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

Ph.D. Thesis

Reconstructing and assessing corynebacterial Gene Regulatory Networks

Ph.D. student: Doglas Parise
SUPERVISORS: Prof. Dr. Vasco Ariston de Carvalho Azevedo
Prof. Dr. Jan Baumbach
CO-SUPERVISOR: Dr. Rodrigo Bentes Kato

BELO HORIZONTE
MAY - 2021

Doglas Parise

Reconstructing and assessing corynebacterial Gene Regulatory Networks

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, Doglas.
Reconstructing and assessing corynebacterial gene regulatory networks [manuscrito] / Doglas Parise. – 2021.
97 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. Elementos Reguladores de Transcrição. 3. Fatores de Transcrição. 4. Corynebacterium. 5. Regulação da 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

Doglas Parise

131/2021
entrada
2º/2016
CPF:
836.962.020-53

Às nove horas do dia **26 de maio de 2021**, reuniu-se, através do aplicativo Zoom, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Reconstructing and assessing corynebacterial Gene Regulatory Networks"**, requisito para obtenção do grau de Doutor 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 ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	Indicação
Dr. Vasco Ariston de Carvalho Azevedo	Universidade Federal de Minas Gerais	Aprovado
Dr. Rodrigo Bentes Kato	Universidade Federal de Minas Gerais	Aprovado
Dr. Jan Baumbach	University of Hamburg	Aprovado
Dr. Richard Röttger	University of Southern Denmark	Aprovado
Dr. Josch Konstantin Pauling	Technical University of Munich	Aprovado
Dra. Raquel Cardoso de Melo Minardi	Universidade Federal de Minas Gerais	Aprovado
Dr. Luiz Eduardo Vieira Del Bem	Universidade Federal de Minas Gerais	Aprovado
Dr. Siomar de Castro Soares	Universidade Federal do Triângulo Mineiro	Aprovado
Dr. Thiago Luiz de Paula Castro	Universidade Federal da Bahia	Aprovado

Pelas indicações, o candidato foi considerado: **Aprovado**

O resultado final foi comunicado publicamente ao candidato 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, 26 de maio de 2021.

Dr. Vasco Ariston de Carvalho Azevedo - Orientador

Universidade Federal de Minas Gerais

Dr. Rodrigo Bentes Kato - Coorientador

Universidade Federal de Minas Gerais

Dr. Jan Baumbach - Coorientador

University of Hamburg

Dr. Richard Röttger

University of Southern Denmark

Dr. Josch Konstantin Pauling

Technical University of Munich

Dra. Raquel Cardoso de Melo Minardi

Universidade Federal de Minas Gerais

Dr. Luiz Eduardo Vieira Del Bem

Universidade Federal de Minas Gerais

Dr. Siomar de Castro Soares

Universidade Federal do Triângulo Mineiro

Dr. Thiago Luiz de Paula Castro

Universidade Federal da Bahia

"If you’re not doing as well as you’d like to be
doing right now, then there’s something you don’t know."

T. Harv Eker

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors Prof. Vasco Azevedo, Prof. Jan Baumbach and Dr. Rodrigo Kato for the opportunity, guidance and support in my work. I am grateful to Prof. Andreas Tauch, Prof. Aristóteles Góes-Neto, Dr. Josch Pauling, Dr. Markus List, Dr. Flávia Aburjaile, Dr. Anne Gomide and Dr. Evans Kataka for their collaboration.

Furthermore, I would like to express gratitude to my wife and work partner for full-time support, discussions, friendship and partnership.

I would also like to thank my colleagues from LGCM, ExBio and LBMCF. I offer a special thanks to Izabela, Olga T., Lion, Drazen, Julian and Marisol for their technical and scientific support during my work development. Additionally, I extend my gratitude to Marcus, Thiago, Sandeep, Arun, Jamal, Francielly, Renata, Rodrigo P., Rodrigo D., Alessandra, Alfonso, Amit, Marisol, Manuela, Gihanna, Nina, Monica, Olga L., Chit Tong, Tobias, Lorenzo, Weilong, Philipp, Mhaned, Zakaria, Amir, Michael, Kundai, Gabriel Q. and Marcelo T. for their friendship.

I would also like to acknowledge the kind support of Sheila, Natália, Fernanda, Tiago, Leonie and Martina.

I am thankful to my family, particularly my parents, Imgrid and Moacir, and my brother Diogo. I am also very grateful to my parents-in-law, Daniela and Elvis, my friends Sílvia, Elizete, Neva, Pedro E., João M., and Carolina B., and my English teacher Samantha for their support.

I want to thank the Brazilian funding agencies CAPES, CNPQ and FAPEMIG for their financial support as well as UFMG and TUM for supporting my work and providing safe and sound work environments.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

GRN	Transcriptional Gene Regulatory Network
TF	Transcription Factor
TG	Target Gene
TFBS	Transcription Factor Binding Site
TRN	Transcriptional Regulatory Network
EMSA	Electrophoretic Mobility Shift Assay
ChIP-chip	Chromatin Immunoprecipitation and DNA microarray
ChIP-seq	Chromatin Immunoprecipitation and sequencing
BBHs	Bi-directional Best BLAST hit
DIAMOND	Double Index AlignMent Of Next-generation sequencing Data
HMM	Hidden Markov Model
PWM	Position Weight Matrices
CMRN	*Corynebacterium*, *Mycobacterium*, *Nocardia* and *Rhodococcus*
UL	Ulcerative Lymphangitis
OSD	Oedematous Skin Disease
CLA	Caseous Lymphadenitis
DEG	Differentially Expressed Gene
HGT	Horizontal Gene Transfer

RESUMO

O gênero *Corynebacterium* é um grupo heterogêneo de organismos de relevância biotecnológica, médica e veterinária. A sua diversidade permite aos organismos viverem em uma grande variedade de ambientes. Os organismos devem ser capazes de ajustar sua maquinaria de expressão gênica para se adaptar a ambientes hostis. A regulação transcricional desempenha um papel crucial nesse processo; permitindo à bactéria mudar rapidamente o conjunto de genes sendo expressos para lidar com os desafios do ambiente. Compreender os mecanismos envolvidos na regulação transcricional é um passo fundamental para compreender como os organismos podem adaptar-se e desenvolver-se. Muitos estudos têm focado em revelar as redes regulatórias gênicas (RRGs) transcricionais de organismos bacterianos; contudo, nosso conhecimento está limitado a poucos organismos modelo com RRGs experimentais. *Escherichia coli* é o organismo bacteriano mais estudado e, mesmo assim, é estimado que menos de 30% das interações regulatórias entre fatores de transcrição (FTs) e genes alvo (GAs) é conhecida. Neste contexto, abordagens computacionais permitem aos pesquisadores revelar as RRGs de muitos organismos com base na conservação evolutiva das RRGs de organismos modelo. Considerando o gênero *Corynebacterium*, pouco se conhece do seu repertório regulatório transcricional. Este gênero é de crucial importância para a medicina, veterinária e biotecnologia por conter patógenos que afetam a saúde humana e animal e também organismos produtores de aminoácidos. CoryneRegNet é o banco de dados de referência em regulação transcricional do gênero *Corynebacterium* desde 2006. Nesta tese, nós estendemos o conhecimento regulatório transcricional do gênero *Corynebacterium* e avaliamos a consistência da RRG de *Corynebacterium glutamicum* com seus dados de expressão gênica. No primeiro artigo científico, nós apresentamos a sétima versão do CoryneRegNet, o qual atualmente armazena RRGs transcricionais de 225 organismos do gênero *Corynebacterium*, representando um aumento de 20 vezes o número de organismos deste gênero com RRGs conhecidas. O conhecimento regulatório gerado combinado com dados oriundos da literatura resultou no primeiro artigo de revisão da RRG transcricional de *Corynebacterium pseudotuberculosis*, na qual nós apresentamos os mecanismos transcricionais conhecidos deste organismo quando submetido a estresse osmótico, ácido, escassez de ferro e térmico. No terceiro artigo científico, nós avaliamos a consistência entre dados de expressão gênica e a RRG de *C. glutamicum* aplicando um modelo de consistência assinada. Nossos resultados mostram que a RRG de *C. glutamicum* não é mais consistente que RRGs aleatórias, sugerindo que dados ômicos de outros elementos regulatórios como regulação pós-transcricional e traducional, deveriam ser integrados em futuros estudos de RRGs neste organismo. Nós concluímos que estamos apenas começando a compreender o panorama regulatório do gênero *Corynebacterium* e que novas camadas de regulação deveriam ser integradas às RRGs para reconstruir redes mais confiáveis para este gênero.

Palavras-chave: redes regulatórias gênicas transcricionais, regulação transcricional, fatores de transcrição, elementos regulatórios, *Corynebacterium*.

ABSTRACT

The *Corynebacterium* genus is a very heterogeneous group of organisms of biotechnological, medical and veterinary relevance. Its diversity allows its organisms to live in a wide range of environments. The organisms must quickly adapt their gene expression machinery to acclimate to hostile environments. Transcriptional regulation plays a crucial role in this process; it allows bacteria to quickly change the set of genes expressed to cope with the environment's challenges. Understanding the mechanisms underlying transcription regulation is a crucial step in understanding how organisms can adapt and thrive. Several studies have focused on unraveling the transcriptional gene regulatory networks (GRNs) of bacterial organisms; however, our knowledge is limited to a few model organisms with experimentally verified GRNs. Note that *Escherichia coli* is the best-studied bacterial organism and yet it is estimated that our knowledge corresponds to less than 30% of regulatory interactions between transcription factors (TFs) and target genes (TGs). In this context, computational approaches allow researchers to unravel the GRNs of many organisms based on the evolutionary conservation of the model organisms' networks. Taken into account the *Corynebacterium* genus, little is known regarding their transcriptional regulatory repertory. This genus is greatly relevant for medicine, veterinary and biotechnology, comprising pathogens that affect human and animal health as well as amino acid producer organisms. CoryneRegNet has been the reference database for corynebacterial transcriptional regulatory knowledge since 2006. Here, we extend the transcriptional regulatory knowledge of the *Corynebacterium* genus and assess the consistency of the *Corynebacterium glutamicum* GRN with its gene expression data. In the first research article, we present the seventh version of CoryneRegNet, which now holds transcriptional GRNs for 225 corynebacterial organisms, increasing by twenty times the number of organisms with known GRNs of this genus. This regulatory knowledge combined with literature research resulted in the first review article of the transcriptional GRN of *Corynebacterium pseudotuberculosis*, in which we present the known transcriptional mechanisms of this organism under osmotic, acid, iron-starvation and thermal stress. In the third research article, we assessed the consistency between gene expression data and the *C. glutamicum* GRN by applying a conservative sign consistency model. Our results show that the *C. glutamicum* GRN is not more consistent than random GRNs, suggesting that omics data concerning other regulatory elements, such as post-transcriptional and translational regulation, should be integrated in future GRN studies for this organism. We conclude that we have just begun to understand the *Corynebacterium* genus' regulatory landscape and that new layers of regulation should be integrated into the GRNs to reconstruct more reliable networks for this genus.

Keywords: transcriptional gene regulatory networks, transcription regulation, transcription factors, regulatory elements, *Corynebacterium*.

I. STRUCTURE

This manuscript is organized into 6 sections and an appendix. Section I presents the structure of this manuscript. In section II, we first introduce gene expression and the transcriptional regulatory mechanisms. We then present the importance of key transcriptional regulatory elements, such as transcription factors (TFs), sigma factors and transcription factor binding sites (TFBSs). We discuss the main experimental and computational approaches to reconstruct transcriptional gene regulatory networks (GRNs) and present databases holding bacterial GRNs. Furthermore, we introduce the *Corynebacterium* genus, the transcriptional GRN reference database for this genus CoryneRegNet and methods to assess the consistency of bacterial GRNs.

Section III presents the objectives of this thesis. In section IV, we present (i) the transcriptional gene regulatory networks of the *Corynebacterium* genus and (ii) the assessment of the consistency between the *C. glutamicum* GRN and its gene expression data. The first presents CoryneRegNet 7, the latest version of the reference database for GRNs of the *Corynebacterium* genus, and a review of the transcriptional GRN of *C. pseudotuberculosis*, which combines literature data and the predicted GRNs from CoryneRegNet. CoryneRegNet 7 is published in the journal Scientific Data and the review is published in the journal Microorganisms. The second assesses the consistency between *C. glutamicum* GRN and its gene expression data available on the GEO database. This research article is published in the journal Network and Systems Medicine.

In section V, we discuss the new functionalities of CoryneRegNet, including the new transfer pipeline and its improvements and limitations. We also discuss the currently available knowledge concerning *C. pseudotuberculosis*' regulatory interactions and the high inconsistency found when assessing the consistency between the *C. glutamicum* GRN and its gene expression data. In section VI, we conclude that (i) the new version of CoryneRegNet extended the knowledge regarding transcriptional GRNs of *Corynebacterium* genus, (ii) little is known concerning the transcriptional regulatory mechanisms of *C. pseudotuberculosis* and (iii) new regulatory layers should be integrated into the GRNs. This section also presents the perspectives of this thesis. Finally, the appendix presents published, submitted and forthcoming research articles.

II. BACKGROUND & MOTIVATION

II.1 Gene expression

Gene expression is the process of making a functional product from the information stored in genes. The main steps in this process are transcription and translation. During the transcription step, the DNA sequence of a gene is transcribed to become an RNA molecule. Then, the RNA molecule is translated to become a protein (BERVOETS; CHARLIER, 2019). In bacteria, transcription and translation are highly regulated and linked in time and space (MCGARY; NUDLER, 2013). This tight regulation is crucial to bacteria because it prevents them from wasting energy and resources; consequently, the organism maintains its potential virulence and fitness in a continually changing and challenging environment (BERVOETS; CHARLIER, 2019; FREYRE-GONZÁLEZ *et al.*, 2013).

The main techniques that identify gene expression include microarrays (SCHENA *et al.*, 1995) and RNA-Seq (NAGALAKSHMI *et al.*, 2008). In the microarrays, the DNA of target organisms is chemically linked to probes on a substrate chip; the fluorescence emission is then measured to obtain the gene expression values of the genes of interest (LUCCHINI; THOMPSON; HINTON, 2001; SCHENA *et al.*, 1995). In the RNA-Seq method, the complete set of RNA of an organism can be unraveled and quantified through converting RNA into cDNA, sequencing it and computationally analyzing the transcript sequences (NAGALAKSHMI *et al.*, 2008).

II.2 Transcriptional regulation

Transcription regulation is considered the main regulatory step for most bacterial genes (BROWNING; BUSBY, 2004, 2016). In this step, bacteria can quickly adjust the molecular machinery in response to changes in the environment (BALLEZA *et al.*, 2009; MADIGAN *et al.*, 2014). The transcription process occurs in three stages: transcription initiation, elongation and termination. The transcription initiation is considered the key step of transcription to keep homeostasis (BROWNING; BUSBY, 2004). In this process, sigma factors and transcription factors (TFs) are key transcriptional regulatory elements (BROWNING; BUSBY, 2016; MADIGAN *et al.*, 2014; PAGET, 2015).

Sigma factors mediate transcription by assisting the RNA Polymerase to identify and open the DNA strands in the promoter region, interacting with TFs, and then acting during the early stages of RNA synthesis. They can be categorized as either essential or alternatives. The essential one is also known as the housekeeping sigma factor and is required to express most genes. Alternative sigma factors are required for the expression of genes in response to stress conditions and developmental processes (BERVOETS; CHARLIER, 2019; DAVIS; KESTHELY; FRANKLIN, 2017; PAGET, 2015). TFs are regulatory proteins that mediate gene expression by recognizing and binding to specific motifs called transcription factor binding sites

(TFBSs) in the promoter region of their target genes (TGs). Their interaction with the TFBSs can activate or repress gene expression (BERVOETS; CHARLIER, 2019; PABO; SAUER, 1992). The activation or repression of a gene or operon may also be modulated by the concentration of its substrate, product or a chemical derivative of either of these (KREBS *et al.*, 2013). Figure 1 presents an overview of TF-driven regulation. For comprehensive reviews see (BALLEZA *et al.*, 2009; BERVOETS; CHARLIER, 2019; BROWNING; BUSBY, 2016).

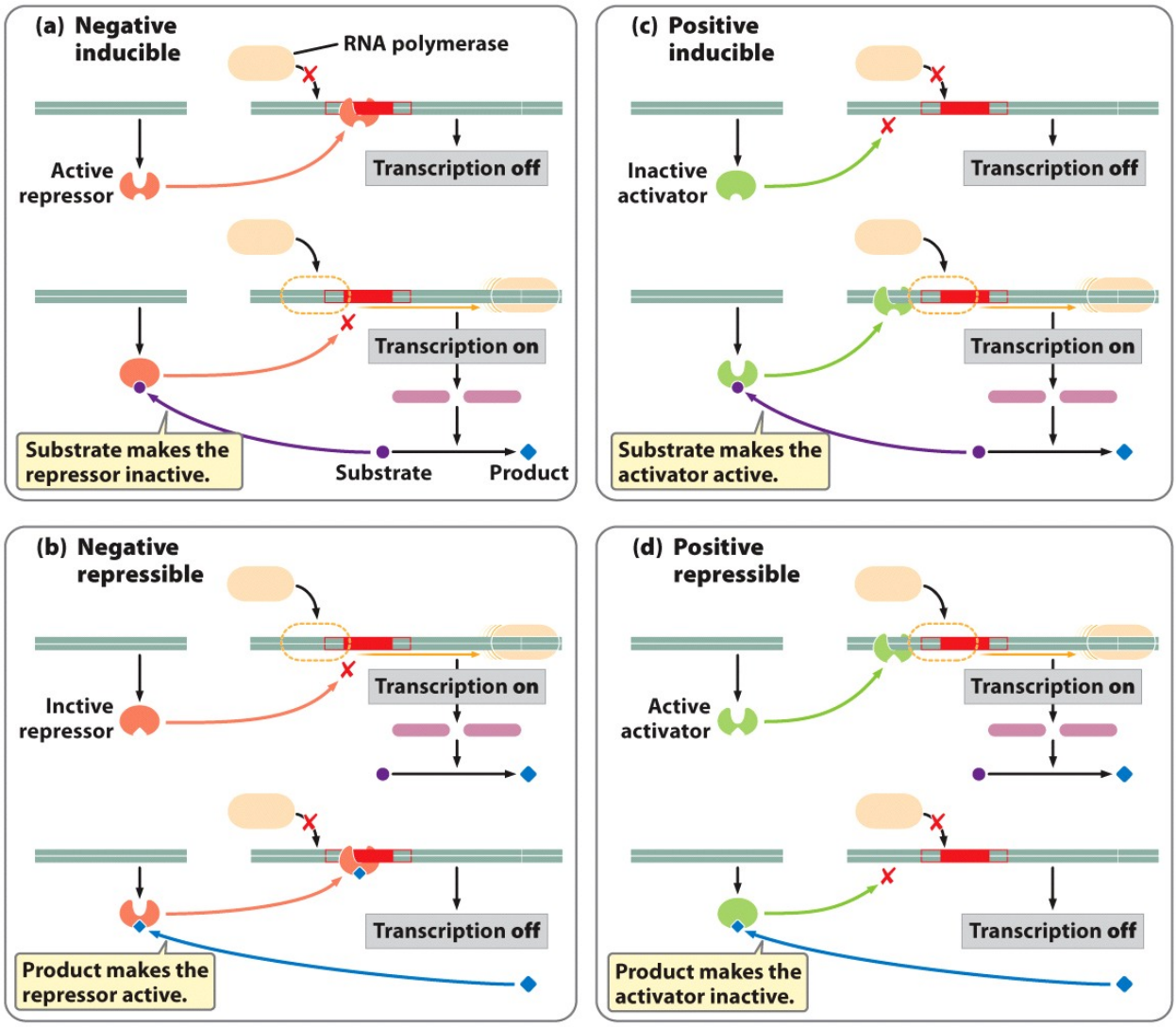


Figure 1 - Positive and negative control with inducible and repressible control. a) Negative inducible: the TF (repressor) prevents RNA polymerase from binding, blocking transcription. The presence of a specific substrate inactivates the TF, initiating transcription. b) Negative repressible: the TF (repressor) is normally inactive; the presence of a specific product activates the TF, blocking transcription. c) Positive inducible: the TF (activator) is normally inactive, blocking transcription and the presence of a specific subtract activates the TF, activating transcription. d) Positive repressible: the TF (activator) is normally active and activating the transcription process; the presence of a specific substrate inactivates the TF, blocking transcription.
Source: Adapted from: (PIERCE, 2013).

In addition to the transcription regulation mechanisms presented here, other mechanisms influence gene expression, such as small RNAs (HÖR; GORSKI; VOGEL, 2018), promoter modifications and methylation (BROWNING; BUSBY, 2016). However, the coactivity of the mentioned molecular entities (RNA polymerase, promoter, sigma factors, TF and TFBS) composes the minimal transcriptional regulatory system (BALLEZA *et al.*, 2009).

II.3 Transcriptional Gene Regulatory Networks

The set of transcriptional regulatory interactions in the cell can be represented as transcriptional Gene Regulatory Networks (GRNs), also known as Transcriptional Regulatory Networks (TRNs), to understand the mechanisms that coordinate the ability of bacteria to quickly adapt to environmental changes and establish cell homeostasis. GRNs are computationally modeled as direct graphs where edges (or arcs) represent regulatory interactions and vertices represent both regulatory proteins (TFs or sigma factors) and TGs (BABU; LANG; ARAVIND, 2009; BAUMBACH; BRINKROLF; CZAJA; *et al.*, 2006; THOMPSON; REGEV; ROY, 2015).

The experimental reconstruction of GRNs is a complex, laborious and expensive task (BAUMBACH; RAHMANN; TAUCH, 2009; GRAINGER; LEE; BUSBY, 2009; LEYN *et al.*, 2016). It consists of identifying both the expression profiles, using techniques such as microarrays and RNA-Seq, and the TF in each regulatory interaction with its corresponding TFBSs using, among others, the following techniques: Electrophoretic Mobility Shift Assay (EMSA) (GARNER; REVZIN, 1986; HELLMAN; FRIED, 2007); Chromatin Immunoprecipitation and DNA microarray (ChIP-chip) (HORAK; SNYDER, 2002); and Chromatin Immunoprecipitation and sequencing (ChIP-seq) (BARSKI *et al.*, 2007). Experimental GRNs are available on online databases such as CoryneRegNet (PAULING; RÖTTGER; TAUCH; *et al.*, 2012) and Abasy Atlas (IBARRA-ARELLANO *et al.*, 2016) for *Corynebacterium glutamicum*, DBTBS (SIERRO *et al.*, 2008) and Subtiwiki (Zhu and Stülke 2018) for *Bacillus subtilis* and RegulonDB (SANTOS-ZAVALETA *et al.*, 2019) for *Escherichia coli*. For comprehensive reviews see (HECKER *et al.*, 2009; MERCATELLI *et al.*, 2020; THOMPSON; REGEV; ROY, 2015).

The computational reconstruction of GRNs allowed researchers to unravel these networks based on the organism's sequence. To achieve this, researchers transferred the knowledge acquired from well-studied organisms (also known as model organisms) to target organisms (BAUMBACH, 2010; KILIÇ; ERILL, 2016; MADAN BABU; TEICHMANN; ARAVIND, 2006). Currently, there are three main approaches to transfer GRNs: (i) conserved genes, (ii) conserved binding sites and (iii) by combining both. The conserved genes approach was the first strategy to transfer the regulatory interactions by assuming that orthologous TFs regulate their orthologous TGs (MADAN BABU; TEICHMANN; ARAVIND, 2006). The orthology detection may be performed using bi-directional best BLAST hit (BBHs) (ALTSCHUL *et al.*, 1997),

OrthoFinder (EMMS; KELLY, 2019), OrthoMCL (LI; STOECKERT; ROOS, 2003), DIAMOND (double index alignment of next-generation sequencing data) (BUCHFINK; XIE; HUSON, 2015) or phylogenetic trees (PRICE; DEHAL; ARKIN, 2007). The conserved binding site approach converts the experimentally identified TFBSs into computational models for profile-based prediction of regulatory interactions. It generally assumes that there is a regulatory interaction if the TFBS of an orthologous TF is conserved. It may lead to the discovery of new regulatory interactions that are not evolutionary conserved between model and target organisms; however, orthologous TFs may regulate orthologous TGs binding different TFBSs (BAUMBACH, 2010; KILIÇ; ERILL, 2016; VENANCIO; ARAVIND, 2009). TFBS prediction is usually performed using Hidden Markov models (HMMs) (EDDY, 1998; KROGH *et al.*, 1994; WHEELER; EDDY, 2013), position weight matrices (PWMs) (BECKSTETTE *et al.*, 2006; GRIBSKOV; MCLACHLAN; EISENBERG, 1987) or using motif discovery algorithms (BAILEY *et al.*, 2009; NOVICHKOV *et al.*, 2010).

