Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

A comparative pan-genomic analysis of 53 *C. pseudotuberculosis* strains based on functional domains

Juliana S. Bernardes, Raphael J. Eberle, Fabio R. J. Vieira & Mônika A. Coronado

To cite this article: Juliana S. Bernardes, Raphael J. Eberle, Fabio R. J. Vieira & Mônika A. Coronado (2021) A comparative pan-genomic analysis of 53 *C. pseudotuberculosis* strains based on functional domains, Journal of Biomolecular Structure and Dynamics, 39:18, 6974-6986, DOI: 10.1080/07391102.2020.1805017

To link to this article: https://doi.org/10.1080/07391102.2020.1805017

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
2021, VOL. 39, NO. 18, 6974–6986
https://doi.org/10.1080/07391102.2020.1805017

A comparative pan-genomic analysis of 53 *C. pseudotuberculosis* strains based on functional domains

Juliana S. Bernardes[a], Raphael J. Eberle[b], Fabio R. J. Vieira[c] and Mônika A. Coronado[b,d]

[a]Laboratoire de Biologie Computationelle et Quantitative, UMR 7238, CNRS, Sorbonne Université, Paris, France; [b]Multiuser Center for Biomolecular Innovation, Department of Physics, Instituto de Biociências, Letras e Ciências Exatas (Ibilce), Universidade Estadual Paulista (UNESP), São Jose do Rio Preto, Brazil; [c]Institut de Biologie de l'École Normale Supérieure (IBENS), Paris, France; [d]Institute of Biological Information Processing (IBI-7: Strucutral Biochemistry), Forschungszentrum Juelich, Juelich, Germany

Communicated by Ramaswamy H. Sarma

ABSTRACT

Corynebacterium pseudotuberculosis is a pathogenic bacterium with great veterinary and economic importance. It is classified into two biovars: *ovis*, nitrate-negative, that causes lymphadenitis in small ruminants and *equi*, nitrate-positive, causing ulcerative lymphangitis in equines. With the explosive growth of available genomes of several strains, pan-genome analysis has opened new opportunities for understanding the dynamics and evolution of *C. pseudotuberculosis*. However, few pan-genomic studies have compared biovars *equi* and *ovis*. Such studies have considered a reduced number of strains and compared entire genomes. Here we conducted an original pan-genome analysis based on protein sequences and their functional domains. We considered 53 *C. pseudotuberculosis* strains from both biovars isolated from different hosts and countries. We have analysed conserved domains, common domains more frequently found in each biovar and biovar-specific (unique) domains. Our results demonstrated that biovar *equi* is more variable; there is a significant difference in the number of proteins per strains, probably indicating the occurrence of more gene loss/gain events. Moreover, strains of biovar *equi* presented a higher number of biovar-specific domains, 77 against only eight in biovar *ovis*, most of them are associated with virulence mechanisms. With this domain analysis, we have identified functional differences among strains of biovars *ovis* and *equi* that could be related to niche-adaptation and probably help to better understanding mechanisms of virulence and pathogenesis. The distribution patterns of functional domains identified in this work might have impacts on bacterial physiology and lifestyle, encouraging the development of new diagnoses, vaccines, and treatments for *C. pseudotuberculosis* diseases.

ARTICLE HISTORY
Received 23 March 2020
Accepted 27 July 2020

KEYWORDS
C. pseudotuberculosis; biovar ovis; biovar equi; functional domains and pan-genomic analyses

Introduction

The gram-positive and intracellular bacterium *Corynebacterium pseudotuberculosis* (*C. pseudotuberculosis*) has veterinary and medical relevances. It belongs to the *Corynebacterium*, *Mycobacterium*, *Nocardia*, and *Rhodococcus* group of the actinobacteria phylum (Baird & Fontaine, 2007; Dorella et al., 2006). These bacteria have specific features in common, such as high DNA G + C content and a specific organization of the cell wall (Dorella et al., 2006). The genus *Corynebacterium* is composed of pathogenic species such as *C. pseudotuberculosis* (Dorella et al., 2006) and *Corynebacterium diphtheriae* (Cerdeno-Tarraga et al., 2003), opportunistic pathogens such as *Corynebacterium jeikeium* (Tauch et al., 2005) and non-pathogenic species such as *Corynebacterium glutamicum* (Kalinowski et al., 2003).

C. pseudotuberculosis 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 (Almeida et al., 2016; Dorella et al., 2006; Trost et al., 2010). *C. pseudotuberculosis* can also affect humans, causing distinct types of lymphadenitis. Contamination occurs through contact with infected animals and consumption of infected food (Dorella et al., 2006; Peel et al., 1997; Trost et al., 2010). *C. pseudotuberculosis* strains are classified into two biovars depending on host preference and nitrate reduction: *ovis* and *equi*. Biovar *ovis* (nitrate negative) is the causative agent of Caseous Lymphadenitis (CLA), a chronic disease affecting goats and sheep (Baird & Fontaine, 2007; Williamson, 2001). *C. pseudotuberculosis* biovar *ovis* has been isolated from cattle (Yeruham et al., 2004), camels (Hawari, 2008), and humans (Baird & Fontaine, 2007; Heggelund et al., 2015; Trost et al., 2010). Since 1960s CLA has also been reported in wildlife species as aardvark (*Orycteropus afer*) (Roth & Vickers, 1966), pronghorn antelope (*Antilocapra americana*) (Clark et al., 1972), white-tailed deer (*Odocoilecus virginianus*) (Stauber

CONTACT Fabio R. J. Vieira fabiorjvieira@gmail.com Institut de Biologie de l'École Normale Supérieure (IBENS), Paris, 75005, France; Mônika A. Coronado m.coronado@fz-juelich.de Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Juelich, Wilhelm-Johnen-Straße, Juelich, 52425, Germany.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2020.1805017.

et al., 1973), cheetah (*Acinonyx jubatus*) (Boomker & Henton, 1980), alpine chamois (*Rupicapra r. rupicapra*) (Bassano et al., 1993), fallow deer (*Dama dama*) (Perez et al., 1996), alpine ibex (*Capra ibex*) (Silinski & Walzer, 2004), Arabian oryx (*Oryx leucorix*) (Tarello & Theneyan, 2008), elk (*Cervus canadensis nelsoni*) (Jane Kelly et al., 2012), spanish ibex (*Capra pyrenaica*) (Colom-Cadena et al., 2014), red deer (*Cervus elaphus*) (Matos et al., 2015), white-tailed gnu (*Connochaetes gnou*) (Müller et al., 2011), and huemul (*Hippocamelus bisulcus*) (Morales et al., 2017). On the other hand, isolates from the biovar *equi* (nitrate positive) can manifest in several forms; Oedematous Skin Disease (OSD) in buffaloes (Selim, 2001) and ulcerative lymphangitis abscesses in internal organs as well as external abscesses in equines (Britz et al., 2014; Foley et al., 2004; Spier & Azevedo, 2017). *Equi* isolates have also been found in cattle (Yeruham et al., 1996, 2004), camels (Tejedor-Junco et al., 2008) and chamois (Domenis et al., 2018), with different manifestation in each host species. Biovar *ovis* has never been found in horses or buffalo and, no sheep or goats have been infected by any strain belonging to the biovar *equi* (Viana et al., 2018).

Strains of biovar *equi* and *ovis* present a significant variation in their molecular characteristics. Pan-genomic studies have helped to reveal the content, organisation and the evolution of *C. pseudotuberculosis* genomes (Oliveira et al., 2016; Soares et al., 2013; Viana et al., 2017). So far, such studies did not assign the different symptoms of *C. pseudotuberculosis* infections to specific genes or proteins. Since protein domains are responsible for a particular function or interaction and contribute to the overall function of a protein, we have investigated if a comparative pan-genomic analysis of *C. pseudotuberculosis* strains, based on functional domains, could explain the differences of the phenotype reported in *C. pseudotuberculosis* biovars. Studies identified genotypic and phenotypic modifications responsible to differentiate biovar *ovis* and *equi* as nitrate reduction, changes in serotype and diseases manifestation, and pathogenicity islands that are specific for each biovar (Barakat et al., 1984; Biberstein et al., 1971; Oliveira et al., 2016; Soares et al., 2013). With domain analysis, we aimed to identify functional differences among strains of biovars *ovis* and *equi*. For that, we comparatively analysed the complete genome sequences of 53 *C. pseudotuberculosis* strains (25 from biovar *ovis* and 28 from biovar *equi*) isolated from different hosts and different countries. Domain sequences and domain properties that are different in both biovars have been highlighted, shedding light on their potential of infection in different animals. We also analyzed conserved and abundant domains shared by biovars *equi* and *ovis*, as well as, specific/unique domains exclusively found in each biovar.

Results and discussion

The main goal of pan-genome studies is to perform a genome comparison of different strains of the same species. With the explosive growth of available bacterial genomes, pan-genome analysis has opened new opportunities for understanding the dynamics and evolution of bacterial genomes, mainly the pathogenic ones. Here we performed an original pan-genomic analysis of 53 *C. pseudotuberculosis* strains focusing on their functional domains. Domain comparisons play an important role in comparative genomics but have been poorly explored in pan-genomics studies (Snipen & Ussery, 2012). We analysed protein/domain properties of strains in both biovars (*equi* and *ovis*) by searching for some divergence and biovar specific-domains.

Proteins and domain identification

A total of 53 completed genomes of *C. pseudotuberculosis* strains were chosen: 25 strains represent the biovar ovis, which were isolated from *Ovies aries* (sheep), *Capra aegagrus hircus* (goat), *Lama glama* (llama), and *Connochaetes taurinus* (wildebeest) and other 28 strains represent biovar *equi*, isolated from *Equus caballus* (horse), *Camelus ferus* (camel), *Bos taurus* (cattle), and *Bubalus bubalis* (buffalo). Table 1 shows the host, country, and the number of proteins for each strain, and Figure 1 shows their phylogenetic relationships.

The complete phylogenetic tree of *Corynebacterium* genus, including the 53 *C. pseudotuberculosis* strains is shown in Figure S1, where we clearly observe two initial phylogenetic branch that separates *C. pseudotuberculosis* from other genus; the tree was generated by Gegenees (Ågren et al., 2012), see Methods. The total number of encoded unique proteins identified in both biovars was 12562 of which 3282 (26.2%) exclusively belong to biovar *ovis*, 7086 (56.4%) to biovar *equi* and, the core proteome is composed of 2194 (17.4%) proteins (Figure 2(A)).

