Peer-reviewed version available at *Genes* **2020**, *11*, 794; doi:10.3390/genes11070794

Article

Co-expression networks for causal gene identification based on RNA-seq data of *Corynebacterium pseudotuberculosis*

Edian F. Franco [1,4], Pratip Rana [2], Ana Lidia Queiroz Cavalcante [1], Artur Silva [1], Adriana R. Carneiro Folador[1], Vasco Azevedo [3], Preetam Ghosh [2‡] and Rommel T. J. Ramos [1,3‡]*

1 Institute of Biological Sciences, Federal University of Para, Belem-PA 66075-110, Brazil
2 Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA
3 Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte-MG 31270-901, Brazil
4 Instituto Tecnologico de Santo Domingo (INTEC), Santo Domingo, Dominican Republic
* Correspondence: rommelthiago@gmail.com
‡ These authors contributed equally to this work.

Version June 11, 2020 submitted to Preprints

Abstract: *Corynebacterium pseudotuberculosis* is a Gram-positive bacterium that causes caseous lymphadenitis, a disease that predominantly affects sheep, goat, cattle, buffalo, and horses, but has also been recognized in other animals. This bacterium generates a severe economic impact on countries producing meat. Gene expression studies using RNA-seq is one of the most commonly used techniques to perform transcriptional experiments. Computational analysis on such data through reverse-engineering algorithms leads to a better understanding of the genome-wide complexity of gene interactomes, enabling the identification of genes having the most significant functions inferred by the activated stress response pathways. In this study, we identified the influential or causal genes from four RNA-seq data-sets from different stress conditions (high iron, low iron, acid, osmosis, and PH) in *C. pseudotuberculosis*, using a consensus-based network inference algorithm called miRsig and identified the causal genes in the network using the miRinfluence tool, which is based on the influence diffusion model. We found that over 50% of the genes identified as influential have some essential cellular functions in the genomes. In the strains analyzed, most of the causal genes have crucial roles or participate in processes associated with response to extracellular stresses, pathogenicity, membrane components, and essential genes. This research brings new insight into the understanding of virulence and infection by *C. pseudotuberculosis*.

Keywords: *Corynebacterium pseudotuberculosis*; RNA-Seq; co-expression networks; influence genes; stress condition

1. Introduction

In the past few years, several genomic, transcriptomic and proteomic studies were performed to understand the biological basis of *Corynebacterium pseudotuberculosis*; these studies allowed the identification and understanding of the genomic mechanisms that contribute to virulence and infection processes used by the bacteria [1–7]. *C. pseudotuberculosis* is a Gram-positive intracellular bacteria and the etiologic agent of caseous lymphadenitis, a chronic, pyogenic, and contagious disease that affects small ruminants causing considerable economic losses for the farmers and meat industries in many countries [5,8]. The gene expression studies, using RNA-Seq under different conditions, can explain which genes inside the genome are responsible for the bacterial maintenance in precarious and restricted situations in order to prevent bacterial infection and propagation [2,3]. These studies generated a significant amount of data from the bacteria, and such data can help perform new bioinformatics studies to understand the genome complexity and the relationship between the genes using network inference algorithms.

Co-expression network inference is a popular computational domain where several algorithms have been developed for predicting genetic interactions. These models are built using statistical techniques on high-throughput experimental gene expression data (such as RNA-seq and/or microarrays) [9–13]. These models allow the construction of co-expression networks (CEN) to describe the correlation and interactions between the transcriptional genes in the organism. This type of network allows the understanding of the regulatory mechanisms and processes in the biological system [11,13]. The CEN shows the relationship between genes and the regulatory processes by following the central dogma of regulatory control; network analysis on the CEN can identify significant genes possessing stronger influence or causality in the network [11,13,14].

The influential genes are the minimal set of causal genes (seed nodes) in the network which, when perturbed initially, leads to influence diffusion to other nodes and finally impacting the maximal number of genes in the network. This concept is based on the popular influence maximization algorithms from the social network domain. The activated seed genes can spread their influence by probabilistically activating their neighboring genes based on their expression levels and the edge weights in the CEN. Such influential genes could play a crucial role in the regulation of the gene expression process under specific conditions, such as stress [15,16].

In this study, we identified the influential or causal genes using four RNA-seq datasets from *C.pseudotuberculosis*, first by using the miRsig pipeline to obtain the predicted gene regulatory network [17] and next applying the miRinfluence tool to identify the influential and causal genes inside the network[15]. We adapted the methodology in these tools to determine the critical genes which play a causal role in the signaling cascade through influence diffusion and hence may regulate the overall gene expression of the entire network.

2. Methodology

2.1. Bacterial strains and growth conditions

In this study were used four strains of *Corynebacterium pseudotuberculosis*, isolated from different animals. *C. pseudotuberculosis* 1002 (CP-1002), is a wild strain belonging to biovar ovis, isolated from a caprine host in Brazil [18]; the *C. pseudotuberculosis* 258 (CP-258), biovar Equi, was isolated from a horse in Belgium [19]. The variation CP-T1 is pathogenic wild-type belonging to biovar ovis, isolated form a from a caseous granuloma found in CLA-affected goats in Brazil [20]; and the strain CP-13 is an iron-acquisition-deficient mutant, and was generated by [20], employing the *in vivo* insertional mutagenesis of the reporter transposon-based system TnFuZ in the strain T1. The molecular characterization of the Cp13 mutant showed that the insertion disrupted the ciuA gene, which encodes a putative-iron transport binding protein of the ciuABCD operon [21,22].

The strains of CP1002 and CP258 were subjected to different stress levels (acid, osmotic, and heat). The bacteria were grown in Petri dishes containing brain heart infusion (BHI) media. For the acid stress condition, the media was supplemented with hydrochloric acid (which changed the pH to 5); Osmotic stress was achieved with 2 M NaCl; thermal stress was induced by pre-heated BHI medium to 50 °C, and in the control condition, the bacteria were grown in BHI medium at a physiological condition [2,3,19].

The Strains CP-T1 and CP-13 were grown individually either in the presence of the iron chelator $2,2'$-dipyridyl-DIP (Low Iron condition), or without it (High Iron condition). The iron-chelated BHI medium was prepared with 250 µM of ferrous iron chelator $2,2'$ -dipyridyl (Sigma Aldrich), which, due to its low aqueous solubility, was prepared with 40% (v/v) of ethanol (0.5 M $2,2'$-dipyridyl stock solution) [22].

2.2. Expression datasets

The cDNA samples from CP-258 and CP-1002 were used to prepare eight individual single-end libraries that were sequenced using the SOLiD^{TM} 3 Plus system platform, to produce 50-nucleotide

RNA reads [3].The datasets using in this study were obtained from ArrayExpress repository with the accession numbers E-MTAB-2017 and E-MTAB-9217.

From CP-T1 and CP-13, the cDNA samples were used to prepare 14 individual single-end libraries, with three replicates per stress condition, to produce fragments with an average size of 100–200 nucleotides. Ion Proton Platform was used to perform two rounds of sequencing[22]. The sequencing data was obtained from the Gene Expression Omnibus (GEO) repository with the accession number GSE114125.

Raw data quality was examined using the FastQC tool v0.11 [23]. Per-base quality filtering was performed with Trimmomatic v0.39 [24] with a sliding window trimming approach, with the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:14 MINLEN:30 for CP-258 and CP-1002; and LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 for CP-13 and CP-T1. The adapters in CP258 and CP-1002 were removed using the Cutadapt [25] using the SOLID Small RNA Adapter sequences.

The RNA-seq expression data was quantified and aligned using the Salmon [26] tool, where the genomes CP-13 (NZ_CP014998), CP-258 (NC_017945.3),CP-T1 (NZ_CP015100.2) and CP-1002B (NC_017300.2) were used as references. The RNA-Seq expression profile was normalized by Transcripts Per Kilobase Million (TPM) [27]. Table 1 shows a summary of the replicon information from NCBI [28]. Table 2 presents the average nucleotide identity (ANI), the nucleotide-level genomic similarity in the coding regions between the *C. pseudotuberculosis* strains, calculated using Pairwise ANI in [29].