The approach combining both orthologous genes and conserved binding sites first searches for orthologous TFs and their orthologous TGs. It then searches for the conserved TFBSs of those TFs in the promoter region of the orthologous TGs. The regulatory interaction is considered conserved if the TF, the TG and the TFBS are preserved in the target genome. Examples of the combined approach include the bioinformatic tool Regulogger (ALKEMA; LENHARD; WASSERMAN, 2004), the work of Espinosa et al. (ESPINOSA *et al.*, 2005) and the work of Baumbach et al. (BAUMBACH; RAHMANN; TAUCH, 2009). Some works have been focused on functional enrichment analysis based on literature search and manual curation that has been applied and combined with homology-based methods (MERCATELLI *et al.*, 2020; MÉTRIS *et al.*, 2017). For comprehensive reviews on computational approaches, see (BAUMBACH, 2010; KILIÇ; ERILL, 2016; MERCATELLI *et al.*, 2020; THOMPSON; REGEV; ROY, 2015).

The application of computational approaches resulted in several online databases that store and make available computational reconstructed GRNs. Among them, we can highlight:

- RegPredict (NOVICHKOV *et al.*, 2010) used a conserved binding sites approach to predict the GRNs of phylogenetically related organisms of representative species from the taxonomic groups available on MicrobesOnline database (DEHAL *et al.*, 2010). This method supports the GRN prediction of up to 15 organisms simultaneously;
- EHECRegNet (PAULING; RÖTTGER; NEUNER; *et al.*, 2012) applied a mixed approach developed by Baumbach and collaborators (BAUMBACH; RAHMANN; TAUCH, 2009). EHECRegNet used *E. coli* k-12 as a model organism to reconstruct the GRNs of 16 human pathogenic *E. coli*;
- CoryneRegNet 6.0 (PAULING; RÖTTGER; TAUCH; *et al.*, 2012) used the same mixed approach to transfer the GRN from *C. glutamicum* ATCC 13032 to 10 organisms of the *Corynebacterium* genus;

- CMRegNet (ABREU *et al.*, 2015) also applied the mixed approach developed by Baumbach and collaborators (BAUMBACH; RAHMANN; TAUCH, 2009), using *C. glutamicum* ATCC 13032 and *M. tuberculosis* H37Rv as model organisms to predict the GRNs of 18 organisms of the *Corynebacterium* and *Mycobacterium* genera;
- SalmoNet (MÉTRIS *et al.*, 2017) combined high throughput data, computational tools and literature data to integrate and generate transcriptional, metabolic and protein-protein interaction networks for 10 strains of *Salmonella enterica*;
- PRODORIC (ECKWEILER *et al.*, 2018) applied a conserved binding site approach to predict the GRNs of 27 organisms; the last version increases its quantity of TFBSs, including new ones derived from high-throughput experiments and the prediction of new TFBSs;
- Abasy Atlas (ESCORCIA-RODRÍGUEZ; TAUCH; FREYRE-GONZÁLEZ, 2020; IBARRA-ARELLANO *et al.*, 2016) is a meta-curated database that uses the experimental regulatory networks of model organisms and literature data to predict the regulatory networks of target organisms.

II.4 *Corynebacterium* genus

The *Corynebacterium*, *Mycobacterium*, *Nocardia* and *Rhodococcus* genera form the CMRN group; this group is part of the order Corynebacteriales. Members of the CMRN group share specific characteristics such as high GC content and cell wall composition (peptidoglycans, arabinogalactans and mycolic acids). The *Corynebacterium* genus is a heterogeneous group of bacteria in terms of shape, oxygen requirements and habitat; additionally, the organisms are of medical, veterinary and biotechnological relevance (DORELLA *et al.*, 2006; GOODFELLOW; JONES, 2015; OLIVEIRA *et al.*, 2017). There are 125 validated *Corynebacterium* species on the DSMZ database (DSMZ, 2021) and, amongst them, more than 50 have their genome sequences available on the NCBI database.

C. glutamicum is a soil bacterium and is considered the model organism of this genus. It has been widely used in the biotechnological industry in the amino acid production for its ability to produce large amounts of L-glutamine and L-lysine by fermentation (GEORGI; RITTMANN; WENDISCH, 2005; IVANOV *et al.*, 2013). The products generated using *C. glutamicum* strains are estimated to have reached a market size of $20.4 billion USD in 2020 (SUN *et al.*, 2019). Due to the importance of its metabolic pathways, the transcriptional regulatory mechanisms have been studied and unraveled (BRINKROLF; BRUNE; TAUCH, 2007; FREYRE-GONZÁLEZ; TAUCH, 2017); its GRN is available on online databases (IBARRA-ARELLANO *et al.*, 2016; PAULING; RÖTTGER; TAUCH; *et al.*, 2012).

Another relevant organism of this genus is the zoonotic pathogen *Corynebacterium pseudotuberculosis*, which causes a great economic impact over livestock farming. This pathogen is the etiological agent of several diseases depending on the biovar. The *equi* biovar

causes pigeon fever or Ulcerative Lymphangitis (UL) in horses, mastitis and ulcerative granulomatous lesions in cattle, and Oedematous Skin Disease (OSD) in buffaloes. The *ovis* biovar causes Caseous Lymphadenitis (CLA) mainly in sheep, goats and humans (DORELLA *et al.*, 2006; SELIM, 2001; YERUHAM *et al.*, 2004). The frequency and spreading of the diseases caused by *C. pseudotuberculosis* have been increasing in the last years mainly in North America (MUÑOZ-BUCIO *et al.*, 2017; PARISE, D. *et al.*, 2018; SPIER; AZEVEDO, 2017), yet there is no efficient treatment to combat this organism (BRUM *et al.*, 2017; COSTA *et al.*, 2011; DROPPA-ALMEIDA; FRANCESCHI; PADILHA, 2018). Despite its zoonotic and agribusiness importance, few studies have analyzed *C. pseudotuberculosis*' transcriptional profile (GOMIDE; DE SÁ; *et al.*, 2018; GOMIDE; IBRAIM; *et al.*, 2018; IBRAIM *et al.*, 2019; PINTO *et al.*, 2014), experimentally studied its regulators (TIWARI *et al.*, 2014) and only three strains have had their GRNs partially unraveled (PAULING; RÖTTGER; TAUCH; *et al.*, 2012). A detailed study of what is known about the *C. pseudotuberculosis* GRN is presented in the second part of Chapter I.

II.5 CoryneRegNet

CoryneRegNet has been the reference GRN database for the *Corynebacterium* genus since 2006. Its first version presented the GRN of the model organism *C. glutamicum*, which was reconstructed based on the integration of experimental studies and literature knowledge (BAUMBACH; BRINKROLF; CZAJA; *et al.*, 2006). CoryneRegNet 2 incorporated binding site prediction analysis and integrated transcriptional data of *C. diphtheriae*, *C. jeikeium* and *C. efficiens* (BAUMBACH; BRINKROLF; WITTKOP; *et al.*, 2006). In the third version, CoryneRegNet integrated the *E. coli* K-12 GRN from RegulonDB and allowed the visualization of a gene's genetic organization in the genome and sequence logos of TFBSs (BAUMBACH *et al.*, 2007). Novelties in the fourth version included gene expression data analysis, homology detection as well as species network comparison and analysis (BAUMBACH, 2007). CoryneRegNet 5 was an internal release in which *C. glutamicum* was used as model organism to transfer its GRN to *C. diphtheriae*, *C. jeikeium* and *C. efficiens*. In the sixth version, the authors transferred the *C. glutamicum* GRN to 10 *Corynebacterium* organisms, thus, holding the GRNs of 11 organisms of the *Corynebacterium* genus and *E. coli* K-12 (PAULING; RÖTTGER; TAUCH; *et al.*, 2012). Chapter I presents CoryneRegNet 7, which holds the GRNs of 4 model organisms (*C. glutamicum*, *E. coli*, *M. tuberculosis* and *B. subtilis*) and of 224 other corynebacteria.

II.6 GRNs evaluation methods

GRNs have been used to understand the regulatory mechanisms in many organisms through various types of applications ranging from medicine to biotechnology (EMMERT-STREIB; DEHMER; HAIBE-KAINS, 2014; SCHRÖDER; TAUCH, 2010). In this context, GRNs must be as accurate as possible to, for instance, design synthetic pathways for amino-acid production or find drug and vaccine target candidates (IVANOV *et al.*, 2013; MADHAMSHETTIWAR *et al.*, 2012; MCCARTY; LEDESMA-AMARO, 2019; OSTERGAARD *et al.*, 2000). Inaccurate GRNs may lead to unsuccessful experimental design or genetic engineering. Despite the importance and need for accurate GRNs, few studies have assessed their consistency with gene expression data (BAUMBACH; APELTSIN, 2008; GUTIÉRREZ-RÍOS *et al.*, 2003; MELAS *et al.*, 2013) and only one of those performed the assessment at a large scale (LARSEN *et al.*, 2019). The evaluation methods employed in the following studies are presented below:

- Gutiérrez-Ríos and collaborators built a rule-based approach to assess the consistency between microarray experiments and the *E. coli* GRN (GUTIÉRREZ-RÍOS *et al.*, 2003).
- Siegel and collaborators developed a sign consistency model providing a mathematical framework to detect inconsistencies between regulatory networks and microarray data (SIEGEL *et al.*, 2006).
- BioQuali (GUZIOLOWSKI *et al.*, 2009) is a Cytoscape app developed to detect inconsistencies and suggest changes in a GRN that would explain the expression data.
- COMA (BAUMBACH; APELTSIN, 2008), another Cytoscape app, detects the inconsistencies in regulatory interactions using a Boolean network model and a single expression study.
- CytoASP (KITTAS *et al.*, 2015) identifies inconsistencies by applying an answering set programming approach and suggesting how to repair them.
- SigNetTrainer (MELAS *et al.*, 2013) identifies and repairs regulatory interactions based on a single gene expression experiment; it also detects insertions and deletions of interactions for a specific set of experiments using an integer linear programming approach.
- Larsen and collaborators applied a conservative sign consistency approach to assess the consistency between *E. coli* GRN and a large gene expression compendium (LARSEN *et al.*, 2019).

Regarding *C. glutamicum*, no evaluation of the consistency between its GRNs and its gene expression data was performed, as per our knowledge. The assessment of *C. glutamicum* GRN and its gene expression data available on the GEO database (BARRETT *et al.*, 2013) is presented in Chapter II.

II.7 Motivation

Understanding the mechanisms that control transcriptional regulation is crucial to understand how organisms are able to adapt and survive to constant changes in the environment. Several experimental assays have been performed to unravel and study these mechanisms, most of them focusing on model organisms, such as *E. coli*, *B. subtilis*, *M. tuberculosis* and *C. glutamicum*. Due to their costs and intensive work, these experiments are not feasible for all organisms; thus, computational approaches have been developed and successfully applied. Both experimental and computational approaches to unravel GRNs have been applied in organisms of biotechnological, medical and veterinary importance, such as those in the *Corynebacterium* genus. Despite the importance of this genus, only a few organisms have their GRNs partially known. To expand the knowledge regarding the regulatory interactions that occur in the cell to adapt to adverse changes in the environment, this work reconstructed the GRNs of all corynebacteria with complete genomes available on the NCBI database (June 2019). Besides the knowledge about the regulatory interactions that form a GRN, it is also crucial that these GRNs are as accurate as possible to deliver precise results in situations such as a drug target being searched using the network. This thesis also analyzed the consistency of the GRN of the model organism *C. glutamicum* to understand its regulatory interactions' accuracy.

III. OBJECTIVES

III.1 Main objective

To extend and assess the transcriptional regulatory knowledge of the *Corynebacterium* genus.

III.2 Specific objectives

- To perform a conservative transfer of *C. glutamicum* ATCC 13032, *E. coli* K-12, *B. subtilis* 168 and *M. tuberculosis* H37Rv transcriptional GRNs to all fully sequenced and annotated corynebacterial genomes available on NCBI;
- To assign statistical significance values to all regulatory interactions of the target organisms;
- To re-implement CoryneRegNet's back- and front-end to meet the requirements of the updated database and support both new features, as well as future database growth;
- To collect, summarize and examine current knowledge of transcriptional regulation of *C. pseudotuberculosis*;
- To assess the consistency between *C. glutamicum* GRN and a large gene expression dataset;
- To compare the consistency of experimental data with *in silico* generated data.

IV. CHAPTERS

IV.1 Chapter I - Transcriptional Gene Regulatory Networks of *Corynebacterium* genus

IV.1.1 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

***Shared first authorship**

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

One of the main objectives of this thesis is to extend transcriptional regulatory knowledge regarding the *Corynebacterium* genus. This article presents CoryneRegNet 7, which stores the largest amount of GRN information available for this genus. It holds the GRNs of 225 organisms belonging to the *Corynebacterium* genus, representing an increase of ~20 times the number of organisms with GRNs available when compared to CoryneRegNet 6. CoryneRegNet 7 holds 9,590 TFs and sigma factors and over 82,000 regulatory interactions. Furthermore, we can highlight the dynamic visualization of organisms' entire networks. The information stored in CoryneRegNet helps the research community to better elucidate knowledge regarding transcriptional mechanisms enabling adaptation, survival and infection.

SCIENTIFIC DATA

Check for updates

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

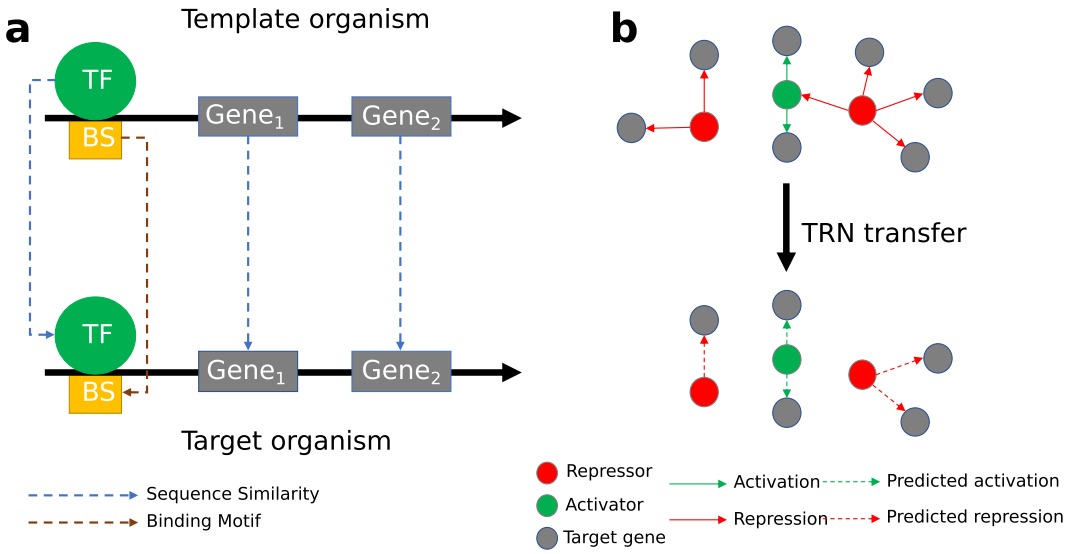


Fig. 1 Overview of the computational reconstruction of TRNs. (**a**) General concept of regulatory conservation. (**b**) TRN transfer scheme: The TRN of a model organism (top) and the predicted TRN of the target organism including all transferred regulations (bottom). In the networks, nodes represent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

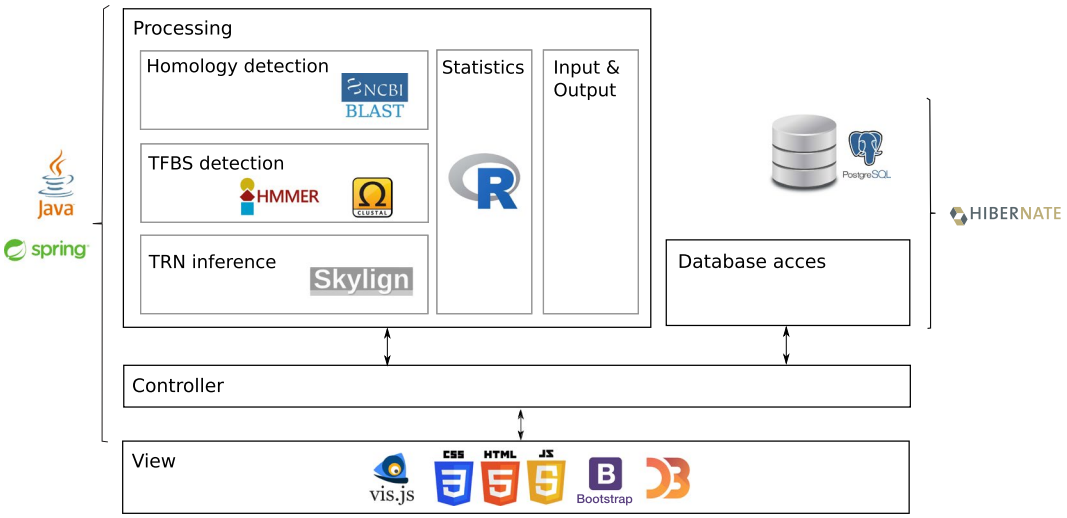


Fig. 2 Schematic of the CoryneRegNet 7 architecture. The processing layer handles the TRN transfer and data parsing. The database access layer manages any query, update, insertion or deletion in the database. The view is responsible for user interaction through the browser. The controller conducts all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

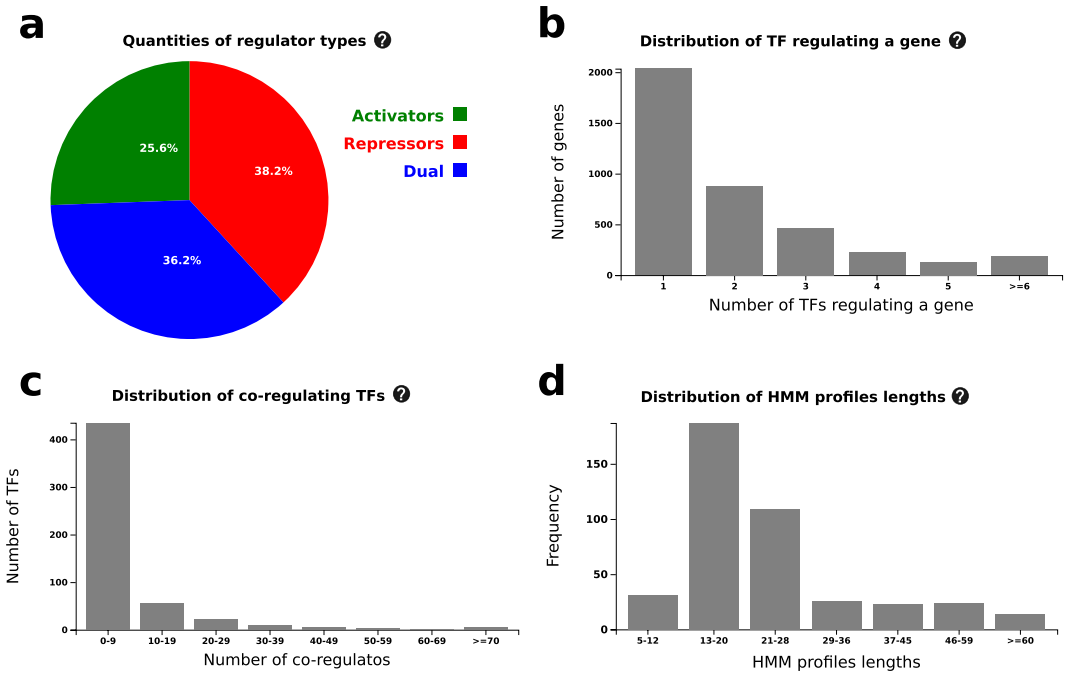


Fig. 3 Overview of main statistics present in CoryneRegNet. In (**a**) the pie chart presents the quantities of regulator types in percentages, (**b**) represents the distribution of the numbers of TFs regulating a gene, (**c**) presents the distribution of co-regulating TFs, and (**d**) demonstrates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

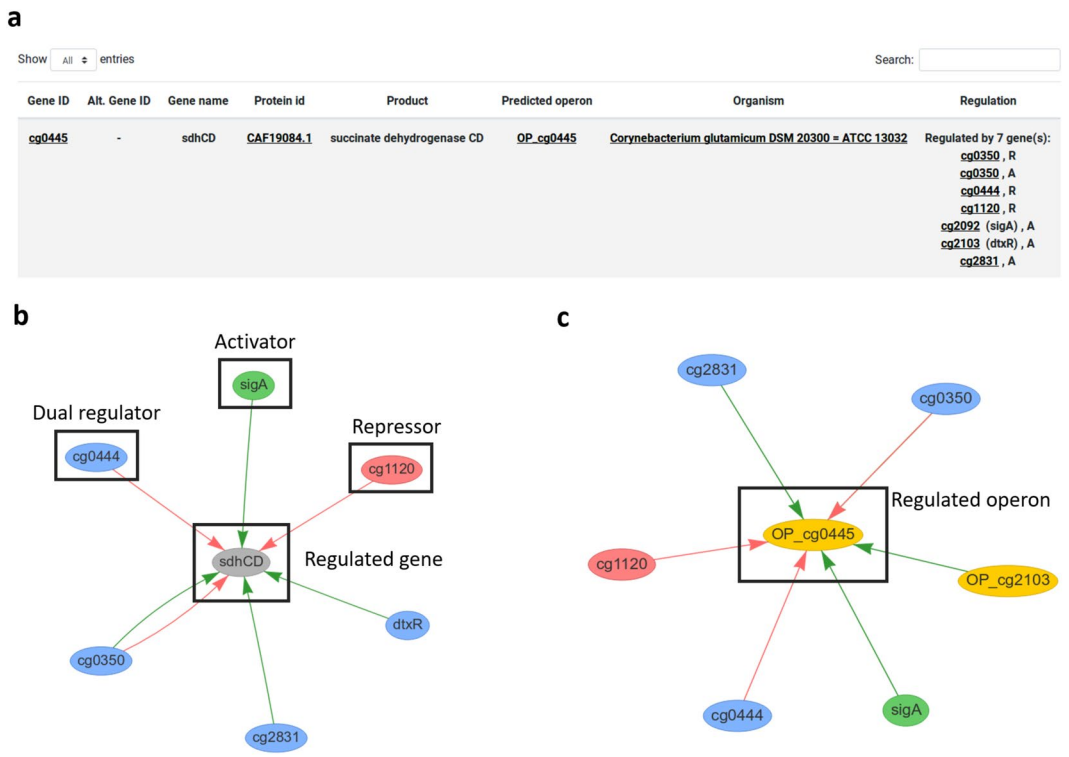


Fig. 4 Example of a RI. (**a**) TG (cg0445) and its regulators in table format. (**b**,**c**) present the same TG as in (**a**), but in a graph format with gene-centered layout (**b**) and an operon-centered layout (**c**).

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.

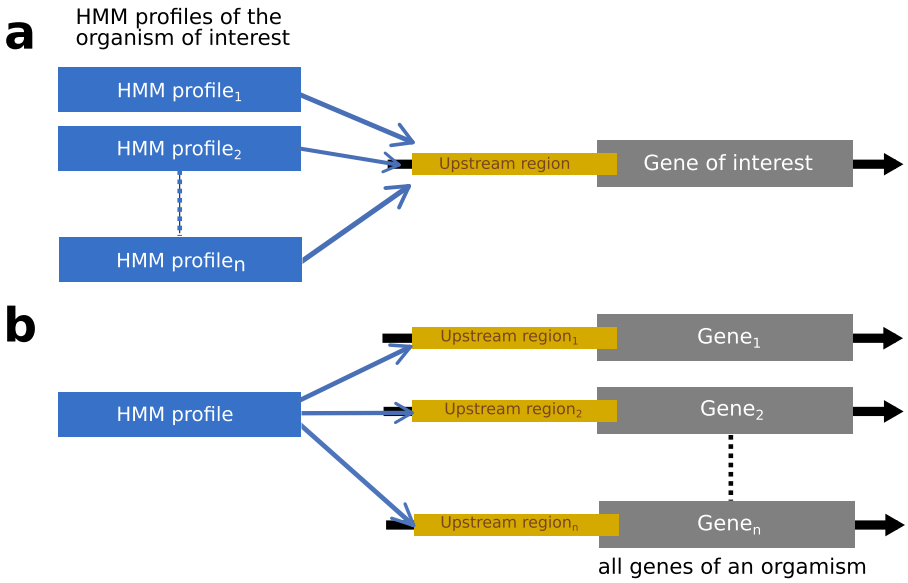


Fig. 6 Kinds of motif search provided in CoryneRegNet 7. (**a**) HMM profiles of one organism being used to identify potential binding sites in the upstream region of a gene of interest. (**b**) HMM profile of interest being used to identify potential binding sites in all genes of an organism.

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.

