as well as simulated, expression data and showed that GRNs alone cannot explain the expression profiles well.

Conclusion: Our results suggest that more sophisticated mechanisms such as a combination of transcriptional, post-transcriptional regulation and signaling should be taken into consideration when analyzing and constructing GRNs.

Keywords: *Corynebacterium glutamicum*; Gene Regulatory Networks; inconsistency assessment; systems biology

Introduction

Bacterial genomes are small and compact; on average, 88% of their genomes consist of coding genes.[1] Consequently, the ability to regulate gene expression in diverse environments is crucial for stabilizing cell homeostasis and adapting to environmental challenges.[2] Computational systems biology uses Gene Regulatory Networks (GRNs) to understand the mechanisms that coordinate the shifts in gene expression and to represent the transcriptional gene regulation of organisms. These networks are consistently expanding our understanding of how the genotype manifests in the phenotype of an organism. Computationally, GRNs are modeled as directed graphs with nodes representing genes and edges or links representing the interactions between regulators, also known as transcription factors (TFs), and their target genes (TGs).[3,4]

Techniques to measure gene expression levels and infer GRNs include microarrays,[5] ChIP-seq,[6] and RNA-Seq.[7] Microarrays measure the expression levels of known genes through the quantification of the fluorescence emitted by chemically marked complementary DNA attached to a solid surface. Microarrays can also determine the binding site of TFs when combined with chromatin immunoprecipitation.[8,9] ChIP-seq is also used to determine the TF binding sites by

[1]TUM School of Life Sciences, Technical University of Munich, Freising-Weihenstephan, Germany.
[2]Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.
[3]Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany.

*Address correspondence to: Doglas Parise, MSc, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Maximus-von-Imhof-Forum 3, Freising-Weihenstephan 385354, Germany, E-mail: doglas@ufmg.br

Transcriptome profile of *Corynebacterium pseudotuberculosis* in response to iron limitation

Izabela Coimbra Ibraim, **Mariana Teixeira Dornelles Parise**, Doglas Parise, Michelle Zibetti Tadra Sfeir, Thiago Luiz de Paula Castro, Alice Rebecca Wattam, Preetam Ghosh, Debmalya Barh, Emannuel Maltempi Souza, Aristóteles Góes-Neto, Anne Cybelle Pinto Gomide and Vasco Azevedo

Contribution: I contributed to the prediction of conserved regulatory interactions in *C. pseudotuberculosis* strains T1, Cp13 or 1002B. I also contributed to the manuscript revision.

Status: Published

Ibraim et al. BMC Genomics (2019) 20:663
https://doi.org/10.1186/s12864-019-6018-1

BMC Genomics

RESEARCH ARTICLE Open Access

Transcriptome profile of *Corynebacterium pseudotuberculosis* in response to iron limitation

Izabela Coimbra Ibraim[1], Mariana Teixeira Dornelles Parise[1], Doglas Parise[1], Michelle Zibetti Tadra Sfeir[2], Thiago Luiz de Paula Castro[3], Alice Rebecca Wattam[4], Preetam Ghosh[5], Debmalya Barh[1], Emannuel Maltempi Souza[2], Aristóteles Góes-Neto[6], Anne Cybelle Pinto Gomide[1†] and Vasco Azevedo[1*†]

Abstract

Background: Iron is an essential micronutrient for the growth and development of virtually all living organisms, playing a pivotal role in the proliferative capability of many bacterial pathogens. The impact that the bioavailability of iron has on the transcriptional response of bacterial species in the CMNR group has been widely reported for some members of the group, but it hasn't yet been as deeply explored in *Corynebacterium pseudotuberculosis*. Here we describe for the first time a comprehensive RNA-seq whole transcriptome analysis of the T1 wild-type and the Cp13 mutant strains of *C. pseudotuberculosis* under iron restriction. The Cp13 mutant strain was generated by transposition mutagenesis of the *ciuA* gene, which encodes a surface siderophore-binding protein involved in the acquisition of iron. Iron-regulated acquisition systems are crucial for the pathogenesis of bacteria and are relevant targets to the design of new effective therapeutic approaches.