Strains of biovar *equi* possess more than twice unique sequences than strains of biovar *ovis*, probably indicating more sequence variability in that biovar. To confirm this hypothesis, an orthologue analysis using EDGAR software (Blom et al., 2009) was carried out. 1138 orthologous groups were detected in all strains of both biovars, despite those, 83 orthologous groups are exclusively for strains of biovar *ovis* and 266 for strains of biovar *equi* (Figure 2(B)). This result confirms lower sequence variability in biovar *ovis* compared to biovar *equi*. The higher diversity of biovar *equi* can also be confirmed by observing the number of proteins of each strain (Figure 3(A) and Table S1). On average, biovar *equi* strains contains 2022 proteins while biovar *ovis* strains contains 2012 proteins. In biovar *ovis*, the strain with the lowest number of proteins is P54B96, isolated from *C. taurinus* in South Africa with 1950 proteins and the strain with the highest number of proteins is 1002, isolated from *C. aegagrus hircus* in Brazil with 2090 proteins that represents a difference of 140 proteins between them. In biovar *equi*, the difference is much more apparent, the strain with fewer proteins is 1_06-A, isolated from *E. caballus*, in USA, with 1860 proteins and the strain having more proteins is 258 with 2126 proteins, also isolated from *E. caballus* but in Belgium, it makes a difference of 266 proteins between them. Strains of biovar *ovis* have more or less the same number of proteins with the standard deviation of 26, while in biovar *equi* we observe an important variation with a standard deviation of 61 (almost 3-fold higher than biovar *ovis*), see Table S1.

Table 1. *C. pseudotuberculosis* strains general information.

Biovar *Equi*				Biovar *Ovis*			
StrainID	Host	Country	Number of Proteins	StrainID	Host	Country	Number of Proteins
262	*B. taurus*	Belgium	2029	PA02	*C. a. hircus*	Brazil	2024
I37	*B. taurus*	Israel	2026	T1	*C. a. hircus*	Brazil	2018
29156	*Bovine*	Israel	2023	1002	*C. a. hircus*	Brazil	2090
I19	*Bovine*	Israel	2021	Cp13	*C. a. hircus*	Brazil	2019
38	*B. bubalis*	Egypt	2061	VD57	*C. a. hircus*	Brazil	2012
31	*B. bubalis*	Egypt	2093	1002B	*C. a. hircus*	Brazil	2016
39	*B. bubalis*	Egypt	2065	MEX9	*C. a. hircus*	Mexico	2022
43	*B. bubalis*	Egypt	2029	PO222_4-1	*C. a. hircus*	Portugal	2023
32	*B. bubalis*	Egypt	2067	PO269-5	*C. a. hircus*	Portugal	2014
46	*B. bubalis*	Egypt	2026	226	*C. a. hircus*	USA	1960
33	*B. bubalis*	Egypt	2064	Ft_2193_67	*C. a. hircus*	Norway	2017
48	*B. bubalis*	Egypt	2068	CS 10	*C. a. hircus*	Norway	2009
34	*B. bubalis*	Egypt	2073	267	*L. glama*	USA	2031
35	*B. bubalis*	Egypt	2069	PAT10	*O. aries*	Argentina	1983
36	*B. bubalis*	Egypt	2064	42_02-A	*O. aries*	Australia	2015
Cp162	*C. ferus*	UK	2031	C231	*O. aries*	Brazil	1993
258	*E. caballus*	Belgium	2126	12C	*O. aries*	Brazil	2009
CIP_52.97	*E. caballus*	Kenya	2044	PA01	*O. aries*	Brazil	2030
MEX31	*E. caballus*	Mexico	2058	E56	*O. aries*	Egypt	1992
MEX30	*E. caballus*	Mexico	2008	E55	*O. aries*	Egypt	1992
316	*E. caballus*	USA	1919	N1	*O. aries*	Egypt	2015
MB20	*E. caballus*	USA	1902	MEX25	*O. aries*	Mexico	2014
1_06-A	*E. caballus*	USA	1860	MEX29	*O. aries*	Mexico	2030
E19	*E. caballus*	Chile	2037	3_99-5	*O. aries*	Scotland	2017
MB11	*E. caballus*	USA	1928	P54B96	*C. taurinus*	South Africa	1950
MB14	*E. caballus*	USA	1965				
MB30	*E. caballus*	USA	1994				
MB66	*E. caballus*	USA	1961				

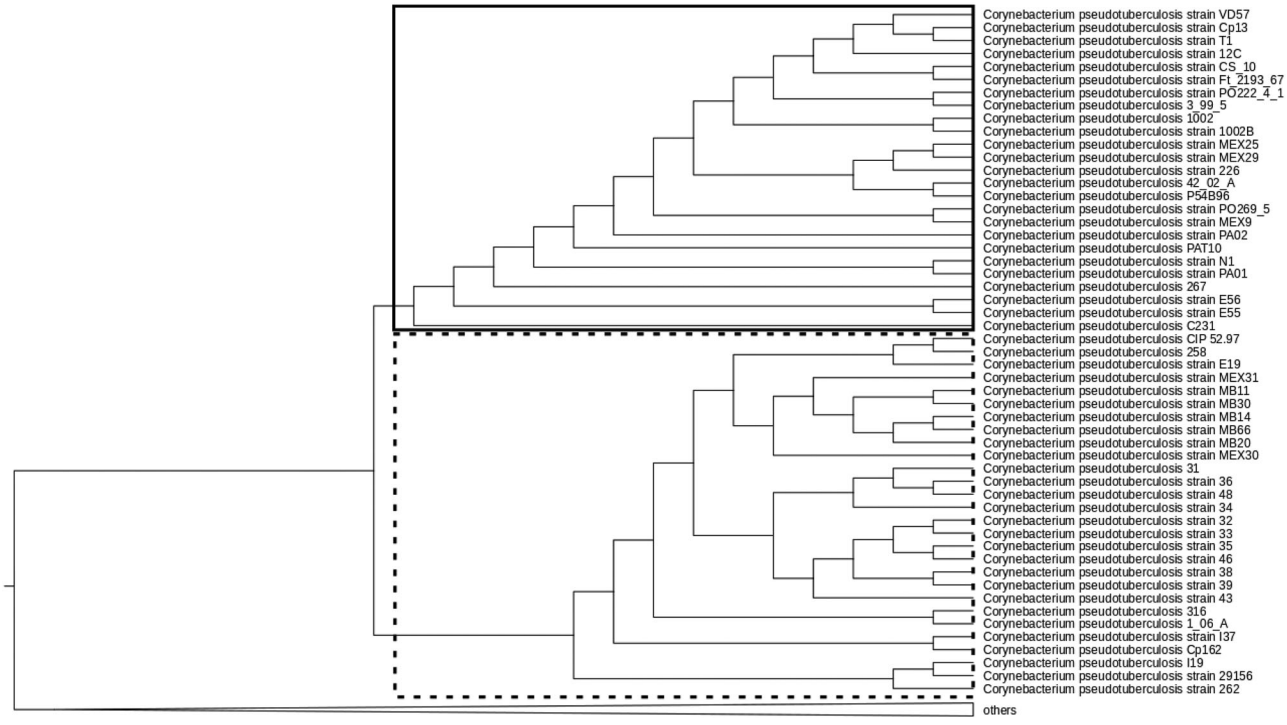

Figure 1. Phylogenomic tree of 25 strains of biovar *ovis* (solid rectangle), 28 strains of biovar *equi* (dotted rectangle) and other *Corynebacterium* species. 168 complete genomes from the *Corynebacterium* genus were retrieved from the NCBI ftp site. Gegenees software was used to derive a variable content for each genome and compute a distance matrix based on percentages of similarities. The distance matrix was used to build a phylogenomic tree with SplitsTree and UPGMA method. Only phylogenetic relationships of *C. pseudotuberculosis* strains are shown, and other species were grouped in the branch "others", the complete tree is found in Figure S1.

Identification of common domains

Functional domains were predicted with CLADE (Bernardes, Zaverucha, et al., 2016) and DAMA (Bernardes, Vieira, et al., 2016), 1694 non-redundant Pfam domains has been identified, of which 1617 (~95%) are shared by both biovars demonstrating high domain conservation between them. Table S2 shows general information of common domains such as: Pfam domain name, accession number, clan,

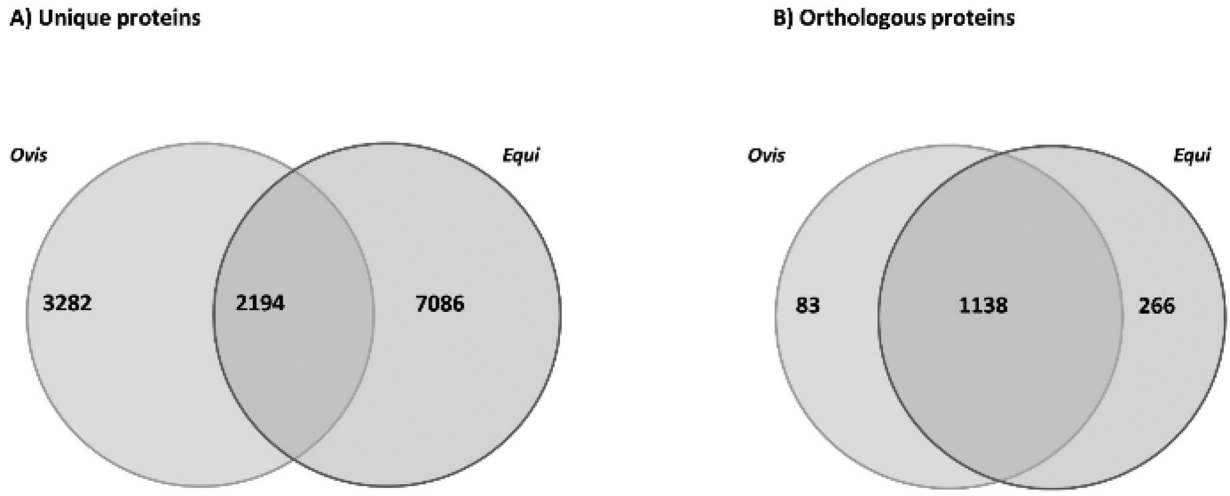

Figure 2. Venn diagram representing unique proteins and orthologous proteins for biovars *equi* and *ovis*. (A) Unique proteins identified in both biovars and exclusively found in each biovar. (B) Orthologous proteins detected in all strains of both biovars, and exclusively found in each biovar.

