# Evidence of episodic positive selection in *Corynebacterium diphtheriae* complex of species and its implementations in identification of drug and vaccine targets

Marcus Vinicius Canário Viana[1,2], Rodrigo Profeta[1], Janaína Canário Cerqueira[1], Alice Rebecca Wattam[3], Debmalya Barh[1,4], Artur Silva[2] and Vasco Azevedo[1]

[1] Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2] Departamento de Genética, Universidade Federal do Pará, Belém, Pará, Brazil
[3] Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, United States
[4] Institute of Integrative Omics and Applied Biotechnology, Nonakuri, West Bengal, India

## ABSTRACT

**Background:** Within the pathogenic bacterial species *Corynebacterium* genus, six species that can produce diphtheria toxin (*C. belfantii*, *C. diphtheriae*, *C. pseudotuberculosis*, *C. rouxii*, *C. silvaticum* and *C. ulcerans*) form a clade referred to as the *C. diphtheria* complex. These species have been found in humans and other animals, causing diphtheria or other diseases. Here we show the results of a genome scale analysis to identify positive selection in protein-coding genes that may have resulted in the adaptations of these species to their ecological niches and suggest drug and vaccine targets.

**Methods:** Forty genomes were sampled to represent species, subspecies or biovars of *Corynebacterium*. Ten phylogenetic groups were tested for positive selection using the PosiGene pipeline, including species and biovars from the *C. diphtheria* complex. The detected genes were tested for recombination and had their sequences alignments and homology manually examined. The final genes were investigated for their function and a probable role as vaccine or drug targets.

**Results:** Nineteen genes were detected in the species *C. diphtheriae* (two), *C. pseudotuberculosis* (10), *C. rouxii* (one), and *C. ulcerans* (six). Those were found to be involved in defense, translation, energy production, and transport and in the metabolism of carbohydrates, amino acids, nucleotides, and coenzymes. Fourteen were identified as essential genes, and six as virulence factors. Thirteen from the 19 genes were identified as potential drug targets and four as potential vaccine candidates. These genes could be important in the prevention and treatment of the diseases caused by these bacteria.

**Subjects** Bioinformatics, Genomics, Microbiology, Molecular Biology
**Keywords** Corynebacterium, Positive selection, Drug target, Vaccine target

Submitted 22 July 2021
Accepted 30 November 2021
Published 16 February 2022

Corresponding author
Vasco Azevedo, vasco@icb.ufmg.br

Academic editor
Joseph Gillespie

Additional Information and Declarations can be found on page 13

DOI **10.7717/peerj.12662**



**How to cite this article** Canário Viana MV, Profeta R, Cerqueira JC, Wattam AR, Barh D, Silva A, Azevedo V. 2022. Evidence of episodic positive selection in *Corynebacterium diphtheriae* complex of species and its implementations in identification of drug and vaccine targets. **PeerJ 10:e12662 DOI 10.7717/peerj.12662**

## INTRODUCTION

The genus *Corynebacterium* are gram-positive bacteria of biotechnological, medical, and veterinary importance (*Bernard & Funke, 2015*). Within the pathogenic species, some can produce diphtheria toxin (DT) after lysogenization by *tox*+ corynephages (*Bernard & Funke, 2015*). Three species that compose a clade were initially described as potential diphtheria toxin (DT) producers: *C. diphtheriae*, *C. ulcerans* and *C. pseudotuberculosis* (*Bernard & Funke, 2015*). The number of species in the clade of potential DT producers increased to six with the inclusion of the recently described *C. belfantii* (*Dazas et al., 2018*), *C. rouxii* (*Badell et al., 2020*) and *C. silvaticum* (*Dangel et al., 2020*). Those six species are described here as the "*C. diphtheria* complex" (*Badell et al., 2020*).

*C. diphtheriae*, *C. belfantii* and *C. rouxii* infect mainly humans (*Bernard & Funke, 2015*; *Dazas et al., 2018*; *Badell et al., 2020*). *C. ulcerans*, *C. pseudotuberculosis* and *C. silvaticum* infect mainly wild and domesticated mammals and/or can cause zoonosis (*Bernard & Funke, 2015*; *Dangel et al., 2020*). *C. belfantii* and *C. rouxii* have recently been reclassified species from some of the *C. diphtheriae* biovar Belfanti strains (*Dazas et al., 2018*; *Badell et al., 2020*). *C. ulcerans*, *C. pseudotuberculosis* and *C. silvaticum* infect mainly wild and domesticated mammals but can also be zoonotic (*Bernard & Funke, 2015*; *Dangel et al., 2020*).

The *C. diphtheria* complex have an impact on public health, and also on the production of animal-based foods. Some of the species contain both DT and strains that lack the toxin. DT-producing *C. diphtheriae* strains cause cutaneous and respiratory diphtheria (*Zasada, 2013*; *Grosse-Kock et al., 2017*). The report of multidrug-resistant strains from Brazil is a new concern (*Zasada, 2014*; *Hennart et al., 2020*). DT-producing *C. pseudotuberculosis* from biovar equi causes Oedematous Skin Disease in buffalos (*Selim et al., 2015*). *C. ulcerans* infects a broad range of mammal species and DT-producing strains have caused diphtheria (*Hacker et al., 2016*). Some non-DT producing strains of *C. diphtheriae* cause endocarditis, septic arthritis, osteomyelitis and sepsis in humans (*Zasada, 2013*; *Grosse-Kock et al., 2017*). Non-DT producing strains of *C. ulcerans* are associated with ulcers in humans (*Hacker et al., 2016*). *C. pseudotuberculosis* also contains non-DT strains, with those in biovar equi causing ulcerative lymphangitis in horses, and those in the biovar ovis causing caseous lymphadenitis in goat and sheep, and lymphadenitis and abscesses in humans (*Selim et al., 2015*).

There are also *C. diphtheria* complex species that never produce DT but do cause disease. *C. belfantii* causes laryngitis and bronchopathy (*Dazas et al., 2018*). *C. rouxii* causes chronic arteritis leading to ulcerations on feet and legs, and peritonitis (*Badell et al., 2020*). *C. silvaticum* has only been isolated from pigs and roe deer to date, causing caseous lymphadenitis (*Dangel et al., 2020*), and is cytotoxic for human epithelial cells (*Möller et al., 2021*).

The host ranges and virulence mechanisms of these species are not entirely known, and better understanding of their biology could be helpful in controlling this group of pathogens. Diphtheria outbreaks were reported globally between 1921 and 2018. The disease is still endemic in some countries, with thousands of annual cases reported in

Asia and Africa. The disease can emerge when the recommended vaccination programs are not applied or sustained (*Sharma et al., 2019*). The current vaccine is based on the DT toxoid (*Rappuoli & Malito, 2014*) but does not prevent the colonization, transmission, and disease manifestation. In addition, the acquired immunity has been found to decrease with time (*Truelove et al., 2020*). Isolation of symptomatic individuals, antitoxin and antibiotics are still essential in the control of these diseases (*Truelove et al., 2020*). Furthermore, the diversity of DT toxin sequences across strains could reduce the effectiveness of diphtheria toxoid–based vaccines and diphtheria antitoxins (*Otsuji et al., 2019*). Another factor to consider in the control of these pathogens is that non-DT producing strains can cause other diseases, such as ulcers and caseous lymphadenitis, the latter associated with the Phospholipase D toxin produced by *C. ulcerans*, *C. pseudotuberculosis* and *C. silvaticum* (*Bernard & Funke, 2015*; *Dangel et al., 2020*).