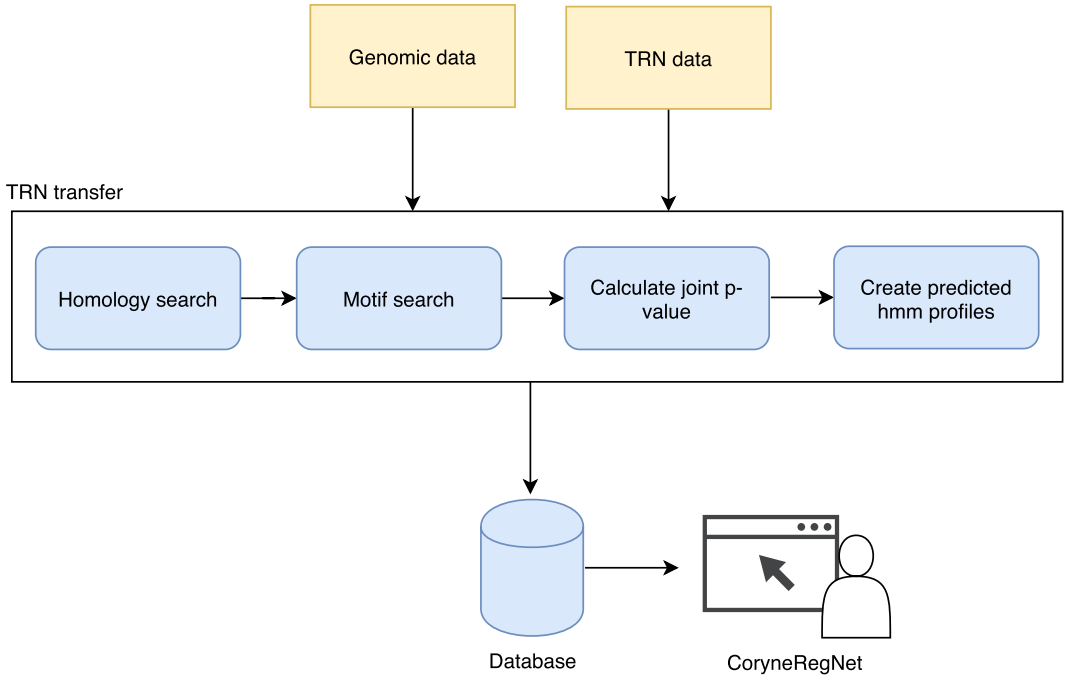


Fig. 7 Schematic overview of the TRN transfer pipeline.

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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

IV.1.2 The Transcriptional Regulatory Network of *Corynebacterium pseudotuberculosis*

Authors: **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 & Jan Baumbach

Status: Published in the journal Microorganisms (impact factor: 4.152)

To achieve the objective of collecting, summarizing and examining the current knowledge of transcriptional regulation of *C. pseudotuberculosis*, we combined a literature analysis with the data and knowledge generated with CoryneRegNet. This review provides an overview of the known regulatory mechanisms of *C. pseudotuberculosis* with a focus on two-component systems, transcription factors and sigma factors. It presents the first panoramic view of the transcriptional GRN of this important zoonotic organism.

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

check for updates

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

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

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

sensing and regulation of virulence, are key transcriptional elements of pathogens [1]. For a comprehensive review of bacterial gene regulation mechanisms in general see, for example, [2].

The main players in gene regulation are transcription factors (TFs), which are regulatory proteins that activate or repress the expression of their target genes (TGs) [3,4]. TF-TG relationships can be experimentally investigated using techniques such as RNA-Seq [5], expression microarrays [6], electrophoretic mobility shift assay (EMSA) [7,8], chromatin immunoprecipitation and DNA microarray (ChIP-chip) [9] and chromatin immunoprecipitation and sequencing (ChIP-seq) [10]. RNA-Seq and expression microarrays are used to measure the transcriptional profile of an organism under different conditions; for a comparison of these techniques see [11] and for comprehensive reviews see [12–15]. EMSA, ChIP-chip and ChIP-seq are used to identify transcription factor binding sites (TFBSs), which are the genomic regions occupied by TFs to regulate gene expression. For comprehensive reviews about ChiP-chip and ChiP-Seq see [16–18].

To have a broader view of these TF-TG relationships in systems biology, we model these as directed graphs in which the nodes represent the TFs or the TGs and the arcs, also called directed edges, represent the regulatory interactions, forming what is called transcriptional regulatory network (TRN) [3,4]. Data generated from both experimental and computational techniques are often available from online TRN databases such as RegulonDB [19] for *Escherichia coli* K12, Subtiwiki [20] for *Bacillus subtilis*, Abasy Atlas [21] for *Corynebacterium glutamicum* and CoryneRegNet [22] for the genus *Corynebacterium*. Such computational models have been utilized to understand the transcriptional mechanisms modulating cellular adaptation, survival and pathogenicity of several bacteria species [23,24]. For instance, several studies have demonstrated the role of TFs in the regulation of virulence in pathogenic bacteria such as *Mycobacterium tuberculosis* [25–29], *Pseudomonas aeruginosa* [30] and *Helicobacter pylori* [31]. Despite the experimental and computational advances, there are several understudied pathogenic bacteria that have no efficient treatment, amongst them we can highlight *Corynebacterium pseudotuberculosis*.

C. pseudotuberculosis is a pathogenic bacterium that stands out for being an organism of medical and veterinary importance that causes great economic impact on agriculture worldwide. This bacterium is a Gram-positive, facultative intracellular organism that is part of the order Corynebacteriales, which includes species of *Corynebacterium*, *Mycobacterium*, *Nocardia* and others. It is also classified into *equi* and *ovis* biovars, each causing different diseases [32]. The first causes ulcerative granulomatous lesions and mastitis in cattle [33,34], ulcerative lymphangitis (UL) or pigeon fever in horses [35] and oedematous skin disease (OSD) in buffaloes [36]. The second mainly affects small ruminants such as sheep and goats, as well as humans, causing caseous lymphadenitis (CLA) [37]. Furthermore, finding an effective vaccine against these bacteria is still a challenge [38,39]. Despite the importance of *C. pseudotuberculosis*, little is known about the mechanisms that control gene expression. A few studies have evaluated its transcriptional profile across different environmental conditions [40–43] and a few others have investigated mutant strains showing the importance of specific genes, such as *plD* [44], the *fagABC* operon [45] and *phoP* [46] in *C. pseudotuberculosis* virulence and pathogenicity. Recently, Parise et al. reconstructed in silico TRNs for all *C. pseudotuberculosis* strains with complete genome sequences and made it available in the seventh version of CoryneRegNet [22].

In this review, we collect, summarize and examine the current knowledge of transcriptional regulation of *C. pseudotuberculosis*. First, we present the TRNs and co-expression networks available for this organism. Then, we discuss single-gene studies and "omic" analyses together with the TRNs regarding the presented genes in order to understand the transcriptional regulation of *C. pseudotuberculosis*; these results are presented by distinct regulator types: two-component signal transduction systems (TCSs), TFs and sigma factors. In this way, we aim to summarize and condense our view on *C. pseudotuberculosis* transcriptional regulation by jointly discussing RNA-seq and mutation assays in the light of network biology.

2. Gene Co-Expression Networks and Transcriptional Regulatory Networks

To understand how genes interact and influence the expression of other genes, TRNs and gene co-expression networks (GCNs) have been used to analyze bacterial genomes [19,20,47–49]. While TRNs present regulatory interactions between regulatory proteins and their targets, GCNs present correlation between the expression of certain genes in the network. Such networks have been used to model how regulatory processes work inside the cell, including amino acid synthesis and virulence mechanisms [50–53]. Franco et al. [54] and Parise et al. [22] performed GCN analysis and TRN transfer, respectively, in *C. pseudotuberculosis*.

Franco et al. inferred the GCNs of four *C. pseudotuberculosis* strains (258, T1, Cp13 and 1002) using RNA-Seq datasets [40–43]. The authors applied the following bioinformatic tools: (i) miRsig [55] to infer the GCNs of all genes and differentially expressed genes (DEGs), (ii) miRinfluence [56] to identify the predicted networks' influential and causal genes and (iii) Online GEne Essentiality (OGEE) database v2 [57] to classify the causal genes as essential, nonessential or conditionally essential [54]. Essential, nonessential and conditionally essential genes demonstrate the consensus of the level of essentiality of a certain gene for bacterial survival, for more details see [57]. The total number of genes and gene-gene interactions of each GCN are presented in Table 1. In the GCN analyses, the following genes were considered causal and categorized as essential: *galU* and *argS* in 258; *pdpB* and *trpC* in T1; and *serC*, *mraY* and *glmS* in Cp13. The above-mentioned genes had not previously been analyzed experimentally in *C. pseudotuberculosis*; however, previous studies in other bacterial species support their relevance for organisms' growth and survival [58–63]. For instance, *galU* increases glycogen and trehalose amounts in *C. glutamicum* [58], and it is also a potential drug target in *M. tuberculosis* [59]. The *argS* gene encodes an arginyl-tRNA synthetase [60], which is induced in the presence of arginine and repressed in the presence of lysine. A study has also proposed that the absence of *argS* is lethal for *Brevibacterium lactofermentum*, because it is not able to synthesize proteins without an arginyl-tRNA synthetase [64]. The *pdpB* gene is homologous of the *lcmF* gene, which is the result of the fusion of the radical B12 enzyme isobutyryl-CoA mutase and its G-protein chaperone [61,65]. In *Francisella*, a deletion mutant of this gene is defective for intracellular bacterial growth [61]. The *trpC* gene, an indole-3-glycerol phosphate synthase [66], is considered essential for growth in both *M. tuberculosis* and *Mycobacterium. bovis* [67], and was identified as a potential drug target in *M. tuberculosis* [59]. The *serC* gene encodes a phosphoserine aminotransferase and is involved in the biosynthesis of L-serine in *C. glutamicum* [68]. In *E. coli* it is induced by cyclic AMP-dependent and together with *aroA* encodes enzymes that lead serine and aromatic amino acids synthesis [69]. The *marY* gene encodes an undecaprenyl-phosphate phospho-*N*-acetylmuramoyl-pentapeptide transferase [70] and is considered essential for cell growth in *E. coli* [63]. It also participates in the building process of the peptidoglycan layer of the cell wall in corynebacteria [71]. The *glmS* gene is a glucosamine-6-phosphate synthase that can enhance *N*-acetylglucosamine synthesis in *C. glutamicum* [62] and was also indicated as a drug target in *M. tuberculosis* in [72,73].

Table 1. Gene co-expression network (GCN) data from Franco et al. [54]. DEGs -differentially expressed genes.

Strain	Technology	GCN from All Genes		GCN from DEGs		Reference
		Genes	Interactions	Genes	Interactions	
Cp13	Ion Proton	2113	86,367	63	46	[43]
T1	Ion Proton	2093	107,202	93	98	[43]
1002	SOLiD	2091	6682	168	155	[40]
258	SOLiD	2064	9376	139	165	[41,42]

Parise et al. analyzed the conservation of transcriptional regulation in the genus *Corynebacterium* [22]. The authors used experimental TRNs of *M. tuberculosis*, *B. subtilis*, *C. glutamicum* and *E. coli* as models for predicting the TRNs of all complete genomes of this genus, including 91 strains from *C. pseudotuberculosis*. This prediction was performed by checking the conservation of the TFs and the TGs with BLAST software [74] and the conservation of the TFBSs with HMMER package [75]. These predictions are publicly available in CoryneRegNet 7, which holds 24,069 regulatory interactions, 2990 TFs and 17,298 TFBSs for this species. Such networks will be discussed and presented alongside literature data regarding *C. pseudotuberculosis* regulations in the following sections.

3. Regulators of Gene Expression

3.1. Two-Component Systems

Two-component signal transduction systems (TCS) detect and mediate the response to external stimuli by means of a series of biochemical signals that result in gene expression changes [46,76]. These processes regulate many processes such as drug resistance, adhesion, sporulation, pilus formation, cell division, nutrient acquisition, nitrogen fixation and virulence [77–80].

One of the most well-known TCS is the PhoPR system, in which a transmembrane sensory histidine kinase protein (PhoR) phosphorylates the receiver domain of the response regulator protein (PhoP). The phosphorylation of PhoP results in the activation of the effector domains, which causes a transcriptional response. Some studies have used *phoP* mutant strains as vaccinal strategies, not only in *C. pseudotuberculosis*, but also in *M. tuberculosis* [46,81]. In *C. pseudotuberculosis*, these mutant strains presented reduced virulence in mice and induced a host cellular immune response [46]. Additionally, the absence of the *phoP* gene resulted in higher levels of IgG antibodies in contrast with the control group.

In CoryneRegNet 7 [22], the predicted TRN of the *phoP* and *phoR* genes in *C. pseudotuberculosis* 1002B strain is shown in Figure 1. The TFs *phoP* and *glxR* jointly activate the phosphate ATP-binding cassette (ABC) transporters *pstB*, *pstC*, *pstS* and Cp1002B_RS04420 (frameshift *pstA*). The *ppiB* gene is repressed by *phoP* and is dually regulated by *glxR*, while *phoR* is activated by *phoP*. Due to its crucial role in bacterial survival, development and adaptation to environmental changes, the PhoPR system is an interesting drug target for future studies [46,82]. In *C. glutamicum*, the *pstSCAB* operon, an ABC transport system for the uptake of phosphate, is induced during phosphate starvation [83]. The *ppiB* gene is a peptidyl-prolyl cis/trans isomerase (PPIase); it is important for growth in both low temperatures and starvation conditions in *Legionella pneumophila* [84] and *B. subtilis* [85]. PPIases were marked as alternative drug targets [86]. Both the *pstSCAB* operon and the PPIases were already indicated as bacterial virulence factors in *L. pneumophila* [86] and *Salmonella* [87–89].

The *hrrA* gene is part of the HrrSA TCS and it was differentially expressed in both *C. pseudotuberculosis* T1 and CP13 under iron starvation [43]. This system both activates the expression of a heme oxygenase (*hmuO* gene) and represses genes acting in heme homeostasis in *C. glutamicum* and *Corynebacterium diphtheriae* [91,92], and is considered the global regulator of heme in *C. glutamicum* [93]. In this system the HrrS is the sensor kinase and the HrrA is the response regulator. Furthermore, Franco et al. [54] identified three TCS genes, namely, *tcsS4*, *mprA_2* and *tcsR3*, as influential genes. However, they remain to be studied; there is no regulatory information in CoryneRegNet 7 for the HrrSA TCS and the three genes found by Franco et al.

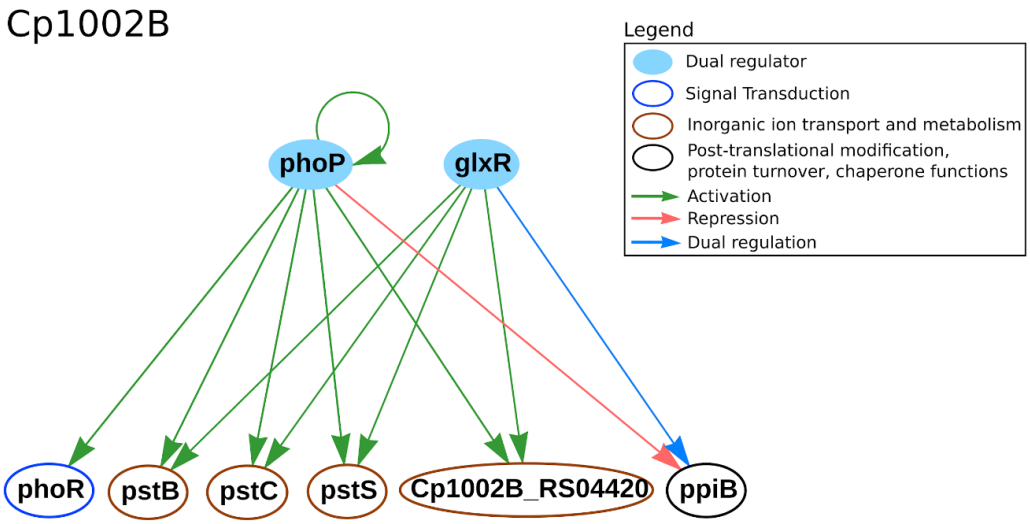


Figure 1. PhoPR transcriptional regulatory network (TRN) retrieved from CoryneRegNet 7 [22] for *C. pseudotuberculosis* 1002B. In the network, nodes represent the genes and arrows represent the regulatory interactions. The functional classification was performed using the database Eggnog (evolutionary genealogy of genes: non-supervised orthologous groups) [90].

3.2. Transcription Factors

TFs can modulate gene expression through activating or repressing transcription by different mechanisms. In all mechanisms, activation enhances the interaction between the RNA polymerase and the promoter, and repression prevents their interaction [94]. In bacteria, the environmental signals are the elements responsible to modulate TFs activity influencing transcription initiation [95]. In this section we describe the TFs that perform key functions in *C. pseudotuberculosis* in the context of the biological mechanism they are involved in.

3.3. Metalloregulation: Iron Uptake

Metal ions are essential for bacterial metabolism; in particular, iron, manganese and zinc are used as cofactors [96]. Iron is an important protein cofactor required for growth and development in virtually all living organisms; the acquisition of host iron is a well-characterized mechanism of infections used by bacterial pathogens to successfully establish the infection in host cells. Besides its importance, an excess of iron can be toxic to the cell; thus, in order to keep iron homeostasis, bacteria developed a tightly-regulated system [43,97].

A recent study used RNA-Seq to analyze *C. pseudotuberculosis*-infected spleens of dairy goats and found many iron-related genes differentially expressed in order to reduce iron availability. The authors hypothesized that *C. pseudotuberculosis* evolved an iron acquisition mechanism to manage this reduction [98]. The expression of DtxR, the master regulator of iron, is downregulated under iron limitation, directly and indirectly influencing the expression of several genes in *C. pseudotuberculosis* [43]. In CoryneRegNet, this TF is predicted to regulate sixteen genes in both Cp13 and CpT1 [22]; such regulatory interactions are presented in Figures 2 and 3, respectively. Some of DtxR's target genes are the *fagABC* operon and the *fagD* gene, which are components of the iron acquisition system and important virulence factors of this organism [40,45]. Studies in *C. diphtheriae* and *C. pseudotuberculosis* indicate that DtxR regulates the *ciuA* gene [99,100]; however, a recent study [43] in *C. pseudotuberculosis* T1 and Cp13 found no difference in the expression of these genes under iron restricted conditions. Likewise, DtxR is not predicted to regulate the gene *ciuA* in CoryneRegNet [22]. The authors of the iron-limitation study [43] used the Ion Proton platform to analyze the transcriptome of the wild-type *C. pseudotuberculosis* strain T1 and its mutant strain, Cp13, which has a disrupted *ciuA* gene. The *ciuA*

gene encodes a protein highly similar to siderophore ABC-type transport systems and has been previously associated with virulence [101]. The lack of this gene resulted in reduced growth [43] and intracellular viability [102] of the mutated strain. In the same study, 77 and 59 DEGs were identified in T1 and Cp13, respectively. Besides that, the authors observed an up-regulation of hemin acquisition systems and down-regulation of iron intracellular utilization in both strains. The expression of hemin uptake systems in Cp13 may indicate the adaptive response of the transcription machinery to iron acquisition from other sources. Hemin uptake genes were found in genomic islands together with many known virulence factors, corroborating previous studies pointing to the association of iron uptake and virulence in *C. pseudotuberculosis* [43,45,102].

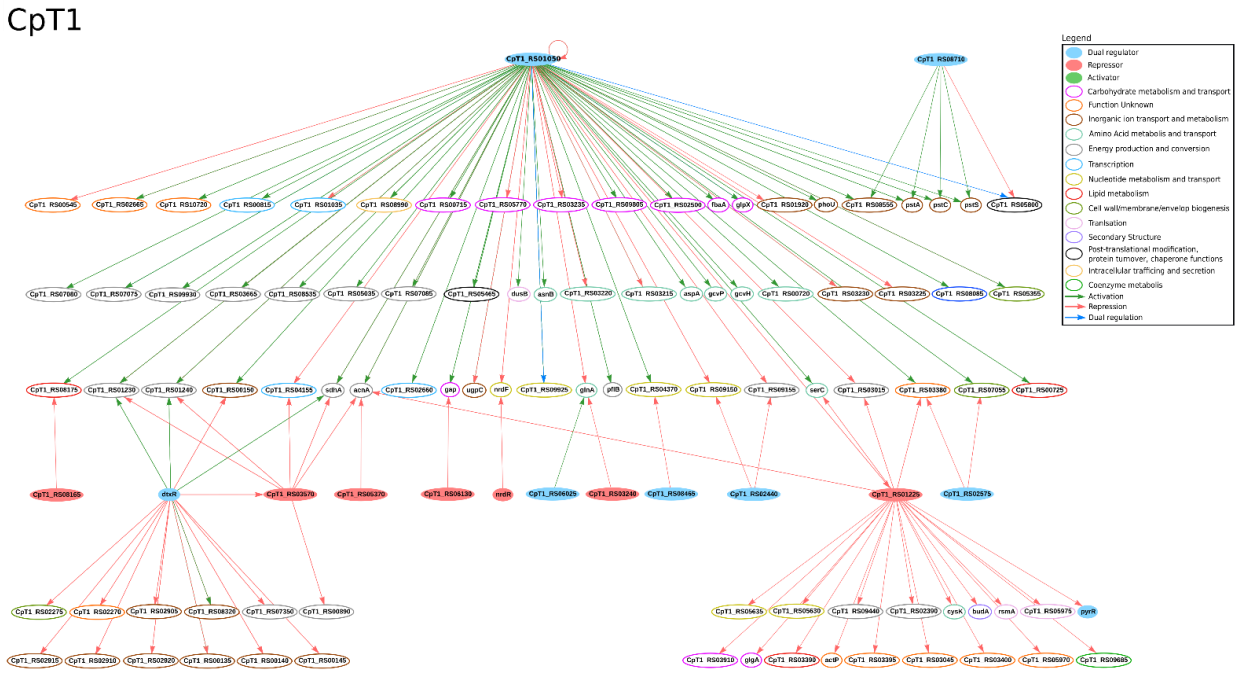


Figure 2. Regulatory interactions taken from CoryneRegNet 7 [22] for *C. pseudotuberculosis* T1 under iron limitation in [43]. In the network, nodes represent the genes and arrows represent the regulatory interactions. The functional classification was performed using the database Eggnog [90].

In the same study, iron restriction increased the expression of genes associated with putative hemin acquisition systems and decreased the expression of genes associated with energy metabolism in both strains. Down-regulated genes involved in the oxidative phosphorylation process and tricarboxylic acid cycle (TCA) cycle were only found in the T1 strain. Amongst them, *sdhA*, *sdhB* and *sdhC* genes encode succinate dehydrogenase iron-sulfur proteins, which compose the respiratory complex II. In this complex, the succinate reduction links the oxidative phosphorylation process with the TCA. These genes are predicted to be jointly regulated by GlxR, DtxR and RipA. Regarding the TFs, Ibraim and collaborators [43] found the following up-regulated genes: *ripA* in both strains; *glpR*, *cspA*, *whiB* and *sufR* in Cp13; and *glxR* in T1. In CoryneRegNet, only *ripA* and *glxR* regulate other genes in these strains [22]. *RipA* encodes a protein that belongs to the AraC family regulators that repress the expression of genes encoding iron-containing proteins. This TF is predicted to be regulated by DtxR and to regulate six genes in both T1 and Cp13 in [22]; the *sdhA*, *sdhB* and *sdhC* genes were identified as differentially expressed in the iron limitation assay [43]. GlxR is a global regulator involved in the regulation of metabolic processes [91,92]. It is predicted to regulate 79 genes in both T1 and Cp13 [22], including two TFs: *ramB* homologue (CpCp13_RS01220 and CpT1_RS01225) and *whiB* homologue (CpCp13_RS01035, CpCp13_RS02650 and CpT1_RS02660) in both strains. The predicted regulatory networks of the DEGs found under iron limitation in T1 and Cp13

are presented in Figures 2 and 3, respectively [22]. The regulators RamB, TetR family protein (CpT1_RS08165), AcnR homologue (CpT1_RS05370), NrdR, AmtR homologue (CpT1_RS03240), Ram*A* homologue (CpT1_RS08465), RbsR homologue (CpT1_RS02440), MtrA homologue (CpT1_RS02575) and PyrR are present in the regulatory network and were not identified in the experimental assays, which indicated that they may be involved in other cellular functions other than metalloregulation and need to be further investigated.