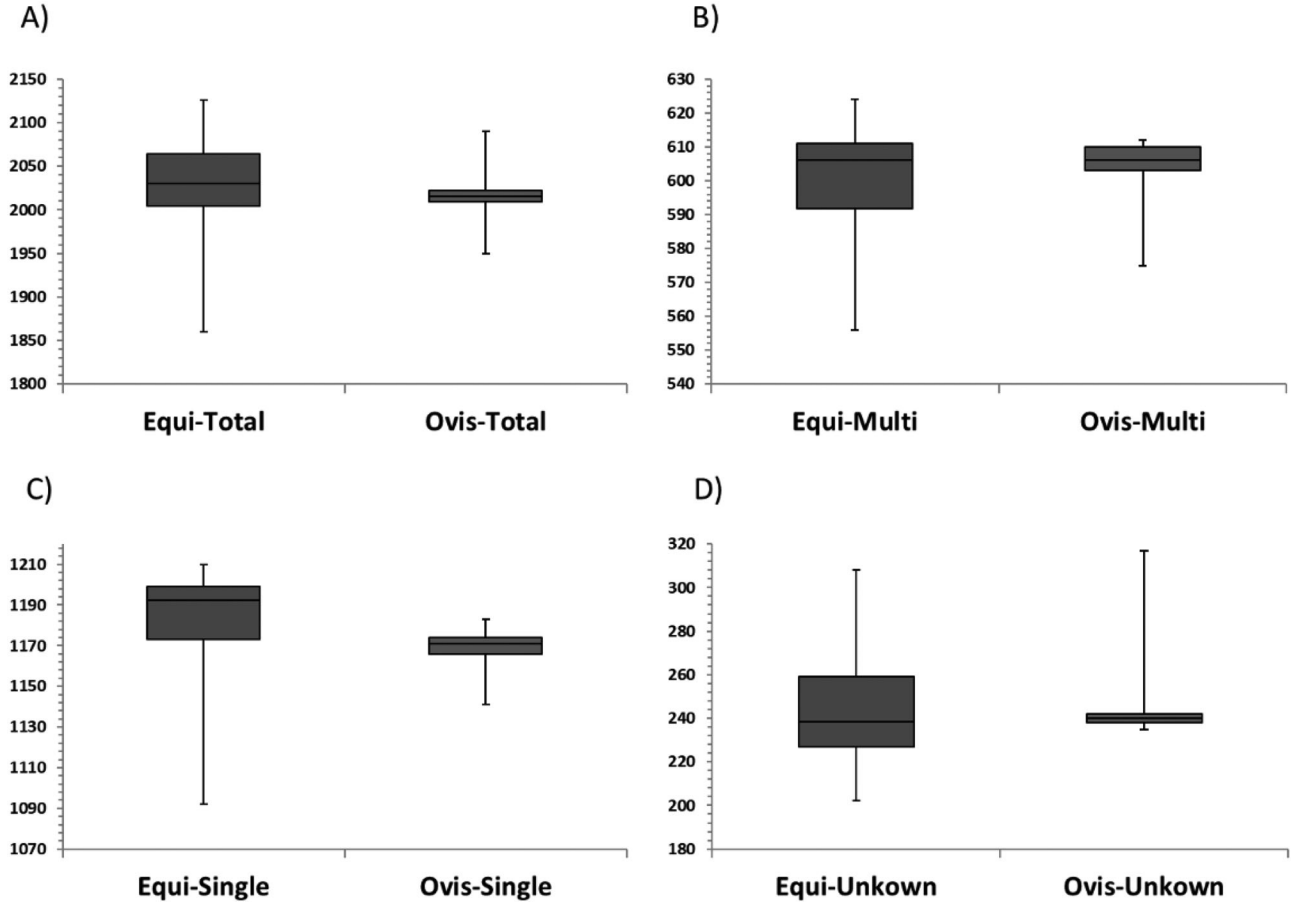

Figure 3. Boxes plots representing sequence and domain properties for biovars *equi* and *ovis*. (A) Number of proteins per strain, p-value = 2.73E-12. (B) Number of proteins having at least two domains, p-value = 1.85E-11. (C) Number of proteins annotated with only one domain, p-value = 0.0043. (D) Number of proteins with no detected domain, p-value = 0.083. We used the t-student statistical analysis to obtain p-values. To highlight the dispersion profile of the boxes we considered only values between the 1st and 3rd quartiles.

description, GO terms (when available), and the number of occurrences for each biovar. We split common domains into three categories: (i) most frequent domains identified in both biovars; (ii) domains mainly identified in strains of biovar *equi* and (iii) domains mainly identified in strains of biovar *ovis*. For category (i) the focus was on common domains with more than ten copies per genome (Table S2i). The most abundant domains are associated to ATPase activity and transmembrane transport and they were found in all strains of both biovars with very similar proportion. BPD_transp_1 is a domain described as bacterial binding protein-dependent transport system and, is involved in the transmembrane transport of inorganic ions, amino acids, sugars, large polysaccharides or even proteins (Higgins et al., 1990). Strains of

biovar *equi* possess on average of 28 copies of BPD_transp_1 domain, while in strains of biovar *ovis* the average is around 26 copies. Another abundant domain is ABC_ATPase; it belongs to the ABC transport clan (accession number CL0023) that was the most frequent clan found in *C. pseudotuberculosis* strains. ATP-binding cassete (ABC) systems are universally distributed among living organisms and it works in many different aspects of bacterial physiology. These transmembrane proteins are involved in the export or import of a wide variety of substrates ranging from small ions to macromolecules (Davidson et al., 2008). A high number of copies of ABC_ATPase domain were found in both biovars, being this domain more frequent in strains of biovar *ovis*, 30 copies against 26 in biovar *equi*. The third most abundant domain is ABC_tran that is associated to ABC_ATPase and belong to the clan CL0023. A mean of 20 copies of ABC_tran domain are present in strains of both biovars. Other domains found in biovar *ovis* and biovar *equi* that are also related to ATP binding activity and membrane transport are AAA_21, cystathionine beta synthase (CBS) and major facilitator superfamily (MFS) domain (Kemp, 2004; Reddy et al., 2012).

Two abundant domains, FecCD and ZnuA, are associated to high-affinity zinc uptake system and iron-citrate transport proteins in the cytoplasmic membrane, respectively. FecCD domain is required in periplasmic-binding-protein-dependent transport mechanism for iron citrate transport (Staudenmaier et al., 1989). Iron acquisition is one of the most important factors for bacterial survival during infection in the host environment (Brown & Holden, 2002). In *C. pseudotuberculosis* the iron uptake is mainly regulated by the diphtheria toxin repressor (DtxR) that modulates the expression of siderophores and other compounds in the iron acquisition pathway (Rodriguez & Smith, 2003; Xu et al., 1994). On average, 11 copies of FecCD domain were detected in all strains of both biovars. ZnuA domain, described as Zinc-uptake complex component and associated with metal ion transport, was found ten times in each proteome of both biovars. Zinc serves as an essential cofactor for many enzymes involved in diverse biological processes but can be toxic at high concentrations. Cells regulate the uptake, distribution and excretion of zinc through several systems, including the Zn^{2+}-specific uptake system (Znu). Znu domain belongs to the ATP-binding cassette (ABC) transporter family (Lee & Helmann, 2007; Schroder et al., 2010). *C. pseudotuberculosis* genomes encode for a copper, zinc- dependent superoxide dismutase (SodC) protective enzyme, where zinc is necessary for the proper activity of the protein. SodC is anchored in the cell membrane (Trost et al., 2010) and the extracellular location of this enzyme suggests that it may protect the surface of *C. pseudotuberculosis* cells against superoxide generated externally by the mammalian host cells (Trost et al., 2010). The protective activity of SodC is continuously associated with virulence in *C. pseudotuberculosis* and other bacteria such as *Neisseria meningitidis* and *Haemophilus ducreyi* (Santana-Jorge et al., 2016; Trost et al., 2010). Another abundant domain identified in biovar *ovis* and biovar *equi* is helicase_C, a domain found in a wide variety of DEAD-box RNA helicase proteins. These enzymes are involved in many critical aspects of RNA metabolism within eukaryotic and prokaryotic organisms. Several studies have shown that these proteins play a key role in mediating microbial pathogenesis and virulence (Heung & Del Poeta, 2005).

Category (ii) clusters the most prominent domains (eight in total) detected in strains of biovar *equi*, see Table S2ii. ILVD_EDD domain belongs to the dehydratase family, which is a group of lyase enzymes, forming double and triple bonds. There are more than 150 different dehydratase enzymes (Cook et al., 1992), making it difficult to assign a specific function in *C. pseudotuberculosis*. The ILVD_EDD domain is present in all strains of biovar *equi*, but in only 11 strains of biovar *ovis*, note that only one copy of ILVD_EDD was found in the proteome of both biovars. UvrD_C_2 domain, described as UvrD-like helicase C- terminal domain, was found in 25 strains of biovar *equi*, and just in one strain of biovar *ovis*. This domain is often found in proteins with a helicase activity. They are tightly integrated (or coupled) components of various macromolecular complexes, which are involved in DNA replication, recombination, and nucleotide excision repair, as well as RNA transcription and splicing (Gwynn et al., 2013). The Soluble-Ligand-Binding β-grasp (SLBB) are known or predicted to bind soluble ligands, like cobalamin or carbohydrates (Burroughs et al., 2007), and is presented in 24 strains of biovar *equi*.