Adaptive mutations for a specific ecological niche can be identified using genomic analyses, including genome-scale positive selection analysis (*Kopac et al., 2014*). At an ecological level, routine selection favors the maintenance of a stable population structure over time, while episodic selection is the effect of a sudden environmental disturbance, such as host change (*Brasier, 1995*). At the molecular level, positive selection can help fix adaptive mutations (*Anisimova & Liberles, 2012*). Episodes of positive selection can act on specific codons at specific times (phylogenetic branches), for which branch-site statistical models were developed (*Zhang, 2005*). Information on the amino acids under selection could be used for drug design (*Farhat et al., 2013*), or even reverse vaccinology if the amino acids are surface exposed (*Goodswen, Kennedy & Ellis, 2018*).

In this work, we used a genome scale positive selection analysis to identify the genes that could be involved in ecological adaptation and identified genes that can be used to develop preventive or therapeutic strategies against this group of important pathogens.

## MATERIALS AND METHODS

### Samples and taxonomy

Episodic positive selection across species or other phylogenetic groups of the diphtheriae group was investigated using a branch-site test (*Zhang, 2005*). This test is more appropriate for inter-specific samples, because it assumes that the observed mutations have already been fixed by selection (*Kryazhimskiy & Plotkin, 2008*; *Anisimova & Liberles, 2012*; *Kosiol & Anisimova, 2012*), and one genome could represent a species. For this reason, we limited the samples to one per species, subspecies or biovar. For the foregrounds (target groups), we selected the six type strains from the *C. diphtheria* complex and other strains to represent biovars and lineages. *C. diphtheriae* biovars could not be used as foregrounds as they are united in a single clade (*Sangal et al., 2014*). As background, we selected 40 total genomes that included 30 representative (*O'Leary et al., 2016*), eight complete and three WGS genomes of *Corynebacterium* species, all of which were available in the Pathosystems Resource Integration Center (PATRIC) (*Davis et al., 2020*). These were annotated by RASTtk (*Brettin et al., 2015*) and downloaded from PATRIC (Table S1).

The taxonomy of the samples was verified using TYGS (*Meier-Kolthoff & Göker, 2019*). TYGS determines the closest related type strains using the MASH algorithm (*Ondov et al., 2016*) for entire genomes and BLASTn for 16S sequences. It calculates the pairwise distances of 16S and genome sequences using GBDP (*Meier-Kolthoff et al., 2013*), followed by inference of 16S and genome phylogenies based on the pairwise GBDP distances using FastME (*Lefort, Desper & Gascuel, 2015*), digital DNA-DNA hybridization (dDDH) using GGDC (*Meier-Kolthoff et al., 2013*), and deviation of G+C content. Genomes with >70% of dDDH and <1% of G+C content deviation are considered to be in the same species (*Meier-Kolthoff & Göker, 2019*).

## Positive selection analysis

A genome-scale positive selection analysis was performed using the PosiGene pipeline (*Sahm et al., 2017*) on the *Corynebacterium* genomes. Ten foreground genomes were tested (Table S2) representing 10 target clades or subclades. Six clades represent the species from the *C. diphtheria* complex (*C. belfantii*, *C. diphtheriae*, *C. pseudotuberculosis*, *C. rouxii*, *C. silvaticum* and *C. ulcerans*), two subclades represent *C. pseudotuberculosis* (biovars equi and ovis), and two subclades representing *C. ulcerans* lineages (lineage 1 and 2).

In the module "create_catalog", homologous genes were identified using BLASTp (*Camacho et al., 2009*) with the best-bidirectional hit criterion (*Altenhoff & Dessimoz, 2009*). In the module "alignments", orthologous genes are identified, gene trees are built, and a species tree is built from the gene trees. In this module, the anchor species is the genome that the orthologous gene sequences are aligned to, and the reference species is the genome from which the gene names are extracted. *C. diphtheriae* NCTC 11397$^{T}$ was selected as both the reference and anchor genome. In the first step, orthologous gene sequences were aligned to the anchor genome sequences using CLUSTALW (*Larkin et al., 2007*), with the parameters' minimum identity and minimum pairs identity set to 40%. The aligned sequences were filtered by GBLOCKS for gaps and unreliable alignment columns (*Jordan & Goldman, 2012*). In the next step, a phylogeny was built for each gene using the parsimony method and jackknifing implemented in DNAPARS from the PHYLIP package (*Felsenstein, 2005*). In the third step, a species tree was built based on the gene trees consensus, using CONSENSE from the PHYLIP package. The species tree is required to test for positive selection along specific lineages (*Yang & Nielsen, 2002*). The consensus tree was visualized using FigTree v1.4.4 and rooted using *C. kroppenstedtii* DSM 44385 as an outgroup. This strain was identified as an outgroup based on another tree generated by the same pipeline including *M. tuberculosis* H37RV to find the correct *Corynebacterium* species to use as an outgroup (Table S1) in that tree. The *M. tuberculosis* tree was not included in the downstream analysis.

In the module "positive_selection", a likelihood ratio test compares the non-synonymous to synonymous substitution rate $\omega = d_N/d_S$ in the foreground and the background. Here, $d_N$ is the number of non-synonymous substitutions per non-synonymous site and $d_S$ is the number of synonymous substitutions per synonymous site. The episodic positive selection model assumes $\omega > 1$ in the foreground and $\omega = 0$ or $\omega < 1$ in the background, while the null model assumes $\omega = 0$ or $\omega < 1$ for foreground and

background (*Yang & Dos Reis, 2011*). We considered positive selection if $p < 0.05$ for False Discovery Rate, as this correction is more suitable for genome wide experiments (*Storey & Tibshirani, 2003*).

Due to an assumption of no recombination by the branch-site models we used (*Yang, 2005*), recombination could cause false positive results (*Anisimova, Nielsen & Yang, 2003*). To avoid that artifact, the genes identified as positively selected by PosiGene were then tested for intragenic recombination using PhiPack (*Bruen, Philippe & Bryant, 2006*) that calculates Pairwise Homoplasy Index method (PHI) (*Bruen, Philippe & Bryant, 2006*), Neighbor Similarity Score (NSS) (*Jakobsen & Easteal, 1996*), and Maximum Chi-Square (*Smith, 1992*). In our analysis, we considered that recombination had occurred when $q < 0.05$ for PHI and at least NSS or Maximum Chi-Square (*Hongo et al., 2015*). Genes identified as recombinant were discarded for downstream analysis.

To minimize the false positive results caused by misalignments, frameshifts and ortholog prediction (*Schneider et al., 2009*; *Markova-Raina & Petrov, 2011*), we visually checked the alignments and checked the homology prediction by comparing the orthologs protein domains using PATRIC's annotation of Local and Global Families (*Davis et al., 2016*) and the Conserved Domain Database (CDD) (*Lu et al., 2020*).

## Gene annotation

Gene function was predicted using annotations from PATRIC and eggNOG-mapper (*Huerta-Cepas et al., 2017*), and InterProScan (*Jones et al., 2014*) was used to examine protein domains. Subcellular localization of the proteins was assessed using SufG+ v1.2.1 (*Barinov et al., 2009*). Protter (*Omasits et al., 2014*) was used to identify the position of the positively selected amino acids in relation to the cytoplasmic, transmembrane, and surface exposed portions of the proteins.

GIPSy v1.1.2 (*Soares et al., 2016*) was used to identify genomic islands (GIs) in *C. diphtheriae* NCTC 11397$^{T}$ using *C. glutamicum* ATCC 1302 (NC_006958.1) as the non-pathogenic reference. The islands were predicted by genes with C+G content and codon usage deviation, presence of transposases, presence of specific genes, non-conservation in comparison to reference genome, and flanking tRNA genes. The islands were classified according to the proportion of specific genes as pathogenicity islands, metabolic islands, resistance islands and symbiotic islands. Finally, prophages were predicted using PHASTER (*Arndt et al., 2016*).