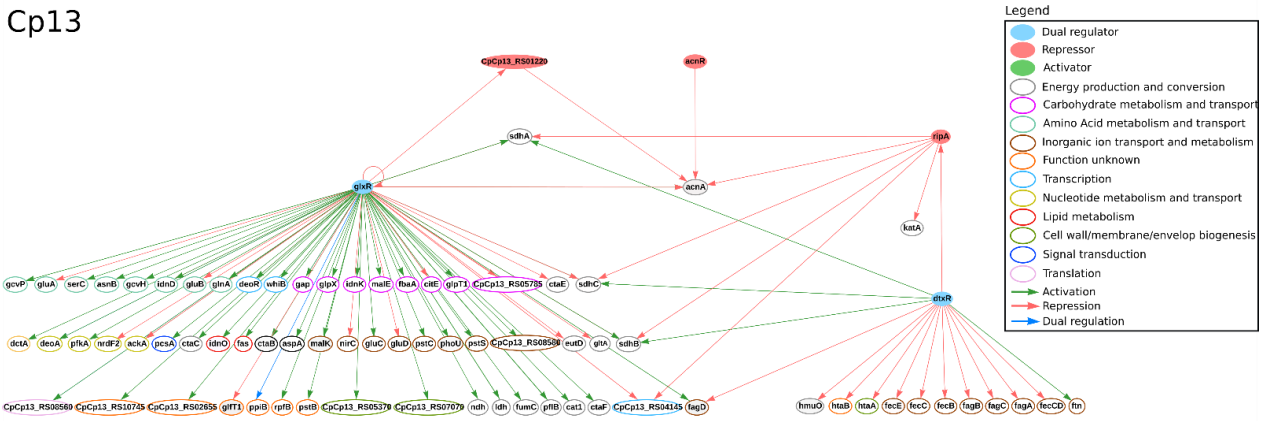


Figure 3. Regulatory interactions taken from CoryneRegNet 7 [22] for *C. pseudotuberculosis* Cp13 under iron limitation in [43]. In the network, nodes represent the genes and arrows represent the regulatory interactions. The functional classification was performed using the database Eggnog [90].

3.4. Response to Osmotic, Thermal and Acid Stress

Bacteria experience stress conditions not only when migrating from the environment to the host, but also when invading and colonizing the host's bloodstream, gastrointestinal and respiratory tracts, mucous membranes and immune system [103–105]. In order to survive these conditions, the organism must assemble a quick protective response at the transcriptional level [105,106]. During environmental changes, a reduction, or even a lack of growth, is considered normal in bacteria [40,107]. Three studies in *C. pseudotuberculosis* point out a reduction of replication of ~23%, ~27% and ~34% in strain 1002 and 16%, 20% and 36% in strain 258 under osmotic, thermal and acid stresses, respectively [40–42]. The first study performed a differential expression analysis under these conditions in *C. pseudotuberculosis* 1002 and identified DEGs involved in oxidoreduction, adhesion and cell division processes [40]. The other two allow us to further understand the transcriptional response induced during these stresses in *C. pseudotuberculosis* 258 [41,42]. In all three studies, the authors performed transcriptome analyses using the SOLiD 3 Plus platform [40–42]. A notable TF identified in these two strains is TetR2, a DEG found across the three stress conditions. This TF belongs to the TetR family that in general regulates the expression of genes involved in drug resistance, biosynthesis of antibiotics, pathogenicity, virulence, quorum sensing and catabolic pathways [40–42,108]. So far, there is no experimentally verified or predicted regulation for this TF in *C. pseudotuberculosis* 1002B and 258.

Under acid stress the following DEGs were highly expressed: *msrB* in strain 1002B, *msrA in* strain 258 and both *dps* and *lysR1* in these two strains [40,42]. In CoryneRegNet, the TF LysR1 is predicted to repress itself and activates the expression of the peroxiredoxin gene *ahpC* in both strains. Gomide et. al. 2018 suggests that this TF plays a modulatory role in *C. pseudotuberculosis* 258; however, no experimentally verified LysR1 regulation is known in these strains [42]. In *P. aeruginosa*, a LysR-family TF regulates the expression of genes related to virulence and stress response modulators [109]. The *msrB* and *msrA* gene products act together resulting in the catalytic activity of the oxidation-reduction of methionine sulfoxide. The *msrA* gene plays a more relevant role in virulence than *msrB* does in bacteria [42,110]; it corroborates the fact that *C. pseudotuberculosis* strains from the *equi* biovar, such as 258, are more virulent than the ones from the *ovis* biovar, such as 1002B.

The *dps* gene protects the bacteria under acid, oxidative and heat stresses, as well as in iron and copper toxicity. The inactivation of this gene in *E. coli* leads to a reduction in the survival rate of the bacteria in an acid environment [42,111,112]. The *dps*, *msrA* and *msrB* genes have no regulatory interactions predicted in CoryneRegNet [22] and are interesting candidates for future experimental assays.

Under thermal shock stress, *hspR*, *dnaK* and *grpE* genes were differentially expressed in both organisms. The *hspR* gene, which encodes a heat shock TF, is known for regulating genes involved in virulence and pathogenicity [40,113]. Additionally, it regulates heat shock operons, which encode genes that maintain the structure of proteins in several cell stresses [114,115]. Likewise, in 1002B and 258 it is predicted to regulate four genes: *dnaK*, *grpE*, *clpB* and *clgR*. The *dnaK* gene improves the immune response in the host and seems to regulate genes encoding virulence factors and bacterial adhesion [40,115]. The *grpE* gene is also a chaperone involved in bacterial virulence and belongs to the same operon as *hspR* and *dnaK* [41]. The *clpB* gene was differentially expressed under both thermal and osmotic stresses in the strain 258. It encodes an ATP-dependent chaperone that is involved in virulence and participates in the stress response system [116,117]. The *clgR* gene regulates the expression of genes acting in DNA repair and proteolysis in *C. glutamicum* [118]. In CoryneRegNet, it is predicted to regulate seven genes in *C. pseudotuberculosis* 1002B and 258, as presented in Figure 4.

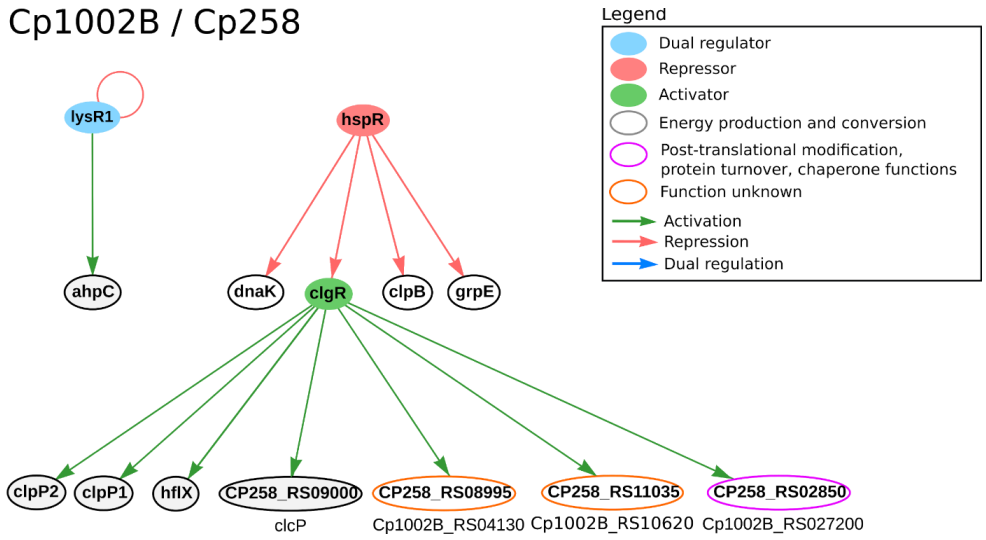


Figure 4. Regulatory interactions from CoryneRegNet 7 for the DEGs of *C. pseudotuberculosis* 258 (Cp1002B) and 1002B (Cp258) under osmotic, acid and heat stress conditions [41,42]. In the network, nodes represent the genes and arrows represent the regulatory interactions. The functional classification was performed using the database Eggnog [90].

Under osmotic stress the following DEGs were highly expressed: *glmU* and *uppP* in strain 258 and *norM* in strains 1002B and 258 [40,42]. The *glmU* gene encodes an enzyme that catalyzes the substrate in the synthesis of bacterial peptidoglycans and lipopolysaccharides of the cell wall [119]; it is predicted to be regulated by the *glk* homologue (CP258_RS07175) in *C. pseudotuberculosis* 258 in CoryneRegNet [22]. A study in *Mycobacterium smegmatis* showed that *glmU* is both a drug target and crucial for bacterial replication [120]. The *uppP* gene encodes an enzyme that is involved in the biosynthesis of both membrane proteins and bacterial cell wall components, a process that is essential for bacterial integrity [121,122]. This also makes the *uppP* gene crucial to bacterial growth and bacterial pathogenicity, making it an interesting drug and vaccine target, as well [42,120]. The *norM* gene is a multidrug efflux pump that belongs to an ABC transporter family, conferring organisms an effective antibiotic resistance [42,123,124]. In *C. pseudotuberculosis* 258, the *srtA* gene was differentially expressed under both osmotic and thermal stresses. The *srtA* gene encodes a sortase, which is a housekeeping gene involved in the pathogenesis and virulence of

Gram-positive bacteria. It contributes to the covalent binding of the peptidoglycan layer and cell surface proteins [41,42,125,126]. There are no regulatory interactions for *uppP*, *norM* and *srtA* genes in CoryneRegNet [22] for these strains, such genes are promising candidates for experimental assays. Figure 4 presents the regulatory interactions from CoryneRegNet of the genes found as DEGs in *C. pseudotuberculosis* 258 and 1002B (Figure 4) strains under osmotic, acid and heat stress conditions [40–42].

3.5. Sigma Factors

In prokaryotes, one of the most important stages of the gene expression regulation is the initiation of the transcription. During this stage, sigma factors are both required to assemble the RNA polymerase holoenzyme and to recognize the promoters [127,128]. Similar to TFs, sigma factors are key players in transcriptional regulation when adapting to stress conditions, such as osmotic, thermal, acid and nutrient starvation stresses [129,130]. These molecules are also known to be involved in the regulation of virulence genes [131,132]. Bacterial sigma factors include SigA, SigB, which are essential and nonessential, respectively, and the alternative sigma factors SigC, SigD, SigE, SigH, SigK and SigM. These alternative sigma factors may belong to the extracytoplasmic factors group [133,134], which is responsible for the regulation of genes involved in the transport, cellular wall adaptation or secretion within the periplasm (Gram-negatives) or extracellular environment [131].

In *C. pseudotuberculosis* 1002 there are eight genes encoding sigma factors [135], whereas in *C. glutamicum*, a nonpathogenic bacteria, there are seven sigma factors [134]. In particular, the *sigK* gene is present only in C. *pseudotuberculosis*, which suggests that it may have a role in the virulence mechanisms of *C. pseudotuberculosis*. The study of Pinto et al. [40] observed the expression changes of some sigma factors in *C. pseudotuberculosis* 1002 in the beginning of the exponential phase under heat, osmotic and acid stresses, simulating host-infection conditions. The sigma factors analyzed in the aforementioned study are shown in Table 2.

Table 2. Fold-change values of the genes encoding sigma factors in osmotic (2 M), heat (50 °C) and acid stresses (pH), from Pinto et al. [40]. °C-degree Celsius. ECF-Extracytoplasmic function.

Sigma Factor	Product	Osmotic Stress		Thermic Stress		Acid Stress	
		Fold-Change	DEG	Fold-Change	DEG	Fold-Change	DEG
sigA	RNA polymerase sigma factor SigA (essential housekeeping sigma factor)	2.1889	Yes	1.4903	No	0.9232	No
sigB	RNA polymerase sigma factor SigB (non-essential SigA-like)	0.6348	No	0.9044	No	2.9154	Yes
sigC	RNA polymerase sigma factor SigC (ECF family)	0.4675	No	0.8031	No	1.7238	No
sigD	RNA polymerase sigma factor SigD (ECF family)	1.5437	No	1.2891	No	0.8654	No
sigE	RNA polymerase sigma factor SigE (ECF family)	0.5483	No	0.9356	No	2.5244	Yes
sigH	RNA polymerase sigma factor SigH (ECF family)	1.8401	No	1.7864	No	3.5832	Yes
sigK	RNA polymerase sigma factor SigK (ECF family)	1.5887	No	1.7415	No	1.6199	No
sigM	RNA polymerase sigma factor SigM (ECF family)	4.7414	Yes	3.5593	Yes	4.4934	Yes

Considering the authors' fold-change threshold of 2×, the following sigma factors were differentially expressed: *sigA* under osmotic stress; *sigB*, *sigE* and *sigH* under acid stress; and *sigM* in all conditions. In *C. pseudotuberculosis* 1002, the *sigA* and *sigH* genes, which encode RpoD and RpoE sigma factors, respectively, were significantly differentially expressed in all three conditions. The *sigA* gene, also known as sigma 70, promotes the binding of the RNA polymerase to specific sites activating the transcription of most essential genes related to the exponential growth in *E. coli* [136]. In *C. pseudotuberculosis* 1002, this

protein conserves the four domains belonging to sigma 70. The *sigB* gene was considered to be induced in the same study and is known to regulate genes involved in the stress response of many Gram-positive bacteria. Both *sigA* and *sigB* were previously associated with virulence in other bacteria; SigA can be specifically required for the expression of virulence genes in *M. tuberculosis* [131,137]. SigB controls the expression of many genes involved in the virulence of pathogens, such as biofilm formation, cellular differentiation, pathogenesis, stress resistance and sporulation in several bacteria [131,138]. Pacheco et al. assessed the role of SigE in *C. pseudotuberculosis* using a *sigE*-mutant strain under different stress conditions, including acid stress. The authors observed a higher in vitro susceptibility of this bacteria in the host-simulated conditions, inferring the importance of SigE in the bacterial maintenance within the unfavorable environment [139].

Interestingly, no significant expression changes in these molecules were observed in both 258, under the same three stress conditions applied to *C. pseudotuberculosis*, and Cp13 and T1, under iron starvation [41–43]. For all sigma factor coding genes, it is necessary to unravel the regulon, as well as unveil all the interaction network coding genes of the sigma factors for a better understanding of the infection process and response modulation in the cell. The predicted regulatory interactions in CoryneRegNet of the sigma factors mentioned in this section are presented in Figure 5A,B for strains 1002B and 258, respectively. Finally, the function of experimentally studied sigma factors in *C. glutamicum* and *M. tuberculosis* help to provide clues on their regulatory roles in *C. pseudotuberculosis* [133,136].

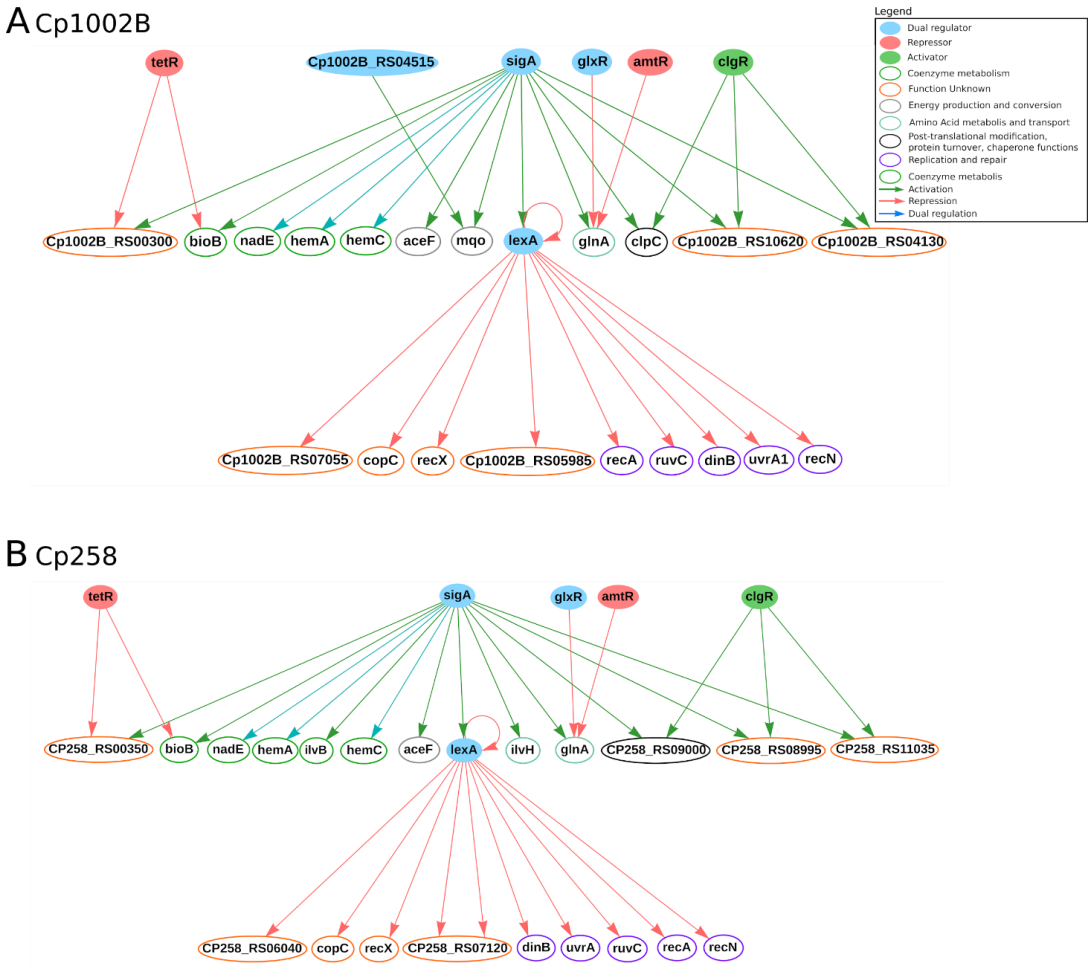


Figure 5. Regulatory interactions of SigA in *C. pseudotuberculosis* 1002B (**A**) and *C. pseudotuberculosis* 258 (**B**) from CoryneRegNet 7. In the network, nodes represent the genes and arrows represent the regulatory interactions. The functional classification was performed using the database Eggnog [90].

4. Conclusions

In this review, we presented the current knowledge of the landscape of *C. pseudotuberculosis* transcriptional regulation. The behavior of this organism under osmotic, acid, iron-starvation and thermal stress was studied exemplarily as well. We conclude that we have just begun to understand the importance of some key transcription factors, such as PhoP, DtxR, Rip*A* and GlxR, as well as of some of the sigma factors. Apart from that, very little is known about the regulatory mechanisms of this organism. New RNA-seq analyses under several conditions and preferably also time-series data combined with other layers of regulatory data are still needed to unravel the pathogenicity, survival and adaptation of *C. pseudotuberculosis* in its diverse range of hosts. Such studies might contribute not only to correctly diagnosing and treating the diseases caused by this organism, but also to identifying better drugs and vaccine candidates based on regulatory pathomechanisms.

Author Contributions: Conceptualization, D.P., M.T.D.P., V.A.d.C.A. and J.B.; writing—original draft preparation, D.P., M.T.D.P. and A.C.P.G.; writing—review and editing, D.P., M.T.D.P., A.C.P.G., F.F.A., R.B.K., M.S.-A., A.T., V.A.d.C.A. and J.B.; supervision, J.B., V.A.d.C.A. and R.B.K.; All authors have read and agreed to the published version of the manuscript.

Funding: J.B. is grateful for support from H2020 grant RepoTrial (no. 777111) and his VILLUM Young Investigator grant (no. 13154). D.P. received support from CAPES (no. 88887.364607/2019-00) and M.T.D.P. from CNPq (no. 201336/2018-9), for their work at TUM in Germany. M.T.D.P.'s work was also supported by the German Research Foundation (under SFB924). V.A.C.A. is grateful for support from his CNPq Research Productivity grant (no. 305093/2015-0), CNPq Universal grant (no. 405233/2016-7) and FAPEMIG grant (no. APQ 02600-17). M.S.-A. received PhD fellowship funding from CONACYT (CVU659273) and the German Academic Exchange Service, DAAD (ref. 91693321). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. González Plaza, J.J. Small RNAs as fundamental players in the transference of information during bacterial infectious diseases. *Front. Mol. Biosci.* **2020**, *7*, 101. [CrossRef]
2. Bervoets, I.; Charlier, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiol. Rev.* **2019**, *43*, 304–339. [CrossRef]
3. 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* **2006**, *7*, 24. [CrossRef] [PubMed]
4. Babu, M.M.; Lang, B.; Aravind, L. Methods to reconstruct and compare transcriptional regulatory networks. *Methods Mol. Biol.* **2009**, *541*, 163–180.
5. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **2008**, *320*, 1344–1349. [CrossRef] [PubMed]
6. Schena, M.; Shalon, D.; Davis, R.W.; Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **1995**, *270*, 467–470. [CrossRef]
7. Hellman, L.M.; Fried, M.G. Electrophoretic mobility shift assay (EMSA) for detecting protein—Nucleic acid interactions. *Nat. Protoc.* **2007**, *2*, 1849–1861. [CrossRef]
8. Garner, M.M.; Revzin, A. The use of gel electrophoresis to detect and study nucleic acid—Protein interactions. *Trends Biochem. Sci.* **1986**, *11*, 395–396. [CrossRef]
9. Horak, C.E.; Snyder, M. ChIP-chip: A genomic approach for identifying transcription factor binding sites. *Guide Yeast Genet. Mol. Cell Biol. Part B* **2002**, *350*, 469–483.
10. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **2007**, *129*, 823–837. [CrossRef]
11. Mantione, K.J.; Kream, R.M.; Kuzelova, H.; Ptacek, R.; Raboch, J.; Samuel, J.M.; Stefano, G.B. Comparing bioinformatic gene expression profiling methods: Microarray and RNA-Seq. *Med. Sci. Monit. Basic Res.* **2014**, *20*, 138–142.

12. Hrdlickova, R.; Toloue, M.; Tian, B. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* **2017**, *8*, e1364. [CrossRef] [PubMed]
13. Marguerat, S.; Bähler, J. RNA-seq: From technology to biology. *Cell. Mol. Life Sci.* **2010**, *67*, 569–579. [CrossRef] [PubMed]
14. Pan, W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **2002**, *18*, 546–554. [CrossRef] [PubMed]
15. Dupuy, A.; Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J. Natl. Cancer Inst.* **2007**, *99*, 147–157. [CrossRef]
16. Furey, T.S. ChIP-seq and beyond: New and improved methodologies to detect and characterize protein—DNA interactions. *Nat. Rev. Genetics* **2012**, *13*, 840–852. [CrossRef]
17. Park, P.J. ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genetics* **2009**, *10*, 669–680. [CrossRef]
18. Buck, M.J.; Lieb, J.D. ChIP-chip: Considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **2004**, *83*, 349–360. [CrossRef]
19. 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.; 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.* **2019**, *47*, D212–D220. [CrossRef] [PubMed]
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.* **2018**, *46*, D743–D748. [CrossRef]
21. 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**, *2016*. [CrossRef]
22. Parise, M.T.D.; Parise, D.; Kato, R.B.; Pauling, J.K.; Tauch, A.; de Azevedo, V.A.C.; Baumbach, J. CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. *Sci. Data* **2020**, *7*, 142. [CrossRef]
23. Kreikemeyer, B.; McIver, K.S.; Podbielski, A. Virulence factor regulation and regulatory networks in Streptococcus pyogenes and their impact on pathogen—Host interactions. *Trends Microbiol.* **2003**, *11*, 224–232. [CrossRef]
24. Pauling, J.; Röttger, R.; Neuner, A.; Salgado, H.; Collado-Vides, J.; Kalaghatgi, P.; Azevedo, V.; Tauch, A.; Pühler, A.; Baumbach, J. On the trail of EHEC/EAEC—Unraveling the gene regulatory networks of human pathogenic *Escherichia coli* bacteria. *Integr. Biol.* **2012**, *4*, 728–733. [CrossRef]
25. Galagan, J.E.; Minch, K.; Peterson, M.; Lyubetskaya, A.; Azizi, E.; Sweet, L.; Gomes, A.; Rustad, T.; Dolganov, G.; Glotova, I.; et al. The Mycobacterium tuberculosis regulatory network and hypoxia. *Nature* **2013**, *499*, 178–183. [CrossRef] [PubMed]
26. Gordon, B.R.G.; Li, Y.; Wang, L.; Sintsova, A.; van Bakel, H.; Tian, S.; Navarre, W.W.; Xia, B.; Liu, J. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5154–5159. [CrossRef]
27. Gonzalo-Asensio, J.; Mostowy, S.; Harders-Westerveen, J.; Huygen, K.; Hernández-Pando, R.; Thole, J.; Behr, M.; Gicquel, B.; Martín, C. PhoP: A missing piece in the intricate puzzle of Mycobacterium tuberculosis virulence. *PLoS ONE* **2008**, *3*, e3496. [CrossRef] [PubMed]
28. Aguilar-Ayala, D.A.; Tilleman, L.; van Nieuwerburgh, F.; Deforce, D.; Palomino, J.C.; Vandamme, P.; Gonzalez-Y-Merchand, J.A.; Martin, A. The transcriptome of Mycobacterium tuberculosis in a lipid-rich dormancy model through RNAseq analysis. *Sci. Rep.* **2017**, *7*, 17665. [CrossRef]
29. Lee, J.; Lee, S.-G.; Kim, K.K.; Lim, Y.-J.; Choi, J.-A.; Cho, S.-N.; Park, C.; Song, C.-H. Characterisation of genes differentially expressed in macrophages by virulent and attenuated Mycobacterium tuberculosis through RNA-Seq analysis. *Sci. Rep.* **2019**, *9*, 4027. [CrossRef] [PubMed]
30. Huang, H.; Shao, X.; Xie, Y.; Wang, T.; Zhang, Y.; Wang, X.; Deng, X. An integrated genomic regulatory network of virulence-related transcriptional factors in Pseudomonas aeruginosa. *Nat. Commun.* **2019**, *10*, 2931. [CrossRef] [PubMed]
31. Danielli, A.; Amore, G.; Scarlato, V. Built shallow to maintain homeostasis and persistent infection: Insight into the transcriptional regulatory network of the gastric human pathogen Helicobacter pylori. *PLoS Pathog.* **2010**, *6*, e1000938. [CrossRef] [PubMed]
32. Dorella, F.A.; Pacheco, L.G.C.; Oliveira, S.C.; Miyoshi, A.; Azevedo, V. *Corynebacterium pseudotuberculosis*: Microbiology, biochemical properties, pathogenesis and molecular studies of virulence. *Vet. Res.* **2006**, *37*, 201–218. [CrossRef]
33. Yeruham, I.; Friedman, S.; Perl, S.; Elad, D.; Berkovich, Y.; Kalgard, Y. A herd level analysis of a *Corynebacterium pseudotuberculosis* outbreak in a dairy cattle herd. *Vet. Dermatol.* **2004**, *15*, 315–320. [CrossRef]
34. Silva, A.; Schneider, M.P.C.; Cerdeira, L.; Barbosa, M.S.; Ramos, R.T.J.; Carneiro, A.R.; Santos, R.; Lima, M.; D'Afonseca, V.; Almeida, S.S.; et al. Complete genome sequence of *Corynebacterium pseudotuberculosis* I19, a strain isolated from a cow in Israel with Bovine Mastitis. *J. Bacteriol.* **2011**, *193*, 323–324. [CrossRef] [PubMed]
35. Spier, S.J.; Azevedo, V. *Corynebacterium pseudotuberculosis* infection in horses: Increasing frequency and spread to new regions of North America. *Equine Vet. Educ.* **2017**, *29*, 436–439. [CrossRef]
36. Selim, S.A. Oedematous skin disease of buffalo in Egypt. *J. Vet. Med. B Infect. Dis. Vet. Public Health* **2001**, *48*, 241–258. [CrossRef] [PubMed]
37. Windsor, P.A.; Bush, R.D. Caseous lymphadenitis: Present and near forgotten from persistent vaccination? *Small Rumin. Res.* **2016**, *142*, 6–10. [CrossRef]