Table 1. Replicon information summary of each *C. pseudotuberculosis* strains.

Strains	Size(Mb)	GC%	Protein	Genes
CP-13	2.34	52.2	2,013	2,135
CP-T1	2.34	52.2	2,008	2,125
CP-258	2.37	52.1	2,038	2,165
CP-1002B	2.34	52.2	2,009	2,124

The whole-genome gene (WEG) expression datasets were selected from all the expressed genes in the differential expression tests. An analysis of differential expression was performed using the GFOLD tool [30] for CP-258 and CP-1002 data, and edgeR [31] for CP-T1 and CP-13 data to generate the differentially expressed genes (DEG) datasets. To select the DEG we used a fold-change of 2 and a p-value < 0.05 [32].

Table 2. Average Nucleotide Identity (ANI) similarity between the *C. pseudotuberculosis* strains.

Strains 1	Strains 2	ANI%
CP-13	**CP-258**	98.9036
CP-13	**CP-1002B**	99.9966
CP-1002B	**CP-258**	98.9051
CP-1002B	**CP-T1**	99.9968
CP-T1	**CP-258**	98.904
CP-T1	**CP-13**	100

2.3. Network Analysis

The predicted gene interaction networks (or equivalently termed as gene coexpression networks in some cases) were built with miRsig [17]. This tool applies seven different network inference algorithms on the gene expression data that individually reverse engineer the interaction scores between the genes. Next, a consensus-based approachx is applied in miRsig to estimate the overall score of every gene-gene interaction using an average ranking based approach to finally infer the gene coexpression network with scores on edges depicting the likelihood of possible genetic interactions between them. This algorithm was adopted in this work to perform the network analyses with the gene expression data.

Peer-reviewed version available at Genes 2020, 11, 794; doi:10.3390/genes11070794

For each strain, we inferred two gene coexpression networks: I) with the whole genome expression and II) only with the differentially expressed genes in all the stressed conditions.

For the consensus ranking scores, we selected 0.85 as the cut-off for the all expressed genes dataset and the differentially expressed genes datasets; these cut-offs ensured high confidence on the edge scores on gene coexpression networks on which the following causal gene identification methods were applied [15].

2.4. Identification of causal genes

To identify the causal genes inside the network, we used the miRfluence algorithm [15]; this algorithm uses network diffusion theory to quantify the influence of individual genes on the signal transaction process in the gene interaction networks across different conditions or stress. The algorithm ranks all the genes in the network according to their influence scores after calculating the optimal coverage (in terms of the number of nodes influenced in the network) with 10,000 Monte Carlo simulation-based random walks.

We selected the top twenty causal genes from the whole genome expression networks while we selected the top ten causal genes for the DEG networks. These tops genes showed a higher influence score in the influence diffusion model of mirInfluence inside the network.

2.5. Network clustering

To analyze the gene interaction network's inherent structure, we performed a cluster analysis using the K-means [33,34] algorithm within ClusterMaker [35] in Cytoscape [36], the distance between the gene were calculated by the Euclidean distance. We used the same tool to discover the optimal number of clusters through the silhouette metric for each network. The K-means was performed to identify the cluster with one or more influential genes present using the betweenness-centrality, degree, and closeness-centrality node attribute as clustering input.

2.6. Sub-Network detection and enrichment analysis

We selected the sub-networks having one or more influential genes present to perform the gene annotation analysis. The enrichment analysis was performed with the GOfeat platform [37]using the gene influence nucleotide sequences and StringApp v.11.0 [38] in Cytoscape-version 3.8.0 [36] using the *Corynebacterium pseudotuberculosis* as reference species, using > 0.70 for the confidence cut-off, and $< 10$for maximum additional interactors score. The enrichment of the clusters' metabolic pathways was performed using the clusterProfiler R package using the enrichKEG with CP-1002 (Kegg ID: cpk) and CP-258 (Kegg ID: coe) as reference organisms [39]. Visual representation was created using ggplot2 v.3.3 [40].0 and ggpubr v.0.3.0 [41].

Results

Read quality assessment was done through FastQC v0.11. Samples CP-1002 and CP-258 presented a Phred score distribution among 16-29 while CP-13 and CP-T1 produced Phred scores between 20-26. Base quality trimming was performed with Trimmomatric; for samples CP-T1 and CP-13 strains, this procedure excluded 1% and 5% of the total reads, respectively; for samples CP-259 and CP-1002 8 to 10% of total reads were excluded from each one. Adapter filtering was performed with CutAdapt in CP-258 and CP-1002. Trimmed datasets produced by this method were mapped to CP-1002B (NZ_CP012837) CP-T1 (NZ_CP015100.2), CP-258 (NC_017945) and CP-13 (NZ_CP014998) reference genomes. For CP-258, the percentage of mapped reads to the reference genome ranged from 57% to 67%; CP-1002 presented mapped reads percentages ranging from 58% to 72.68%; CP-13 strain mapping covered between 63.48% to 90.59% of reads, and CP-T1 mapped between 54.55% and 67.60% of trimmed reads.

Differential expression analyses were performed with edgeR [4] Cp-13 and CP-T1 samples. A total of 93 genes for CP-T1 and 62 genes in CP-13 were found differentially expressed between control

and high iron conditions. We utilized GFold [5] to perform differential analyses in CP-258 and CP-1002; this method yielded 167 differentially expressed genes in CP-1002 and 138 genes in CP-258.

Using the miRsig tool, we built gene coexpression networks with four data sets, two for each *C. pseudotuberculosis* strains; Table 1 shows the sizes of these data-sets. For the whole genes expressed data-sets, we predicted a network with 86,367 gene-gene interactions for Cp-13, 9,376 interactions for CP-258, 6,682 for CP-1002 and 107,202 for CP-T1 (Figure 1). For the differentially expressed networks, we predicted a total of 46 gene-gene interactions for CP-13, 165 interactions for CP-258, 155 for CP-1002, and 98 for CP-T1. The Whole-genome and differentially expressed genes networks are provided in the supplementary material (File 1: datasheet 1 and datasheet 2).

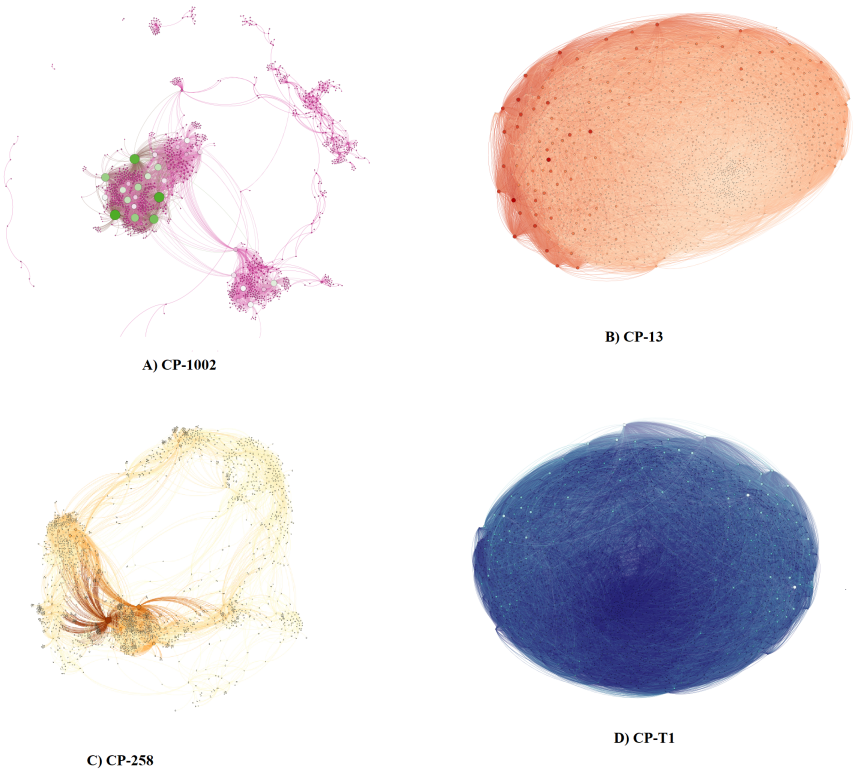

Figure 1. Whole expressed gene data-set networks. **A**- The network generated by the data isolated from *C. pseudotuberculosis 1002*. **B**- The network generated by the data from *C. pseudotuberculosis strain CP-13*. **C**- The network generated by the data from C. pseudotuberculosis 258. **D**- The network generated by the data from *C. pseudotuberculosis* strain T1.