## Prediction of drug and vaccine targets

Virulence genes and drug targets were predicted using the Pipeline Builder for Identification of Targets (PBIT) (*Shende et al., 2016*). PBIT predicted drug targets among the cytoplasmic proteins by a subtractive approach, identifying sequences of interest using BLASTp and specific databases in the following way. First, homologs to the human proteome, anti-targets and gut microbiota proteomes were filtering out to avoid cross-reactivity of drugs. Then the essential genes were identified using the Database of Essential Genes (*Zhang, 2004*) and virulence genes using the Virulence Factor Database (*Chen et al., 2005*). The druggability of the remaining candidates was predicted by

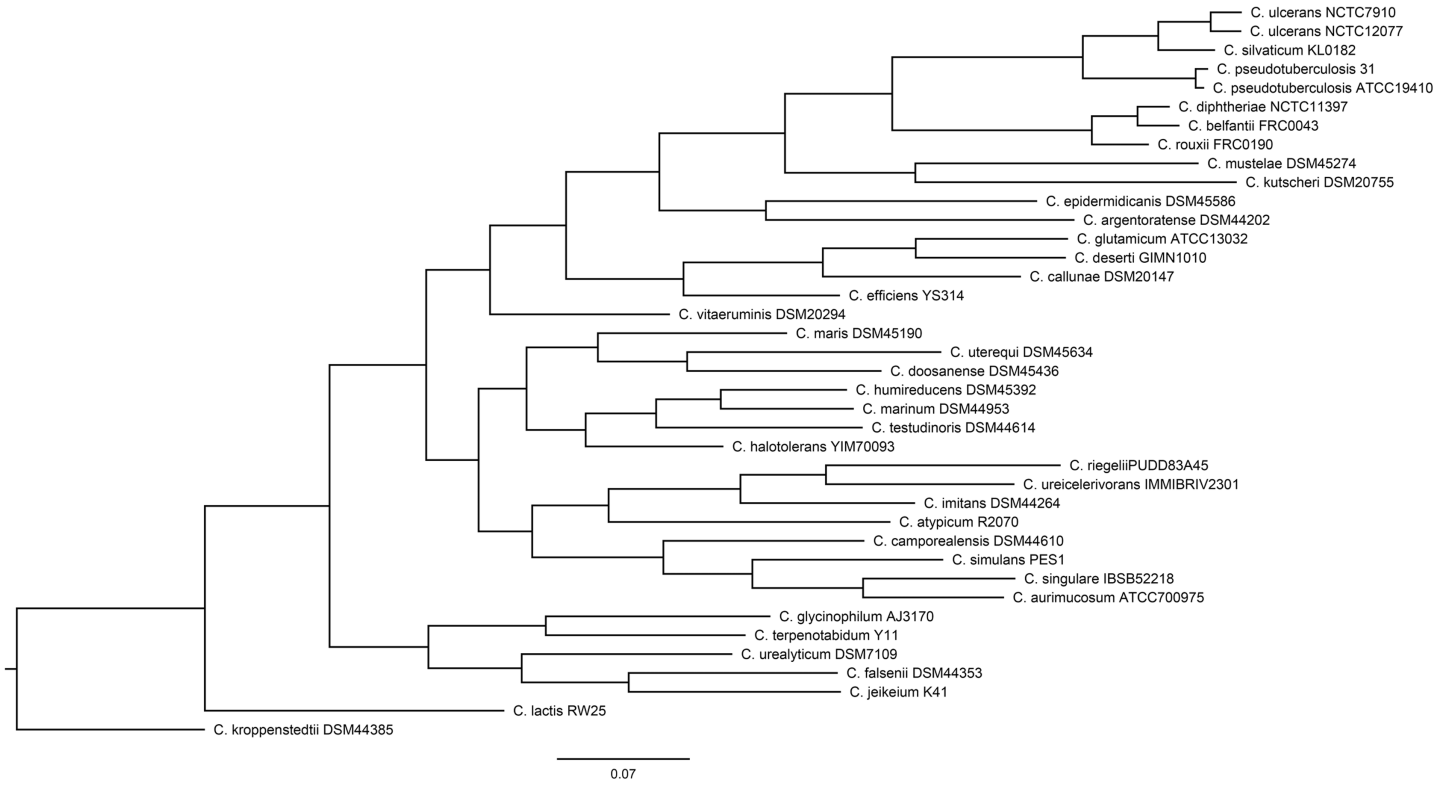


**Figure 1** **The *Corynebacterium* species tree that was generated by the PosiGene pipeline, using CONSENSE from the PHYLIP package.**
Full-size DOI: 10.7717/peerj.12662/fig-1

similarity to experimentally validated druggable targets from the Therapeutic Target Database (*Li et al., 2018*).

Vaccine targets were predicted from the transmembrane, putative surface exposed and secreted proteins predicted using SurfG+ and Vaxign (*He, Xiang & Mobley, 2010*) with prediction of human MHC Class I and II epitopes.

## RESULTS

Of the 40 genomes examined, the genome of *C. casei* LMG S-19264 was discarded due to its being identified as *Brevibacterium linens* based on the TYGS pipeline (Data S1). For this reason, only 39 *Corynebacterium* genomes were used for the downstream analysis. Figure 1 shows the phylogeny of the *Corynebacterium* genomes built by the PosiGene pipeline.

The PosiGene analysis showed zero to nine genes under positive selection ($p < 0.05$ for FDR) depending upon the 10 foregrounds that were used (Cb, Cd, Cp, Cpequi, Cpovis, Cr, Cs, Cul, Cul1, Cul2 (Tables S3–S12)), with 22 total genes shown to be under selection. Additional analysis of the 22 genes using PhiPack ($q < 0.05$ for PHI and at least NSS or Maximum Chi-Square tests) showed recombination in two out of four genes when Cd was the foreground, and one out of nine genes when Cp was used (Table S13). Manual curation of the 19 remaining proteins did not reveal any false positives caused by misalignments and ortholog prediction by comparison of protein domains (Table S14).

SurfG+ predicted 15 cytoplasmic, two membrane, one putative surface exposed and one secreted protein (Table S15). Thirty-five genomic islands and two prophages were predicted in *C. diphtheriae* NCTC 11397$^{T}$ (Tables S16 and S17).

From the 19 genes that were identified as positively selected across the species, two were identified in *C. diphtheriae*, 10 in *C. pseudotuberculosis*, one in *C. rouxii*, and six in *C. ulcerans* (Table 1 and Table S15). The COG categories of those genes were shown to be involved in defense, translation, energy production, and transport and metabolism of carbohydrates, amino acids, nucleotides, and coenzymes (Fig. 2). Based on our *in silico* prediction, 14 genes were found to be essential, six were virulence factors, and three were found to be in genomic islands (Table 1 and Table S15).

Thirteen of the genes were identified as potential drug targets by different analyses. Three genes were predicted based on our pipeline (Tables 1 and 2, and Table S15). The other 10 genes were not tagged as potential targets by the homology or druggability filters of the pipeline, but were included due to the possibility of targeting them with other methods (see Discussion section). For vaccine targets, four genes were predicted based on our pipeline (Tables 1 and 2, and Table S15).