38. Brum, A.A.; de Rezende, A.F.S.; Brilhante, F.S.; Collares, T.; Begnine, K.; Seixas, F.K.; Collares, T.V.; Dellagostin, O.A.; Azevedo, V.; Santos, A.; et al. Recombinant esterase from *Corynebacterium pseudotuberculosis* in DNA and subunit recombinant vaccines partially protects mice against challenge. *J. Med. Microbiol.* **2017**, *66*, 635–642. [CrossRef] [PubMed]
39. Droppa-Almeida, D.; Franceschi, E.; Padilha, F.F. Immune-informatic analysis and design of peptide vaccine from multi-epitopes against *Corynebacterium pseudotuberculosis*. *Bioinform. Biol. Insights* **2018**, *12*, 117793221875533. [CrossRef]
40. Pinto, A.C.; de Sá, P.H.C.G.; Ramos, R.T.J.; Barbosa, S.; Barbosa, H.P.M.; Ribeiro, A.C.; Silva, W.M.; Rocha, F.S.; Santana, M.P.; de Paula Castro, T.L.; et al. Differential transcriptional profile of *Corynebacterium pseudotuberculosis* in response to abiotic stresses. *BMC Genomics* **2014**, *15*, 14. [CrossRef] [PubMed]
41. Gomide, A.C.P.; de Sá, P.G.; Cavalcante, A.L.Q.; de Jesus Sousa, T.; Gomes, L.G.R.; Ramos, R.T.J.; Azevedo, V.; Silva, A.; Folador, A.R.C. Heat shock stress: Profile of differential expression in *Corynebacterium pseudotuberculosis* biovar Equi. *Gene* **2018**, *645*, 124–130. [CrossRef]
42. Gomide, A.C.P.; Ibraim, I.C.; Alves, J.T.C.; de Sá, P.G.; de Oliveira Silva, Y.R.; Santana, M.P.; Silva, W.M.; Folador, E.L.; Mariano, D.C.B.; de Paula Castro, T.L.; et al. Transcriptome analysis of *Corynebacterium pseudotuberculosis* biovar Equi in two conditions of the environmental stress. *Gene* **2018**, *677*, 349–360. [CrossRef]
43. 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.; et al. Transcriptome profile of *Corynebacterium pseudotuberculosis* in response to iron limitation. *BMC Genomics* **2019**, *20*, 663. [CrossRef]
44. McKean, S.C.; Davies, J.K.; Moore, R.J. Expression of phospholipase D, the major virulence factor of *Corynebacterium pseudotuberculosis*, is regulated by multiple environmental factors and plays a role in macrophage death. *Microbiology* **2007**, *153*, 2203–2211. [CrossRef] [PubMed]
45. Billington, S.J.; Esmay, P.A.; Glenn Songer, J.; Helen Jost, B. Identification and role in virulence of putative iron acquisition genes from *Corynebacterium pseudotuberculosis*. *FEMS Microbiol. Lett.* **2002**, *208*, 41–45. [CrossRef] [PubMed]
46. Tiwari, S.; da Costa, M.P.; Almeida, S.; Hassan, S.S.; Jamal, S.B.; Oliveira, A.; Folador, E.L.; Rocha, F.; de Abreu, V.A.C.; Dorella, F.; et al. *C. pseudotuberculosis* Phop confers virulence and may be targeted by natural compounds. *Integr. Biol.* **2014**, *6*, 1088–1099. [CrossRef] [PubMed]
47. Turkarslan, S.; Peterson, E.J.R.; Rustad, T.R.; Minch, K.J.; Reiss, D.J.; Morrison, R.; Ma, S.; Price, N.D.; Sherman, D.R.; Baliga, N.S. A comprehensive map of genome-wide gene regulation in *Mycobacterium tuberculosis*. *Sci. Data* **2015**, *2*, 150010. [CrossRef]
48. Luo, F.; Yang, Y.; Zhong, J.; Gao, H.; Khan, L.; Thompson, D.K.; Zhou, J. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. *BMC Bioinform.* **2007**, *8*, 299. [CrossRef]
49. Liu, W.; Li, L.; Long, X.; You, W.; Zhong, Y.; Wang, M.; Tao, H.; Lin, S.; He, H. Construction and Analysis of gene co-expression networks in *Escherichia coli*. *Cells* **2018**, *7*, 19. [CrossRef] [PubMed]
50. Brinkrolf, K.; Brune, I.; Tauch, A. The transcriptional regulatory network of the amino acid producer *Corynebacterium glutamicum*. *J. Biotechnol.* **2007**, *129*, 191–211. [CrossRef]
51. Kohl, T.A.; Tauch, A. The GlxR regulon of the amino acid producer *Corynebacterium glutamicum*: Detection of the corynebacterial core regulon and integration into the transcriptional regulatory network model. *J. Biotechnol.* **2009**, *143*, 239–246. [CrossRef] [PubMed]
52. Balasubramanian, D.; Schneper, L.; Kumari, H.; Mathee, K. A dynamic and intricate regulatory network determines *Pseudomonas aeruginosa* virulence. *Nucleic Acids Res.* **2013**, *41*. [CrossRef] [PubMed]
53. Graham, M.R.; Smoot, L.M.; Migliaccio, C.A.L.; Virtaneva, K.; Sturdevant, D.E.; Porcella, S.F.; Federle, M.J.; Adams, G.J.; Scott, J.R.; Musser, J.M. Virulence control in group A Streptococcus by a two-component gene regulatory system: Global expression profiling and in vivo infection modeling. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13855–13860. [CrossRef]
54. Franco, E.F.; Rana, P.; Queiroz Cavalcante, A.L.; da Silva, A.L.; Cybelle Pinto Gomide, A.; Carneiro Folador, A.R.; Azevedo, V.; Ghosh, P.; Ramos, R.T.J. Co-expression networks for causal gene identification based on RNA-Seq data of *Corynebacterium pseudotuberculosis*. *Genes* **2020**, *11*, 794. [CrossRef] [PubMed]
55. Nalluri, J.J.; Barh, D.; Azevedo, V.; Ghosh, P. miRsig: A consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures. *Sci. Rep.* **2017**, *7*, 39684. [CrossRef] [PubMed]
56. Nalluri, J.J.; Rana, P.; Barh, D.; Azevedo, V.; Dinh, T.N.; Vladimirov, V.; Ghosh, P. Determining causal miRNAs and their signaling cascade in diseases using an influence diffusion model. *Sci. Rep.* **2017**, *7*, 8133. [CrossRef]
57. Chen, W.-H.; Lu, G.; Chen, X.; Zhao, X.-M.; Bork, P. OGEE v2: An update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res.* **2017**, *45*, D940–D944. [CrossRef]
58. Padilla, L.; Morbach, S.; Krämer, R.; Agosin, E. Impact of heterologous expression of *Escherichia coli* UDP-glucose pyrophosphorylase on trehalose and glycogen synthesis in *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* **2004**, *70*, 3845–3854. [CrossRef]
59. Anishetty, S.; Pulimi, M.; Pennathur, G. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput. Biol. Chem.* **2005**, *29*, 368–378. [CrossRef]
60. Sharp, P.M.; Mitchell, K.J. *Corynebacterium glutamicum* arginyl-tRNA synthetase. *Mol. Microbiol.* **1993**, *8*, 200. [CrossRef] [PubMed]
61. de Bruin, O.M.; Duplantis, B.N.; Ludu, J.S.; Hare, R.F.; Nix, E.B.; Schmerk, C.L.; Robb, C.S.; Boraston, A.B.; Hueffer, K.; Nano, F.E. The biochemical properties of the Francisella pathogenicity island (FPI)-encoded proteins IglA, IglB, IglC, PdpB and DotU suggest roles in type VI secretion. *Microbiology* **2011**, *157*, 3483–3491. [CrossRef] [PubMed]

62. Deng, C.; Lv, X.; Liu, Y.; Li, J.; Lu, W.; Du, G.; Liu, L. Metabolic engineering of *Corynebacterium glutamicum* S9114 based on whole-genome sequencing for efficient N-acetylglucosamine synthesis. *Synth. Syst. Biotechnol.* **2019**, *4*, 120–129. [CrossRef] [PubMed]
63. Boyle, D.S.; Donachie, W.D. mraY is an essential gene for cell growth in *Escherichia coli*. *J. Bacteriol.* **1998**, *180*, 6429–6432. [CrossRef]
64. Oguiza, J.A.; Malumbres, M.; Eriani, G.; Pisabarro, A.; Mateos, L.M.; Martin, F.; Martín, J.F. A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum: Regulation of argS-lysA cluster expression by arginine. *J. Bacteriol.* **1993**, *175*, 7356–7362. [CrossRef] [PubMed]
65. Cracan, V.; Padovani, D.; Banerjee, R. IcmF is a fusion between the radical B12 enzyme isobutyryl-CoA mutase and its G-protein chaperone. *J. Biol. Chem.* **2010**, *285*, 655–666. [CrossRef] [PubMed]
66. Hayashi, M.; Mizoguchi, H.; Shiraishi, N.; Obayashi, M.; Nakagawa, S.; Imai, J.-I.; Watanabe, S.; Ota, T.; Ikeda, M. Transcriptome analysis of acetate metabolism in *Corynebacterium glutamicum* using a newly developed metabolic array. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1337–1344. [CrossRef]
67. Yang, Y.; Zhang, M.; Zhang, H.; Lei, J.; Jin, R.; Xu, S.; Bao, J.; Zhang, L.; Wang, H. Purification and characterization of *Mycobacterium tuberculosis* indole-3-glycerol phosphate synthase. *Biochemistry* **2006**, *71*, S38–S43. [CrossRef] [PubMed]
68. Peters-Wendisch, P.; Stolz, M.; Etterich, H.; Kennerknecht, N.; Sahm, H.; Eggeling, L. Metabolic engineering of *Corynebacterium glutamicum* for L-serine production. *Appl. Environ. Microbiol.* **2005**, *71*, 7139–7144. [CrossRef]
69. Lim, C.J.; Hwang, W.; Park, E.H.; Fuchs, J.A. Cyclic AMP-dependent expression of the *Escherichia coli* serC-aroA operon. *Biochim. Biophys. Acta* **1994**, *1218*, 250–253. [CrossRef]
70. Ikeda, M.; Wachi, M.; Jung, H.K.; Ishino, F.; Matsuhashi, M. The *Escherichia coli* mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide: Undecaprenyl-phosphate phospho-N-acetylmuramoyl-pentapeptide transferase. *J. Bacteriol.* **1991**, *173*, 1021–1026. [CrossRef]
71. Burkovski, A. Cell envelope of corynebacteria: Structure and influence on pathogenicity. *ISRN Microbiol.* **2013**, *2013*, 935736. [CrossRef] [PubMed]
72. Chandrasekaran, S.; Price, N.D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 17845–17850. [CrossRef]
73. Moraes, G.L.; Gomes, G.C.; de Monteiro Sousa, P.R.; Alves, C.N.; Govender, T.; Kruger, H.G.; Maguire, G.E.M.; Lamichhane, G.; Lameira, J. Structural and functional features of enzymes of *Mycobacterium tuberculosis* peptidoglycan biosynthesis as targets for drug development. *Tuberculosis* **2015**, *95*, 95–111. [CrossRef]
74. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]
75. Wheeler, T.J.; Eddy, S.R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **2013**, *29*, 2487–2489. [CrossRef] [PubMed]
76. Pao, G.M.; Saier, M.H., Jr. Response regulators of bacterial signal transduction systems: Selective domain shuffling during evolution. *J. Mol. Evol.* **1995**, *40*, 136–154. [CrossRef]
77. Blanco, A.G.; Sola, M.; Gomis-Rüth, F.X.; Coll, M. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* **2002**, *10*, 701–713. [CrossRef]
78. Hoch, J.A. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **2000**, *3*, 165–170. [CrossRef]
79. López-Goñi, I.; Guzmán-Verri, C.; Manterola, L.; Sola-Landa, A.; Moriyón, I.; Moreno, E. Regulation of Brucella virulence by the two-component system BvrR/BvrS. *Vet. Microbiol.* **2002**, *90*, 329–339. [CrossRef]
80. Matsushita, M.; Janda, K.D. Histidine kinases as targets for new antimicrobial agents. *Bioorg. Med. Chem.* **2002**, *10*, 855–867. [CrossRef]
81. Cardona, P.J.; Asensio, J.G.; Arbués, A.; Otal, I.; Lafoz, C.; Gil, O.; Caceres, N.; Ausina, V.; Gicquel, B.; Martin, C. Extended safety studies of the attenuated live tuberculosis vaccine SO2 based on phoP mutant. *Vaccine* **2009**, *27*, 2499–2505. [CrossRef]
82. Capra, E.J.; Laub, M.T. Evolution of two-component signal transduction systems. *Ann. Rev. Microbiol.* **2012**, *66*, 325–347. [CrossRef]
83. Sorger-Herrmann, U.; Taniguchi, H.; Wendisch, V.F. Regulation of the pstSCAB operon in *Corynebacterium glutamicum* by the regulator of acetate metabolism RamB. *BMC Microbiol.* **2015**, *15*, 113. [CrossRef] [PubMed]
84. Söderberg, M.A.; Cianciotto, N.P. A Legionella pneumophila Peptidyl-Prolyl cis-trans Isomerase Present in Culture Supernatants is necessary for optimal growth at low temperatures. *Appl. Environ. Microbiol.* **2008**, *74*, 1634–1638. [CrossRef]
85. Göthel, S.F.; Scholz, C.; Schmid, F.X.; Marahiel, M.A. Cyclophilin and Trigger factor from *Bacillus subtilis* catalyze in Vitro protein folding and are necessary for viability under starvation conditions†. *Biochemistry* **1998**, *37*, 13392–13399. [CrossRef] [PubMed]
86. Unal, C.M.; Steinert, M. Microbial Peptidyl-Prolyl cis/trans Isomerases (PPIases): Virulence factors and potential alternative drug targets. *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 544–571. [CrossRef] [PubMed]
87. Choi, S.; Choi, E.; Cho, Y.-J.; Nam, D.; Lee, J.; Lee, E.-J. The Salmonella virulence protein MgtC promotes phosphate uptake inside macrophages. *Nat. Commun.* **2019**, *10*. [CrossRef] [PubMed]
88. Eriksson, S.; Lucchini, S.; Thompson, A.; Rhen, M.; Hinton, J.C.D. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol. Microbiol.* **2003**, *47*, 103–118. [CrossRef]
89. Heithoff, D.M.; Shimp, W.R.; House, J.K.; Xie, Y.; Weimer, B.C.; Sinsheimer, R.L.; Mahan, M.J. Intraspecies variation in the emergence of hyperinfectious bacterial strains in nature. *PLoS Pathog.* **2012**, *8*, e1002647. [CrossRef]

90. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [CrossRef]
91. Bibb, L.A.; Kunkle, C.A.; Schmitt, M.P. The ChrA-ChrS and HrrA-HrrS signal transduction systems are required for activation of the hmuO promoter and repression of the hemA promoter in *Corynebacterium diphtheriae*. *Infect. Immun.* **2007**, *75*, 2421–2431. [CrossRef] [PubMed]
92. Frunzke, J.; Gätgens, C.; Brocker, M.; Bott, M. Control of heme homeostasis in *Corynebacterium glutamicum* by the two-component system HrrSA. *J. Bacteriol.* **2011**, *193*, 1212–1221. [CrossRef]
93. Keppel, M.; Hünnefeld, M.; Filipchyk, A.; Viets, U.; Davoudi, C.-F.; Krüger, A.; Mack, C.; Pfeifer, E.; Polen, T.; Baumgart, M.; et al. HrrSA orchestrates a systemic response to heme and determines prioritization of terminal cytochrome oxidase expression. *Nucleic Acids Res.* **2020**, *48*, 6547–6562. [CrossRef] [PubMed]
94. Browning, D.F.; Busby, S.J.W. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2004**, *2*, 57–65. [CrossRef]
95. Browning, D.F.; Butala, M.; Busby, S.J.W. Bacterial transcription factors: Regulation by pick "N" Mix. *J Mol. Biol.* **2019**, *431*, 4067–4077. [CrossRef]
96. Merchant, S.S.; Helmann, J.D. Elemental economy: Microbial strategies for optimizing growth in the face of nutrient limitation. *Adv. Microb. Physiol.* **2012**, *60*, 91–210.
97. Andrews, S.C.; Robinson, A.K.; Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **2003**, *27*, 215–237. [CrossRef]
98. Fu, M.; Su, H.; Su, Z.; Yin, Z.; Jin, J.; Wang, L.; Zhang, Q.; Xu, X. Transcriptome analysis of *Corynebacterium pseudotuberculosis*-infected spleen of dairy goats. *Microb. Pathog.* **2020**, *147*, 104370. [CrossRef]
99. Kunkle, C.A.; Schmitt, M.P. Analysis of a DtxR-regulated iron transport and siderophore biosynthesis gene cluster in *Corynebacterium diphtheriae*. *J. Bacteriol.* **2005**, *187*, 422–433. [CrossRef] [PubMed]
100. Trost, E.; Ott, L.; Schneider, J.; Schröder, J.; Jaenicke, S.; Goesmann, A.; Husemann, P.; Stoye, J.; Dorella, F.A.; Rocha, F.S.; et al. The complete genome sequence of *Corynebacterium pseudotuberculosis* FRC41 isolated from a 12-year-old girl with necrotizing lymphadenitis reveals insights into gene-regulatory networks contributing to virulence. *BMC Genomics* **2010**, *11*, 728. [CrossRef]
101. Dorella, F.A.; Estevam, E.M.; Pacheco, L.G.C.; Guimarães, C.T.; Lana, U.G.P.; Gomes, E.A.; Barsante, M.M.; Oliveira, S.C.; Meyer, R.; Miyoshi, A.; et al. In vivo insertional mutagenesis in *Corynebacterium pseudotuberculosis*: An efficient means to identify DNA sequences encoding exported proteins. *Appl. Environ. Microbiol.* **2006**, *72*, 7368–7372. [CrossRef]
102. Ribeiro, D.; de Rocha, F.S.; Leite, K.M.C.; de Soares, S.C.; Silva, A.; Portela, R.W.D.; Meyer, R.; Miyoshi, A.; Oliveira, S.C.; Azevedo, V.; et al. An iron-acquisition-deficient mutant of *Corynebacterium pseudotuberculosis* efficiently protects mice against challenge. *Vet. Res.* **2014**, *45*, 28. [CrossRef]
103. Henderson, A.G.; Ehre, C.; Button, B.; Abdullah, L.H.; Cai, L.-H.; Leigh, M.W.; DeMaria, G.C.; Matsui, H.; Donaldson, S.H.; Davis, C.W.; et al. Cystic fibrosis airway secretions exhibit mucin hyperconcentration and increased osmotic pressure. *J. Clin. Invest.* **2014**, *124*, 3047–3060. [CrossRef] [PubMed]
104. Wood, J.M. Bacterial responses to osmotic challenges. *J. Gen. Physiol.* **2015**, *145*, 381–388. [CrossRef]
105. Janakiraman, A.; Lesser, C.F. How to manage stress: Lessons from an intracellular pathogen. *Virulence* **2017**, *8*, 359–361. [CrossRef] [PubMed]
106. Fang, F.C.; Frawley, E.R.; Tapscott, T.; Vázquez-Torres, A. Bacterial stress responses during host infection. *Cell Host Microbe* **2016**, *20*, 133–143. [CrossRef] [PubMed]
107. Jozefczuk, S.; Klie, S.; Catchpole, G.; Szymanski, J.; Cuadros-Inostroza, A.; Steinhauser, D.; Selbig, J.; Willmitzer, L. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* **2010**, *6*, 364. [CrossRef]
108. Ramos, J.L.; Martínez-Bueno, M.; Molina-Henares, A.J.; Terán, W.; Watanabe, K.; Zhang, X.; Gallegos, M.T.; Brennan, R.; Tobes, R. The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 326–356. [CrossRef] [PubMed]
109. McCarthy, R.R.; Mooij, M.J.; Reen, F.J.; Lesouhaitier, O.; O'Gara, F. A new regulator of pathogenicity (bvlR) is required for full virulence and tight microcolony formation in *Pseudomonas aeruginosa*. *Microbiology* **2014**, *160*, 1488–1500. [CrossRef]
110. Sasindran, S.J.; Saikolappan, S.; Dhandayuthapani, S. Methionine sulfoxide reductases and virulence of bacterial pathogens. *Future Microbiol.* **2007**, *2*, 619–630. [CrossRef]
111. Choi, S.H.; Baumler, D.J.; Kaspar, C.W. Contribution of dps to acid stress tolerance and oxidative stress tolerance in *Escherichia coli* O157: H7. *Appl. Environ. Microbiol.* **2000**, *66*, 3911–3916. [CrossRef]
112. Calhoun, L.N.; Kwon, Y.M. Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in *Escherichia coli*: A review. *J. Appl. Microbiol.* **2011**, *110*, 375–386. [CrossRef] [PubMed]
113. Das Gupta, T.; Bandyopadhyay, B.; Das Gupta, S.K. Modulation of DNA-binding activity of *Mycobacterium tuberculosis* HspR by chaperones. *Microbiology* **2008**, *154*, 484–490. [CrossRef] [PubMed]
114. Lindquist, S.; Craig, E.A. The heat-shock proteins. *Ann. Rev. Genetics* **1988**, *22*, 631–677. [CrossRef]
115. Bandyopadhyay, B.; Das Gupta, T.; Roy, D.; Das Gupta, S.K. DnaK dependence of the mycobacterial stress-responsive regulator HspR is mediated through its hydrophobic C-terminal tail. *J. Bacteriol.* **2012**, *194*, 4688–4697. [CrossRef]
116. Ishikawa, M.; Okamoto-Kainuma, A.; Matsui, K.; Takigishi, A.; Kaga, T.; Koizumi, Y. Cloning and characterization of clpB in *Acetobacter* pasteurianus NBRC 3283. *J. Biosci. Bioeng.* **2010**, *110*, 69–71. [CrossRef] [PubMed]