The most abundant domains observed in strains of biovar *ovis*, category (iii), present a higher frequency when compared to biovar *equi*, for instance, Glyco_hydro_15 domain is present in all strains of biovar *ovis* against only six for biovar *equi* (Table S2iii). Note that two copies of Glyco_hydro_15 were found in both biovars. Glycoside hydrolases are a widespread group of enzymes that catalyse the glycosidic linkage between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety, existing more than 100 different families. Glycoside hydrolase family 15 comprises enzymes with several known activities; glucoamylase, alpha-glucosidase, and glucodextranase (Davies & Henrissat, 1995; Henrissat et al., 1995; Henrissat & Bairoch, 1996). Another example is Cu-oxidase 2 (multicopper oxidase) domain found in all biovar *ovis* strains and only seven strains of biovar *equi*. Interestingly, copper is utilised by macrophages to kill pathogenic bacteria (Achard et al., 2012; Wolschendorf et al., 2011) and, multicopper oxidases are important in copper detoxification in many bacteria and play a crucial role in the virulence of pathogens (Achard et al., 2010; Rowland & Niederweis, 2013). Phospholipase D (PLD) domain, which is associated to the major virulence determinant, was identified in all strains of both biovars, see line 1104 and 1115 of Table S2i. PLD play a critical role in dissemination of the bacteria from the site of infection to the lymph nodes (McKean et al., 2007). The PLD exotoxin is the trigger to increased vascular permeability through catalysis that dissociate the sphingomyelin, allowing the spread of pathogen into a new host (Sá et al., 2013).

Identification of biovar-specific domains

Remarkable is the difference in the number of biovar-specific (unique) domains that was identified in both biovars, Table

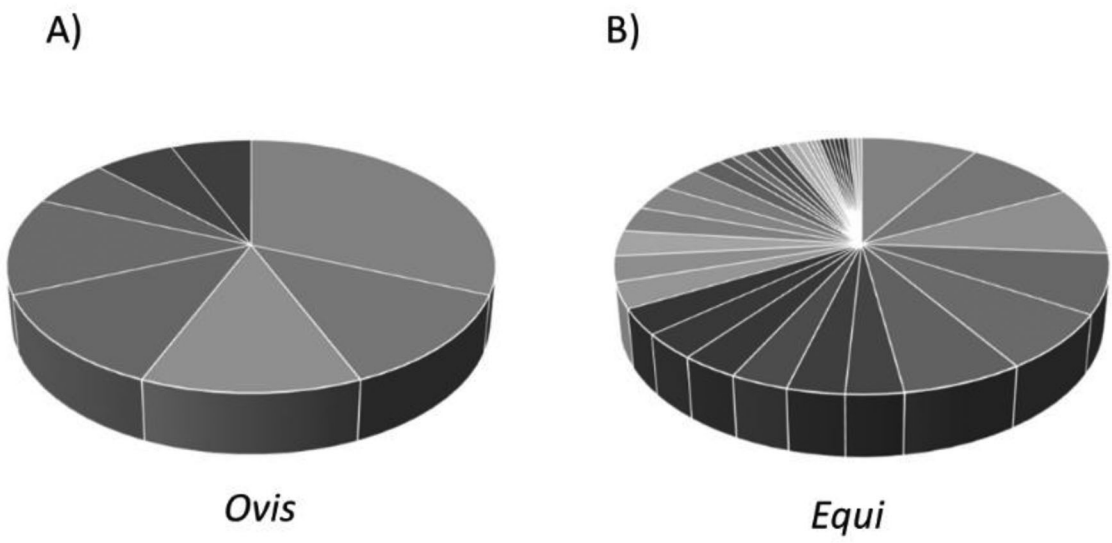

Figure 4. Pie charts showing the distribution of biovar-specific (unique) domains. (A) Unique domains found in strains of biovar *ovis*. (B) Unique domains found in strains of biovar *equi*.

S3; Biovar *equi* possess a very high diversity with 77 unique domains (Table S3ii) against eight in biovar ovis (Table S3i), see also Figures 4–6. Unique domains of biovar *equi* were grouped according to their protein function defined in the Pfam database or their GO terms (Table S3iii). Note that nine unique domains detected in strains of biovar *equi* were not considered because they present similar function with those found in both biovars (Table S3iii). In biovar *ovis* unique domains were found in a few number of strains (just five strains), while in biovar *equi* half part of unique domains (38/77 ~ 50%) was constantly found in at least 20 strains, Table S3ii - line 44. Figures 5 and 6 show the distribution of the unique domains according to the strains, country and hosts of biovars *equi* and *ovis*, respectively. A black circle represents the presence of a domain in a given strain, while a white circle denotes the absence; strains are grouped by host. In biovar *equi*, we observed a significant difference among strains depending on the hosts and the regions where the samples were collected. For instance, more unique domains were revealed in *Bubalus bubalis* (buffalo) when compared to *Bos taurus* (cattle), *Equus caballus* (horse) and *Camelus ferus* (camel).

One of the most abundant domains found in strains of biovar *equi* (Figure 5) were related to (CRISPR) associated proteins - Clustered Regularly Interspaced Short Palindromic Repeats; among them, Cas_Cas2CT1978 and Cas_Cas1 domain were identified in 25 strains of biovar *equi*, see Table S3ii. Other frequent unique domains in biovar *equi* are acetyl/methyl transferase (GCD14 and Acetyltransf_9 domain), molybdopterin (Mob_synth_C domain), and phage (Phage_capsid domain); these domains were detected in 25 strains. The proteins harbour these domains are important for maintaining cell metabolism and defence mechanisms, for example, CRISPR associated proteins play a crucial role in the antiviral protection system of prokaryotes (Barrangou, 2015). The GCD14 domain is part of the tRNA methyltransferase complex (Anderson et al., 1998) and, Acetyltransf_9 domain is involved in transferase activities like chloramphenicol and amino acid acetylation (Caesar et al., 2006; Shaw et al., 1979). Molybdopterin is a cofactor and plays an essential role in the function of molybdoenzymes, such as sulfite oxidase, nitrate reductase, and dimethyl sulfoxide reductase, all of which play indispensable roles in bacterial metabolism (Feirer & Fuqua, 2017). The biological function of bacterial lipoproteins is mainly ascribed to bacterial cell growth and triggering the host innate immune response (Nakayama et al., 2012).

Interestingly, we observe the occurrence of many phage domains in the strains of biovar *equi* (Phage_capsid, Phage_int_SAM_5, Phage_r1t_holin, Phage_CI_repr, Sipho_tail, Phage_portal, ANT, Phage_int_SAM_3, and Phage_GP20); it can be due to lysogenic phage infecting the strains (Table S3ii–iii). It is already demonstrated that bacteriophage might to control bacterial virulence since several toxin genes are phage encoded (Wagner & Waldor, 2002). Furthermore, phages can alter host bacterial properties relevant to all stages of the infectious process, including bacterial adhesion, colonization, invasion, spread through host tissues, resistance to immune defenses, exotoxin production, sensitivity to antibiotics, and transmissibility among hosts (Wagner & Waldor, 2002). The variation in the phage domain frequency in strains of biovar *equi* might influence the pathogen virulence (Table S3ii–iii). Accordingly, one of the characteristics of biovar *equi* strains is the reduction of nitrogen (Selim, 2001) and respiratory nitrate reductase domains (Nitr_red_bet_C and Nitr_red_alph_N) were identified in 24 strains of biovar *equi*. The domain abundances of unique domains in strains of biovar *equi* is shown in Table S3ii.

Geographical location seems to play an important role when comparing unique domains of biovar *equi* strains (Figure 5). For instance, all strains from *B. bubalis* isolated in Egypt present a very similar domain context; the same for strains of *E. caballus* isolated in Mexico. On the other hand, strains isolated from the same host, but in different countries

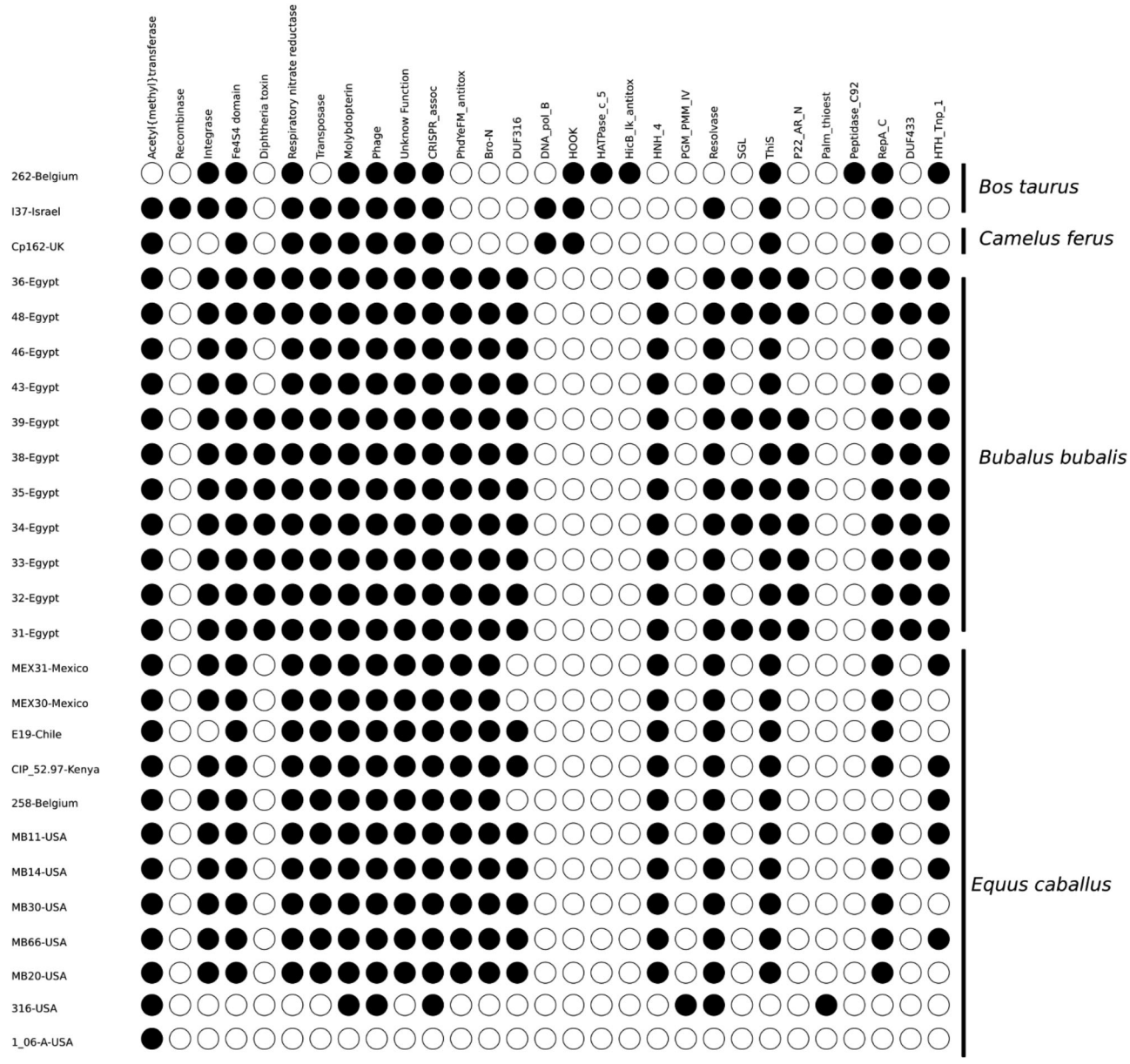

Figure 5. Unique domains of biovar *equi*. The distribution of biovar *equi* unique domains according to the strains, country and hosts. A black circle represents the presence of a domain in a given strain, while a white circle denotes the absence; strains are grouped by host.

present remarkable differences, compare *B. taurus* strains isolated in Belgium and Israel, and *E. caballus* strains isolated in USA, Mexico, Chile and Kenya (Figure 5).