Table 3. Sizes of the gene expression datasets.

Strains	CP-13	CP-258	CP-1002	CP-T1
Whole Genomes	2,113	2,064	2,091	2,093
Differentially Expressed Genes	63	139	168	93

We selected the influential genes using the coexpression network as input to the miRinfluence tool. From the output set of ranked genes based on their respective influence scores, we selected the top 20 for the whole genes expressed network and the top 10 for the DEG network. Tables 4 and 5 show the gene annotation for each network 75 % of these genes showed a high degree distribution compared with the average of the other nodes in the network.

From the selected list of genes in the whole gene expressed datasets, we found that 15% of the genes in Cp-258, 25% in CP-13, 10% in CP-T1 and 25% in Cp-1002 are considered as essential

Table 4. Products generated by the top 20 influential genes identified in the networks built with the whole gene expression data from Cp-13, Cp-258 and Cp-1002.

Genes Cp-13	Product	Genes Cp-258	Product	Genes Cp-1002	Product	Genes Cp-T1	Product
mca	Mycothiol S-conjugate amidase	**CP258_RS02980**	Methylmalonyl-CoA carboxyltransferase 12S subunit	**Cp1002B_RS03590**	Hypothetical protein	***pknL***	Serine/threonine protein kinase
hisN	Histidinol-phosphatase	**CP258_RS03105**	LPxTG domain-containing protein	***mcbR***	TetR family transcriptional regulator	***fhs***	Formate–tetrahydrofolate ligase
rbsK	Ribokinase	**CP258_RS03110**	Aminotransferase	***ureE***	Urease accessory protein UreE	***ogt***	Methylated-DNA–protein-cysteine methyltransferase
pepC2	M18 family aminopeptidase	***tcsS4***	Two component system sensor kinase protein	***uvrD3***	DNA helicase	**Cp13_RS09405**	Oxidoreductase
tcsR3	Two-component system transcriptional regulatory protein	**CP258_RS04420**	Hypothetical protein	**Cp1002B_RS02755**	Abi family protein	***pbpB***	Penicillin binding protein transpeptidase
Cp13_RS10225	Antimicrobial peptide ABC transporter ATPase	***rimJ***	Ribosomal-protein-alanine acetyltransferase	***cmtB***	Trehalose corynomycolyl transferase B	**nodI**	Nod factor export ATP-binding protein I
ald	Alanine dehydrogenase	**CP258_RS03575**	ABC transporter ATP-binding protein	***whiB***	Transcriptional regulator WhiB	**Cp13_RS04005**	Hypothetical protein
mrpD	Na(+)/H(+) antiporter subunit D	***tehA***	C4-dicarboxylate transporter/malic acid transport protein	***rplO***	50S ribosomal protein L15	***pafA2***	Pup–protein ligase
serC	Phosphoserine transaminase	**CP258_RS03580**	Antibiotic biosynthesis monooxygenase	**Cp1002B_RS02920**	LPxTG domain-containing protein	**Cp13_RS01180**	Secreted hydrolase
mgtA	Glycosyl transferase group 1	***echA6***	Enoyl-CoA hydratase echA6	***rluC***	Ribosomal pseudouridine synthase	***trpC***	N-(5′-phosphoribosyl)anthranilate isomerase
Cp13_RS01600	ABC-type metal ion transport system, periplasmic component/surface adhesin	***rmlD***	dTDP-4-dehydrorhamnose reductase	***yhcL***	Cryptic C4-dicarboxylate membrane transporter dcuD	**Cp13_RS08505**	Acetyltransferase
sdaA	L-serine dehydratase	**CP258_RS07545**	Hypothetical protein	**Cp1002B_RS01655**	ABC transporter inner membrane protein	***nagB***	Glucosamine-6-phosphate deaminase
Cp13_RS07170	Glyoxalase/bleomycin resistance protein/dioxygenase	***galU***	UTP–glucose-1-phosphate uridylyltransferase	**Cp1002B_RS10840**	LPxTG domain-containing protein	**Cp13_RS03565**	Hypothetical protein
Cp13_RS01385	Hypothetical protein	***cstA***	Response regulator	**Cp1002B_RS01705**	Hypothetical protein	**Cp13_RS05420**	ABC transporter ATP-binding protein
lpdA	Flavoprotein disulfide reductase	***argS***	Arginine–tRNA ligase	**Cp1002B_RS10170**	ABC transporter	***yvrC***	ABC transporter substrate-binding protein
mraY	Phospho-N-acetylmuramoyl pentapeptide-transferase	***hpf***	Ribosome hibernation promoting factor	***fadF***	Protein fadF	***fagA***	Hypothetical protein
Cp13_RS08655	Hemolysin III-like protein	***deoA***	Thymidine phosphorylase	***udgA***	UDP-glucose 6-dehydrogenase	***mnmA***	tRNA-specific 2-thiouridylase MnmA
copD	Copper resistance D domain-containing protein/Cytochrome c oxidase caa3 assembly factor (Caa3_CtaG)	***mprA_2***	Two component system response regulator	***gltT***	Sodium/glutamate symporte	**Cp13_RS06640**	Hypothetical protein
glmS	Glutamine–fructose-6-phosphate aminotransferase [isomerizing]	***oppC2***	Oligopeptide transport system permease OppC	**Cp1002B_RS01200**	Serine proteases of the peptidase family S9A	**Cp13_RS07860**	Phosphoglycerate dehydrogenase
Cp13_RS10090	NYN domain-containing protein	***scpB***	Segregation and condensation protein B	***uvrA***	UvrABC system protein A	***Cp13_RS03800***	Putative secreted protein

and conditionally essential genes with variable essentiality statuses across datasets according to the OGEE [42] dataset used from *Mycobacterium tuberculosis*, the phylogenetically close organism with *C. pseudotuberculosis* available in the dataset and Rocha et al. reference genes study [43].

In the whole expressed genes list, the genes *galU* and *argS* are categorized as essential and gene rmlD is conditionally essential in strain CP-258. In CP-T1 the genes *pdpB* and *trpC* were classified as essential genes. In strain Cp-1002, the genes *uvrD3, whiB, rplO, udgA* and *uvrA* are conditionally essential genes. *serC, mraY* and *glmS* were listed as essential and the genes *sdaA* and *lpdA* were classified as conditionally essential genes.

For the DEG list, the gene metX in CP-13, dnaK in CP-1002 and lysA2 in CP-T1 were labeled as essential genes. cdd in CP-1002 and cstA in CP-258 were classified as conditionality essential.