117. Chastanet, A.; Derre, I.; Nair, S.; Msadek, T. clpB, a novel member of the Listeria monocytogenes CtsR regulon, is involved in virulence but not in general stress tolerance. *J. Bacteriol.* **2004**, *186*, 1165–1174. [CrossRef]
118. Engels, S.; Ludwig, C.; Schweitzer, J.-E.; Mack, C.; Bott, M.; Schaffer, S. The transcriptional activator ClgR controls transcription of genes involved in proteolysis and DNA repair in *Corynebacterium glutamicum*. *Mol. Microbiol.* **2005**, *57*, 576–591. [CrossRef] [PubMed]
119. Doig, P.; Boriack-Sjodin, P.A.; Dumas, J.; Hu, J.; Itoh, K.; Johnson, K.; Kazmirski, S.; Kinoshita, T.; Kuroda, S.; Sato, T.-O.; et al. Rational design of inhibitors of the bacterial cell wall synthetic enzyme GlmU using virtual screening and lead-hopping. *Bioorg. Med. Chem.* **2014**, *22*, 6256–6269. [CrossRef]
120. Zhang, W.; Jones, V.C.; Scherman, M.S.; Mahapatra, S.; Crick, D.; Bhamidi, S.; Xin, Y.; McNeil, M.R.; Ma, Y. Expression, essentiality, and a microtiter plate assay for mycobacterial GlmU, the bifunctional glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphate uridyltransferase. *Int. J. Biochemistry Cell Biol.* **2008**, *40*, 2560–2571. [CrossRef]
121. Kim, J.K.; Lee, H.J.; Kikuchi, Y.; Kitagawa, W.; Nikoh, N.; Fukatsu, T.; Lee, B.L. Bacterial cell wall synthesis gene uppP is required for Burkholderia colonization of the Stinkbug Gut. *Appl. Environ. Microbiol.* **2013**, *79*, 4879–4886. [CrossRef] [PubMed]
122. Zhao, H.; Sun, Y.; Peters, J.M.; Gross, C.A.; Garner, E.C.; Helmann, J.D. Depletion of undecaprenyl pyrophosphate phosphatases disrupts cell envelope biogenesis in *Bacillus subtilis*. *J. Bacteriol.* **2016**, *198*, 2925–2935. [CrossRef] [PubMed]
123. Braibant, M.; Guilloteau, L.; Zygmunt, M.S. Functional characterization of *Brucella melitensis* NorMI, an efflux pump belonging to the multidrug and toxic compound extrusion family. *Antimicrob. Agents Chemother.* **2002**, *46*, 3050–3053. [CrossRef]
124. Morita, Y.; Kodama, K.; Shiota, S.; Mine, T.; Kataoka, A.; Mizushima, T.; Tsuchiya, T. NorM, a putative multidrug efflux protein, of Vibrio parahaemolyticus and its homolog in *Escherichia coli*. *Antimicrob. Agents Chemother.* **1998**, *42*, 1778–1782. [CrossRef] [PubMed]
125. Spirig, T.; Weiner, E.M.; Clubb, R.T. Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **2011**, *82*, 1044–1059. [CrossRef]
126. Paterson, G.K.; Mitchell, T.J. The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. *Microbes Infect.* **2006**, *8*, 145–153. [CrossRef]
127. Paget, M.S. Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules* **2015**, *5*, 1245–1265. [CrossRef]
128. Davis, M.C.; Kesthely, C.A.; Franklin, E.A.; MacLellan, S.R. The essential activities of the bacterial sigma factor. *Can. J. Microbiol.* **2017**, *63*, 89–99. [CrossRef]
129. Adcock, I.M.; Caramori, G. Transcription factors. *Asthma COPD* **2009**, 373–380.
130. Feklístov, A.; Sharon, B.D.; Darst, S.A.; Gross, C.A. Bacterial sigma factors: A historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* **2014**, *68*, 357–376. [CrossRef]
131. Kazmierczak, M.J.; Wiedmann, M.; Boor, K.J. Alternative sigma factors and their roles in bacterial virulence. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 527–543. [CrossRef] [PubMed]
132. Riordan, J.T.; Mitra, A. Regulation of *Escherichia coli* pathogenesis by alternative sigma factor N. *EcoSal Plus* **2017**, *7*. [CrossRef] [PubMed]
133. Pátek, M.; Nešvera, J. Sigma factors and promoters in *Corynebacterium glutamicum*. *J. Biotechnol.* **2011**, *154*, 101–113. [CrossRef] [PubMed]
134. Taniguchi, H.; Busche, T.; Patschkowski, T.; Niehaus, K.; Pátek, M.; Kalinowski, J.; Wendisch, V.F. Physiological roles of sigma factor SigD in *Corynebacterium glutamicum*. *BMC Microbiol.* **2017**, *17*, 158. [CrossRef]
135. Ruiz, J.C.; D'Afonseca, V.; Silva, A.; Ali, A.; Pinto, A.C.; Santos, A.R.; Rocha, A.A.M.C.; Lopes, D.O.; Dorella, F.A.; Pacheco, L.G.C.; et al. Evidence for reductive genome evolution and lateral acquisition of virulence functions in two *Corynebacterium pseudotuberculosis* strains. *PLoS ONE* **2011**, *6*, e18551. [CrossRef]
136. Helmann, J.D.; Chamberlin, M.J. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* **1988**, *57*, 839–872. [CrossRef]
137. Wu, S.; Howard, S.T.; Lakey, D.L.; Kipnis, A.; Samten, B.; Safi, H.; Gruppo, V.; Wizel, B.; Shams, H.; Basaraba, R.J.; et al. The principal sigma factor sigA mediates enhanced growth of *Mycobacterium tuberculosis* in vivo. *Mol. Microbiol.* **2004**, *51*, 1551–1562. [CrossRef]
138. Rodriguez Ayala, F.; Bartolini, M.; Grau, R. The stress-responsive alternative sigma factor SigB of *Bacillus subtilis* and its relatives: An old friend with new functions. *Front. Microbiol.* **2020**, *11*, 228. [CrossRef]
139. Pacheco, L.; Castro, T.; Carvalho, R.; Moraes, P.; Dorella, F.; Carvalho, N.; Slade, S.; Scrivens, J.; Feelisch, M.; Meyer, R.; et al. A role for Sigma Factor σ^E in *Corynebacterium pseudotuberculosis* resistance to nitric oxide/peroxide stress. *Front. Microbiol.* **2012**, *3*, 126. [CrossRef]

IV.2 Chapter II - Assessing the consistency between the *C. glutamicum* GRN and its gene expression data

IV.2.1 On the consistency between gene expression and the gene regulatory network of *Corynebacterium glutamicum*

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

Status: Published in the journal Network and Systems Medicine

One of this thesis's main objectives was to assess the consistency between the *C. glutamicum* GRN and a large gene expression dataset and, thus, compare the results with *in silico* generated data. To accomplish that, we first applied a sign consistency approach to assess the consistency between the *C. glutamicum* GRN and its gene expression data available on the GEO database. Then, gene expression data from an *in silico* GRN were simulated and assessed with the same method. Lastly, the consistency results were compared and plotted. This work shows that the *C. glutamicum* GRN is inconsistent with its gene expression data, which means that the *C. glutamicum* GRN cannot explain the gene expression data available on the GEO database. It also suggests that other mechanisms such as post-transcriptional regulation and signaling should be considered when reconstructing GRNs.

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

© Doglas Parise *et al.*, 2021; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sequencing DNA fragments that are bound to the TFs during chromatin immunoprecipitation, and is then used to map them to a reference genome.[6] RNA-Seq is used to quantify the entire set of RNA in a biological sample at a particular moment through high-throughput sequencing.[7] Several methods to infer GRNs from such gene expression data have been developed and evaluated[10–12]; these networks are commonly modeled as Boolean and Bayesian networks.[13–16] These methods have been applied to reconstruct experimental GRNs and have resulted in multiple databases and online platforms for the analysis of model organisms. Such resources include RegulonDB,[17] Subtiwiki,[18] CoryneRegNet,[19] and Abasy Atlas,[20] for *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, and several organisms, respectively. Previous evaluations of gene expression-based methods to infer GRNs demonstrated a moderate performance on experimental microarray data and a better performance on *in silico*-generated gene expression data.[10,11] Inference methods developed for both bulk and single-cell data were evaluated with single-cell transcriptomic data, and reached the conclusion that the algorithms performed poorly using both experimental or *in silico*-generated data.[21,22]

Despite the importance of GRNs and the low performance of GRN inference methods, few studies systematically evaluated the consistency of these networks with gene expression data. In 2003, Gutiérrez-Ríos et al.[23] assessed the consistency of the *E. coli* GRN in a few well-studied genes. Later, Siegel et al.[24] developed a mathematical framework evaluating the consistency in interaction graphs; they used an *in silico* experiment derived from experimental literature data of gene and metabolic networks to test their approach. Based on this framework, Guziolowski et al.[25] analyzed the consistency of the *E. coli* GRN using three independent microarray datasets and ascertained the inconsistency of the network. Guziolowski et al. developed BioQuali,[26] a Cytoscape app for detecting inconsistencies in GRNs and suggesting changes that would restore the network consistency with the user-provided expression data. Other Cytoscape apps such as contradictions in microarrays,[27] CytoASP,[28] and SigNetTrainer[29] assess the consistency of interaction networks and expression data from a single study. The first makes use of Boolean network models to detect inconsistencies in interaction networks. CytoASP[28] uses logical roles through Answering Set Programming to identify inconsistencies and to suggest how to repair them. SigNetTrainer[29] uses Integer Linear Programming to detect and remove inconsistencies from the networks. Collectively, these pioneer works allowed researchers to evaluate the consistency of the existing GRNs and their gene expression data. Furthermore, some of them pointed to the inconsistency of the GRNs, or gene sets, when evaluated with small sets of regulatory data.

Recently, Larsen et al.[30] studied *E. coli* GRNs and found that they are inconsistent when evaluated with gene expression data. The authors used a conservative sign consistency approach on a large microarray data compendium. Here, we analyzed the consistency of *C. glutamicum* GRN using a similar approach and also included RNA-seq gene expression data to obtain an exhaustive data compendium. In general, we assume that activations should increase the expression of the TGs when the TF is upregulated. Likewise, repressions should reduce the TGs' expression, when the TF is also upregulated. Our results show a positive correlation in both cases, contradicting our current understanding of the role of TF regulation. The consistency model assessment indicates that the *C. glutamicum* GRN is even less consistent than random GRNs, implying both that additional research is needed to further refine GRNs and additional factors have to be considered to explain gene expression.

Materials and Methods

C. glutamicum GRN and gene expression data

The experimental GRN of *C. glutamicum* was downloaded from CoryneRegNet 7.0.[19] The gene expression compendium was retrieved from Gene Expression Omnibus (GEO)[31] and consisted of microarray and RNA-seq data with a total of 429 samples (see Supplementary Data S1 for more information about the datasets).

Gene expression data normalization

All gene expression datasets were normalized using limma.[32] Microarray data were background corrected before applying lowess (two-color microarrays) and quantile normalization. Similarly, RNA-seq data were quantile normalized using the voom method.[33] Finally, to have all the data on the same scale, we combined and z-score normalized all gene expression data using the survJamda package.[34] The z-normalized distribution of the expression data is given in Supplementary Figure S1.

In silico data generation

The *in silico* GRN and gene expression data were generated using GeneNetWeaver.[35] This software

uses an experimental network as the model to create an *in silico* network with similar topology and to simulate gene expression data for the novel network.

Inconsistency detection and assessment

We used the same method applied by Larsen et al.[30] to assess *E. coli* GRN. In the first step, the method identifies genes that are up- or downregulated in each experiment (contrasts). This is performed by computing the contrasts as the difference between the expression of the reference and the case(s) in the experiment. Next, it uses a conservative sign consistency model similar to the one applied in COMA[27] and BioQuali.[26] For each contrast, the model labels the vertices as: upregulated, downregulated, or unchanged. The labels are attributed based on the expression differences between the control and each case in the contrast, and on a threshold _*t*_. Then the labels are compared with the role of the regulatory interactions in the GRN to determine the consistency. Finally, we compared the experimental network with random data by applying two perturbation methods. The first method shuffles the expression profiles and keeps the network topology, whereas the second shuffles the network topology and keeps the node degrees and gene expression profiles. For more details about this method, refer to the work from Larsen et al.[30]

Statistical analyses

The mean correlation (*mc*) was calculated by first adding the correlation of each TF and TG/operon pair (*cPairs*), and then dividing the sum by the total number of pairs (*nPairs*).

$$mc = \left(\sum_{i=1}^{nPairs} cPairs_i\right)/nPairs$$

This same process was applied to compute the means for Pearson's and Spearman's correlation. We computed the mean of the global inconsistency load (*mGlobal*) of the two applied perturbation methods by adding the global inconsistency load (*global*) of each iteration and then dividing the sum by the number of iterations (*nIteration*).

$$mGlobal = \left(\sum_{i=1}^{nIteration} global_i\right)/nIteration$$

We determined the mean edge inconsistency (*mEdge*) through the sum of the number of inconsistencies of each TF and TG/operon pair (*nIncons*), which we then divided by the total number of pairs (*nPairs*).

$$mEdge = \sum_{i=1}^{nIncons} nIncons_i)/nPairs$$

The significance of the comparison between the inconsistency load in contrasts with and without perturbation was computed using the Mann–Whitney *U*-test.[36] We used the same test to compute the significance of the number of up- or downregulated genes in each contrast, with and without perturbation.

Results

Correlation between gene regulatory interactions and gene expression profiles

To assess the correlation between expression profiles of known regulatory interactions (e.g., a TF regulating a TG), we applied Pearson's correlation coefficient (Fig. 1). We analyzed the distribution of correlations between known TF–TG pairs and of all possible TF–TG pairs. The mean correlation of known TF–TG pairs is 0.09, whereas for all the possible TF–TG pairs it is −0.0003 (Fig. 1A). For activations, we expect a positive correlation (e.g., increased TF expression enhances TG expression), whereas repressions should result in a negative correlation (e.g., increased TF expression reduces TG expression). However, separately analyzing the distribution of the correlations of known TF–TG pairs by interaction role, we see a very low mean correlation in both cases: 0.11 for activating interactions and to 0.07 for repressing interactions. Complex regulations where multiple TFs control the same TG could influence the results observed in Figure 1B. Taking this into account, we also analyzed the TF–TG pairs where the TG is regulated by a single TF (Fig. 1C). Of interest, the results are similar (0.10 and 0.04 for activating and repressing interactions, respectively).

The same process was repeated using an *in silico* network and gene expression data to demonstrate the extent of the correlations expected from a GRN that is consistent with the observed expression data. The *in silico* data were generated based on the *C. glutamicum* GRN using GeneNetWeaver.[35]

The mean correlation of known interactions was −0.06 and the mean correlation of all possible TF–TG pairs was 0.002 (Fig. 1D). The correlation of the interactions (Fig. 1E) separated by the interaction role presented a distinct partition between them: the mean correlation was 0.22 and −0.28 for activation

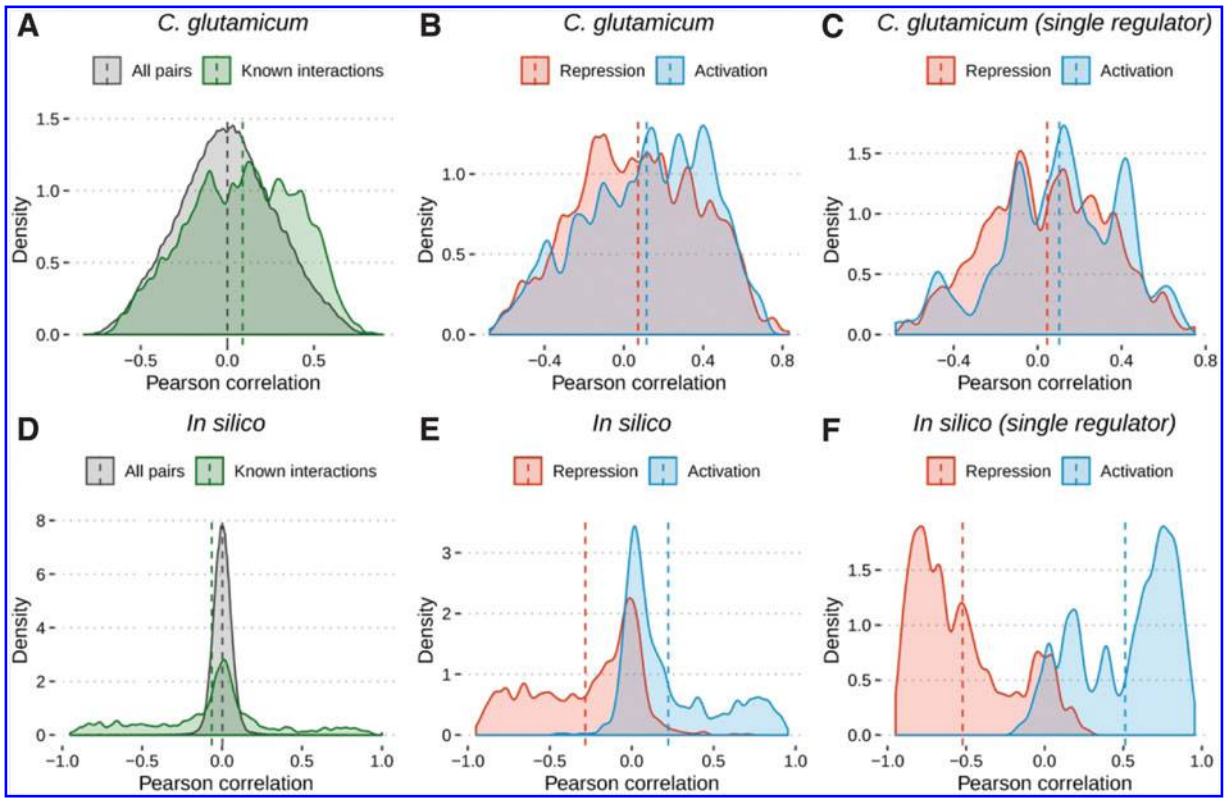


FIG. 1. Distribution of Pearson correlation for TF and TG/operon pairs. Comparison between the correlation of all possible TF–TG pairs and all known TF–TG pairs. [**(A)** *Corynebacterium glutamicum* and **(D)** *in silico*]. Comparison between the correlation of known TF–TG pairs separated by interaction role: activation and repression [**(B)** *C. glutamicum* and **(E)** *in silico*]. Comparison between the correlation of known TF–TG pairs where each TG has only one regulator [**(C)** *C. glutamicum* and **(F)** *in silico*]. Dashed vertical lines show the mean correlation for each TF and TG/operon pair. TF, transcription factor; TG, target gene.

and repression interactions, respectively (Fig. 1E). Upon analyzing only the single regulators, an even stronger separation between them was noticed: mean 0.51 and −0.52 for activation and repression interactions, respectively (Fig. 1F).

Consistency assessment between the regulatory networks and the expression profiles

To assess the consistency between the GRN and gene expression, the model we applied makes use of a threshold. Our threshold considers any changes in the expression data (upregulation, downregulation, or unchanged) and assumes ∼50% of the contrasts to be up- or downregulated, as previously explained by Larsen et al.[30] This resulted in a threshold of ±0.25 for the experimental data and ±0.82 for the *in silico* data. We compared the global inconsistency load of the experimental and *in silico* GRNs against randomly perturbed GRNs and expression profiles. The global inconsistency load (Fig. 2A) of the experimental network (31,922 inconsistencies) was higher than the mean of the perturbed data (31,030.30 and 31,733.00 inconsistencies in swapped edges and swapped expression profiles, respectively). These numbers indicate that the consistency between the experimental network and the expression data are not more significant than the consistency in the random networks. In contrast, the original *in silico* network (Fig. 2D) had fewer inconsistencies (20,538) than the mean of the perturbed data (34,133.40 and 33,180.00 in swapped edges and swapped expression profiles, respectively).

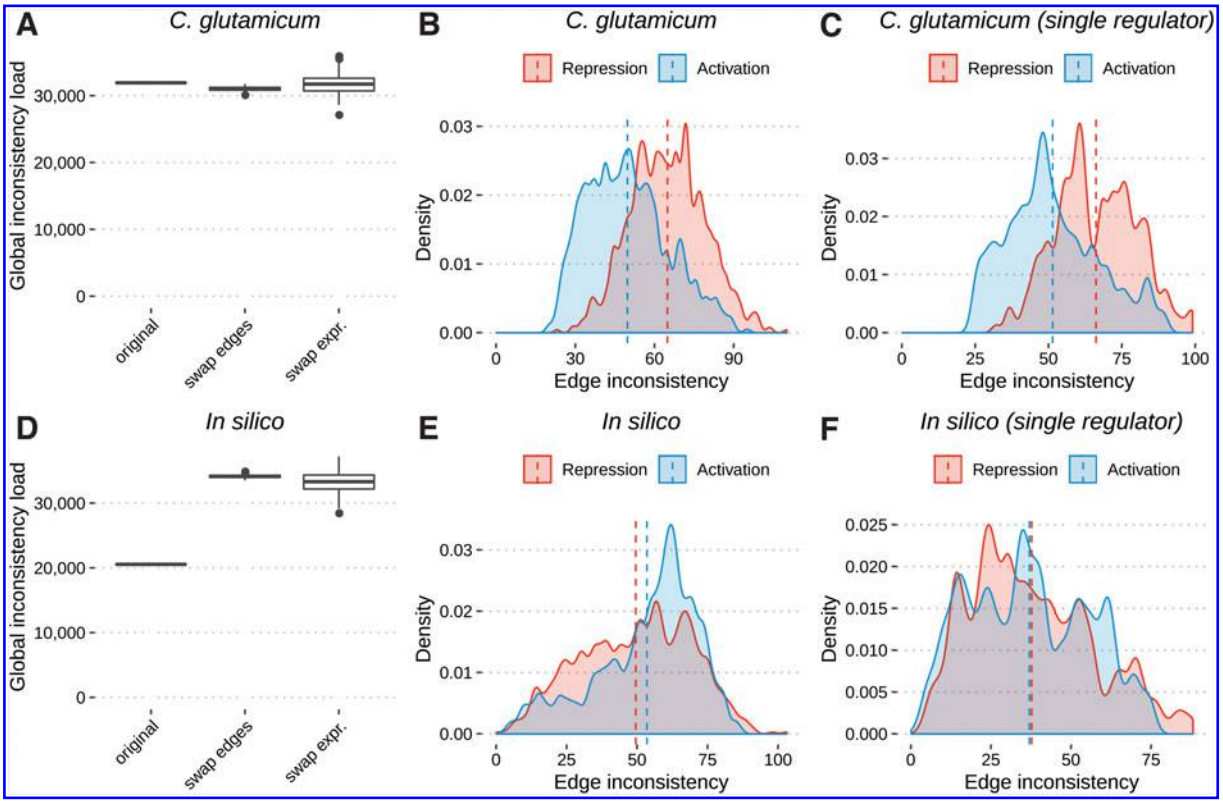


FIG. 2. Evaluation of the inconsistency load in the GRN and perturbed GRN models. Comparison among the global inconsistency load (total number of inconsistent cases) in the GRNs with two random GRN models. The experiments were repeated 200 times for the random models [**(A)** *Corynebacterium glutamicum* and **(D)** *in silico*]. The edge inconsistency distribution split by interaction role: repression and activation [**(B)** *C. glutamicum* and **(E)** *in silico*]. The edge inconsistency distribution is split by role: repression and activation where a single TF regulates the TG/operon **(C, F)**. In [**(B, C)** *C. glutamicum*, **(E, F)** *in silico*] dashed vertical lines show the mean inconsistency for each pair TF and TG/operon. GRN, Gene Regulatory Network.

Analyzing the edge inconsistencies separated by the interaction role revealed that the number of inconsistencies was larger for repressions (mean 64.8) than for activations (mean 49.7; Fig. 2B). The analysis of single regulators presented similar results, but slightly larger means (mean 66.1 and 51.3 for repressions and activations, respectively) in both cases (Fig. 2C). For the *in silico* data, the inconsistencies separated by the interaction role (Fig. 2E) presented similar numbers in both cases (mean of 49.5 and 53.4 for repressions and activations, respectively). The results for single regulators (Fig. 2F) were smaller and even more similar (mean 37.7 and 37 for repressions and activations, respectively).