Results: Transcriptome analyses showed differential expression in 77 genes within the wild-type parental T1 strain and 59 genes in Cp13 mutant under iron restriction. Twenty-five of these genes had similar expression patterns in both strains, including up-regulated genes homologous to the hemin uptake *hmu* locus and two distinct operons encoding proteins structurally like hemin and Hb-binding surface proteins of *C. diphtheriae*, which were remarkably expressed at higher levels in the Cp13 mutant than in the T1 wild-type strain. These hemin transport protein genes were found to be located within genomic islands associated with known virulent factors. Down-regulated genes encoding iron and heme-containing components of the respiratory chain (including *ctaCEF* and *qcrCAB* genes) and up-regulated known iron/DtxR-regulated transcription factors, namely *ripA* and *hrrA*, were also identified differentially expressed in both strains under iron restriction.

Conclusion: Based on our results, it can be deduced that the transcriptional response of *C. pseudotuberculosis* under iron restriction involves the control of intracellular utilization of iron and the up-regulation of hemin acquisition systems. These findings provide a comprehensive analysis of the transcriptional response of *C. pseudotuberculosis*, adding important understanding of the gene regulatory adaptation of this pathogen and revealing target genes that can aid the development of effective therapeutic strategies against this important pathogen.

Keywords: *Corynebacterium pseudotuberculosis*, Iron homeostasis, Iron-regulated transcriptional factors, Heme acquisition, Differential gene expression

* Correspondence: vasco@icb.ufmg.br
†Anne Cybelle Pinto Gomide and Vasco Azevedo shared senior authorship
[1]Laboratório de Genética Molecular e Celular, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
Full list of author information is available at the end of the article

Global Regulator of Rubber Degradation in *Gordonia polyisoprenivorans* VH2: Identification and Involvement in the Regulation Network

Jan de Witt, Sylvia Oetermann, **Mariana Parise**, Doglas Parise, Jan Baumbach and Alexander Steinbüchel

Contribution: I contributed to the prediction of TFBSs for the global regulator CRP_{VH2} protein (CRPVH2) in *Gordonia polyisoprenivorans* VH2. I also contributed to the revision of the original manuscript.

Status: Published

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Global Regulator of Rubber Degradation in *Gordonia polyisoprenivorans* VH2: Identification and Involvement in the Regulation Network

Jan de Witt,[a] Sylvia Oetermann,[a] Mariana Parise,[b] Doglas Parise,[b] Jan Baumbach,[b] Alexander Steinbüchel[a,c]

[a]Institut für Molekulare Mikrobiologie und Biotechnologie, Westfälische Wilhelms-Universität, Münster, Germany
[b]Chair of Experimental Bioinformatics at TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[c]Environmental Sciences Department, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT A cAMP receptor protein (CRP_{VH2}) was detected as a global regulator in *Gordonia polyisoprenivorans* VH2 and was proposed to participate in the network regulating poly(*cis*-1,4-isoprene) degradation as a novel key regulator. CRP_{VH2} shares a sequence identity of 79% with GlxR, a well-studied global regulator of *Corynebacterium glutamicum*. Furthermore, CRP_{VH2} and GlxR have a common oligomerization state and similar binding motifs, and thus most likely have similar functions as global regulators. Size exclusion chromatography of purified CRP_{VH2} confirmed the existence as a homodimer with a native molecular weight of 44.1 kDa in the presence of cAMP. CRP_{VH2} bound to the $TGTGAN_6TCACT$ motif within the 131-bp intergenic region of divergently oriented $lcp1_{VH2}$ and $lcpR_{VH2}$, encoding a latex clearing protein and its putative repressor, respectively. DNase I footprinting assays revealed the exact operator size of CRP_{VH2} in the intergenic region (25 bp), which partly overlapped with the proposed promoters of $lcpR_{VH2}$ and $lcp1_{VH2}$. Our findings indicate that CRP_{VH2} represses the expression of $lcpR_{VH2}$ while simultaneously directly or indirectly activating the expression of $lcp1_{VH2}$ by binding the competing promoter regions. Furthermore, binding of CRP_{VH2} to upstream regions of additional putative enzymes of poly(*cis*-1,4-isoprene) degradation was verified *in vitro*. *In silico* analyses predicted 206 CRP_{VH2} binding sites comprising 244 genes associated with several functional categories, including carbon and peptide metabolism, stress response, etc. The gene expression regulation of several subordinated regulators substantiated the function of CRP_{VH2} as a global regulator. Moreover, we anticipate that the novel *lcpR* regulation mechanism by CRPs is widespread in other rubber-degrading actinomycetes.