In *B. taurus*, the strain 262-Belgium presents fewer domains compared to strain I37-Israel. In 262-Belgium strain, we highlighted the absence of Transposase and Acetyl methyltransferase, the most abundant domains uniquely detected in biovar *equi*. In *E. caballus*, we highlighted the strains 1_06-A isolated in USA, on which only one group of unique domains, those associated to Acetylmethyltransferase function, was observed. Intriguing, 1_06-A presents the fewest number of proteins suggesting an important loss of genes during the evolution. The strain 316 isolated from *E. caballus* also presents a different domain context when compared to other strains of the same host. Note that any unique domain was found in strains 29156 and I19 both isolated from *B. taurus* in Israel. The great diversity of unique domains in biovar *equi*, when compared to biovar *ovis*, could be because unique domains belong to unique proteins. For instance, the 68 unique domains belong to 159 proteins exclusively found in strains of biovar *equi*. In addition, these 159 proteins do not have any orthologous with strains of biovar *ovis*.

Only nine strains of biovar *ovis* presented unique domains (Figure 6 and Table S3i). Unlike the biovar *equi*, domain patterns in biovar *ovis* seem to be related more to the hosts. For strains isolated from *O. aries*, two domain patterns were observed, strains E55 and E56 both isolated in Egypt, and MEX29 in Mexico, share the domain DUF2326 with unknown function, while strains N1 isolated in Equatorial Guinea and PA01 in Brazil, share the domain Oxidored_q5_N defined in Pfam as oxidoreductase (NADH-ubiquinone oxidoreductase). The sodium-pumping NADH ubiquinone oxidoreductase (Na+-NQR) is the principal ion pump and the primary entry site for electrons into the

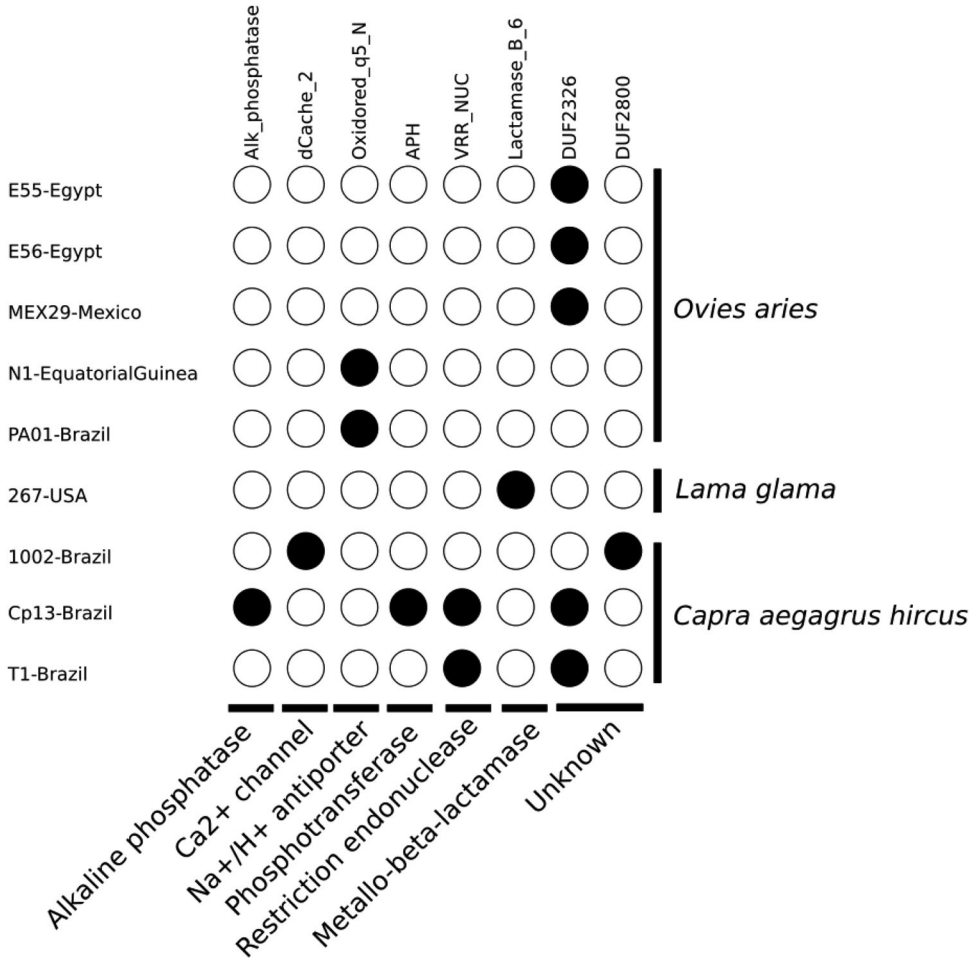

Figure 6. Unique domains of biovar *Ovis*. The distribution of biovar *ovis* unique domains according to the strains, country and hosts. A black circle represents the presence of a domain in a given strain, while a white circle denotes the absence; strains are grouped by host.

respiratory chain of many different types of pathogenic bacteria. This enzymatic complex creates a transmembrane gradient of sodium that is used by the cell to sustain ionic homeostasis, nutrient transport, ATP synthesis, flagellum rotation and other essential processes (Reyes-Prieto et al., 2014).

No unique domain was detected in strains of biovar *ovis* isolated in Argentina, Australia, Scotland and South Africa. For the host *L. glama* just one unique domain was observed, which is associated to the Metallo-beta-lactamase function, in the strain 267, isolated in USA. Metallo-β-lactamases are bacterial enzymes that catalyses the hydrolysis of β-lactam antibiotics like: penicillins, cephalosporins, carbapenems, and monobactams. Consequently, these proteins provide antibiotic resistance for the bacteria (Palzkill, 2013). From the 12 strains isolated from *C. aegagrus hircus* only three strains, which were isolated in Brazil, presented unique domains. However, they present a very different domain pattern with only two domains in common: DUF2326, marked as unknown function, and VRR_NUC that is associated with members of the PD-(D/E)XK nuclease superfamily, which constitute a large and diverse superfamily of enzymes involved in numerous nucleic acid cleavage events essential for various cellular processes (Kinch et al., 2005).

Conclusion

Corynebacterium pseudotuberculosis is a veterinary pathogenic bacterium with a worldwide distribution and considerable economic importance. The bacterium has been identified in Europe, Australia, North and South America, Africa, and the Middle East. It infects a variety of hosts, including camelids, cattle, horses, buffaloes, wild ruminants and humans. Once successfully established within the host, this pathogen will easily evade the immune system causing chronic infections in most cases. *C. pseudotuberculosis* strains are classified into two biovars depending on host preference and nitrate reduction. Biovar *ovis* (nitrate negative) causes CLA, a disease that affects the lymphatic system, resulting in abscesses in the lymph nodes and internal organs. Biovar *equi* (nitrate positive) causes many infections, such as ulcerative lymphangitis, which mainly affects horses, resulting in significant losses to agribusinesses. The availability of the genome for several strains of both biovars opens new opportunities to identify novel gene sequences (functional domains) that may encode virulence factors. Here we have conducted an original pan-genome study centered in domain comparisons rather than in the whole genome data as usually done in classical pan-genome studies. Domains are often linked to protein function, and several works demonstrated a close relationship between domain content and functional repertoire of a

genome. We have compared domain functional annotations of 28 strains of biovar *equi* against 25 strains of biovar *ovis*, focusing on domain conservation/variations and domain content of all proteins in both biovars.

We observed a more significant variation in the number of proteins per strain in biovar *equi* (Figure 3(A) and Table S1) probably indicating more gain or loss gene events. 88% of the coding genes of both biovars contained one or more Pfam domains; 30% of proteins are multi-domain (Figure 3(B)), while single domain proteins were the most abundant protein type with a percentage of 58% (Figure 3(C)); no-domain proteins account for 12% of proteins on average (Figure 3(D)). Strains of biovar *equi* present a higher number of biovar-specific (unique) domains, 77 against only 8 in biovar *ovis* (Figures 5 and 6), in biovar *ovis,* unique domains were simultaneously found in few strains (just nine), while in biovar *equi* half part of unique domains (39/77 ∼ 50%) was constantly found in more than 20 strains, see line 44 in S3ii. For both biovars, the number of unique domains varies according to the host type and their geographic locations. In strains of biovar *equi*, most of biovar-specific domains play a role in the virulence mechanism and immune response such as CRISPR, phage domains, GCD14, etc. In biovar *ovis*, the most observed unique domain was DUF2326, an uncharacterized domain detected in five strains.

Domain characterisation of biovar *equi* and biovar *ovis* is important for achieving a better understanding of *C. pseudotuberculosis*, once it might pinpoint relevant features not yet identified. This characteristic enables different strains to survive in diverse environmental niches. Our results could provide a better understanding of this organism and its mechanisms of virulence and pathogenesis, as well as to help in develop new diagnoses, vaccines, and treatments.