Association between inconsistency load in contrast with and without perturbation

To check if perturbed experiments result in a higher level of inconsistencies, we analyzed the inconsistency load across the cases in each experiment, resulting in 239 contrasts. It resulted in a range from 5 to 306 inconsistencies with a mean of 133.56 (Fig. 3A). The number of inconsistencies was slightly more substantial in the contrasts with perturbed conditions than with the unperturbed ones (means of 145.31 and 127.88, respectively; Fig. 3B). Perturbed conditions in the experimental gene expression profiles include stress, overexpressed genes, knockout, and double-knockout genes. Genes

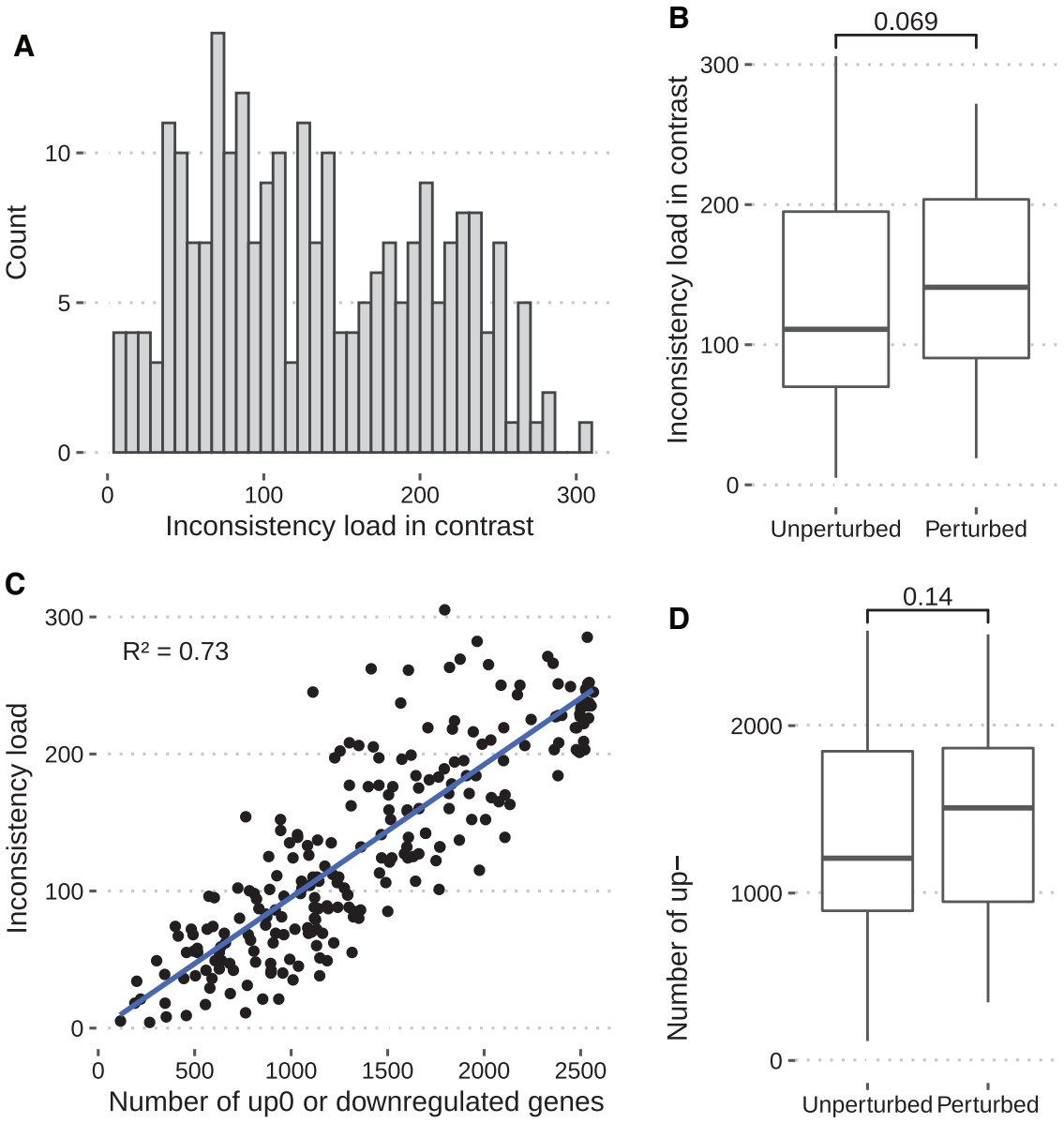


FIG. 3. Evaluation of the inconsistency load of *Corynebacterium glutamicum* across contrasts. Distribution of the number of inconsistencies across the 239 contrasts **(A)**. Perturbations (e.g., stress conditions) increase the inconsistency load in contrasts when compared with nonperturbations **(B)**. Relationship between the number of deregulated genes (up or down) and inconsistency load in contrasts **(C)**. Comparison between the number of deregulated genes in the model for the contrasts with and without perturbation **(D)**. The *p*-values in **(B)** and **(D)** were computed using the Mann–Whitney *U*-test.

considered to be up- or downregulated are associated with a higher level of inconsistencies in the network (Fig. 3C). Finally, a higher number of genes on average were considered up- or downregulated within the perturbed contrasts (Fig. 3D).

Discussion

In this work, we applied microarray and RNA-seq data to investigate the widely accepted assumption in which changes in the expression of gene regulators affect the expression of their TGs. The regulation type (activation

or repression) determines the effect of expression changes. Of interest, our results show that this is not the case for *C. glutamicum*, particularly at the transcriptional level. We found a positive mean correlation between the TFs and the TGs, even for repression interactions, which are supposed to have a negative correlation. When multiple TFs regulate the same TG, we cannot expect all regulations to be consistent. However, when analyzing the TGs that are regulated by a single regulator, a positive mean correlation between the TFs and the TGs was also observed. Larsen et al. found a similar behavior in the *E. coli* GRN[30]; considering it is more complete than *C. glutamicum* GRN,[37] it is not surprising that *C. glutamicum* GRN is also not consistent with its expression data. Pearson's correlation analysis may not identify nonlinear relationships between the expression of TFs and TGs. We, thus, also applied Spearman's rank correlation (Supplementary Fig. S2) and observed a slightly positive interaction when analyzing, first, all known interactions at once (mean correlation of 0.09), then those split by interaction role (activations and repressions, mean correlation of 0.12 and 0.08, respectively) and, finally, for single regulators (mean correlation of 0.11 and 0.05 for activations and repressions, respectively). It demonstrates that nonlinearity does not affect our results.

The analysis of the global inconsistency load (Fig. 2A) shows that the experimental network is slightly more inconsistent than the perturbed ones. This may imply that the experimental GRN is not explained by the expression data retrieved from the GEO database. Because the chosen threshold could have influenced these results, we repeated the analysis using thresholds that consider ~33% and ~66% of the contrasts as up- or downregulated, as previously suggested by Larsen et al.,[30] with no noticeable effect on the inconsistency between the GRN and the gene expression data (Supplementary Figs. S3–S6). Supplementary Figures S7–S9 demonstrate that network perturbation increases the global inconsistency of the in silico network while it has little effect on *C. glutamicum* GRN.

The high inconsistency between the *C. glutamicum* network and the expression data may be owing to technical or biological reasons. For example, the methods used to generate the GRNs from the experimental data may have performed poorly.[10,11,21,22] Another explanation may be the unavailability of adequate time-series experimental data, which would reveal regulatory interactions at different time points.[38] In addition, other regulatory mechanisms such as post-translational modifications,[39,40] inactive conformation of the TFs and metabolites[23] are neither identified by microarrays nor by RNA-seq data. Moreover, a low correlation between transcriptome and proteome data exists.[41,42]

Conclusion

Our results corroborate previous studies that analyzed the consistency between *E. coli* GRN and gene expression data, and further showed that, when considering the static GRN, the networks are inconsistent.[23,25,30] These results suggest that the traditional methods to reconstruct GRNs may not be able to fully represent the complexity of gene expression regulation. In the case of *C. glutamicum*, an accurate and consistent GRN is essential for the development of various robust strains that are required to meet its industrial demand in the production of biomolecules, such as amino acids.[43] Although the accurate construction of *C. glutamicum* GRNs may provide an understanding of possible biosynthetic routes for these molecules, the inconsistency of GRNs and gene expression data suggest that extra caution should still prevail when using GRNs to elucidate probable amino acid biosynthetic routes in the biotechnology industry.

In our view, the incorporation of multi-omics time-series data and robust statistical approaches, as well as the performance of multiple perturbations within biological systems, are necessary to model the GRNs accurately. Moreover, a single GRN may not be adequate to capture the regulatory landscape across all possible conditions, requiring the development of condition-specific and dynamic GRNs in the future.

Authors' Contributions

D.P. downloaded the data and ran the analysis. D.P., M.T.D.P., and E.K. did the data curation. D.P. and M.T.D.P. wrote the article. D.P., J.B., and V.A.de.C.A. designed the project. All the authors reviewed and approved the article.

Disclaimer

The article has been submitted solely to this Journal and is not published, in press, or submitted elsewhere.

Availability of Data and Materials

All data used in this work are publicly available on the GEO database (https://www.ncbi.nlm.nih.gov/geo). See Supplementary Table S1 in Supplementary Data S2 for more details.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

J.B. is grateful for support from H2020 grant RepoTrial (No. 777111) and his VILLUM Young Investigator grant (No. 13154). D.P. received support from CAPES (No. 88887.364607/2019-00) and M.T.D.P. from CNPq (No. 201336/2018-9), for their work at TUM in Germany. M.T.D.P.'s work was also supported by the German Research Foundation (under SFB924). V.A.de.C.A. is grateful for support from his CNPq Research Productivity grant (No. 305093/2015-0), CNPq Universal grant (No. 405233/2016-7) and FAPEMIG grant (No. APQ 02600-17). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Supplementary Material

Supplementary Data S1
Supplementary Data S2
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Figure S7
Supplementary Figure S8
Supplementary Figure S9
Supplementary Table S1

References

1. Land M, Hauser L, Jun SR, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics. 2015;15:141–161.
2. López-Maury L, Marguerat S., Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9:583–593.
3. Babu MM, Lang B, Aravind L. Methods to reconstruct and compare transcriptional regulatory networks. Methods Mol Biol. 2009;541:163–180.
4. Baumbach J, Brinkrolf K, Czaja LF, et al. CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. BMC Genomics. 2006;7:24.
5. Schena M, Shalon D, Davis RW, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270:467–470.
6. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129:823–837.
7. Nagalakshmi U, Wang Z, Waern K, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320:1344–1349.
8. Horak CE, Snyder M. ChIP-chip: a genomic approach for identifying transcription factor binding sites. Methods Enzymol. 2002;350:469–483.
9. Bumgarner R. Overview of DNA microarrays: types, applications, and their future. Curr Protoc Mol Biol. 2013;Chapter 22:Unit 22.1.
10. Marbach D, Costello JC, Küffner R, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 2012;9:796–804.
11. Madhamshettiwar PB, Maetschke SR, Davis MJ, et al. Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. Genome Med. 2012;4:41.
12. De Smet R, Marchal K. Advantages and limitations of current network inference methods. Nat Rev Microbiol. 2010;8:717–729.
13. Faure A, Naldi A, Chaouiya C, et al. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics. 2006;22:e124–e131.
14. Lovrics A, Gao Y, Juhász B, et al. Boolean modelling reveals new regulatory connections between transcription factors orchestrating the development of the ventral spinal cord. PLoS One. 2014;9:e111430.
15. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *Proceedings of the Fourth Annual International Conference on Computational Molecular Biology—RECOMB'00*. Tokyo, Japan. April 2000. pp. 127–135.
16. Yu J, Smith VA, Wang PP, et al. Using Bayesian network inference algorithms to recover molecular genetic regulatory networks. Int Conf Syst Biol 2002;2002:ICSB02.
17. Santos-Zavaleta A, Salgado H, Gama-Castro S, 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. 2019;47:D212–D220.
18. Zhu B, Stülke J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. Nucleic Acids Res. 2018;46:D743–D748.
19. Parise MTD, Parise D, Kato RB, et al. CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. Sci Data. 2020;7:142.
20. Escorcia-Rodríguez JM, Tauch A, Freyre-González JA. Abasy Atlas v2.2: the most comprehensive and up-to-date inventory of meta-curated, historical, bacterial regulatory networks, their completeness and system-level characterization. Comput Struct Biotechnol J. 2020;18:1228–1237.
21. Chen S, Mar JC. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. BMC Bioinformatics. 2018;19:232.
22. Pratapa A, Jalihal AP, Law JN, et al. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods. 2020:17:147–154.
23. Gutiérrez-Ríos RM, Rosenblueth DA, Loza JA, et al. Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles. Genome Res. 2003;13:2435–2443.
24. Siegel A, Radulescu O, Le Borgne M, et al. Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks. Biosystems. 2006;84:153–174.
25. Guziolowski C, Veber P, Le Borgne M, et al. Checking consistency between expression data and large scale regulatory networks: a case study. J Biol Phys Chem. 2007;7:37–43.
26. Guziolowski C, Bourdé A, Moreews F, et al. BioQuali Cytoscape plugin: analysing the global consistency of regulatory networks. BMC Genomics. 2009;10:244.
27. Baumbach J, Apeltsin L. Linking Cytoscape and the corynebacterial reference database CoryneRegNet. BMC Genomics. 2008;9:184.
28. Kittas A, Barozet A, Sereshti J, et al. CytoASP: a Cytoscape app for qualitative consistency reasoning, prediction and repair in biological networks. BMC Syst. Biol. 2015;9:34.
29. Melas IN, Samaga R, Alexopoulos LG, et al. Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs. PLoS Comput Biol. 2013;9:e1003204.
30. Larsen SJ, Röttger R, Schmidt HHHW, et al. *E. coli* gene regulatory networks are inconsistent with gene expression data. Nucleic Acids Res. 2019;47:85–92.
31. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41:D991–D995.
32. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
33. Law CW, Chen Y, Shi W, et al. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15:R29.
34. Yasrebi H. SurvJamda: an R package to predict patients' survival and risk assessment using joint analysis of microarray gene expression data. Bioinformatics. 2011;27:1168–1169.
35. Schaffter T, Marbach D, Floreano D. GeneNetWeaver: *in silico* benchmark generation and performance profiling of network inference methods. Bioinformatics. 2011;27:2263–2270.
36. McKnight PE, Najab J. Mann-Whitney U Test. The Corsini Encyclopedia of Psychology. (Weiner IB, Craighead WE; eds). John Wiley & Sons, Inc., Hoboken, NJ. 2010.

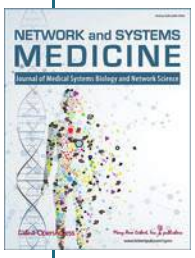
37. Röttger R, Rückert U, Taubert J, et al. How little do we actually know? On the size of gene regulatory networks. IEEE/ACM Trans Comput Biol Bioinform. 2012;9:1293–1300.
38. Yu H, Luscombe NM, Qian J, et al. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends Genet. 2003; 19:422–427.
39. Hunter T, Karin M. The regulation of transcription by phosphorylation. Cell. 1992;70:375–387.
40. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol Sci. 2014;35:76–85.
41. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett. 2009;583:3966–3973.
42. Ghazalpour A, Bennett B, Petyuk VA, et al. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet. 2011;7:e1001393.
43. Ivanov K, Stoimenova A, Obreshkova D, et al. Biotechnology in the production of pharmaceutical industry ingredients: amino acids. Biotechnol Biotechnol Equipment. 2013;27:3620–3626.

Cite this article as: Parise D, Parise MTD, Kataka E, Kato RB, List M, Tauch A, Azevedo VAdeC, Baumbach J (2021) On the consistency between gene expression and the gene regulatory network of *Corynebacterium glutamicum*, *Network and Systems Medicine* 4:1, 51–59, DOI: 10.1089/nsm.2020.0014

Abbreviations Used

GEO = Gene Expression Omnibus
GRNs = Gene Regulatory Networks
TFs = transcription factors
TGs = target genes

Publish in Network and Systems Medicine



- Immediate, unrestricted online access
- Rigorous peer review
- Compliance with open access mandates
- Authors retain copyright
- Highly indexed
- Targeted email marketing

liebertpub.com/nsm

V. DISCUSSION

Understanding the organisms' transcriptional regulatory mechanisms is an essential step to understand their ability to quickly adapt to challenging and constantly changing environments. In the context of the *Corynebacterium* genus, CoryneRegNet is the reference database for transcriptional GRNs. The seventh version of CoryneRegNet presents an entirely redesigned back- and front-end to support the new functionalities and database growth. It provides the most extensive data collection available for the *Corynebacterium* genus. CoryneRegNet allows users to download the GRN of each organism as a sif file, allowing them to visualize, modify and enrich the networks locally using third-party software. The new transfer pipeline presents two main improvements: (i) the use of profile HMMs and (ii) the results' interpretability. Given the low evolutionary conservation of TFBSs between species (BAUMBACH, 2010), the use of profile HMMs provides a more robust and flexible approach to predict TFBSs when compared with position weight matrices, due to its ability to model insertions, deletions and length variation in model TFBSs (DELORENZI; SPEED, 2002; RIVA, 2012; WHEELER; EDDY, 2013). Furthermore, we now assign p-values to the regulatory interactions, representing the likelihood of observing such interactions conserved by chance and thus aiding in the results' interpretability.

The combination of transcriptional regulatory data collected, generated and stored in CoryneRegNet 7 and literature search allowed us to present the first review of transcriptional regulatory mechanisms in *C. pseudotuberculosis*. In this pathogenic bacteria, we can highlight the following two-component systems: (i) PhoPR, a drug target predicted to regulate the expression of virulence factors, and (ii) HrrSA, the global regulator of heme in *C. glutamicum* (KEPPEL *et al.*, 2020). The first has already been studied experimentally, while the biological function of the second remains to be better investigated; both are great targets to be studied in the light of transcriptional regulation. DtxR, the master regulator of iron in *C. glutamicum* (BRINKROLF *et al.*, 2010), is downregulated under iron restriction media (IBRAIM *et al.*, 2019). This TF regulates 16 genes in *C. pseudotuberculosis* T1 and Cp13, including the ones encoding hemin uptake systems. These specific genes are located in the genome together with several known virulence factors in genomic islands, which may indicate both their potential acquisition through horizontal gene transfer (HGT) events (BONHAM; WOLFE; DUTTON, 2017; XIONG *et al.*, 2015) and the role of these systems in *C. pseudotuberculosis*' virulence (BILLINGTON *et al.*, 2002; RIBEIRO *et al.*, 2014). The TF TetR2 was the only gene differentially expressed in both *C. pseudotuberculosis* 1002 and 258 under osmotic, thermal and acid stress (GOMIDE; DE SÁ; *et al.*, 2018; GOMIDE; IBRAIM; *et al.*, 2018; PINTO *et al.*, 2014). TetR2 belongs to a TetR family of transcriptional regulators, which are well-known for regulating genes associated with drug resistance, biofilm formation, pathogenicity and virulence (LIU *et al.*, 2017; RAMOS *et al.*, 2005); however, there are no known regulatory interactions for this gene in both studied strains. There is little information available regarding sigma factor regulations in *C. pseudotuberculosis*. In *C. pseudotuberculosis* 1002, SigM was the only sigma factor differentially expressed in osmotic, thermal and acid stresses; the housekeeping SigA was found as a differentially expressed gene (DEG) only in the first condition and; SigB, SigE and

SigH were found as DEGs in the third condition (PINTO *et al.*, 2014). Collectively, these results present the first attempt to unravel and understand the regulatory landscape of *C. pseudotuberculosis* in the light of GRNs.

Note that the approach applied to transfer GRNs also presents a few limitations: (i) limited availability of experimental data regarding bacterial GRNs; (ii) a high dependence on known regulatory interactions of model organisms; and (iii) genome sequence similarity between model and target organisms. When considering organisms with diverse life-styles, such as the *Corynebacterium* genus, the amount of experimentally validated GRNs becomes especially relevant when considering possible regulatory interactions acquired by HGT. It has already been demonstrated that HGT events can co-transfer neighbor regulators with their TGs (PRICE; DEHAL; ARKIN, 2008) and that known virulence factors are located in the mobile genetic elements of organisms such as in Enterohemorrhagic *E. coli* and *C. pseudotuberculosis* (PAULING; RÖTTGER; NEUNER; *et al.*, 2012; SOARES *et al.*, 2012).

Another point to be considered is the accuracy of the experimentally validated GRNs of the model organisms; their accuracy will directly influence both the results of the analysis and experiments that rely on them, as well as the quality of the predicted GRNs. The assessment of the consistency of *C. glutamicum* GRN with its gene expression data showed that there is a large number of inconsistencies in the network and a low correlation between the expression of regulators and their TGs. When considering repression relationships, the increase in the expression of the regulator should result in the decrease of the expression of the TG, which is represented by a negative correlation in Pearson's analysis. However, it is not the case for the *C. glutamicum* GRN, which presents a slightly positive correlation in both cases when: (i) analyzing all repression interactions together and (ii) the TG has a single regulator. This can be partially explained by the fact that the set of transcriptional regulatory interactions of *C. glutamicum* is still incomplete (ESCORCIA-RODRÍGUEZ; TAUCH; FREYRE-GONZÁLEZ, 2020; FREYRE-GONZÁLEZ; TAUCH, 2017) and, thus, the expression of the genes regulated by a single TF in the network may be influenced by more regulators.

The similar degree of inconsistencies found in both experimental and random networks suggests that *C. glutamicum* GRN cannot explain the gene expression data better than random models. To reconstruct more accurate and consistent experimental GRNs, one should consider incorporating multi-omic time-series data and robust statistical approaches. As a first step, we recently collected sRNA data from literature combined with the prediction of new sRNAs and their regulatory interactions, thus adding a new layer of regulation in 6 corynebacterial organisms in CoryneRegNet (Parise *et al.*, submitted).

VI. CONCLUSIONS AND OUTLOOK

VI.1 Conclusions

CoryneRegNet's updated network-transfer method allowed us to transfer the regulatory networks of four model organisms to 224 organisms belonging to the *Corynebacterium* genus; this increased the number of organisms of this genus with GRN available by ~20 times. The new interface presents each network as a dynamic graph and allows the user to visualize only the regulations of a gene or operon of interest. Combining the predicted GRNs from CoryneRegNet 7 (PARISE, M. T. D. *et al.*, 2020) with literature data allowed us to present the first glimpse of the transcriptional regulatory mechanisms of *C. pseudotuberculosis*. Transcription factors, sigma factors and two-component systems play a crucial role in organism adaptation capacity and thus are relevant drug and vaccine targets. Future experimental assays may contribute to further understand these mechanisms. In this context, DtxR and HrrSA are good targets for experimental investigation, taking into account that both regulate genes involved in hemin uptake systems. Furthermore, we assessed the consistency between the *C. glutamicum* GRN and the gene expression data available on the GEO database. These results indicate a high level of inconsistencies and a lack of correlation between the expression of regulatory proteins and their target genes. It indicates that more omics data combined with robust statistics methods should be integrated into the network.

Combining our results highlighted a need for (i) unraveling experimentally verified transcriptional regulatory networks for more species of the *Corynebacterium* genus and (ii) integrating other regulatory layers, such as post-transcriptional and translational, into these networks through multi-omics approaches. A comprehensive view of GRNs is necessary for reconstructing more consistent and accurate networks that could provide even more valuable guidance for future experimental assays. More accurate, experimentally verified networks for more species could significantly improve the quantity and quality of GRNs for the *Corynebacterium* genus in future genus-scale *in silico* GRN reconstruction. This information would aid in biotechnological, veterinary and medical research by providing more trustworthy predicted regulatory interactions.

VI.2 Outlook

As perspectives to future work, we intend:

- To develop CoryneRegNet version 8 with transcriptional and post-transcriptional GRNs for all complete genomes of the *Corynebacterium* genus available on the NCBI database;
- To integrate protein-protein interaction (PPI) networks as a new regulatory layer in CoryneRegNet;

- To assess the consistency of *E. coli* GRN after integrating sRNA regulation and PPI networks into the GRN;
- To study conserved regulatory interactions considering the life-style of the members of a genus, thus identifying life-style conserved regulatory interactions.

VII. BIBLIOGRAPHY

ABREU, V. A. C. *et al.* CMRegNet-An interspecies reference database for corynebacterial and mycobacterial regulatory networks. *BMC genomics*, v. 16, p. 452, 11 jun. 2015.

ALKEMA, W. B. L.; LENHARD, B.; WASSERMAN, W. W. Regulog analysis: detection of conserved regulatory networks across bacteria: application to Staphylococcus aureus. *Genome research*, v. 14, n. 7, p. 1362–1373, jul. 2004.

ALTSCHUL, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, v. 25, n. 17, p. 3389–3402, 1 set. 1997.

BABU, M. M.; LANG, B.; ARAVIND, L. Methods to reconstruct and compare transcriptional regulatory networks. *Methods in molecular biology* , v. 541, p. 163–180, 2009.

BAILEY, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, v. 37, n. Web Server issue, p. W202–8, jul. 2009.

BALLEZA, E. *et al.* Regulation by transcription factors in bacteria: beyond description. *FEMS microbiology reviews*, v. 33, n. 1, p. 133–151, jan. 2009.

BARRETT, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids research*, v. 41, n. Database issue, p. D991–5, jan. 2013.

BARSKI, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell*, v. 129, n. 4, p. 823–837, 18 maio 2007.

BAUMBACH, J.; BRINKROLF, K.; WITTKOP, T.; *et al.* CoryneRegNet 2: An Integrative Bioinformatics Approach for Reconstruction and Comparison of Transcriptional Regulatory Networks in Prokaryotes. *Journal of integrative bioinformatics*, v. 3, n. 2, p. 1–13, 1 dez. 2006.

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*, v. 129, n. 2, p. 279–289, 30 abr. 2007.

BAUMBACH, J. CoryneRegNet 4.0--A reference database for corynebacterial gene regulatory networks. *BMC bioinformatics*, v. 8, n. 1, p. 429, 2007.

BAUMBACH, J.; BRINKROLF, K.; CZAJA, L. F.; *et al.* CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. *BMC genomics*, v. 7, p. 24, 14 fev. 2006.

BAUMBACH, J. On the power and limits of evolutionary conservation—unraveling bacterial gene regulatory networks. *Nucleic acids research*, v. 38, n. 22, p. 7877–7884, 1 dez. 2010. Acesso em: 27 set. 2019.

BAUMBACH, J.; APELTSIN, L. Linking Cytoscape and the corynebacterial reference database CoryneRegNet. *BMC genomics*, v. 9, p. 184, 21 abr. 2008.