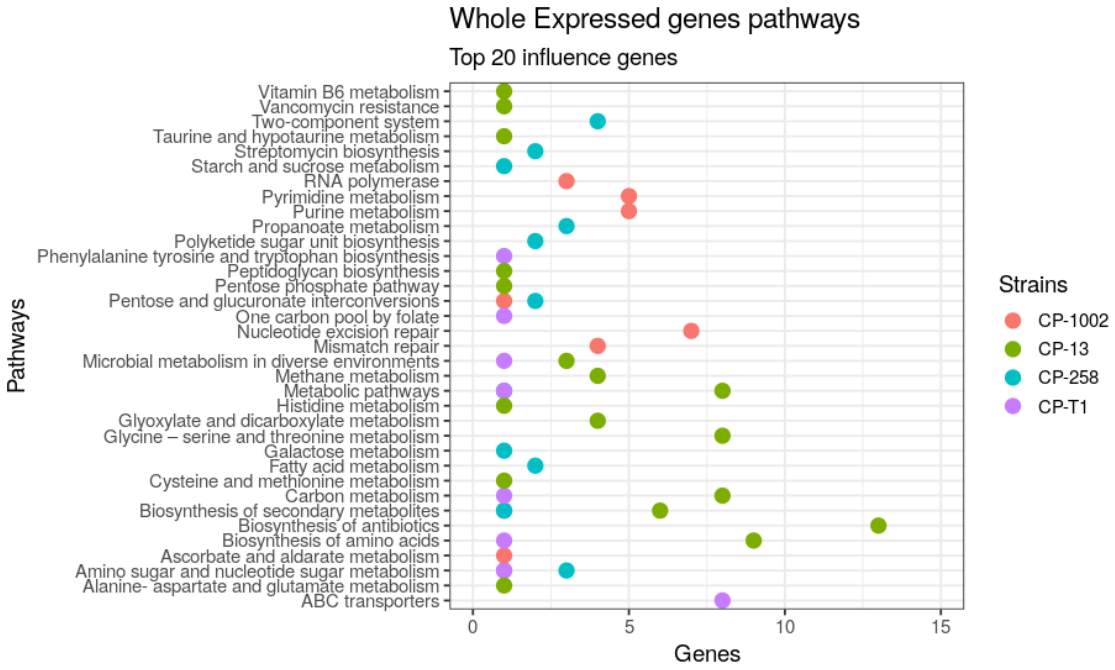

Figure 2. The pathways that involve the top 20 genes identified in the whole expressed genes network in CP-258, CP-13, CP-T1 and CP-1002.

We made the functional annotation using Gofeat [37] and Cytoscape StringApp [38]; these tools allowed the characterization and functional annotation of the causal genes groups present in each of the studied genomes in the whole gene expression and differential gene expression data-sets.

Figure 2 shows the pathways of the genes in the whole expressed genes data-sets. The pathways with more genes in CP-13 were the biosynthesis of antibiotics with 13 genes and Biosynthesis of Amino Acids with nine genes; the Nucleotide excision repair and pyrimidine metabolism were the more active pathways with seven genes and five genes respectively in the CP-1002; for CP-258 two-component system pathway with four genes and propanoate metabolism pathway with three genes were activated. In CP-T1, the ABC transporter was the more activated pathway.

Figure 3 shows the Kegg pathways found in the DEG data-sets. We found 11 pathways in four *C. pseudotuberculosis* strains; the pathway with most genes relate to porphyrin and chlorophyll metabolism, expressed in CP-1002 and CP-T1. Other important pathways were metabolic pathways, fatty acid metabolism, citrate cycle (TCA cycle), biosynthesis of unsaturated fatty acids and biosynthesis of amino acid; these pathways are involved in the cell walls component and protect the bacteria from environmental stress [44].

Supplmetary File 2- Figures S1-S3 show the results of the gene ontology analysis for the top 20 causal genes. The cell adhesion, SOS response, cell division, carbohydrate metabolism process, transmembrane transport, methylation, regulation of transcription DNA-templated, were the biological

Table 5. Proteins produced by the top 10 influential genes identified in the networks built with the differential expressed genes data from Cp-258 and Cp-1002.

Genes Cp-13	Product	Genes Cp-258	Product	Genes Cp-1002	Product	Genes Cp-T1	Product
desA3	Stearoyl-CoA 9-desaturase	**CP258_RS02710**	Putative secreted protein	***cynT***	Carbonic anhydrase	***pyc***	Pyruvate carboxylase
htaA	Cell-surface hemin receptor	**CP258_RS02905**	Hypothetical protein	**Cp1002B_RS03070**	HtaA domain-containing protein	***fecE***	Fe(3+) dicitrate transport ATP-binding protein FecE
Cp13_RS02285	Hypothetical protein	***vapI***	Virulence-associated protein	**Cp1002B_RS02920**	LPxTG domain-containing protein	***hmuT***	Hemin-binding periplasmic protein
lutB	Putative iron-sulfur protein	***tetR3***	TetR family transcriptional regulator	***cdd***	Cytidine deaminase	**Cp13_RS04105**	LUD_dom domain-containing protein
Cp13_RS04105	LUD_dom domain-containing protein	***rshA***	Anti-sigma factor	**Cp1002B_RS03075**	Hypothetical protein	**Cp13_RS02285**	Hypothetical protein
htaB	Cell-surface hemin receptor	***cstA***	Response regulator	**Cp1002B_RS03455**	Oxidoreductase	***lysA2***	Diaminopimelate decarboxylase
Cp13_RS02610	Stearoyl-CoA 9-desaturase electron transfer partner	**gluC**	Glutamate transport system permease protein gluC	**Cp1002B_RS03180**	Hypothetical protein	***icd***	Isocitrate dehydrogenase [NADP]
Cp13_RS09955	Flavin reductase	***odhI***	Oxoglutarate dehydrogenase inhibitor	***ftn***	Ferritin	***sprT***	Trypsin
metX	Homoserine O-acetyltransferase	**yecS**	ABC transporter domain-containing protein	**dnaK**	Chaperone protein DnaK	***gluA***	Glutamine ABC transporter ATP-binding protein
ripA_2	HTH-type transcriptional repressor of iron protein A	***ykoE***	HMP/thiamine permease protein ykoE	**glpQ_1**	Glycerophosphoryl diester phosphodiesterase	***lutB***	Putative iron-sulfur protein

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0144.v1

Peer-reviewed version available at Genes 2020, 11, 794; doi:10.3390/genes11070794

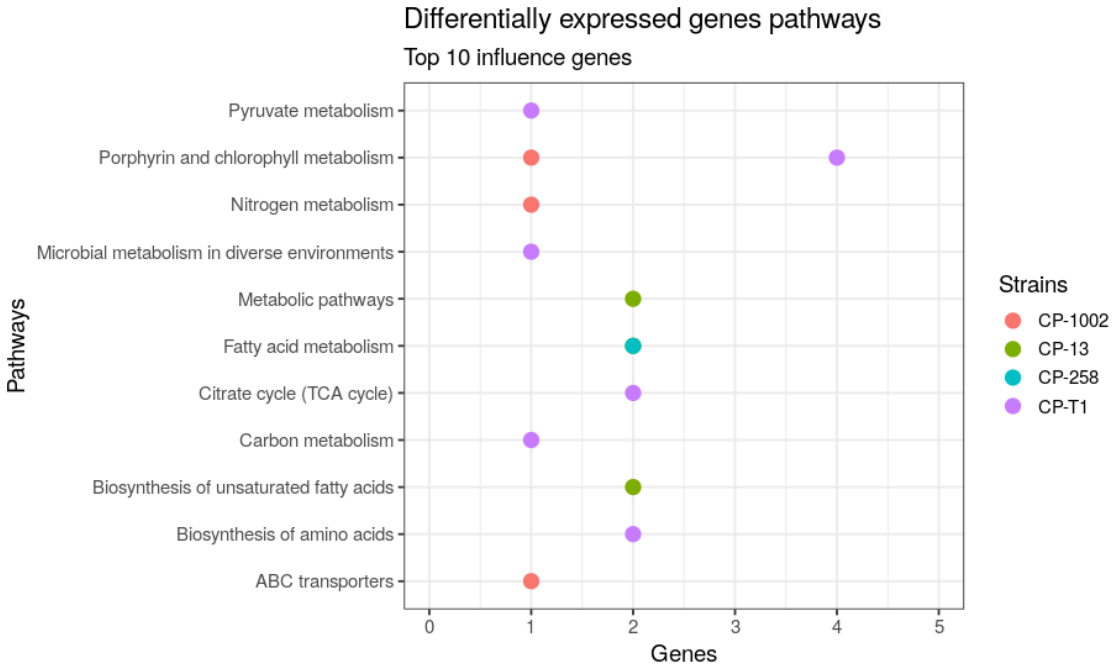

Figure 3. Pathways involved with the top 10 genes identified in the differential expression network in Cp-258 and Cp-1002.

processes in which the most genes participate in the four studied strains. Concerning cellular components, most of the causal genes are part of the cytoplasm, integral component membrane, and plasma membrane in the four strains. The majority of causal genes in these genomes participate in molecular functions such as ATP binding, ATPase activity, metal ion binding, DNA, and transferase activity.