IMPORTANCE In order to develop efficient microbial recycling strategies for rubber waste materials, it is required that we understand the degradation pathway of the polymer and how it is regulated. However, only little is known about the transcriptional regulation of the rubber degradation pathway, which seems to be upregulated in the presence of the polymer. We identified a novel key regulator of rubber degradation (CRP_{VH2}) that regulates several parts of the pathway in the potent rubber-degrader *G. polyisoprenivorans* VH2. Furthermore, we provide evidence for a widespread involvement of CRP regulators in the degradation of rubber in various other rubber-degrading actinomycetes. Thus, these novel insights into the regulation of rubber degradation are essential for developing efficient microbial degradation strategies for rubber waste materials by this group of actinomycetes.

KEYWORDS cAMP receptor protein, CRP, global regulator, *Gordonia*, latex clearing protein, poly(*cis*-1,4-isoprene) degradation

Citation de Witt J, Oetermann S, Parise M, Parise D, Baumbach J, Steinbüchel A. 2020. Global regulator of rubber degradation in *Gordonia polyisoprenivorans* VH2: identification and involvement in the regulation network. Appl Environ Microbiol 86:e00774-20. https://doi.org/10.1128/AEM.00774-20.

Editor Shuang-Jiang Liu, Chinese Academy of Sciences

Address correspondence to Alexander Steinbüchel, steinbu@uni-muenster.de.

Received 31 March 2020
Accepted 17 May 2020
Accepted manuscript posted online 22 May 2020
Published 20 July 2020

Re-sequencing and optical mapping reveals misassemblies and real inversions on *Corynebacterium pseudotuberculosis* genomes

Thiago de Jesus Sousa, Doglas Parise, Rodrigo Profeta, **Mariana Teixeira Dornelles Parise**, Anne Cybelle Pinto Gomide, Rodrigo Bentos Kato, Felipe Luiz Pereira, Henrique Cesar Pereira Figueiredo, Rommel Ramos, Bertram Brenig, Artur Luiz da Costa da Silva, Preetam Ghosh, Debmalya Barh, Aristóteles Góes-Neto and Vasco Azevedo

Contribution: I contributed to the assembly, annotation and synteny analysis processes as well as the revision of the manuscript.

Status: Published

SCIENTIFIC REPORTS
nature research

OPEN

Re-sequencing and optical mapping reveals misassemblies and real inversions on *Corynebacterium pseudotuberculosis* genomes

Thiago de Jesus Sousa[1], Doglas Parise[1], Rodrigo Profeta[1], Mariana Teixeira Dornelles Parise[1], Anne Cybelle Pinto Gomide[1], Rodrigo Bentos Kato[1], Felipe Luiz Pereira[2], Henrique Cesar Pereira Figueiredo[2], Rommel Ramos[3], Bertram Brenig[4], Artur Luiz da Costa da Silva[3], Preetam Ghosh[5], Debmalya Barh[6], Aristóteles Góes-Neto[1] & Vasco Azevedo[1*]

The number of draft genomes deposited in Genbank from the National Center for Biotechnology Information (NCBI) is higher than the complete ones. Draft genomes are assemblies that contain fragments of misassembled regions (gaps). Such draft genomes present a hindrance to the complete understanding of the biology and evolution of the organism since they lack genomic information. To overcome this problem, strategies to improve the assembly process are developed continuously. Also, the greatest challenge to the assembly progress is the presence of repetitive DNA regions. This article highlights the use of optical mapping, to detect and correct assembly errors in *Corynebacterium pseudotuberculosis*. We also demonstrate that choosing a reference genome should be done with caution to avoid assembly errors and loss of genetic information.