## DISCUSSION

We identified 19 genes under positive selection, 13 potential drug targets and four potential vaccine targets. From the 13 potential drug targets, 10 were includes despite not passing the homology or druggability filters of the PBIT pipeline (Tables 1 and 2, and Table S15). The problem of having homology with the human proteome or human gut microbiota proteome can be solved by screening compounds that selectively inhibit the pathogen protein (*Arya et al., 2015*). The druggability prediction of the PBIT pipeline is based on sequence similarity to experimentally validated targets (*Shende et al., 2016*). So, the lack of predicted druggability by that pipeline can be solved by prediction of druggable pockets of a protein based on its own structure (*Volkamer et al., 2012*). Additionally, some of those proteins are known drug targets in other species.

### *C. diphtheriae*

In *C. diphtheriae* (foreground Cd), we identified two genes encoding proteins predicted to be essential and involved in translation, amino acid transport and metabolism (Table 1 and Table S15). The gene *ansB* is in GI10 and encodes secreted L-asparaginase type II which is a high-affinity enzyme that catalyzes the conversion of L-asparagine to L-aspartate and ammonia. The *E. coli*, *Dickeya dadantii* and human homologs to this gene are used for leukemia treatment, where the consequent low L-asparagine levels in plasma leads to apoptosis of the leukemia cells (*Lubkowski & Wlodawer, 2021*). This gene was suggested as a candidate vaccine target due to its classification as a secreted protein and predicted epitopes (Table 2). The second protein, SSU ribosomal protein S3p (*rpsC*), is a 30S ribosomal subunit that binds to the initiator Met-tRNA (*Burd & Dreyfuss, 1994*). Possible reasons for the selective pressure on these genes could be the effects on L-aspartate uptake (*ansB*) and translation efficiency (*rpsC*). *rpsC* was suggested as a drug target but

**Table 1** **Characterization and possible application of 19 genes under positive selection in different species of the *Corynebacterium diphtheria* complex.**

| *n* | Product (Gene) | PS sites | PS positions | Local | COG | Essential | VF | Target | GenBank ID |
|---|---|---|---|---|---|---|---|---|---|
| | ***C. diphtheriae*** | | | | | | | | |
| 1 | L-asparaginase, type II (EC 3.5.1.1) (*ansB*) | 3 | 69, 182, 339 | S | EJ | Yes | No | Va[2] | ERS451417_00414 |
| 2 | SSU ribosomal protein S3p (S3e) (*rpsC*) | 1 | 89 | C | J | Yes | No | Dr[4] | ERS451417_00402 |
| | ***C. pseudotuberculosis*** | | | | | | | | |
| 3 | ABC transporter, permease protein (*mntC*) | 1 | 111 | M | P | Yes | Yes | Va[2] | ERS451417_00548 |
| 4 | Adenosine deaminase (*add*) | 1 | 25 | C | F | – | No | – | ERS451417_00570 |
| 5 | Adhesin SpaE (*spaE*) | 20 | 23, 33, 35, 108, 119, 122, 125, 223, 232, 238, 239, 243, 244, 247, 251, 252, 253, 255, 257, 259 | SE | – | No | Yes | Va[2] | ERS451417_00159 |
| 6 | Dihydropteroate synthase 2 (nonfunctional) (*folP*) | 2 | 135, 158 | C | H | Yes | No | Dr[1,3] | ERS451417_00887 |
| 7 | HNH endonuclease | 4 | 48, 111, 284, 352 | C | V | Yes | No | Dr[4] | ERS451417_00880 |
| 8 | Peptide chain release factor 1 (*prfA*) | 1 | 60 | C | J | Yes | No | Dr[3] | ERS451417_00951 |
| 9 | Putative oxidoreductase | 8 | 33, 42, 48, 52, 109, 209, 226, 285 | C | CH | No | Yes | – | ERS451417_02135 |
| 10 | Putative phosphoglycerate mutase (*pgmB*) | 2 | 69, 143 | C | G | Yes | No | Dr[1,3] | ERS451417_02267 |
| | ***C. pseudotuberculosis* equi** | | | | | | | | |
| 11 | Methionine aminopeptidase (EC 3.4.11.18) (*mapB*) | 2 | 96, 97 | C | E | Yes | No | Dr[3] | ERS451417_01521 |
| 12 | Tyrosyl-tRNA synthetase (EC 6.1.1.1) (*tyrS*) | 4 | 23, 58, 59, 403 | V | J | Yes | Yes | Dr[3] | ERS451417_01169 |
| | ***C. rouxii*** | | | | | | | | |
| 13 | Hypothetical protein | 4 | 5, 74, 95, 154 | M | S | – | No | Va[2] | ERS451417_00470 |
| | ***C. ulcerans*** | | | | | | | | |
| 14 | Serine hydroxymethyltransferase (EC 2.1.2.1) (*glyA*) | 1 | 385 | C | E | Yes | No | Dr[4] | ERS451417_00836 |
| | ***C. ulcerans* lineage 1** | | | | | | | | |
| 15 | Hypothetical protein | 1 | 32 | C | – | – | – | Dr[4] | ERS451417_00635 |
| 16 | Phosphoenolpyruvate-dihydroxyacetone phosphotransferase, dihydroxyacetone binding subunit DhaK (*dnaK*) | 4 | 248, 251, 255, 256 | C | G | Yes | Yes | Dr[4] | ERS451417_02360 |
| 17 | Similar to citrate lyase beta chain, 3 (*citE*) | 1 | 235 | C | G | Yes | Yes | Dr[1] | ERS451417_00750 |
| | ***C. ulcerans* lineage *2*** | | | | | | | | |
| 18 | DNA polymerase III epsilon subunit (EC 2.7.7.7) (*dnaQ*) | 1 | 142 | C | L | Yes | No | Dr[4] | ERS451417_00985 |
| 19 | Precorrin-6A reductase (EC 1.3.1.54) (*cobK*) | 1 | 206 | C | H | Yes | No | Dr[4] | ERS451417_01234 |

**Notes:**
Columns: VF, virulence factor; COG, Clusters of Orthologous Groups; PS sites, positively selected sites; SE, surface exposed (sites).
Column Local: C, cytoplasm; M, membrane; SE, surface exposed; S, secreted.
Column COG: C, energy production and conversion; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; J, translation, ribosomal structure and biogenesis; S, function unknown; V, defense mechanisms.
Column Target: Dr, drug target; Va, vaccine target; [1]–predicted by PBIT pipeline, [2]–predicted by essentiality, local and Vaxign, [3]–described in literature for other species, [4]–suggested despite not attending one or more pipeline filters.