Material and methods

In this section, we describe methods and tools used to perform a pan-genomic analysis based on functional domains of *C. pseudotuberculosis* strains (biovars *ovis* and *equi*). Such analyses estimate the functional domain content of *C. pseudotuberculosis* strains. For that, we first downloaded a set of strains genome sequences isolated from different hosts in several countries. Next, we performed sophisticated annotation tools to obtain as many as possible domain regions for each protein of all strains. To enrich the analyses, we also identified orthologous groups and the most specific Gene Ontology terms.

Genome sequences

The complete genomes of 53 *C. pseudotuberculosis* strains were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/genbank/ in January 2017); 25 strains belong to biovar *ovis* and 28 to biovar *equi*. Strains of biovar *ovis* were isolated from *Ovies aries* (sheep), *Capra aegagrus hircus* (goat), *Lama glama* (llama), and *Connochaetes taurinus* (wildebeest), while strains of biovar *equi* from *Equus caballus* (horse), *Camelus ferus* (camel), *Bos taurus* (cattle), and *Bubalus bubalis* (buffalo). The strains originate from several countries: USA, Mexico, Chile, Argentina, Brazil, Israel, Egypt, Kenya, Equatorial Guinea, Belgium, Portugal, Scotland, Norway, and Australia. Table 1 shows host, country, and number of proteins for each strain.

Domain annotation

Domain fusion/shuffling is one of the most critical events in the evolution of proteins (Ye & Godzik, 2004). The majority of proteins (especially in complex organisms) contain multiple domains (modules); Domains are sequence fragments that can be independently stable and folded. They have a shape, a function, can occur alone or in groups, and are the building blocks of all proteins. The function of a protein to a large extent structure is determined by the arrangement of its constituent domains, termed as domain architecture. Thus, domain identification is crucial for protein function analyses. Several databases group domain types into families. The most popular are Pfam (El-Gebali et al., 2019), SMART (Schultz et al., 1998), PRODOM (Servant et al., 2002), and INTERPRO (Mulder et al., 2003). From this, domains annotation tools use sophisticated inference methods such as probabilistic models and machine learning to detect potential domains in proteins. Here we used CLADE (Bernardes, Zaverucha, et al., 2016), a sensitive tool for domain annotation. CLADE is a multi-source approach where several hundred probabilistic models are used to represent each domain in the Pfam database (El-Gebali et al., 2019), instead of just a single model as traditional annotation methods. CLADE provides domains annotations based on the Pfam database, version 30, containing 16306 different domain families, was used in this work. In order to reduce the false predictions, CLADE computes a specific E-value cut-off for each Pfam domain family rather than a fixed cut-off. Specific e-values are computed for minimizing the false discovery rate. This strategy generates better results by allowing filter out a large number of false positives due to suitable domain specific cut-offs. To analyse the 53 *C. pseudotuberculosis* strains, we first identify their unique sequences, a total of 12582 proteins. Unique sequences were then scanned by CLADE to identify a set of potential Pfam domains. Since a protein sequence might contain overlapped domains, we used DAMA (Bernardes, Vieira, et al., 2016), a multi-objective optimization approach, to provide the most likely domain combination or domain architecture for each unique protein.

Gene ontologies

The gene ontology (GO) initiative (Gene Ontology Consortium, 2007) maintains and develops a controlled vocabulary of gene and gene product attributes. The gene ontology terms (GO terms) (Ashburner et al., 2000) is a machine-readable vocabulary that provides a standard output for functional predictions, avoiding the ambiguity of natural language. GO terms describe three aspects of gene product function: molecular function, biological process, and cellular location. Since GO terms are hierarchical, we only

retained the most specific ones; those placed on the leaves of the GO tree. Specific GO terms were only retained if their hierarchical level is above three. If multiple GO terms with identical hierarchy were found, all were kept. To obtain only the most significant terms, we retained those with large deviations, that is, its frequency is at least three standard deviations greater than the arithmetic average of all considered specific GO terms.

Orthology information

The software EDGAR (version 2.3) (Blom et al., 2009) was used to compare multiple-strain genomes providing orthology information for *C. pseudotuberculosis* strains. EDGAR uses a specific BLAST (Altschul et al., 1997) cut off that is automatically adjusted to detect orthologous groups in a set of analyzed genomes. We directly downloaded orthologous groups for the 53 strains of *C. pseudotuberculosis* from EDGAR web interface, located at http://edgar.cebitec.uni-bielefeld.de.

Phylogenetic tree construction

We built a phylogenetic tree for all species of *Corynebacterium* genus, including the 53 strains of *C. pseudotuberculosis*. To end this, we used Gegenees (version 1.1.4) (Ågren et al., 2012) to construct a distance matrix for 168 *Corynebacterium* species, downloaded from NCBI ftp site, see the list of species and amount of strains in Table S4. First, Gegenees divided each genome into small sequences, and performed an all-versus-all similarity search to determine the minimum content shared among all genomes. Then, the minimum content was subtracted from each genome to obtain the variable contents. Next, the variable contents were compared to generate the percentage of similarity for each genome pair and derive a distance matrix. Finally, the distance matrix was used as an input for the SplitsTree (version 4.12.6) (Kloepper & Huson, 2008) to generate the phylogenetic tree using the UPGMA method (Unweighted Pair Grouping Methodwith Arithmetic-mean).

Author contributions

JSB and FRJV carried out the studies and performed the computational experiments. RJE and MAC analyzed and interpreted the results. JSB and MAC wrote the manuscript, FRJV and MAC designed and supervised the project. All authors read and approved the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by grants from FAPESP (2016/12904-0, 2018/07572-3).

ORCID

Juliana S. Bernardes http://orcid.org/0000-0003-1341-4256
Raphael J. Eberle http://orcid.org/0000-0002-8763-3884
Mônika A. Coronado http://orcid.org/0000-0002-6518-6497

References

Achard, M. E. S., Stafford, S. L., Bokil, N. J., Chartres, J., Bernhardt, P. V., Schembri, M. A., Sweet, M. J., & McEwan, A. G. (2012). Copper redistribution in murine macrophages in response to salmonella infection. *The Biochemical Journal*, *444*(1), 51–57. https://doi.org/10.1042/BJ20112180

Achard, M. E. S., Tree, J. J., Holden, J. A., Simpfendorfer, K. R., Wijburg, O. L. C., Strugnell, R. A., Schembri, M. A., Sweet, M. J., Jennings, M. P., & McEwan, A. G. (2010). The multi-copper-ion oxidase CueO of salmonella enterica serovar typhimurium is required for systemic virulence. *Infection and Immunity*, *78*(5), 2312–2319. https://doi.org/10.1128/IAI.01208-09

Ågren, J., Sundström, A., Håfström, T., & Segerman, B. (2012). Gegenees: Fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PloS One*, *7*(6), e39107. https://doi.org/10.1371/journal.pone.0039107

Almeida, S., Tiwari, S., Mariano, D., Souza, F., Jamal, S. B., Coimbra, N., Raittz, R. T., Dorella, F. A., de Carvalho, A. F., Pereira, F. L., Soares de, C., Leal, C. A., Barh, D., Ghosh, P., Figueiredo, H., Moura-Costa, L. F., Portela, R. W., Meyer, R., Silva, A., & Azevedo, V. (2016). The genome anatomy of *Corynebacterium pseudotuberculosis* VD57 a highly virulent strain causing caseous lymphadenitis. *Standards in Genomic Sciences*, *11*, 29. https://doi.org/10.1186/s40793-016-0149-7

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Anderson, J., Phan, L., Cuesta, R., Carlson, B. A., Pak, M., Asano, K., Bjork, G. R., Tamame, M., & Hinnebusch, A. G. (1998). The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes & Development*, *12*(23), 3650–3662. https://doi.org/10.1101/gad.12.23.3650

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Baird, G. J., & Fontaine, M. C. (2007). *Corynebacterium pseudotuberculosis* and its role in ovine caseous lymphadenitis. *Journal of Comparative Pathology*, *137*(4), 179–210. https://doi.org/10.1016/j.jcpa.2007.07.002

Barakat, A. A., Selim, S. A., Atef, A., Saber, M. S., Nafie, E. K., & EL-Edeeby, A. ;A. (1984). Two serotypes of *Corynebacterium pseudotuberculosis* isolated from different animal species. *Revue Scientifique et Technique de L'oie*, *3*(1), 151–163. https://doi.org/10.20506/rst.3.1.147

Barrangou, R. (2015). The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Current Opinion in Immunology*, *32*, 36–41. https://doi.org/10.1016/j.coi.2014.12.008

Bassano, B., Peracino, V., Bossi, D., Schr€Oder, C., & Guarda, F. (1993). Vertebral osteomyelitis with medullary compression in chamois. *Journal of Mountain Ecology*, *1*, 31e33.

Bernardes, J., Zaverucha, G., Vaquero, C., & Carbone, A. (2016). Improvement in protein domain identification is reached by breaking consensus, with the agreement of many profiles and domain co-occurrence. *PLoS Computational Biology*, *12*(7), e1005038. https://doi.org/10.1371/journal.pcbi.1005038

Bernardes, J. S., Vieira, F. R. J., Zaverucha, G., & Carbone, A. (2016). A multi-objective optimization approach accurately resolves protein domain architectures. *Bioinformatics (Oxford, England)*, *32*(3), 345–353. https://doi.org/10.1093/bioinformatics/btv582

Biberstein, E. L., Knight, H. D., & Jang, S. (1971). Two biotypes of *Corynebacterium pseudotuberculosis*. *The Veterinary Record*, *89*(26), 691–692. https://doi.org/10.1136/vr.89.26.691

Blom, J., Albaum, S. P., Doppmeier, D., Pühler, A., Vorhölter, F. J., Zakrzewski, M., & Goesmann, A. (2009). Edgar: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics*, *10*, 154. https://doi.org/10.1186/1471-2105-10-154

Boomker, J., & Henton, M. M. (1980). Pseudotuberculosis in a cheetah (*Acinonyx jubatus*). *South African Journal of Wildlife Research*, *10*, 63e66.