Next Generation Sequencing (NGS) platforms provide an exponential increase in the amount of data produced in a single assay (high-throughput data). This approach provided the scientific community with the ability to sequence more genomes at reduced costs. The NGS platforms perform the sequencing through different technologies, which were developed by different companies, such as 454 GS FLX system (Roche)[1]; Hiseq paired-end (Illumina)[2]; Ion Torrent PGM (Life Technologies)[3]; PacBio sequel system(Pacific Biosciences); and MinION (Oxford Nanopore)[4]. From these, thousands of genomic projects were created to sequence Bacteria, Archaea, and Eukarya species, viruses, and metagenomes[5].

The main database of these sequences is GenBank maintained by the National Center for Biotechnology Information (NCBI), which in September 2018, contained 153,992 bacterial genomes, most of these being drafts, and only 11,103 sequences (7%) were complete genome sequences. Furthermore, the complete sequences still might have misassemblies due to the presence of repetitive regions, such as ribosomal RNA (rRNA), transposases, phage regions, and plasmids[6]. These errors bias future studies and inferences, such as in comparative genomic or structural genomic analyses, and even ordering of phylogenetically related genomes[7]. Thus, obtaining a more precise and accurate complete genome sequence of an organism is fundamental to understanding its biological and evolutionary characteristics[7].

The assembly problem persists even with the increase in the reads size, sequencing quality, and updates of *de novo* assembly algorithms. Another limiting factor to the increase of complete sequences is the lack of trained professionals. However, approaches to support this process have been gaining prominence[8]. For example, the use of SSPACE[9] software to use paired-end reads to create a consensus sequence and perform scaffolding of contigs. Similarly, MapRepeat[10] and riboSeed[11] try to solve the repetitive region's problem.

[1]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [2]National Reference Laboratory for Aquatic Animal Diseases (AQUACEN) of Ministry of Agriculture, Livestock and Food Supply, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [3]Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil. [4]Institute of Veterinary Medicine, University Göttingen, Göttingen, Germany. [5]Department of Computer Science, Virginia Commonwealth University, Richmond, United States. [6]Institute of Integrative Omics and Applied Biotechnology, Nonakuri West Bengal, India. *email: vasco@icb.ufmg.br

Analysis of the microarray gene expression for breast cancer progression after the application modified logistic regression

Francielly Morais-Rodrigues, Rita Silvério-Machado, Rodrigo Bentes Kato, Diego Lucas Neres Rodrigues, Juan Valdez-Baez, Vagner Fonseca, Emmanuel James San, Lucas Gabriel Rodrigues Gomes, Roselane Gonçalves dos Santos, Marcus Vinicius Canário Viana, Joyce da Cruz Ferraz Dutra, **Mariana Teixeira Dornelles Parise**, Doglas Parise, Frederico F. Campos, Sandro J. de Souza, José Miguel Ortega, Debmalya Barh, Preetam Ghosh, Vasco A. C. Azevedo and Marcos A. dos Santos

Contribution: I contributed to the manuscript revision.

Status: Published

Gene 726 (2020) 144168

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Analysis of the microarray gene expression for breast cancer progression after the application modified logistic regression

Francielly Morais-Rodrigues[a,*,1], Rita Silvério-Machado[a], Rodrigo Bentes Kato[a], Diego Lucas Neres Rodrigues[a], Juan Valdez-Baez[a], Vagner Fonseca[a,b], Emmanuel James San[b], Lucas Gabriel Rodrigues Gomes[a], Roselane Gonçalves dos Santos[a], Marcus Vinicius Canário Viana[a,c], Joyce da Cruz Ferraz Dutra[a], Mariana Teixeira Dornelles Parise[a], Doglas Parise[a], Frederico F. Campos[d], Sandro J. de Souza[e], José Miguel Ortega[a], Debmalya Barh[f], Preetam Ghosh[g], Vasco A.C. Azevedo[a], Marcos A. dos Santos[d]