BAUMBACH, J.; RAHMANN, S.; TAUCH, A. Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms. *BMC systems biology*, v. 3, p. 8, 15 jan. 2009.

BECKSTETTE, M. *et al.* Fast index based algorithms and software for matching position specific scoring matrices. *BMC bioinformatics*, v. 7, p. 389, 24 ago. 2006.

BERVOETS, I.; CHARLIER, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. *FEMS microbiology reviews*, v. 43, n. 3, p. 304–339, 1 maio 2019.

BILLINGTON, S. J. *et al.* Identification and role in virulence of putative iron acquisition genes from Corynebacterium pseudotuberculosis. *FEMS microbiology letters*, v. 208, n. 1, p. 41–45, 19 fev. 2002.

BONHAM, K. S.; WOLFE, B. E.; DUTTON, R. J. Extensive horizontal gene transfer in cheese-associated bacteria. *eLife*, v. 6, 23 jun. 2017. Disponível em: <http://dx.doi.org/10.7554/eLife.22144>.

BRINKROLF, K. *et al.* The transcriptional regulatory repertoire of Corynebacterium glutamicum: reconstruction of the network controlling pathways involved in lysine and glutamate production. *Journal of biotechnology*, v. 149, n. 3, p. 173–182, 1 set. 2010.

BRINKROLF, K.; BRUNE, I.; TAUCH, A. The transcriptional regulatory network of the amino acid producer Corynebacterium glutamicum. *Journal of biotechnology*, v. 129, n. 2, p. 191–211, 30 abr. 2007.

BROWNING, D. F.; BUSBY, S. J. The regulation of bacterial transcription initiation. *Nature reviews. Microbiology*, v. 2, n. 1, p. 57–65, jan. 2004.

BROWNING, D. F.; BUSBY, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature reviews. Microbiology*, v. 14, n. 10, p. 638–650, out. 2016.

BRUM, A. A. *et al.* Recombinant esterase from Corynebacterium pseudotuberculosis in DNA and subunit recombinant vaccines partially protects mice against challenge. *Journal of medical microbiology*, v. 66, n. 5, p. 635–642, maio 2017.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods*, v. 12, n. 1, p. 59–60, jan. 2015.

COSTA, M. P. *et al.* Molecular characterization of the Corynebacterium pseudotuberculosis hsp60-hsp10 operon, and evaluation of the immune response and protective efficacy induced by hsp60 DNA vaccination in mice. *BMC research notes*, v. 4, p. 243, 20 jul. 2011.

DAVIS, M. C.; KESTHELY, C. A.; FRANKLIN, E. A. The essential activities of the bacterial sigma factor. *Canadian Journal of*, 2017. Disponível em: <https://cdnsciencepub.com/doi/abs/10.1139/cjm-2016-0576>.

DEHAL, P. S. *et al.* MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic acids research*, v. 38, n. Database issue, p. D396–400, jan. 2010.

DELORENZI, M.; SPEED, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* , v. 18, n. 4, p. 617–625, abr. 2002.

DORELLA, F. A. *et al.* Corynebacterium pseudotuberculosis: microbiology, biochemical properties, pathogenesis and molecular studies of virulence. *Veterinary research*, v. 37, n. 2, p. 201–218, mar. 2006.

DROPPA-ALMEIDA, D.; FRANCESCHI, E.; PADILHA, F. F. Immune-Informatic Analysis and Design of Peptide Vaccine From Multi-epitopes Against. *Bioinformatics and biology insights*, v. 12, p. 1177932218755337, 14 maio 2018.

DSMZ. *DSMZ*. Disponível em: <https://lpsn.dsmz.de/genus/corynebacterium>. Acesso em: 16 mar. 2021.

ECKWEILER, D. *et al.* PRODORIC2: the bacterial gene regulation database in 2018. *Nucleic acids research*, v. 46, n. D1, p. D320–D326, 4 jan. 2018.

EDDY, S. R. Profile hidden Markov models. *Bioinformatics* , v. 14, n. 9, p. 755–763, 1998.

EMMERT-STREIB, F.; DEHMER, M.; HAIBE-KAINS, B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in cell and developmental biology*, v. 2, p. 38, 19 ago. 2014.

EMMS, D. M.; KELLY, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, v. 20, n. 1, p. 238, 14 nov. 2019.

ESCORCIA-RODRÍGUEZ, J. M.; TAUCH, A.; FREYRE-GONZÁLEZ, J. A. Abasy Atlas v2.2: The most comprehensive and up-to-date inventory of meta-curated, historical, bacterial regulatory networks, their completeness and system-level characterization. *Computational and structural biotechnology journal*, v. 18, p. 1228–1237, 1 jan. 2020.

ESPINOSA, V. *et al.* Comparative studies of transcriptional regulation mechanisms in a group of eight gamma-proteobacterial genomes. *Journal of molecular biology*, v. 354, n. 1, p. 184–199, 18 nov. 2005.

FREYRE-GONZÁLEZ, J. A. *et al.* Lessons from the modular organization of the transcriptional regulatory network of Bacillus subtilis. *BMC systems biology*, v. 7, p. 127, 16 nov. 2013.

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. *Journal of biotechnology*, v. 257, p. 199–210, 10 set. 2017.

GARNER, M. M.; REVZIN, A. The use of gel electrophoresis to detect and study nucleic acid—protein interactions. *Trends in biochemical sciences*, v. 11, n. 10, p. 395–396, 1986.

GEORGI, T.; RITTMANN, D.; WENDISCH, V. F. Lysine and glutamate production by Corynebacterium glutamicum on glucose, fructose and sucrose: roles of malic enzyme and fructose-1,6-bisphosphatase. *Metabolic engineering*, v. 7, n. 4, p. 291–301, jul. 2005.

GOMIDE, A. C. P.; DE SÁ, P. G.; *et al.* Heat shock stress: Profile of differential expression in Corynebacterium pseudotuberculosis biovar Equi. *Gene*, v. 645, p. 124–130, 1 mar. 2018.

GOMIDE, A. C. P.; IBRAIM, I. C.; *et al.* Transcriptome analysis of Corynebacterium pseudotuberculosis biovar Equi in two conditions of the environmental stress. *Gene*, v. 677, p. 349–360, 30 nov. 2018.

GOODFELLOW, M.; JONES, A. L. Corynebacteriales ord. nov. *Bergey's Manual of Systematics of*, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118960608.obm00009>.

GRAINGER, D. C.; LEE, D. J.; BUSBY, S. J. W. Direct methods for studying transcription regulatory proteins and RNA polymerase in bacteria. *Current opinion in microbiology*, v. 12, n. 5, p. 531–535, out. 2009.

GRIBSKOV, M.; MCLACHLAN, A. D.; EISENBERG, D. Profile analysis: detection of distantly related proteins. *Proceedings of the National Academy of Sciences of the United States of America*, v. 84, n. 13, p. 4355–4358, jul. 1987.

GUTIÉRREZ-RÍOS, R. M. *et al.* Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles. *Genome research*, v. 13, n. 11, p. 2435–2443, nov. 2003.

GUZIOLOWSKI, C. *et al.* BioQuali Cytoscape plugin: analysing the global consistency of regulatory networks. *BMC genomics*, v. 10, p. 244, 26 maio 2009.

HECKER, M. *et al.* Gene regulatory network inference: data integration in dynamic models-a review. *Bio Systems*, v. 96, n. 1, p. 86–103, abr. 2009.

HELLMAN, L. M.; FRIED, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols*, v. 2, n. 8, p. 1849–1861, 2007.

HORAK, C. E.; SNYDER, M. ChIP-chip: A genomic approach for identifying transcription factor binding sites. In: GUTHRIE, C.; FINK, G. R. (Org.). . *Methods in Enzymology*. [S.l.]: Academic Press, 2002. v. 350. p. 469–483.

HÖR, J.; GORSKI, S. A.; VOGEL, J. Bacterial RNA Biology on a Genome Scale. *Molecular cell*, v. 70, n. 5, p. 785–799, 7 jun. 2018.

IBARRA-ARELLANO, M. A. *et al.* Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria. *Database: the journal of biological databases and curation*, v. 2016, 30 maio 2016. Disponível em: <http://dx.doi.org/10.1093/database/baw089>.

IBRAIM, I. C. *et al.* Transcriptome profile of Corynebacterium pseudotuberculosis in response to iron limitation. *BMC genomics*, v. 20, n. 1, p. 663, 20 ago. 2019.

IVANOV, K. *et al.* Biotechnology in the Production of Pharmaceutical Industry Ingredients: Amino Acids. *Biotechnology, biotechnological equipment*, v. 27, n. 2, p. 3620–3626, 1 jan. 2013.

KEPPEL, M. *et al.* HrrSA orchestrates a systemic response to heme and determines prioritization of terminal cytochrome oxidase expression. *Nucleic acids research*, v. 48, n. 12, p. 6547–6562, 9 jul. 2020.

KILIÇ, S.; ERILL, I. Assessment of transfer methods for comparative genomics of regulatory networks in bacteria. *BMC bioinformatics*, v. 17 Suppl 8, p. 277, 31 ago. 2016.

KITTAS, A. *et al.* CytoASP: a Cytoscape app for qualitative consistency reasoning, prediction and repair in biological networks. *BMC systems biology*, v. 9, p. 34, 11 jul. 2015.

KREBS, J. E. *et al. Lewin's GENES XI*. [S.l.]: Jones & Bartlett Publishers, 2013.

KROGH, A. *et al.* Hidden Markov models in computational biology. Applications to protein modeling. *Journal of molecular biology*, v. 235, n. 5, p. 1501–1531, 4 fev. 1994.

LARSEN, S. J. *et al.* E. coli gene regulatory networks are inconsistent with gene expression data. *Nucleic acids research*, v. 47, n. 1, p. 85–92, 10 jan. 2019.

LEYN, S. A. *et al.* Comparative genomics and evolution of transcriptional regulons in Proteobacteria. *Microbial genomics*, v. 2, n. 7, p. e000061, jul. 2016.

LI, L.; STOECKERT, C. J., Jr; ROOS, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research*, v. 13, n. 9, p. 2178–2189, set. 2003.

LIU, J. *et al.* TetR Family Regulator brpT Modulates Biofilm Formation in Streptococcus sanguinis. *PloS one*, v. 12, n. 1, p. e0169301, 3 jan. 2017.

LUCCHINI, S.; THOMPSON, A.; HINTON, J. C. D. Microarrays for microbiologists. *Microbiology*, v.

147, n. Pt 6, p. 1403–1414, jun. 2001.

MADAN BABU, M.; TEICHMANN, S. A.; ARAVIND, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *Journal of molecular biology*, v. 358, n. 2, p. 614–633, 28 abr. 2006.

MADHAMSHETTIWAR, P. B. *et al.* Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome medicine*, v. 4, n. 5, p. 41, 1 maio 2012.

MADIGAN, M. T. *et al. Brock Biology of Microorganisms – 14th edition*. [S.l.]: Pearson, 2014. v. 14th edition.

MCCARTY, N. S.; LEDESMA-AMARO, R. Synthetic Biology Tools to Engineer Microbial Communities for Biotechnology. *Trends in biotechnology*, v. 37, n. 2, p. 181–197, fev. 2019.

MCGARY, K.; NUDLER, E. RNA polymerase and the ribosome: the close relationship. *Current opinion in microbiology*, v. 16, n. 2, p. 112–117, abr. 2013.

MELAS, I. N. *et al.* Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs. *PLoS computational biology*, v. 9, n. 9, p. e1003204, 5 set. 2013.

MERCATELLI, D. *et al.* Gene regulatory network inference resources: A practical overview. *Biochimica et Biophysica Acta, Gene Regulatory Mechanisms*, v. 1863, n. 6, p. 194430, jun. 2020.

MÉTRIS, A. *et al.* SalmoNet, an integrated network of ten Salmonella enterica strains reveals common and distinct pathways to host adaptation. *NPJ systems biology and applications*, v. 3, p. 31, 18 out. 2017.

MUÑOZ-BUCIO, A. V. *et al.* Identification of Corynebacterium pseudotuberculosis isolated from muscular abscesses in two horses: First report in Mexico. *Equine veterinary education*, v. 29, n. 8, p. 431–435, 6 ago. 2017.

NAGALAKSHMI, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, v. 320, n. 5881, p. 1344–1349, 6 jun. 2008.

NOVICHKOV, P. S. *et al.* RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic acids research*, v. 38, n. Web Server issue, p. W299–307, jul. 2010.

OLIVEIRA, A. *et al.* Insight of Genus Corynebacterium: Ascertaining the Role of Pathogenic and Non-pathogenic Species. *Frontiers in microbiology*, v. 8, p. 1937, 12 out. 2017.

OSTERGAARD, S. *et al.* Increasing galactose consumption by Saccharomyces cerevisiae through metabolic engineering of the GAL gene regulatory network. *Nature biotechnology*, v. 18, n. 12, p. 1283–1286, dez. 2000.

PABO, C. O.; SAUER, R. T. Transcription factors: structural families and principles of DNA recognition. *Annual review of biochemistry*, v. 61, p. 1053–1095, 1992.

PAGET, M. S. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*, v. 5, n. 3, p. 1245–1265, 26 jun. 2015.

PARISE, D. *et al.* First genome sequencing and comparative analyses of strains from Mexico. *Standards in genomic sciences*, v. 13, p. 21, 10 out. 2018.

PARISE, M. T. D. *et al.* CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. *Scientific data*, v. 7, n. 1, p. 142, 11 maio 2020.

PAULING, J.; RÖTTGER, R.; TAUCH, A.; *et al.* CoryneRegNet 6.0--Updated database content, new analysis methods and novel features focusing on community demands. *Nucleic acids research*, v. 40, n. Database issue, p. D610–4, jan. 2012.

PAULING, J.; RÖTTGER, R.; NEUNER, A.; *et al.* On the trail of EHEC/EAEC—unraveling the gene regulatory networks of human pathogenic Escherichia coli bacteria. *Integrative Biology*, v. 4, n. 7, p. 728–733, 25 jun. 2012. Acesso em: 14 dez. 2019.

PIERCE, B. A. *Genetics: A Conceptual Approach*. [S.l.]: W. H. Freeman, 2013.

PINTO, A. C. *et al.* Differential transcriptional profile of Corynebacterium pseudotuberculosis in response to abiotic stresses. *BMC genomics*, v. 15, p. 14, 9 jan. 2014.

PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli. *Genome biology*, v. 9, n. 1, p. R4, 7 jan. 2008.

PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. Orthologous transcription factors in bacteria have different functions and regulate different genes. *PLoS computational biology*, v. 3, n. 9, p. 1739–1750, set. 2007.

RAMOS, J. L. *et al.* The TetR family of transcriptional repressors. *Microbiology and molecular biology reviews: MMBR*, v. 69, n. 2, p. 326–356, jun. 2005.

RIBEIRO, D. *et al.* An iron-acquisition-deficient mutant of Corynebacterium pseudotuberculosis efficiently protects mice against challenge. *Veterinary research*, v. 45, p. 28, 6 mar. 2014.

RIVA, A. The MAPPER2 Database: a multi-genome catalog of putative transcription factor binding sites. *Nucleic acids research*, v. 40, n. Database issue, p. D155–61, jan. 2012.

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 research*, v. 47, n. D1, p. D212–D220, 8 jan. 2019.

SCHENA, M. *et al.* Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, v. 270, n. 5235, p. 467–470, 20 out. 1995.

SCHRÖDER, J.; TAUCH, A. Transcriptional regulation of gene expression in Corynebacterium glutamicum: the role of global, master and local regulators in the modular and hierarchical gene regulatory network. *FEMS microbiology reviews*, v. 34, n. 5, p. 685–737, set. 2010.

SELIM, S. A. Oedematous skin disease of buffalo in Egypt. *Journal of veterinary medicine. B, Infectious diseases and veterinary public health*, v. 48, n. 4, p. 241–258, maio 2001.

SIEGEL, A. *et al.* Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks. *Bio Systems*, v. 84, n. 2, p. 153–174, maio 2006.

SIERRO, N. *et al.* DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic acids research*, v. 36, n. Database issue, p. D93–6, jan. 2008.

SOARES, S. C. *et al.* PIPS: pathogenicity island prediction software. *PloS one*, v. 7, n. 2, p. e30848, 15

fev. 2012.

SPIER, S. J.; AZEVEDO, V. Corynebacterium pseudotuberculosis infection in horses: Increasing frequency and spread to new regions of North America. *Equine veterinary education*, v. 29, n. 8, p. 436–439, 19 ago. 2017.

SUN, D. *et al.* Metabolic engineering of Corynebacterium glutamicum by synthetic small regulatory RNAs. *Journal of industrial microbiology & biotechnology*, v. 46, n. 2, p. 203–208, fev. 2019.

THOMPSON, D.; REGEV, A.; ROY, S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annual review of cell and developmental biology*, v. 31, p. 399–428, 3 set. 2015.

TIWARI, S. *et al.* C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds. *Integrative biology: quantitative biosciences from nano to macro*, v. 6, n. 11, p. 1088–1099, nov. 2014.

VENANCIO, T. M.; ARAVIND, L. Reconstructing prokaryotic transcriptional regulatory networks: lessons from actinobacteria. *Journal of biology*, v. 8, n. 3, p. 29, 15 abr. 2009.

WHEELER, T. J.; EDDY, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* , v. 29, n. 19, p. 2487–2489, 1 out. 2013.

XIONG, J. *et al.* Genome of the facultative scuticociliatosis pathogen Pseudocohnilembus persalinus provides insight into its virulence through horizontal gene transfer. *Scientific reports*, v. 5, p. 15470, 21 out. 2015.

YERUHAM, I. *et al.* A herd level analysis of a Corynebacterium pseudotuberculosis outbreak in a dairy cattle herd. *Veterinary dermatology*, v. 15, n. 5, p. 315–320, out. 2004.

VIII. APPENDIX

A. Published, submitted and forthcoming research articles

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

Contribution: This paper is the result of my Master's project. In this study, I worked through all steps of the process: assembly, annotation and genome analysis. I also wrote the original manuscript and worked on the reviewer's requests.

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 Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

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: In this collaboration article, I produced the assembly, mapping of the contigs in the images generated by the optical map and annotation steps. Additionally, I contributed to 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

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: In this collaboration article, I contributed to the prediction and analysis of the regulatory interactions among the genes of interest. To predict the regulatory interactions, we checked for the conservation of the TF, the TG and the TFBS between the model organisms (*C. glutamicum*, *M. tuberculosis* or *C. diphtheriae*) and the target organisms (*C. pseudotuberculosis* T1, Cp13 or 1002B). Furthermore, I contributed to the revision of the manuscript.

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

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

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: In this collaboration article, I contributed to the prediction of TFBSs for cAMP receptor protein (CRP_{VH2}). I also aided in writing and revising the prediction of TFBSs methodology.

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

Copyright © 2020 American Society for Microbiology. All Rights Reserved.

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

Downloaded from http://aem.asm.org/ on March 15, 2021 by guest

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

Francielly Morais-Rodrigues, Rita Silv́erio-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: In this collaboration article, I contributed to the revision of the manuscript.

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/ © 2019 Elsevier B.V. All rights reserved.

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

Mariana Teixeira Dornelles Parise, **Doglas Parise**, Flavia F. Aburjaile, Anne C. Pinto Gomide, Rodrigo B. Kato, Martin Raden, Rolf Backofen, Vasco A. Azevedo and Jan Baumbach

Contribution: In this collaboration article, I contributed to the update of CoryneRegNet's front-end, writing the original manuscript, generating the figures and answering the reviewers' requests.

Status: Under review

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:
21 Jan 2021

Journal website link:
www.frontiersin.org

In review

Comparative genomics and in silico gene evaluation involved in the probiotic potential of *Bifidobacterium longum* 51A

Jéssica Gabrielle Vidal da Silva, Angélica Thomaz Vieira, Thiago J Sousa, Marcus Vinicius Canário Viana, **Doglas Parise**, Bruna Sampaio, Alessandra Lima da Silva, Luís Cláudio Lima de Jesus, Pedro Kássio Ribeiro Matos Loureiro de Carvalho, Letícia de Castro Oliveira, Flavia Figueira Aburjaile, Flaviano S Martins, Jacques Robert Nicoli, Preetam Ghosh, Bertram Brenig, Vasco A Azevedo and Anne Cybelle Pinto Gomide.

Contribution: In this collaboration article, I ran the software BPGA for pangenomic analysis and contributed to the manuscript's revision and reviewers' requests.

Status: Under review

Comparative genomics and in silico gene evaluation involved in the probiotic potential of *Bifidobacterium longum* 51A

Jéssica Gabrielle Vidal da Silva, Angélica Thomaz Vieira, Thiago J Sousa, Marcus Vinicius Canário Viana, Doglas Parise, Bruna Sampaio, Alessandra Lima da Silva, Luís Cláudio Lima de Jesus, Pedro Kássio Ribeiro Matos Loureiro de Carvalho, Letícia de Castro Oliveira, Flavia Figueira Aburjaile, Flaviano S Martins, Jacques Robert Nicoli, Preetam Ghosh, Bertram Brenig, Vasco A Azevedo and Anne Cybelle Pinto Gomide.

Abstract

The 51A strain of Bifidobacterium longum isolated from feces of a healthy child, has demonstrated probiotic properties by in vivo and in vitro studies, which may be assigned to its production of metabolites such as acetate. Thus, through the study of comparative genomics, the present work sought to identify unique genes that might be related to the production of acetate. To perform the study, the DNA strain was sequenced using Illumina HiSeq technology, followed by assembly and manual curation of coding sequences. Comparative analysis was performed including 19 complete B. longum genomes, available in Genbank. In the phylogenetic analysis, the CECT 7210 and 157F strains of B. longum subsp. infantis aggregated within the subsp. longum cluster, suggesting that their taxonomic classification should be reviewed. The strain 51A of B. longum has 26 unique genes, six of which are possibly related to carbohydrate metabolism and acetate production. The phosphoketolase pathway from B. longum 51A showed a difference in acetyl-phosphate production. This result seems to corroborate the analysis of their unique genes, whose presence suggests the strain may use different sources of carbohydrates that allow a greater production of acetate and consequently offer benefits to the host health.

1. Introduction

Firmicutes, Bacteroidetes and Actinobacteria constitute the most abundant phyla in the microbiota of the human gastrointestinal tract (GIT). Firmicutes and Bacteroidetes are predominant in adults, and Actinobacteria in breast-fed infants, where the genus Bifidobacterium can exceed 90% of the total bacterial population, contributing to the health and well-being of the newborn (Turroni et al., 2009b) Bifidobacterium species are comprised of Gram-positive, obligate anaerobic, immobile and generally non-pathogenic bacteria (Hidalgo-Cantabrana et al., 2017).

Well controlled clinical trials showed that, when administered as probiotics, bifidobacteria produce several beneficial effects, such as prevention of antibiotic associated diarrhea (Corrêa et al., 2005) and improvement of symptoms in lactose-intolerant patients (He et al., 2008), women with irritable bowel syndrome (Whorwell et al., 2006), children and adolescents with constipation (Guerra et al., 2011), patients with ulcerative colitis (Ishikawa et al., 2011) and child

***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

*First authorship shared

Contribution: In this article, I contributed to normalizing the RNA-seq data and writing 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

Exploring sRNAs in the *Corynebacterium* genus

Mariana T. D. Parise*, **Doglas Parise***, Flavia Alburjaile, Rodrigo Kato, Vasco Azevedo and Jan Baumbach

* This authors contributed equally

Contribution: In this review article, I am contributing to searching for bacterial sRNA databases and writing the manuscript.

Status: Forthcoming

Title: Exploring sRNAs in the *Corynebacterium* genus

Authors: Mariana T. D. Parise [1,2*], Doglas Parise [1,2*], Flavia Alburjaile [3], Olga Zolotareva [1], 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.

* This authors contributed equally

Abstract:

Gene expression regulation allows bacteria to quickly adapt to changes in the environment. Bacterial small regulatory RNAs (sRNAs) are one of the key players in post-transcriptional control and greatly influence several processes such as response to stress conditions, virulence and iron homeostasis. Despite the medical, veterinary and economic importance of the Corynebacterium genus, a few studies have been conducted concerning its sRNAs and their mechanisms of action. Here, we present the first review article regarding sRNAs in the Corynebacterium genus, covering both experimentally verified and predicted sRNAs. sRNAs involved in cell division control, plasmid copy-number control, glucose uptake and glutamate production as well as sRNA families conserved across several corynebacterial species are the focal point of this review. We conclude that sRNA knowledge regarding the Corynebacterium genus is still hindered by the lack of experimental and predicted investigations; therfronefore, 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, which is the model organism of this genus, has been extensively used in L-glutamate and L-Lysine production (Georgi, Rittmann and Wendisch, 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).

Unravelling *Anoxybacillus* diversity, phylogeny and function

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

Contribution: In this collaboration article, I am performing phylogenetic and phylogenomic analysis, and writing the phylogenetic and phylogenomic methodology.

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.5; 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).