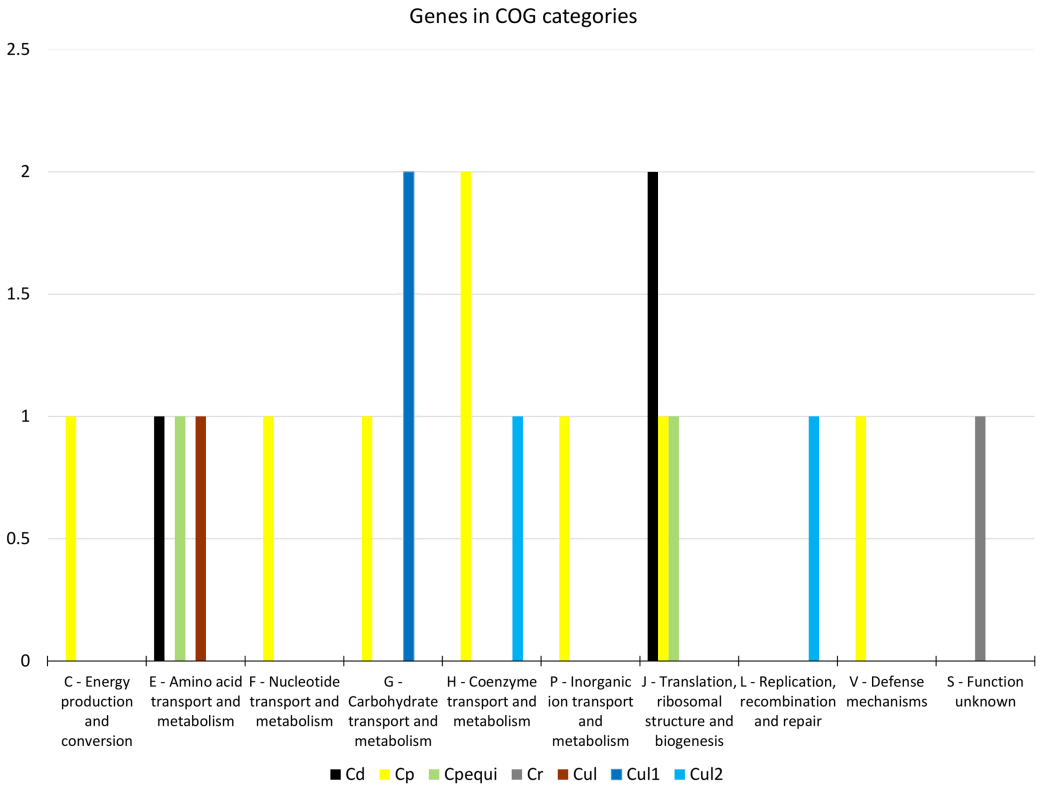


**Figure 2 Distribution of 19 genes under positive selection in COG categories.** Target groups: Cd, *C. diphtheriae*; Cp, *C. pseudotuberculosis*; Cpequi, *C. pseudotuberculosis* biovar equi; Cr, *C. rouxii*; Cul, *C. ulcerans*; Cul1, *C. ulcerans* lineage 1; Cul2, *C. ulcerans* lineage 2. COG categories: C-Q, metabolism; J-L, information storage and processing; M-V, cellular processes and signaling; S, poorly characterized.
Full-size DOI: 10.7717/peerj.12662/fig-2

has homology to a protein in the human gut microbiota proteome, requiring compounds that can selectively inhibit it.

## *C. pseudotuberculosis*

When *C. pseudotuberculosis* was the foreground (Cp), eight positively selected genes were identified. Five of the genes were tagged as essential, and three as virulence factors. They are involved in translation, coenzyme transport and metabolism, inorganic ion transport, defense from foreign DNA, nucleotide transport and metabolism and adhesion (Table 1 and Table S15). Among the essential genes, an ABC transporter permease protein (*mntC*) plays a role in the transport of $Mn^{2+}$ and $Zn^{2+}$ (*Claverys, 2001*) and was classified as a virulence factor. The Dihydropteroate synthase 2 (*folP2*) is nonfunctional according to PATRIC annotation but functional genes with this annotation are essential for the *de novo* synthesis of folate (*Bertacine Dias et al., 2018*). The HNH endonucleases degrade foreign DNA, and can also be involved in DNA repair, replication and recombination (*Wu, Lin & Yuan, 2020*). Peptide chain release factor 1 (*prfA*) recognizes the stop codons UAA and UAG, promoting the end of translation (*Scolnick et al., 1968*). The Putative phosphoglycerate mutase (*pgmB*) is capable of interconverting

**Table 2** **Final drug and vaccine target candidates for *Corynebacterium* species based on a positive selection analysis by foreground, application, and priority.**

| Foreground | Application | Priority | Product (Gene) | PS sites (exposed sites) | Local | GenBank ID |
|---|---|---|---|---|---|---|
| Cd | Drug target | 1 (gut microbiota homolog) | SSU ribosomal protein S3p (S3e) (*rpsC*) | 89 | C | ERS451417_00402 |
| Cd | Vaccine | 1 | L-asparaginase (*ansB*) | 69, 182, 339 (69, 182, 339) | S | ERS451417_00414 |
| Cp | Drug target | 1 | Putative phosphoglycerate mutase (*pgmB*) | 69, 143 | C | ERS451417_02267 |
| Cp | Drug target | 2 (predicted as nonfunctional) | Dihydropteroate synthase 2 (nonfunctional) (*folP2*) | 135, 158 | C | ERS451417_00887 |
| Cp | Drug target | 3 (human and gut microbiota homolog, no predicted druggability) | Peptide chain release factor 1 (*prfA*) | 60 | C | ERS451417_00951 |
| Cp | Drug target | 4 (no predicted druggability) | HNH endonuclease | 48, 111, 284, 352 | C | ERS451417_00880 |
| Cp | Vaccine | 1 | Adhesin SpaE (*spaE*) | 23, 33, 35, 108, 119, 122, 125, 223, 232, 238, 239, 243, 244, 247, 251, 252, 253, 255, 257, 259 (23, 33, 35, 108, 119, 122, 125, 223) | SE | ERS451417_00159 |
| Cp | Vaccine | 2 (no exposed PS site) | ABC transporter, permease protein (*mntC*) | 111 | M | ERS451417_00548 |
| Cpequi | Drug target | 1 (virulence factor, more PS sites, gut microbiota homolog, target in literature) | Tyrosyl-tRNA synthetase (*tyrS*) | 23, 58, 59, 403 | C | ERS451417_01169 |
| Cpequi | Drug target | 2 (less PS sites, gut microbiota homolog, target in literature) | Methionine aminopeptidase (*mapB*) | 96, 97 | C | ERS451417_01521 |
| Cr | Vaccine | 1 | Hypothetical protein | 5, 74, 95, 154 (154) | M | ERS451417_00470 |
| Cul | Drug target | 1 (human and gut microbiota homolog) | Serine hydroxymethyltransferase (*glyA*) | 385 | C | ERS451417_00836 |
| Cul1 | Drug target | 1 (virulence factor) | Similar to citrate lyase beta chain (*citE*) | 235 | C | ERS451417_00750 |
| Cul1 | Drug target | 2 (virulence factor, gut microbiota homolog) | Phosphoenolpyruvate-dihydroxyacetone phosphotransferase, dihydroxyacetone binding subunit DhaK (*dhaK*) | 248, 251, 255, 256 | – | ERS451417_02360 |
| Cul1 | Drug target | 3 (hypothetical protein, no predicted druggability) | Hypothetical protein | 32 | C | ERS451417_00635 |
| Cul2 | Drug target | Equal. No predicted druggability | DNA polymerase III epsilon subunit (*dnaQ*) | 142 | C | ERS451417_00985 |
| Cul2 | Drug target | Equal. No predicted druggability | Precorrin-6A reductase (*cobK*) | 206 | C | ERS451417_01234 |

**Note:**
PS sites, positively selected sites.

2- and 3-phosphoglycerate in glycolysis (*Rigden, 2008*), although this particular gene is annotated as putative.

The other three identified genes (*add, spaE* and the putative oxidoreductase) are not characterized as essential, so may not be suitable drug targets. Adenosine deaminase (*add*) catalyzes the hydrolytic deamination of adenosine into inosine (*Chang et al., 1991*). *spaE*, from the operon *spaDEF*, encodes the minor pilin SpaE in *C. diphtheriae* (*Mandlik et al., 2008*) and is in GI5. It's ortholog in *C. pseudotuberculosis* also encodes the minor pilin and was first described as *spaB* from the operon *spaABC* (*Trost et al., 2010*). The putative oxidoreductase is a flavoenzyme with a "FAD-binding domain, ferredoxin reductase-type" (IPR017927), but its specific reaction is unknown.