[a] Institute of Biological Sciences, Federal University of Minas Gerais, Brazil. Av. Antônio Carlos, 6627, Belo Horizonte, MG 31270-901, Brazil
[b] KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), College of Health Sciences, University of KwaZulu-Natal, Durban 4001, South Africa
[c] Federal University of Pará, UFPA, Brazil
[d] Department of Computer Science, Federal University of Minas Gerais, Brazil Av Antônio Carlos, 6627, Belo Horizonte, MG 31270-901, Brazil
[e] Brain Institute, Federal University of Rio Grande d oNorte, Brazil
[f] Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal 721172, India
[g] Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA

ARTICLE INFO

Keywords:
Tumor classification
Samples
New logistic regression-based model
GRN
TFs
MCF-7
Oncogenic

ABSTRACT

Methods based around statistics and linear algebra have been increasingly used in attempts to address emerging questions in microarray literature. Microarray technology is a long-used tool in the global analysis of gene expression, allowing for the simultaneous investigation of hundreds or thousands of genes in a sample. It is characterized by a low sample size and a large feature number created a non-square matrix, and by the incomplete rank, that can generate countless more solution in classifiers. To avoid the problem of the 'curse of dimensionality' many authors have performed feature selection or reduced the size of data matrix. In this work, we introduce a new logistic regression-based model to classify breast cancer tumor samples based on microarray expression data, including all features of gene expression and without reducing the microarray data matrix. If the user still deems it necessary to perform feature reduction, it can be done after the application of the methodology, still maintaining a good classification. This methodology allowed the correct classification of breast cancer sample data sets from Gene Expression Omnibus (GEO) data series GSE65194, GSE20711, and GSE25055, which contain the microarray data of said breast cancer samples. Classification had a minimum performance of 80% (sensitivity and specificity), and explored all possible data combinations, including breast cancer subtypes. This methodology highlighted genes not yet studied in breast cancer, some of which have been observed in Gene Regulatory Networks (GRNs). In this work we examine the patterns and features of a GRN composed of transcription factors (TFs) in MCF-7 breast cancer cell lines, providing valuable information regarding breast cancer. In particular, some genes whose $\alpha i *$ associated parameter values revealed extreme positive and negative values, and, as such, can be identified as breast cancer prediction genes. We indicate that the PKN2, MKL1, MED23, CUL5 and GLI genes demonstrate a tumor suppressor profile, and that the MTR, ITGA2B, TELO2, MRPL9, MTTL1, WIPI1, KLHL20, PI4KB, FOLR1 and SHC1 genes demonstrate an oncogenic profile. We propose that these may serve as potential breast cancer prediction genes, and should be prioritized for further clinical studies on breast cancer. This new model allows for the assignment of values to the $\alpha i *$ parameters associated with gene

Abbreviations: BC, Breast cancer; BGRMI, Bayesian Gene Regulatory Model Inference; EGF, Epidermal growth factor; GEO, Gene Expression Omnibus; GRNs, Gene Regulatory Networks; HER2, Human Epidermal Growth Factor Receptor 2; HRG, Cells stimulated with heregulin; LumA, Luminal A; LumB, Luminal B; NCBI, National Center for Biotechnology Information; RNA-Seq, RNA sequencing; TFs, Transcription factors; TNBC, Triple Negative Breast Cancer

* Corresponding author.
E-mail address: franrodriguesdacosta@gmail.com (F. Morais-Rodrigues).
[1] First author.

https://doi.org/10.1016/j.gene.2019.144168
Received 5 July 2019; Received in revised form 21 September 2019; Accepted 11 October 2019
Available online 21 November 2019
0378-1119/

***In silico* identification of biofilm regulators and their co-expression networks in *Leptospira biflexa* using RNA-seq data**

Mariana Parise, Doglas Parise, Artur Cancio, Rodrigo Kato, Paula Ristow and Vasco Azevedo

Contribution: I contributed to the running of the co-expression analysis and the preparation of the original manuscript.