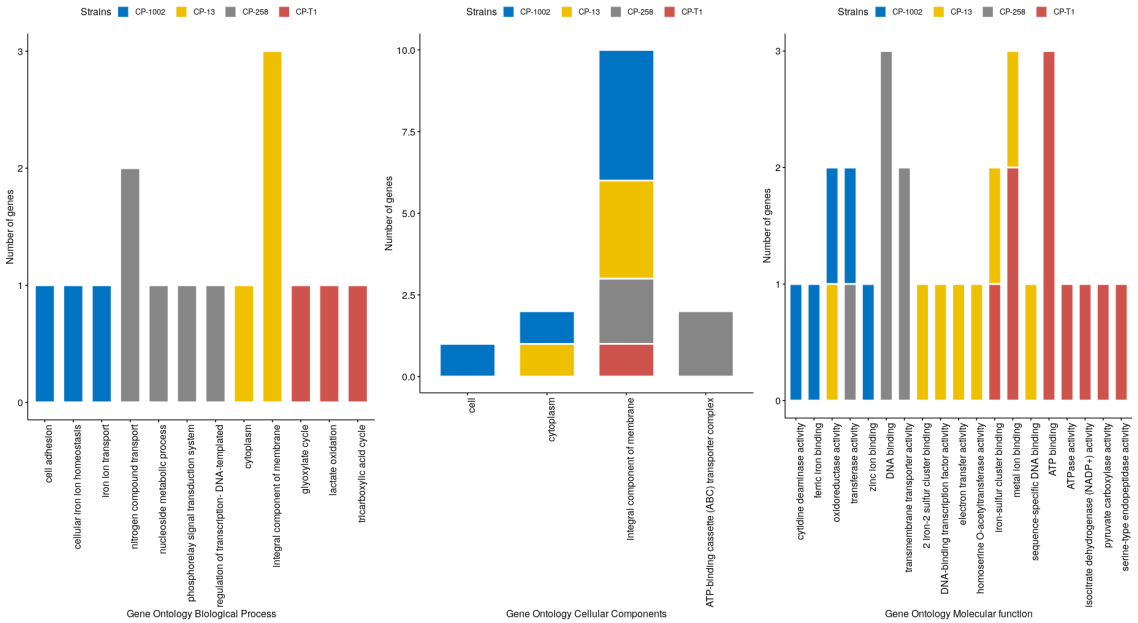

Figure 4. Gene Ontology results of deferentially expressed genes in CP-1002, CP-258, CP-13, CP-T1. Left: Biological process; Center: Cellular Components and Right: Molecular function

Considering the top 10 causal genes identified in the differentially expressed datasets (Figure 4), most of these genes participating in molecular functions such as iron-sulfur clusters binding,

transmembrane transport activity, DNA binding, transferase activity, and oxidoreductase activity. In the cellular components ontology, these genes are part of the integral component of the membrane and the cytoplasm. Considering the biological processes, the top 10 DE genes relate to processes such as cell adhesion, cellular iron ion homeostasis, nitrogen compound transport nucleoside metabolic process, phosphorelay signal transduction system, and other processes.

We performed the K-means algorithm in all the coexpression networks. For whole expressed genes in Cp-13, the silhouette metric predicted 24 clusters, and the causal genes were present in 12 of these clusters. In Cp-258, the metric identified 21 clusters with all the causal genes present in 4 clusters. In the CP-1002 strain, it identified 16 clusters, and the casual genes were distributed in 4 clusters. Finally, in the CP-T1 we identified 30 clusters and causal genes were assigned to 5 clusters (Supplementary Files 3- Datasheet 1).

In the DE genes network for CP-258, the silhouette metric split the network into 6 clusters, and the causal genes were distributed into 2 clusters. For CP-1002, we detected 6 clusters, and all the influential genes were represented in 4 clusters. In the CP-13 strain, the network was divided into 4 clusters, and the causal genes were present in 3 clusters. Finally, the network of CP-T1 was split into 5 clusters, and the causal genes were assigned to 3 clusters (Supplementary Files 3- Datasheet 2.

In the clusters with causal genes, we performed gene enrichment analysis using the clusterProfiler package to identify the Keggs pathways in these clusters. Figure 5 shows the pathways involved in the clusters from the whole expressed genes network in CP-13 and CP-1002, the other clusters figures are in supplementary file 2, Figures S4 (CP-258) and S5 (CP-13). The more representative pathways in the clusters in all the strains were Metabolic pathways and Biosynthesis of secondary metabolites, which are present in almost all the clusters of the studied genomes. Other interesting pathways activated by the causal genes in these clusters are Biosynthesis of amino acids, Microbial metabolism in diverse environments, quorum sensing, ribosome, and carbon metabolism metabolites.

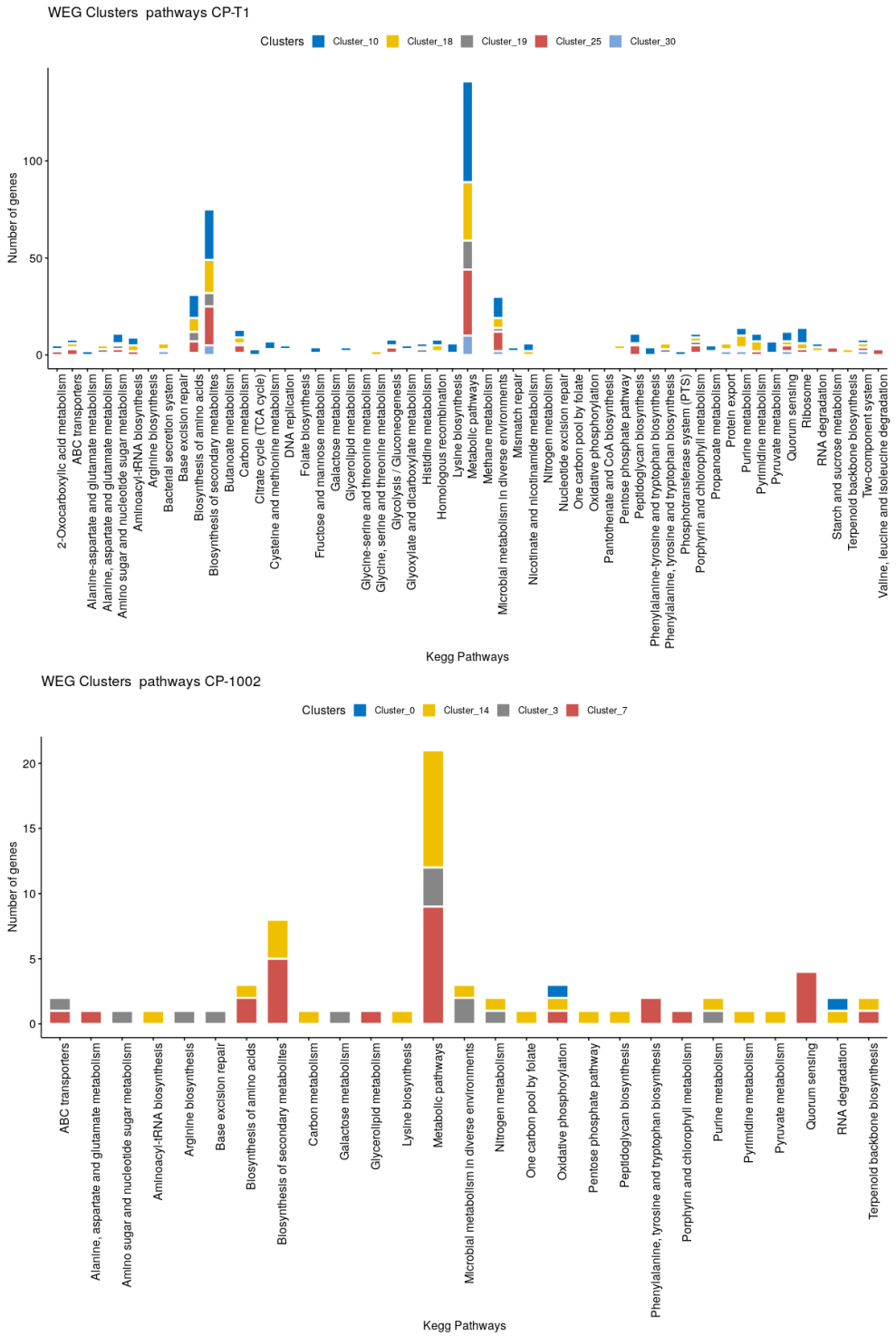

Figure 5. Pathways of the clusters where the influential genes are present in the whole expressed genes network. Up: CP-T1 and Down:C-CP1002.