Why would these particular genes be under positive selection? One could hypothesize that there would be more efficient manganese uptake (*mntC*), tissue adhesion on a new host range (*spaE*), improved efficiency for defense against foreign DNA (HNH endonuclease), translation (*prfA*), and metabolism of nucleotides (*add*) and carbohydrates (*pgmB*).

The vaccine targets (*mntC* and *spaE*) were indicated due to either their membrane location, their predicted epitopes, and that they might have surface exposed sites that are under positive selection. There were four drug targets (Table 2). *pgmB* was predicted as a target by PBIT and is this same gene is a drug target in helminth parasites (*Timson, 2016*). *folP2* was also predicted and is a well know target of sulfa and imidazole derivatives in human pathogens such as *Staphylococcus aureus*, *M. tuberculosis*, *Bacillus anthracis*, *Streptococcus pneumoniae*, *Burkholderia cenocepacia* and *Yersinia pestis* (*Bertacine Dias et al., 2018*). Although it was annotated as non-functional, the evidence of positive selection in this protein suggests an unknown adaptation due to specific amino acids that could be targeted. *prfA* has homology to human and human gut microbiota proteome and has no predicted druggability, but it is a known target of the drug Apidaecin in gram negative bacteria (*Matsumoto et al., 2017*). The HNH endonuclease had no predicted druggability.

## *C. pseudotuberculosis* biovar equi

When *C. pseudotuberculosis* biovar equi was used as the foreground (Cpequi), two genes were identified as being under positive selection, and were also characterized as essential. These two genes (*mapB* and *tyrS*) are both involved in translation (Table 1 and Table S15). Methionine aminopeptidase (*mapB*) cleaves the initiator methionine from newly synthesized polypeptides (*Helgren et al., 2016*; *Pillalamarri et al., 2021*). Tyrosyl-tRNA synthetase (*tyrS*) attaches the amino acid tyrosine to the appropriate tRNA (*Hughes et al., 2020*; *Othman et al., 2021*). These two genes could be under positive selection as it could affect translation efficiency. Both genes were predicted as homologs to human gut microbiota and have been identified as drug targets in other studies (Table 2). *tyrS* was predicted as a virulence factor and an ortholog from *Pseudomonas aeruginosa* was found to be targeted by four drug-like compounds (*Hughes et al., 2020*), and a *S. aureus* ortholog to this gene was targeted by new pyrazolone and dipyrazolotriazine derivatives (*Othman et al., 2021*). *mapB* from *M. tuberculosis* and

*S. pneumoniae* were shown to be selectively targeted despite homology to human protein (*Krátký et al., 2012*; *Arya et al., 2015*).

### *C. rouxii*

A single hypothetical protein (ERS451417_00470) was identified when *C. rouxii* was used as the foreground (Cr). It had no predicted domains, but it could be a vaccine target candidate (Table 2) due to its transmembrane location and the surface exposed sites under positive selection.

### *C. ulcerans*

A single essential gene (*glyA*) was identified when *C. ulcerans* was used as the foreground (Cul). *glyA* encodes a serine hydroxymethyltransferase enzyme (Table 1 and Table S15). This gene is known to participate in the one-carbon metabolism of serine/glycine interconversion and also in the folate/methionine cycle (*Batool et al., 2020*), which could explain its being under selective pressure. This gene was also identified as a potential drug target (Table 2), but it does have homology to human and human gut microbiota proteome. It has been shown to play a key role in lysostaphin resistance in *Staphylococcus aureus* (*Batool et al., 2020*).

### *C. ulcerans* lineage 1

Three genes were identified when the *C. ulcerans* lineage 1 genome was used as the foreground (Cul1). Two of them were essential, involved in carbohydrate transport and metabolism (Table 1 and Table S15). The Phosphoenolpyruvate-dihydroxyacetone phosphotransferase, dihydroxyacetone binding subunit DhaK (*dhaK*) gene is in GI34. This enzyme phosphorylates ketones and short chain aldoses using adenosine triphosphate (ATP) (*Peiro et al., 2019*). The second gene is annotated in PATRIC as "Similar to citrate lyase beta chain, 3" (*citE*) and is probably one of the catalytic subunits of citrate lyase, the enzyme that catalyzes the cleavage of citrate to acetate and oxaloacetate during citrate fermentation (*Schneider, Dimroth & Bott, 2000*). Both proteins were predicted as virulence factors by PBIT. The third gene encodes a hypothetical protein (ERS451417_00635) with no predicted domain or cellular localization.

The two genes with predicted function (*dhaK* and *citE*) appear to be related to metabolism inside the host. In *Listeria monocytogenes*, Phosphoenolpyruvate-dihydroxyacetone phosphotransferase (DhaK and other subunits) is required to utilize carbon sources for amino acid synthesis inside murine macrophages (*Eylert et al., 2008*). In *Enterococcus faecalis*, mutants of citrate fermentation genes (*citE* and others) were less pathogenic for the model *Galleria mellonella* (*Martino et al., 2018*). These same two genes are possible drug targets (Table 2). *citE* was predicted as a virulence factor and drug target candidate, while *dhaK* was predicted as a virulence factor and suggested despite the homology to a protein in the gut microbiota homology.

### *C. ulcerans* lineage 2

Two essential genes (Table 1 and Table S15) were identified when the *C. ulcerans* lineage 2 was used as the foreground (Cul2). The DNA polymerase III epsilon subunit (*dnaQ*) has a

domain with 3′–5′ exonuclease proofreading activity (*Raia, Delarue & Sauguet, 2019*). The other gene, *cobK*, encodes Precorrin-6A reductase which is involved in part I of the cobalamin cofactor (vitamin B12) biosynthesis pathway (*Kipkorir et al., 2021*). The mutations seen in these genes could provide the organism with a more efficient means of DNA replication (*dnaQ*) and biosynthesis of the essential cofactor cobalamin (*cobK*). Neither of these genes had any predictable druggability.

### Probable adaptations across groups

It is reported that most of the genes identified as being under positive selection are exposed on the surface and are involved in host colonization, and resistance to phage and antibiotics (*Petersen et al., 2007*; *Anisimova & Liberles, 2012*). Those under positive selection that are not surface exposed have been shown to be involved in metabolism (*Petersen et al., 2007*; *Rao, Sivakumar & Jayakumar, 2019*) or gene regulation (*Zhang et al., 2011*). Considering the function of the identified genes, most of the probable adaptations appear to be related to metabolism. A notable exception is in *C. pseudotuberculosis*, where the pilin SpaE could have enhanced adhesion to different host species tissues. Although the specific adaptations are not clear, an amino acid fixed by positive selection is an attractive target for a therapeutic molecule, as a non-synonymous mutation that could avoid interaction would decrease fitness. These genes could be used for reverse vaccinology and *in silico* drug targeting methods.

## CONCLUSION

In this analysis, we predicted 19 genes with non-synonymous mutations that are probably involved in adaptations found in the pathogens *C. diphtheriae*, *C. pseudotuberculosis*, *C. rouxii* and *C. ulcerans*. Based on our pipeline and literature data, 13 genes are candidate drug targets and four are potential vaccine targets, but their effectiveness would require experimental validation.

## ADDITIONAL INFORMATION AND DECLARATIONS


### Funding

This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil.
Conselho Nacional de Desenvolvimento Científico e Tecnológico.
Fundação de Amparo à Pesquisa de Minas Gerais.

## Competing Interests

Debmalya Barh and Vasco Azevedo are Academic Editors for PeerJ.

## Author Contributions

- Marcus Vinicius Canário Viana conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Rodrigo Profeta conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Janaína Canário Cerqueira conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Alice Rebecca Wattam conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Debmalya Barh conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Artur Silva conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Vasco Azevedo conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

The raw data are the gene nucleotide sequences listed in Table S1 and Tables 1 and 2.

## Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12662#supplemental-information.

# REFERENCES

**Altenhoff AM, Dessimoz C. 2009.** Phylogenetic and functional assessment of orthologs inference projects and methods. *PLOS Computational Biology* **5**:e1000262 DOI 10.1371/journal.pcbi.1000262.

**Anisimova M, Liberles DA. 2012.** Detecting and understanding natural selection. In: Cannarozzi GM, Schneider A, eds. *Codon Evolution: Mechanisms and Models*. Oxford: Oxford University Press, 73–96.

**Anisimova M, Nielsen R, Yang Z. 2003.** Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**:1229–1236 DOI 10.1093/bioinformatics/btn086.

**Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016.** PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* **44**:W16–W21 DOI 10.1093/nar/gkw387.

**Arya T, Reddi R, Kishor C, Ganji RJ, Bhukya S, Gumpena R, McGowan S, Drag M, Addlagatta A. 2015.** Identification of the molecular basis of inhibitor selectivity between the

human and streptococcal type I methionine aminopeptidases. *Journal of Medicinal Chemistry* **58**:2350–2357 DOI 10.1021/jm501790e.

**Badell E, Hennart M, Rodrigues C, Passet V, Dazas M, Panunzi L, Bouchez V, Carmi-Leroy A, Toubiana J, Brisse S. 2020.** Corynebacterium rouxii sp. nov., a novel member of the diphtheriae species complex. *Research in Microbiology* **171**:122–127 DOI 10.1016/j.resmic.2020.02.003.

**Barinov A, Loux V, Hammani A, Nicolas P, Langella P, Ehrlichh D, Maguin E, van Guchte MD. 2009.** Prediction of surface exposed proteins in Streptococcus pyogenes, with a potential application to other Gram-positive bacteria. *Proteomics* **9**:61–73 DOI 10.1002/pmic.200800195.

**Batool N, Ko KS, Chaurasia AK, Kim KK. 2020.** Functional identification of serine hydroxymethyltransferase as a key gene involved in lysostaphin resistance and virulence potential of staphylococcus aureus strains. *International Journal of Molecular Sciences* **21**:9135 DOI 10.3390/ijms21239135.

**Bernard AL, Funke G. 2015.** Corynebacterium. In: *Bergey's Manual of Systematic of Archaea and Bacteria (Online)*. London: John Wiley & Sons, Bergey's Manual Trust, 1–70.

**Bertacine Dias MV, Santos JC, Libreros-Zúñiga GA, Ribeiro JA, Chavez-Pacheco SM. 2018.** Folate biosynthesis pathway: mechanisms and insights into drug design for infectious diseases. *Future Medicinal Chemistry* **10**:935–959 DOI 10.4155/fmc-2017-0168.

**Brasier CM. 1995.** Episodic selection as a force in fungal microevolution, with special reference to clonal speciation and hybrid introgression. *Canadian Journal of Botany* **73**:1213–1221 DOI 10.1139/b95-381.

**Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015.** RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports* **5**:1–6 DOI 10.1038/srep08365.

**Bruen TC, Philippe HH, Bryant D. 2006.** A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**:2665–2681 DOI 10.1534/genetics.105.048975.

**Burd CG, Dreyfuss G. 1994.** Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**:615–621 DOI 10.1126/science.8036511.

**Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421 DOI 10.1186/1471-2105-10-421.

**Chang ZY, Nygaard P, Chinault AC, Kellems RE. 1991.** Deduced amino acid sequence of Escherichia coli adenosine deaminase reveals evolutionarily conserved amino acid residues: implications for catalytic function. *Biochemistry* **30**:2273–2280 DOI 10.1021/bi00222a033.

**Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005.** VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* **33**:D325–D328 DOI 10.1093/nar/gki008.

**Claverys JP. 2001.** A new family of high-affinity ABC manganese and zinc permeases. *Research in Microbiology* **152**:231–243 DOI 10.1016/s0923-2508(01)01195-0.

**Dangel A, Berger A, Rau J, Eisenberg T, Kämpfer P, Margos G, Contzen M, Busse H-J, Konrad R, Peters M, Sting R, Sing A. 2020.** Corynebacterium silvaticum sp. nov., a unique group of NTTB corynebacteria in wild boar and roe deer. *International Journal of Systematic and Evolutionary Microbiology* **70(6)**:3614–3624 DOI 10.1099/ijsem.0.004195.

**Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR, Yoo H. 2016.** PATtyFams: protein families for the microbial genomes in the PATRIC database. *Frontiers in Microbiology* **7**:1–12 DOI 10.3389/fmicb.2016.00118.

**Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler RRM, Butler RRM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020.** The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Research* **48**:D606–D612 DOI 10.1093/nar/gkz943.

**Dazas M, Badell E, Carmi-Leroy A, Criscuolo A, Brisse S. 2018.** Taxonomic status of Corynebacterium diphtheriae biovar Belfanti and proposal of Corynebacterium belfantii sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **68**:3826–3831 DOI 10.1099/ijsem.0.003069.

**Eylert E, Schär J, Mertins S, Stoll R, Bacher A, Goebel W, Eisenreich W. 2008.** Carbon metabolism of Listeria monocytogenes growing inside macrophages. *Molecular Microbiology* **69**:1008–1017 DOI 10.1111/j.1365-2958.2008.06337.x.

**Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PKC, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. 2013.** Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nature Genetics* **45**:1183–1189 DOI 10.1038/ng.2747.

**Felsenstein J. 2005.** PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by Author. Seattle: Department of Genome Sciences, University of Washington. *Available at http://evolution.genetics.washington.edu/phylip.html*.

**Goodswen SJ, Kennedy PJ, Ellis JT. 2018.** A gene-based positive selection detection approach to identify vaccine candidates using toxoplasma gondii as a test case protozoan pathogen. *Frontiers in Genetics* **9**:1–16 DOI 10.3389/fgene.2018.00332.

**Grosse-Kock S, Kolodkina V, Schwalbe EC, Blom J, Burkovski A, Hoskisson PA, Brisse S, Smith D, Sutcliffe IC, Titov L, Sangal V. 2017.** Genomic analysis of endemic clones of toxigenic and non-toxigenic Corynebacterium diphtheriae in Belarus during and after the major epidemic in 1990s. *BMC Genomics* **18**:1–10 DOI 10.1186/s12864-017-4276-3.

**Hacker E, Antunes CA, Mattos-Guaraldi AL, Burkovski A, Tauch A. 2016.** Corynebacterium ulcerans, an emerging human pathogen. *Future Microbiology* **11**:1191–1208 DOI 10.2217/fmb-2016-0085.

**He Y, Xiang Z, Mobley HLT. 2010.** Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Journal of Biomedicine and Biotechnology* **2010**:297505 DOI 10.1155/2010/297505.

**Helgren TR, Wangtrakuldee P, Staker BL, Hagen TJ. 2016.** Advances in bacterial methionine aminopeptidase inhibition. *Current Topics in Medicinal Chemistry* **16**:397–414 DOI 10.2174/1568026615666150813145410.

**Hennart M, Panunzi LG, Rodrigues C, Gaday Q, Baines SL, Barros-Pinkelnig M, Carmi-Leroy A, Dazas M, Wehenkel AM, Didelot X, Toubiana J, Badell E, Brisse S. 2020.** Population genomics and antimicrobial resistance in Corynebacterium diphtheriae. *Genome Medicine* **12**:107 DOI 10.1186/s13073-020-00805-7.