Britz, E., Spier, S. J., Kass, P. H., Edman, J. M., & Foley, J. E. (2014). The relationship between *Corynebacterium pseudotuberculosis* biovar equi phenotype with location and extent of lesions in horses. *The Veterinary Journal*, *200*(2), 282–286. https://doi.org/10.1016/j.tvjl.2014.03.009

Brown, J. S., & Holden, D. W. (2002). Iron acquisition by gram-positive bacterial pathogens. *Microbes* and Infection, *4*(11), 1149–1156. https://doi.org/10.1016/S1286-4579(02)01640-4

Burroughs, A. M., Balaji, S., Iyer, L. M., & Aravind, L. (2007). A novel superfamily containing the beta-grasp fold involved in binding diverse soluble ligands. *Biology Direct*, *2*, 4. https://doi.org/10.1186/1745-6150-2-4

Caesar, R., Warringer, J., & Blomberg, A. (2006). Physiological importance and identification of novel targets for the n-terminal acetyltransferase NatB. *Eukaryotic Cell*, *5*(2), 368–378. https://doi.org/10.1128/EC.5.2.368-378.2006

Cerdeno-Tarraga, A. M., Efstratiou, A., Dover, L. G., Holden, M. T. G., Pallen, M., Bentley, S. D., Besra, G. S., Churcher, C., James, K. D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M. A., … Parkhill, J. (2003). The complete genome sequence and analysis of *Corynebacterium diphtheriae* nctc13129. *Nucleic Acids Research*, *31*(22), 6516–6523. https://doi.org/10.1093/nar/gkg874

Clark, K. A., Robinson, R. M., Weishuhn, L. L., Litton, G. W., & Marburger, R. G. (1972). Caseous lymphadenitis in pronghorns (*Antilocapra americana*). *Journal of Wildlife Diseases*, *8*(1), 67–71. https://doi.org/10.7589/0090-3558-8.1.67

Colom-Cadena, A., Velarde, R., Salinas, J., Borge, C., García-Bocanegra, I., Serrano, E., Gassó, D., Bach, E., Casas-Díaz, E., López-Olvera, J. R., Lavín, S., León-Vizcaíno, L., & Mentaberre, G. (2014). Management of a caseous lymphadenitis outbreak in a new Iberian ibex (*Capra pyrenaica*) stock reservoir. *Acta Veterinaria Scandinavica*, *56*(1), 1e11. https://doi.org/10.1186/s13028-014-0083-x

Cook, L., Nagi, M. N., Suneja, S. K., Hand, A. R., & Cinti, D. L. (1992). Evidence that β-hydroxyacyl-CoA dehydrase purified from rat liver microsomes is of peroxisomal origin. *Biochemical Journal*, *287*(1), 91–100. https://doi.org/10.1042/bj2870091

Davidson, A. L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews: MMBR*, *72*(2), 317–364. https://doi.org/10.1128/MMBR.00031-07

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure (London, England: 1993)*, *3*(9), 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9

Domenis, L., Spedicato, R., Pepe, E., Orusa, R., & Robetto, S. (2018). Caseous lymphadenitis caused by *Corynebacterium pseudotuberculosis* in Alpine Chamois (*Rupicapra r. rupicapra*): A review of 98 cases. *Journal of Comparative Pathology*, *161*, 11–19. https://doi.org/10.1016/j.jcpa.2018.04.003

Dorella, F. A., Pacheco, L. G. C., Oliveira, S. C., Miyoshi, A., & Azevedo, V. (2006). *Corynebacterium pseudotuberculosis*: Microbiology, biochemical properties, pathogenesis and molecular studies of virulence. *Veterinary Research*, *37*(2), 201–218. https://doi.org/10.1051/vetres:2005056

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Feirer, N., & Fuqua, C. (2017). Pterin function in bacteria. *Pteridines*, *28*(1), 23–36. https://doi.org/10.1515/pterid-2016-0012

Foley, J. E., Spier, S. J., Mihalyi, J., Drazenovich, N., & Leutenegger, C. M. (2004). Molecular epidemiologic features of *Corynebacterium pseudotuberculosis* isolated from horses. *American Journal of Veterinary Research*, *65*(12), 1734–1737. https://doi.org/10.2460/ajvr.2004.65.1734

Gene Ontology Consortium. (2007). The gene ontology project in 2008. *Nucleic Acids Research*, *36*(Database issue), D440–D444.

Gwynn, E. J., Smith, A. J., Guy, C. P., Savery, N. J., McGlynn, P., & Dillingham, M. S. (2013). The conserved c-terminus of the PcrA/UvrD helicase interacts directly with RNA polymerase. *PLoS One*, *8*(10), e78141. https://doi.org/10.1371/journal.pone.0078141

Hawari, A. D. (2008). *Corynebacterium pseudotuberculosis* infection (caseous lymphadenitis) in camels (*Camelus dromedarius*) in Jordan. *American Journal of Animal and Veterinary Sciences*, *3*(3), 68–72. https://doi.org/10.3844/ajavsp.2008.68.72

Heggelund, L., Gaustad, P., Havelsrud, O. E., Blom, J., Borgen, L., Sundset, A., Sørum, H., & Frøland, S. S. (2015). *Corynebacterium pseudotuberculosis* pneumonia in a veterinary student infected during laboratory work. *Open Forum Infectious Diseases*, *2*, ofv053.

Henrissat, B., & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*, *316*(2), 695–696. https://doi.org/10.1042/bj3160695

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., & Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(15), 7090–7094. https://doi.org/10.1073/pnas.92.15.7090

Heung, L. J., & Del Poeta, M. (2005). Unlocking the dead-box: A key to cryptococcal virulence? *The Journal of Clinical Investigation*, *115*(3), 593–595. https://doi.org/10.1172/JCI24508

Higgins, C. F., Hyde, S. C., Mimmack, M. M., Gileadi, U., Gill, D. R., & Gallagher, M. P. (1990). Binding protein-dependent transport systems. *Journal of Bioenergetics and Biomembranes*, *22*(4), 571–592. https://doi.org/10.1007/BF00762962

Jane Kelly, E., Rood, K. A., & Skirpstunas, R. (2012). Abscesses in captive elk associated with *Corynebacterium pseudotuberculosis*, Utah, USA. *Journal of Wildlife Diseases*, *48*(3), 803–805. https://doi.org/10.7589/0090-3558-48.3.803

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Krämer, R., Linke, B., McHardy, A. C., Meyer, F., Möckel, B., Pfefferle, W., … Tauch, A. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of l-aspartate-derived amino acids and vitamins. *Journal of Biotechnology*, *104*(1–3), 5–25. https://doi.org/10.1016/s0168-1656(03)00154-8

Kemp, B. E. (2004). Bateman domains and adenosine derivatives form a binding contract. *Journal of Clinical Investigation*, *113*(2), 182–184. https://doi.org/10.1172/JCI200420846

Kinch, L. N., Ginalski, K., Rychlewski, L., & Grishin, N. V. (2005). Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acids Research*, *33*(11), 3598–3605. https://doi.org/10.1093/nar/gki676

Kloepper, T. H., & Huson, D. H. (2008). Drawing explicit phylogenetic networks and their integration into splitstree. *BMC Evolutionary Biology*, *8*, 22. https://doi.org/10.1186/1471-2148-8-22

Lee, J. W., & Helmann, J. D. (2007). Functional specialization within the Fur family of metalloregulators. *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine*, *20*(3–4), 485–499. https://doi.org/10.1007/s10534-006-9070-7

Matos, A. C., Dias, A. P., Morais, M., Figueira, L., Martins, M. H., Matos, M., Pinto, M. L., & Coelho, A. C. (2015). Granuloma co-infection with *Mycobacterium bovis*, *Mycobacterium avium subsp. paratuberculosis*, and *Corynebacterium pseudotuberculosis* in five hunted red deer (Cervus elaphus) in Portugal. *Journal of Wildlife Diseases*, *51*(3), 793–794. https://doi.org/10.7589/2014-09-240

McKean, S. C., Davies, J. K., & Moore, R. J. (2007). 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 (Reading, England)*, *153*(Pt 7), 2203–2211. https://doi.org/10.1099/mic.0.2007/005926-0

Morales, N., Aldridge, D., Bahamonde, A., Cerda, J., Araya, C., Muñoz, R., Saldías, M. E., Lecocq, C., Fresno, M., Abalos, P., & Retamal, P. (2017). *Corynebacterium pseudotuberculosis* infection in Patagonian Huemul (*Hippocamelus bisulcus*). *Journal of Wildlife Diseases*, *53*(3), 621–624. https://doi.org/10.7589/2016-09-213

Müller, B., de Klerk-Lorist, L.-M., Henton, M. M., Lane, E., Parsons, S., Gey van Pittius, N. C., Kotze, A., van Helden, P. D., & Tanner, M. (2011). Mixed infections of *Corynebacterium pseudotuberculosis* and non-tuberculous mycobacteria in South African antelopes presenting with tuberculosislike lesions. *Veterinary Microbiology*, *147*(3–4), 340–345. https://doi.org/10.1016/j.vetmic.2010.07.017

Mulder, N. M., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., … Zdobnov, E. M. (2003). The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Research*, *31*(1), 315–318. https://doi.org/10.1093/nar/gkg046

Nakayama, H., Kurokawa, K., & Lee, B. L. (2012). Lipoproteins in bacteria: Structures and biosynthetic pathways. *The FEBS Journal*, *279*(23), 4247–4268. https://doi.org/10.1111/febs.12041

Oliveira, A., Teixeira, P., Azevedo, M., Jamal, S. B., Tiwari, S., Almeida, S., Silva, S., Barh, D., Dorneles, E. M. S., Haas, D. J., Heinemann, M. B., Ghosh, P., Lage, A. P., Figueiredo, H., Ferreira, R. S., & Azevedo, V. (2016). *Corynebacterium pseudotuberculosis* may be under anagenesis and biovar equi forms biovar ovis: A phylogenic inference from sequence and structural analysis. *BMC Microbiology*, *16*, 100. https://doi.org/10.1186/s12866-016-0717-4

Palzkill, T. (2013). Metallo-β-lactamase structure and function. *Annals of the New York Academy of Sciences*, *1277*, 91–104. https://doi.org/10.1111/j.1749-6632.2012.06796.x