For the top 10 influential genes from the DE genes networks, we found that the metabolic pathway was the more activated in all strains. Other pathways with more representation in the strains are Biosynthesis of antibiotics, microbial metabolism in diverse environments, Biosynthesis of secondary metabolites, ABC transporters, and Carbon metabolism. The figures are in Supplementary File 2-Figures S6-S9

Peer-reviewed version available at Genes 2020, 11, 794; doi:10.3390/genes11070794

Discussion

Intracellular pathogens have mechanisms of response to conditions of harmful extracellular stresses caused by the host. Among the types of stresses that a bacterium can face, are the drastic change in temperature, pH alteration, sudden changes in osmolarity, presence of Reactive Oxygen Species, among others. In this study, we identified the influential genes referring to the stresses mentioned above from experiments performed by [3] and [22]. In the strains analyzed, the genes that control most interactions within the co-expression network are related to response to extracellular stresses, pathogenicity, membrane components, and essential genes.

The stress conditions faced by *C. pseudotuberculosis* during the infectious process are diverse, from entry into the host, through the lymphatic system, to intracellular replication in macrophages, and the establishment of lesions within the organs [45]. The strains CP-13 (mutant) and CP-T1 (wild) were subjected to restriction of iron, an essential micronutrient for the proliferation of pathogens. In the strains, CP-1002 and CP-258, three types of stresses were applied: acidic, thermal, and osmotic. These conditions simulate the environment found by the bacteria during infection in the host.

Between the top 20 causal genes of the CP-13, a mutant with the disrupted ciuA gene that encodes a putative iron transport binding protein from the ciuABCD operon may be involved in the stress response. These genes are *mca, Hisn, rbsK, pepC2, tcsR3, MRPD, ALD, LPDA, COPD,* and Cp13_RS10225 Cp13_RS07170 (Table 2). The products of these genes showed a relationship with Reactive Oxygen Species (ROS). ROS can condition oxidative stress, which is a condition of imbalance between ROS production and its removal through systems (enzymatic or non-enzymatic) that remove or repair the damage caused by them [46]. Bacteria are regularly exposed to free radicals present in the extracellular environment.

Actinobacteria, the phylum that *C. pseudotuberculosis* is included in, are capable of resisting extracellular ROSs present in the phagocytic environment, produced by macrophages against pathogen invasion [47]. In response to ROSs, neutralization mechanisms, called antioxidants, are used. For example, the mca gene is involved in the mycothiol metabolic process (Figure 4). This compound is believed to act as an antioxidant like glutathione, keeping the intracellular medium free of alkylating agents and other toxins in Gram-positive bacteria [48] . The mycothiol loss in Mycobacteria is associated with slow growth and increased sensitivity to reactive oxygen species and antibiotics [49]. Moreover, the *hisN* gene encodes a protein involved in histidine biosynthesis, which in Actinobacteria is converted into the antioxidant Ergotionein [50,51].

In the differential expression analysis, the genes Cp13_RS02610 and Cp13_RS09955 were repressed in the iron limitation condition (supplementary material). They are involved in redox processes, in which they usually occur during cellular respiration or in oxidative stress conditions producing free radicals [46]. These two genes could be participating in the production of ROS, but they may have been repressed with the expression of antioxidants. Other genes identified in CP-13 encode proteins related to cell adhesion.

The T1 wild-type strain also showed genes that may be involved in the stress response, such as *ogt*, Cp13_RS09405, Cp13_RS07860 genes. The Cp13_RS09405 and Cp13_RS07860 genes are related to redox reactions. The *ogt* gene is involved in the DNA methylation and repair process (Figure 4). DNA repair pathways are essential to maintain the integrity and stability of the genome. Pathogenic bacteria are constantly under external pressure, from the environment, and from interaction with the host, which can cause damage to bacterial DNA [52]. Consequently, the presence of this gene among the influential ones in the co-expression network determines its importance in the face of possible stresses that can destabilize the bacterium's genome, impairing its survival.

In the differential expression analysis of the CP-T1 strain, the Cp13_RS02285 gene (uncharacterized protein) was induced six times. It is an integral membrane protein. Furthermore, the *hmuT* and *sprT* genes were induced three times (supplementary file 4). The *hmuT* gene encodes a periplasmic protein that plays a role in the acquisition and transport of hemin. This gene is identified as an influencer and corroborates the results found by [22], who found these genes to be involved in the transport of hemin.

The strains CP-1002 and CP-258, showed causal genes related to stresses *ureE*, *uvrA* (CP-1002) and *tcsS4*, *mprA_2* (CP-258). The Urease (ure) has a route in which bacteria try to alkalize their environment during acidic stress. The neutralization of acids results from the production of ammonia (NH_3), which combines with a proton from the cytoplasm to produce ammonium (NH_4^+), thus raising the internal pH [53]. The urease enzyme promotes the hydrolysis of urea, which acts as an H^+ ion receptor, generating a neutral pH inside the bacteria, which gives, for example, *H. pylori* resistance to gastric acidity. Most of the urease synthesized by the bacterium is located in its cytoplasm [54]. The *uvrA* gene also has an adaptive response to acidity (Hanna et al., 2001).

The *tcsS4* and *mprA_2* genes encode an osmosensor kinase and a two-component system response standard (Table 2). Two-component systems may be involved in response to osmotic stress. Osmosensors that regulate the expression of genes that encode osmoregulators, constituting two-component systems: the sensor located on the membrane has a histidine kinase domain that in the presence of the stimulus, transmits the information via phosphorylation to the response regulators. In *E. coli*, the EnvZ-OmpR system, which regulates the expression of the *OmpC* and *OmpF* porins, facilitates the diffusion of hydrophilic molecules. In response to an increase in osmotic pressure, the expression of *OmpF* is decreased, and *OmpC* has its expression increased [55].

In addition, coders were identified in relation to cell adhesion: the Cp1002B_RS02920 (CP-1002), Cp1002B_RS10840 and CP258_RS03105 (CP-258) genes. These genes encode adhesive proteins that are important for binding the receptor to host cells during pathogenesis. For example, *E. coli PapG* adhesin from pilus P is necessary for binding to the human renal receptor during the pathogenesis of Pyelonephritis [56].

In the analysis of differential expression, the Cp1002B_RS02920 gene was repressed in osmotic stress (supplementary material). The *vapI* gene (CP-258) encodes a protein associated with virulence (induced in osmotic stress). Furthermore, the *rshA* gene (CP-258) produces an anti-sigma factor and contains a CXXC motif like a thiol-disulfide redox switch [57]. This gene was induced in acid stress (supplementary material), confirming that this condition is essential for the analysis of genes of *C. pseudotuberculosis* that may be involved in the host's infection, mainly related to infection of phagocytic cells.

The biosynthesis of antibiotics pathway, which is regulated by phosphate (figure 2) is activated for the environment's nutritional stress as carbon or nitrogen limitation. The antibiotics produced for this pathway can destroy or inhibit the growth of other bacteria in the microfilms in the nutrients competition; furthermore, some antibiotics may have inter-cellular communication functions in the communities [58,59]. Other critical pathways were microbial metabolism in a diverse environment related to the stress response [44]; Biosynthesis of Amino Acids pathways, is connected with central carbon, nitrogen, and sulfur metabolism [60], and Nucleotide excision repair can help repair DNA that has been damaged by different stresses [61].

Through these pathways identified by the genes clusters in WGE and DEG datasets, we could infer that the causal genes identified in the network allow the bacteria to respond to the stress' stimulus, activating, and co-regulating the expression of their neighbors. These influential genes could send some alert signals to activate the regulatory factors in these pathways to stimulate or inhibit the translation, transcription or the expression of other genes, to generate a physiological and biochemical adaptation in response to the environmental stress [3]. An essential point in the network analysis is the highest degree rates represented for the clusters with more influential genes, which could indicate the regulating role to control their other neighbors whose expressions are correlated in the network [14].