Status: Forthcoming

In silico identification of biofilm regulators and their co-expression networks in *Leptospira biflexa* using RNA-seq data

Mariana Parise*, Doglas Parise*, Artur Cancio, Rodrigo Kato, Paula Ristow, Vasco Azevedo

*First authorship shared

1 Laboratory of Cellular and Molecular Genetics (LGCM), Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
2 Laboratory of Bacteriology and Health (LABACS), Institute of Biology, Federal University of Bahia, Salvador, Bahia, Brazil.
3 Postgraduate Program in Microbiology, Institute of Biology, Federal University of Bahia, Salvador, Brazil.
E-mail: paula.ristow@ufba.br (PR)

Abstract:
Biofilms improve the survival of microorganisms in hostile environments and are related to various medical conditions. Leptospires form biofilms *in vitro* and in the environment; the possible involvement of the biofilm in pathogenesis has already been identified. However, biofilm regulatory mechanisms in leptospires are notably unknown. Here, we identified transcriptional regulators involved in biofilm formation in *Leptospira biflexa* and described the co-expression networks for these regulators. We detected 13 regulators influencing mature biofilm and 12 regulators influencing late biofilm. Amongst them, ten regulators of the TetR family were differentially expressed in the biofilm condition. Our results provide an initial comprehension of the regulatory mechanisms in Leptospira biofilm and give directions for future experimental assays.

Keywords: transcriptional regulators, regulatory networks, leptospirosis

Unravelling *Anoxybacillus* diversity, phylogeny and function

Júnia Schultz, **Mariana T. D. Parise**, Doglas Parise, Laenne Medeiros, Thiago J. Sousa, Fabrício Araújo, Jan Baumbach, Rommel T. J. Rammos, Bertram Brening, Vasco A. C. Azevedo, Aristóteles Góes-Neto and Alexandre Rosado

Contribution: I have been contributing to the performance of phylogenetic and phylogenomics analyses.

Status: Forthcoming

Unravelling *Anoxybacillus* diversity, phylogeny and function

SCHULTZ, J.1,2; PARISE, M. T. D.3; PARISE, D.3; MEDEIROS, L. G.2; SOUZA, T. J.3, ARAUJO, F.4; BAUMBACH, J.5; RAMOS, R. T. J.4; BRENIG, B.6; AZEVEDO, V. A. C.3; GÓES NETO, A.3*; ROSADO, A. S.1,2*

1 Microbial Ecogenomics and Biotechnology Laboratory, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Makkah, 23955, Saudi Arabia
2 Laboratory of Molecular Microbial Ecology, Institute of Microbiology, Federal University of Rio de Janeiro, Rio de Janeiro, 21941-902, Brazil
3 Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, 31270-901, Brazil
4 Institute of Biological Sciences, Federal University of Pará, Belém, Brazil
5 Chair of Computational Systems Biology, University of Hamburg, 22607 Hamburg, Germany
6 Department of Molecular Biology of Livestock, Institute of Veterinary Medicine, Georg August University, Göttingen, 37077, Germany

Corresponding authors:
* alexandre.rosado@kaust.edu.sa
* arigoesneto@gmail.com

ABSTRACT
Keywords: Anoxybacillus; Comparative genomics; Phylogeny; Antarctica; Polar volcano.

1 INTRODUCTION

The genus *Anoxybacillus* belongs to the *Bacillaceae* family and Firmicutes phylum. In contrast to *Bacillus* and *Geobacillus*, *Anoxybacillus* is a relatively new genus that was proposed in the year 2000 (Pikuta et al., 2000). *Anoxybacillus flavithermus* was the first species of the *Anoxybacillus* genus and was formerly known as *Bacillus flavithermus*. The strain was discovered in a New Zealand hot spring and was characterized as a gram-positive and endospore-forming, facultative anaerobe microbe, presenting intense yellow pigmentation and capacity to grow in a thermotolerant range (37-70°C) (Heinen et al., 1982). Decades later, Pikuta and colleagues (2000) isolated a new anaerobic strain (K1T) from animal manure and, based on the phenotypic characteristics (16S rRNA, DNA–DNA hybridization analyses), *Anoxybacillus* was proposed as a new genus of the *Bacillaceae* family and the strain was named *Anoxybacillus pushchinensis* KT1. When both were phylogenetically compared, *A. pushchinensis* K1T and *B. flavithermus* clustered together and were distinct from *Bacillus*. Therefore, *B. flavothermus* was reclassified as *Anoxybacillus flavithermus* (Pikuta et al., 2000).