Peel, M. M., Palmer, G. G., Stacpoole, A. M., & Kerr, T. G. (1997). Human lymphadenitis due to *Corynebacterium pseudotuberculosis*: Report of ten cases from Australia and review. *Clinical Infectious Diseases*, *24*(2), 185–191. https://doi.org/10.1093/clinids/24.2.185

Perez, L., León, L., Cubero, P., & Ganz_Alez, M. (1996). A propos de deux cases de pseudotuberculosis chez le daime (Dama dama). 14eme Reunion du Groupe d'Etudes sur l'Ecopathologie de la faune Sauvage de Montagne.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I., & Saier, M. H. (2012). The major facilitator superfamily (MFS) revisited. *The FEBS Journal*, *279*(11), 2022–2035. https://doi.org/10.1111/j.1742-4658.2012.08588.x

Reyes-Prieto, A., Barquera, B., & Juarez, O. (2014). Origin and evolution of the sodium -pumping NADH: ubiquinone oxidoreductase. *PLoS One*, *9*(5), e96696https://doi.org/10.1371/journal.pone.0096696

Rodriguez, G. M., & Smith, I. (2003). Mechanisms of iron regulation in mycobacteria: Role in physiology and virulence. *Molecular Microbiology*, *47*(6), 1485–1494. https://doi.org/10.1046/j.1365-2958.2003.03384.x

Roth, H. H., & Vickers, D. B. (1966). Pseudotuberculosis in an antbear (*Orycteropus afer*) in Rhodesia. *British Veterinary Journal*, *122*(7), 296–300. https://doi.org/10.1016/S0007-1935(17)40506-9

Rowland, J. L., & Niederweis, M. (2013). A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. *Journal of Bacteriology*, *195*(16), 3724–3733. https://doi.org/10.1128/JB.00546-13

Sá, M. C. A., Gouveia, G. V., Krewer, C. d C., Veschi, J. L. A., de Mattos-Guaraldi, A. L., & da Costa, M. M. (2013). Distribution of PLD and FagA, B, C and D genes in *Corynebacterium pseudotuberculosis* isolates from sheep and goats with caseus lymphadenitis. *Genetics and Molecular Biology*, *36*(2), 265–268. https://doi.org/10.1590/S1415-47572013005000013

Santana-Jorge, K. T. O., Santos, T. M., Tartaglia, N. R., Aguiar, E. L., Souza, R. F. S., Mariutti, R. B., Eberle, R. J., Arni, R. K., Portela, R. W., Meyer, R., & Azevedo, V. (2016). Putative virulence factors of *Corynebacterium pseudotuberculosis* FRC41: Vaccine potential and protein expression. *Microbial Cell Factories*, *15*, 83. https://doi.org/10.1186/s12934-016-0479-6

Schroder, J., Jochmann, N., Rodionov, D. A., & Tauch, A. (2010). The Zur regulon of *Corynebacterium glutamicum* ATCC 13032. *BMC Genomics*, *11*, 12. https://doi.org/10.1186/1471-2164-11-12

Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P. (1998). Smart, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(11), 5857–5864. https://doi.org/10.1073/pnas.95.11.5857

Selim, S. A. (2001). Oedematous skin disease of buffalo in Egypt. *Journal of Veterinary Medicine. B, Infectious Diseases and Veterinary Public Health*, *48*(4), 241–258. https://doi.org/10.1046/j.1439-0450.2001.00451.x

Servant, F., Bru, C., Carrere, S., Courcelle, E., Gouzy, J., Peyruc, D., & Kahn, D. (2002). ProDom: Automated clustering of homologous domains. *Briefings in Bioinformatics*, *3*(3), 246–251. https://doi.org/10.1093/bib/3.3.246

Shaw, W. V., Packman, L. C., Burleigh, B. D., Dell, A., Morris, H. R., & Hartley, B. S. (1979). Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. *Nature*, *282*(5741), 870–872. https://doi.org/10.1038/282870a0

Silinski, S., Walzer, C. (2004). Caseous lymphadenitis in a captive group of alpine ibex (*Capra ibex*) e management and implications. *Proceeding of the 5th Scientific Meeting of the European Association of Zoo and Wildlife Veterinarians* (p. 321).

Snipen, L. G., & Ussery, D. W. (2012). A domain sequence approach to pangenomics: applications to *Escherichia coli*. *F1000Research*, *1*, 19. https://doi.org/10.12688/f1000research.1-19.v1

Soares, S. C., Silva, A., Trost, E., Blom, J., Ramos, R., Carneiro, A., Ali, A., Santos, A. R., Pinto, A. C., Diniz, C., Barbosa, E. G., Dorella, F. A., Aburjaile, F., Rocha, F. S., Nascimento, K. K., Guimarães, L. C., Almeida, S., Hassan, S. S., Bakhtiar, S. M., … Azevedo, V. (2013). The pan-genome of the animal pathogen *Corynebacterium pseudotuberculosis* reveals differences in genome plasticity between the biovar ovis and equi strains. *PLoS One*, *8*(1), e53818. https://doi.org/10.1371/journal.pone.0053818

Spier, S. J., & Azevedo, V. (2017). *Corynebacterium pseudotuberculosis* infection in horses: Increasing frequency and spread to new regions of North America. *Equine Veterinary Education*, *29*(8), 436–439. https://doi.org/10.1111/eve.12589

Stauber, E., Armstrong, P., Chamberlain, K., & Gorgen, B. (1973). Caseous lymphadenitis in a white-tailed deer. *Journal of Wildlife Diseases*, *9*(1), 56–57. https://doi.org/10.7589/0090-3558-9.1.56

Staudenmaier, H., Van Hove, B., Yaraghi, Z., & Braun, V. (1989). Nucleotide sequences of the fecBCDE genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. *Journal of Bacteriology*, *171*(5), 2626–2633. https://doi.org/10.1128/jb.171.5.2626-2633.1989

Tarello, W., & Theneyan, M. (2008). *Corynebacterium pseudotuberculosis* and *Corynebacterium renale* isolated from two Arabian oryx (*Oryx leucoryx*). *The Veterinary Record*, *162*(26), 862–863. https://doi.org/10.1136/vr.162.26.862

Tauch, A., Kaiser, O., Hain, T., Goesmann, A., Weisshaar, B., Albersmeier, A., Bekel, T., Bischoff, N., Brune, I., Chakraborty, T., Kalinowski, J., Meyer, F., Rupp, O., Schneiker, S., Viehoever, P., & Pühler, A. (2005). Complete genome sequence and analysis of the multiresistant nosocomial pathogen *Corynebacterium jeikeium* K411, a lipid-requiring bacterium of the human skin flora. *Journal of Bacteriology*, *187*(13), 4671–4682. https://doi.org/10.1128/JB.187.13.4671-4682.2005

Tejedor-Junco, M. T., Lupiola, P., Schulz, U., & Gutierrez, C. (2008). Isolation of nitrate-reductase positive *Corynebacterium pseudotuberculosis* from dromedary camels. *Tropical Animal Health and Production*, *40*(3), 165–167. https://doi.org/10.1007/s11250-007-9077-2

Trost, E., Ott, L., Schneider, J., Schröder, J., Jaenicke, S., Goesmann, A., Husemann, P., Stoye, J., Dorella, F. A., Rocha, F. S., Soares, S. d C., D'Afonseca, V., Miyoshi, A., Ruiz, J., Silva, A., Azevedo, V., Burkovski, A., Guiso, N., Join-Lambert, O. F., Kayal, S., & Tauch, A. (2010). 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*, *11*, 728. https://doi.org/10.1186/1471-2164-11-728

Viana, M. C. V., Sahm, A., Neto, A. G., Figueiredo, H. C. P., Wattam, A. R., & Azevedo, V. (2018). Rapidly evolving changes and gene loss associated with host switching in *Corynebacterium pseudotuberculosis*. *PloS ONE*, *13*(11), e0207304. https://doi.org/10.1371/journal.pone.0207304

Viana, M. V. C., Figueiredo, H., Ramos, R., Guimaraes, L. C., Pereira, F. L., Dorella, F. A., Selim, S. A. K., Salaheldean, M., Silva, A., Wattam, A. R., & Azevedo, V. (2017). Comparative genomic analysis between *Corynebacterium pseudotuberculosis* strains isolated from buffalo. *PLoS One*, *12*(4), e0176347. https://doi.org/10.1371/journal.pone.0176347

Wagner, P. L., & Waldor, M. K. (2002). Bacteriophage control of bacterial virulence. *Infection and Immunity*, *70*(8), 3985–3993. https://doi.org/10.1128/iai.70.8.3985-3993.2002

Williamson, L. H. (2001). Caseous lymphadenitis in small ruminants. *Veterinary Clinics of North America: Food Animal Practice*, *17*(2), 359–371. https://doi.org/10.1016/s0749-0720(15)30033-5

Wolschendorf, F., Ackart, D., Shrestha, T. B., Hascall-Dove, L., Nolan, S., Lamichhane, G., Wang, Y., Bossmann, S. H., Basaraba, R. J., & Niederweis, M. (2011). Copper resistance is essential for virulence of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1621–1626. https://doi.org/10.1073/pnas.1009261108

Xu, T., Schiering, N., Zeng, H. Y., Ringe, D., & Murphy, J. R. (1994). Iron, DtxR, and the regulation of diphtheria toxin expression. *Molecular Microbiology*, *14*(2), 191–197. https://doi.org/10.1111/j.1365-2958.1994.tb01280.x

Ye, Y., & Godzik, A. (2004). Comparative analysis of protein domain organization. *Genome Research*, *14*(3), 343–353. https://doi.org/10.1101/gr.1610504

Yeruham, I., Braverman, Y., Shpigel, N. Y., Chizov-Ginzburg, A., Saran, A., & Winkler, M. (1996). Mastitis in dairy cattle caused by *Corynebacterium pseudotuberculosis* and the feasibility of transmission by houseflies I. *The Veterinary Quarterly*, *18*(3), 87–89. https://doi.org/10.1080/01652176.1996.9694623

Yeruham, I., Friedman, S., Perl, S., Elad, D., Berkovich, Y., & Kalgard, Y. (2004). A herd level analysis of a *Corynebacterium pseudotuberculosis* outbreak in a dairy cattle herd. *Veterinary Dermatology*, *15*(5), 315–320. https://doi.org/10.1111/j.1365-3164.2004.00388.x