Conclusion

In this work, we have developed a co-expression network analysis to identify the ranked influential and causal genes set using the gene expression data-sets from four strains of *C. pseudotuberculosis* under different stress conditions. The network analyses were performed using computational methods based on the information diffusion concept to identify these genes. For

the analysis, we used both cases of considering all the expressed genes, and only the deferentially expressed genes to compare the detected genes in both networks. These causal genes were shown to play a critical role in activating other genes to generate the bacterial response against the stress conditions in the environment.

The influential genes identified in the differential expression data-sets are correlated with basal cellular processes inside the genome. Moreover, essential genes and pathways were identified, which are related to the defense and adaptation of bacteria to different stresses; this can mean that these influential genes play a critical role in the synthesis of proteins and the survival process of the bacteria under the different stress conditions.

Author Contributions: Study conceptualization, P.G., R.T.J.R, EF,and PR .; methodology, P.R and E.F.; software, P.R.; validation, E.F., P.R ,A.R.F.C., A.L.Q.C. and R.T.J.R.; resources, V.A., R.T.J.R. and P.G.; data curation and prepossessing , E.F., P.R.; writing–original draft preparation, E.F. and A.L.Q.C; writing–review and editing, A.R.F.C.,R.T.J.R.,P.R.,P.G. and A.L.Q.C.; visualization,E.F. and P.R.; supervision, V.A.,R.T.J.R.,P.G.; funding acquisition, V.A, R.T.J.R, P.G

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and The Brazilian National Council for Scientific and Technological Development(CNPq) #307584/2018 − 6

Acknowledgments: The present study was conducted with support of the Brazilian Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES) and the Brazilian National Council for Scientific, Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq) and Pró-reitoria de Pesquisa e Pós-graduação(PROPESP) - UFPA. Biological Engineering Laboratory, Federal University of Pará (Universidade Federal do Pará – UFPA) and the Biological Networks Lab at Virginia Commonwealth University-VA.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CEM	co-expression networks
TPM	Transcripts Per Kilobase Million
WEG	Whole-genome gene
DE	Differentially expressed genes

Supplementary materials and data

Files 1-Networks files

Contains the files with the interaction networks obtained by the miRsig algorithm.

File 2-Supplementary Figures

Contains the supplementary figures of the paper.

Files 3- Network analysis and Clusters

Contains the network analyzes performed by Cytoscape and the results of the clustering analysis.

Files 4-Differential expressed genes results

Contains the results of differential expression analyzes performed with the Gfold and edgeR tools

References

1. Silva, W.M.; Dorella, F.A.; Soares, S.C.; Souza, G.H.; Castro, T.L.; Seyffert, N.; Figueiredo, H.; Miyoshi, A.; Le Loir, Y.; Silva, A.; others. A shift in the virulence potential of Corynebacterium pseudotuberculosis biovar ovis after passage in a murine host demonstrated through comparative proteomics. BMC microbiology **2017**, 17, 55.

2. de Sá, P.H.; Veras, A.A.; Carneiro, A.R.; Baraúna, R.A.; Guimarães, L.C.; Pinheiro, K.C.; Pinto, A.C.; Soares, S.C.; Schneider, M.P.; Azevedo, V.; others. Corynebacterium pseudotuberculosis RNA-seq data from abiotic stresses. Data in brief **2015**, 5, 963–966.
3. Pinto, A.C.; de Sá, P.H.C.G.; Ramos, R.T.; Barbosa, S.; Barbosa, H.P.M.; Ribeiro, A.C.; Silva, W.M.; Rocha, F.S.; Santana, M.P.; de Paula Castro, T.L.; others. Differential transcriptional profile of Corynebacterium pseudotuberculosis in response to abiotic stresses. BMC genomics **2014**, 15, 14.
4. Silva, W.M.; Seyffert, N.; Santos, A.V.; Castro, T.L.; Pacheco, L.G.; Santos, A.R.; Ciprandi, A.; Dorella, F.A.; Andrade, H.M.; Barh, D.; others. Identification of 11 new exoproteins in Corynebacterium pseudotuberculosis by comparative analysis of the exoproteome. Microbial pathogenesis **2013**, 61, 37–42.
5. Santana-Jorge, K.T.; Santos, T.M.; Tartaglia, N.R.; Aguiar, E.L.; Souza, R.F.; Mariutti, R.B.; Eberle, R.J.; Arni, R.K.; Portela, R.W.; Meyer, R.; others. Putative virulence factors of Corynebacterium pseudotuberculosis FRC41: vaccine potential and protein expression. Microbial cell factories **2016**, 15, 83.
6. Cerdeira, L.T.; Schneider, M.P.C.; Pinto, A.C.; de Almeida, S.S.; dos Santos, A.R.; Barbosa, E.G.V.; Ali, A.; Aburjaile, F.F.; de Abreu, V.A.C.; Guimarães, L.C.; others. Complete Genome Sequence of Corynebacterium pseudotuberculosis Strain CIP 52.97, Isolated from a Horse in Kenya. Journal of Bacteriology **2011**, 193, 7025–7026, [https://jb.asm.org/content/193/24/7025.full.pdf]. doi:10.1128/JB.06293-11.
7. Barh, D.; Gupta, K.; Jain, N.; Khatri, G.; León-Sicairos, N.; Canizalez-Roman, A.; Tiwari, S.; Verma, A.; Rahangdale, S.; Shah Hassan, S.; others. Conserved host–pathogen PPIs Globally conserved inter-species bacterial PPIs based conserved host-pathogen interactome derived novel target in C. pseudotuberculosis, C. diphtheriae, M. tuberculosis, C. ulcerans, Y. pestis, and E. coli targeted by Piper betel compounds. Integrative Biology **2013**, 5, 495–509, [https://academic.oup.com/ib/article-pdf/5/3/495/27302530/c2ib20206a.pdf]. doi:10.1039/c2ib20206a.
8. Morales, N.; Aldridge, D.; Bahamonde, A.; Cerda, J.; Araya, C.; Muñoz, R.; Saldías, M.E.; Lecocq, C.; Fresno, M.; Abalos, P.; others. Corynebacterium pseudotuberculosis infection in Patagonian Huemul (Hippocamelus bisulcus). Journal of wildlife diseases **2017**, 53, 621–624.
9. Chaitankar, V.; Ghosh, P.; Perkins, E.; Gong, P.; Zhang, C. Time lagged information theoretic approaches to the reverse engineering of gene regulatory networks. BMC Bioinform. **2010**, 11:S19.
10. Chaitankar, V.; Ghosh, P.; Perkins, E.; Gong, P.; Deng, Y.; Zhang, C. A novel gene network inference algorithm using predictive minimum description length approach. BMC Systems Biology **2010**, 4:S7.
11. Chai, L.E.; Loh, S.K.; Low, S.T.; Mohamad, M.S.; Deris, S.; Zakaria, Z. A review on the computational approaches for gene regulatory network construction. Computers in biology and medicine **2014**, 48, 55–65.
12. Marbach, D.; Costello, J.C.; Küffner, R.; Vega, N.M.; Prill, R.J.; Camacho, D.M.; Allison, K.R.; Aderhold, A.; Bonneau, R.; Chen, Y.; others. Wisdom of crowds for robust gene network inference. Nature methods **2012**, 9, 796.
13. Tieri, P.; Farina, L.; Petti, M.; Astolfi, L.; Paci, P.; Castiglione, F. Network Inference and Reconstruction in Bioinformatics. In Encyclopedia of Bioinformatics and Computational Biology; Ranganathan, S.; Gribskov, M.; Nakai, K.; Schönbach, C., Eds.; Academic Press: Oxford, 2019; pp. 805 – 813. doi:https://doi.org/10.1016/B978-0-12-809633-8.20290-2.
14. Suratanee, A.; Chokrathok, C.; Chutimanukul, P.; Khrueasan, N.; Buaboocha, T.; Chadchawan, S.; Plaimas, K. Two-State Co-Expression Network Analysis to Identify Genes Related to Salt Tolerance in Thai Rice. Genes **2018**, 9, 594.
15. 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. Scientific reports **2017**, 7, 8133.
16. Guille, A.; Hacid, H.; Favre, C.; Zighed, D.A. Information diffusion in online social networks: A survey. ACM Sigmod Record **2013**, 42, 17–28.
17. Nalluri, J.J.; Barh, D.; Azevedo, V.; Ghosh, P. miRsig: a consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures. Scientific reports **2017**, 7, 39684.
18. Ruiz, J.C.; D'Afonseca, V.; Silva, A.; Ali, A.; Pinto, A.C.; Santos, A.R.; Rocha, A.A.; Lopes, D.O.; Dorella, F.A.; Pacheco, L.G.; others. Evidence for reductive genome evolution and lateral acquisition of virulence functions in two Corynebacterium pseudotuberculosis strains. PloS one **2011**, 6.

19. Gomide, A.C.P.; de Sa, 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.
20. Dorella, F.A.; Estevam, E.M.; Pacheco, L.G.; Guimarães, C.T.; Lana, U.G.; Gomes, E.A.; Barsante, M.M.; Oliveira, S.C.; Meyer, R.; Miyoshi, A.; others. In vivo insertional mutagenesis in Corynebacterium pseudotuberculosis: an efficient means to identify DNA sequences encoding exported proteins. Appl. Environ. Microbiol. **2006**, 72, 7368–7372.
21. Ribeiro, D.; de Souza Rocha, F.; Leite, K.M.C.; de Castro Soares, S.; Silva, A.; Portela, R.W.D.; Meyer, R.; Miyoshi, A.; Oliveira, S.C.; Azevedo, V.; others. An iron-acquisition-deficient mutant of Corynebacterium pseudotuberculosis efficiently protects mice against challenge. Veterinary research **2014**, 45, 28.
22. 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.; others. Transcriptome profile of Corynebacterium pseudotuberculosis in response to iron limitation. BMC genomics **2019**, 20, 663.
23. Andrews, S.; others. FastQC: a quality control tool for high throughput sequence data, 2010.
24. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014**, 30, 2114–2120.
25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal **2011**, 17, 10–12.
26. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods **2017**, 14, 417.
27. Wagner, G.P.; Kin, K.; Lynch, V.J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in biosciences **2012**, 131, 281–285.
28. Wheeler, D.L.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; others. Database resources of the national center for biotechnology information. Nucleic acids research **2007**, 36, D13–D21.
29. Markowitz, V.M.; Chen, I.M.A.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; others. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic acids research **2012**, 40, D115–D122.
30. Feng, J.; Meyer, C.A.; Wang, Q.; Liu, J.S.; Shirley Liu, X.; Zhang, Y. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics **2012**, 28, 2782–2788.
31. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **2010**, 26, 139–140.
32. Warden, C.D.; Yuan, Y.C.; Wu, X. Optimal calculation of RNA-Seq fold-change values. International Journal of Computational Bioinformatics and In Silico Modeling **2013**, 2, 285–292.
33. MacQueen, J.; others. Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Oakland, CA, USA, 1967, Vol. 1, pp. 281–297.
34. Pavlopoulos, G.A.; Secrier, M.; Moschopoulos, C.N.; Soldatos, T.G.; Kossida, S.; Aerts, J.; Schneider, R.; Bagos, P.G. Using graph theory to analyze biological networks. BioData mining **2011**, 4, 10.
35. Morris, J.H.; Apeltsin, L.; Newman, A.M.; Baumbach, J.; Wittkop, T.; Su, G.; Bader, G.D.; Ferrin, T.E. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC bioinformatics **2011**, 12, 436.
36. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research **2003**, 13, 2498–2504.
37. Araujo, F.A.; Barh, D.; Silva, A.; Guimarães, L.; Ramos, R.T.J. GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. Scientific reports **2018**, 8, 1794.
38. Doncheva, N.T.; Morris, J.H.; Gorodkin, J.; Jensen, L.J. Cytoscape stringApp: Network analysis and visualization of proteomics data. Journal of proteome research **2018**, 18, 623–632.
39. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology **2012**, 16, 284–287.
40. Wickham, H. ggplot2: elegant graphics for data analysis; Springer, 2016.
41. Kassambara, A. ggpubr:"ggplot2" based publication ready plots. R package version 0.1 **2017**, 6.

42. 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 research **2016**, p. gkw1013.
43. Rocha, D.J.; Santos, C.S.; Pacheco, L.G. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. Antonie Van Leeuwenhoek **2015**, 108, 685–693.
44. Radkov, A.D.; Hsu, Y.P.; Booher, G.; VanNieuwenhze, M.S. Imaging bacterial cell wall biosynthesis. Annual review of biochemistry **2018**, 87, 991–1014.
45. McKean, S.; Davies, J.; Moore, R. Identification of macrophage induced genes of Corynebacterium pseudotuberculosis by differential fluorescence induction. Microbes and infection **2005**, 7, 1352–1363.
46. Merkamm, M.; Guyonvarch, A. Cloning of the sodA Gene fromCorynebacterium melassecola and Role of Superoxide Dismutase in Cellular Viability. Journal of bacteriology **2001**, 183, 1284–1295.
47. den Hengst, C.D.; Buttner, M.J. Redox control in actinobacteria. Biochimica et Biophysica Acta (BBA)-General Subjects **2008**, 1780, 1201–1216.
48. Newton, G.L.; Av-Gay, Y.; Fahey, R.C. A novel mycothiol-dependent detoxification pathway in mycobacteria involving mycothiol S-conjugate amidase. Biochemistry **2000**, 39, 10739–10746.
49. Rawat, M.; Newton, G.L.; Ko, M.; Martinez, G.J.; Fahey, R.C.; Av-Gay, Y. Mycothiol-deficient Mycobacterium smegmatis mutants are hypersensitive to alkylating agents, free radicals, and antibiotics. Antimicrobial agents and chemotherapy **2002**, 46, 3348–3355.
50. Seebeck, F.P. In vitro reconstitution of Mycobacterial ergothioneine biosynthesis. Journal of the American Chemical Society **2010**, 132, 6632–6633.
51. Cheah, I.K.; Halliwell, B. Ergothioneine; antioxidant potential, physiological function and role in disease. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease **2012**, 1822, 784–793.
52. Ciccia, A.; Elledge, S.J. The DNA damage response: making it safe to play with knives. Molecular cell **2010**, 40, 179–204.
53. Satorhelyi, P. Microarray-analyse der pH-stressantwort von Listeria monocytogenes und Corynebacterium glutamicum. PhD thesis, Technische Universität München, 2005.
54. Weeks, D.L.; Sachs, G. Sites of pH regulation of the urea channel of Helicobacter pylori. Molecular microbiology **2001**, 40, 1249–1259.
55. Heermann, R.; Jung, K. Structural features and mechanisms for sensing high osmolarity in microorganisms. Current opinion in microbiology **2004**, 7, 168–174.
56. Dodson, K.W.; Pinkner, J.S.; Rose, T.; Magnusson, G.; Hultgren, S.J.; Waksman, G. Structural basis of the interaction of the pyelonephritic E. coli adhesin to its human kidney receptor. Cell **2001**, 105, 733–743.
57. Park, J.H.; Roe, J.H. Mycothiol regulates and is regulated by a thiol-specific antisigma factor RsrA and σR in Streptomyces coelicolor. Molecular microbiology **2008**, 68, 861–870.
58. Martín, J.F. Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: an unfinished story. Journal of bacteriology **2004**, 186, 5197–5201.
59. Liu, G.; Chater, K.F.; Chandra, G.; Niu, G.; Tan, H. Molecular regulation of antibiotic biosynthesis in Streptomyces. Microbiol. Mol. Biol. Rev. **2013**, 77, 112–143.
60. Bromke, M.A. Amino acid biosynthesis pathways in diatoms. Metabolites **2013**, 3, 294–311.
61. Kisker, C.; Kuper, J.; Van Houten, B. Prokaryotic nucleotide excision repair. Cold Spring Harbor perspectives in biology **2013**, 5, a012591.