**Hongo JA, de Castro GM, Cintra LC, Zerlotini A, Lobo FP. 2015.** POTION: an end-to-end pipeline for positive Darwinian selection detection in genome-scale data through phylogenetic comparison of protein-coding genes. *BMC Genomics* **16**:567 DOI 10.1186/s12864-015-1765-0.

**Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017.** Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution* **34**:2115–2122 DOI 10.1093/molbev/msx148.

**Hughes CA, Gorabi V, Escamilla Y, Dean FB, Bullard JM. 2020.** Two forms of tyrosyl-tRNA synthetase from pseudomonas aeruginosa : characterization and discovery of inhibitory compounds. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **25**:1072–1086 DOI 10.1177/2472555220934793.

**Jakobsen IB, Easteal S. 1996.** A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *Computer Applications in the Biosciences* **12**:291–295 DOI 10.1093/bioinformatics/12.4.291.

**Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014.** InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240 DOI 10.1093/bioinformatics/btu031.

**Jordan G, Goldman N. 2012.** The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular Biology and Evolution* **29**:1125–1139 DOI 10.1093/molbev/msr272.

**Kipkorir T, Mashabela GT, de Wet TJ, Koch A, Wiesner L, Mizrahi V, Warner DF. 2021.** De novo cobalamin biosynthesis, transport, and assimilation and cobalamin-mediated regulation of methionine biosynthesis in mycobacterium smegmatis. *Journal of Bacteriology* **203**:e00620 DOI 10.1128/JB.00620-20.

**Kopac S, Wang Z, Wiedenbeck J, Sherry J, Wu M, Cohan FM. 2014.** Genomic heterogeneity and ecological speciation within one subspecies of bacillus subtilis. *Applied and Environmental Microbiology* **80**:4842–4853 DOI 10.1128/AEM.00576-14.

**Kosiol C, Anisimova M. 2012.** Selection on the protein-coding genome. In: Anisimova M, ed. *Evolutionary Genomics*. New York, Dordrecht, Heidelberg, London: Humana Press, 113–140.

**Kryazhimskiy S, Plotkin JB. 2008.** The population genetics of dN/dS. *PLOS Genetics* **4**:e1000304 DOI 10.1371/journal.pgen.1000304.

**Krátký M, Vinšová J, Novotná E, Mandíková J, Wsól V, Trejtnar F, Ulmann V, Stolaříková J, Fernandes S, Bhat S, Liu JO. 2012.** Salicylanilide derivatives block Mycobacterium tuberculosis through inhibition of isocitrate lyase and methionine aminopeptidase. *Tuberculosis* **92**:434–439 DOI 10.1016/j.tube.2012.06.001.

**Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007.** Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948 DOI 10.1093/bioinformatics/btm404.

**Lefort V, Desper R, Gascuel O. 2015.** FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution* **32**:2798–2800 DOI 10.1093/molbev/msv150.

**Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, Fu T, Zhang X, Cui X, Tu G, Zhang Y, Li S, Yang F, Sun Q, Qin C, Zeng X, Chen Z, Chen YZ, Zhu F. 2018.** Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Research* **46**:D1121–D1127 DOI 10.1093/nar/gkx1076.

**Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020.** CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* **48**:D265–D268 DOI 10.1093/nar/gkz991.

**Lubkowski J, Wlodawer A. 2021.** Structural and biochemical properties of L-asparaginase. *The FEBS Journal* **288**:4183–4209 DOI 10.1111/febs.16042.

**Mandlik A, Swierczynski A, Das A, Ton-That H. 2008.** Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends in Microbiology* **16**:33–40 DOI 10.1016/j.tim.2007.10.010.

**Markova-Raina P, Petrov D. 2011.** High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. *Genome Research* **21**:863–874 DOI 10.1101/gr.115949.110.

**Martino GP, Perez CE, Magni C, Blancato VS. 2018.** Implications of the expression of Enterococcus faecalis citrate fermentation genes during infection. *PLOS ONE* **13**:e0205787 DOI 10.1371/journal.pone.0205787.

**Matsumoto K, Yamazaki K, Kawakami S, Miyoshi D, Ooi T, Hashimoto S, Taguchi S. 2017.** In vivo target exploration of apidaecin based on acquired resistance induced by gene overexpression (ARGO assay). *Scientific Reports* **7**:12136 DOI 10.1038/s41598-017-12039-6.

**Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013.** Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**:60 DOI 10.1186/1471-2105-14-60.

**Meier-Kolthoff JP, Göker M. 2019.** TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications* **10**:2182 DOI 10.1038/s41467-019-10210-3.

**Möller J, Busch A, Berens C, Hotzel H, Burkovski A. 2021.** Newly isolated animal pathogen corynebacterium silvaticum is cytotoxic to human epithelial cells. *International Journal of Molecular Sciences* **22**:3549 DOI 10.3390/ijms22073549.

**Omasits U, Ahrens CH, Müller S, Wollscheid B. 2014.** Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**:884–886 DOI 10.1093/bioinformatics/btt607.

**Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016.** Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* **17**:132 DOI 10.1186/s13059-016-0997-x.

**Othman IMM, Gad-Elkareem MAM, Hassane Anouar E, Aouadi K, Snoussi M, Kadri A. 2021.** New substituted pyrazolones and dipyrazolotriazines as promising tyrosyl-tRNA synthetase and peroxiredoxin-5 inhibitors: design, synthesis, molecular docking and structure-activity relationship (SAR) analysis. *Bioorganic Chemistry* **109**:104704 DOI 10.1016/j.bioorg.2021.104704.

**Otsuji K, Fukuda K, Ogawa M, Saito M. 2019.** Mutation and diversity of diphtheria toxin in Corynebacterium ulcerans. *Emerging Infectious Diseases* **25**:2122–2123 DOI 10.3201/eid2511.181455.

**O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016.** Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**:D733–D745 DOI 10.1093/nar/gkv1189.

**Peiro C, Millard P, de Simone A, Cahoreau E, Peyriga L, Enjalbert B, Heux S. 2019.** Chemical and metabolic controls on dihydroxyacetone metabolism lead to suboptimal growth of Escherichia coli. *Applied and Environmental Microbiology* **85**:1–17 DOI 10.1128/AEM.00768-19.

**Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. 2007.** Genes under positive selection in Escherichia coli. *Genome Research* **17**:1336–1343 DOI 10.1101/gr.6254707.

**Pillalamarri V, Reddy CG, Bala SC, Jangam A, Kutty VV, Addlagatta A. 2021.** Methionine aminopeptidases with short sequence inserts within the catalytic domain are differentially inhibited: Structural and biochemical studies of three proteins from Vibrio spp. *European Journal of Medicinal Chemistry* **209**:112883 DOI 10.1016/j.ejmech.2020.112883.

**Raia P, Delarue M, Sauguet L. 2019.** An updated structural classification of replicative DNA polymerases. *Biochemical Society Transactions* **47**:239–249 DOI 10.1042/BST20180579.

**Rao RT, Sivakumar N, Jayakumar K. 2019.** Analyses of livestock-associated staphylococcus aureus pan-genomes suggest virulence is not primary interest in evolution of its genome. *OMICS: A Journal of Integrative Biology* **23**:224–236 DOI 10.1089/omi.2019.0005.

**Rappuoli R, Malito E. 2014.** History of diphtheria vaccine development. In: Burkovski A, ed. *Corynebacterium Diphtheriae and Related Toxigenic Species*. Dordrecht: Springer Netherlands, 225–238.

**Rigden DJ. 2008.** The histidine phosphatase superfamily: structure and function. *The Biochemical journal* **409**:333–348 DOI 10.1042/BJ20071097.

**Sahm A, Bens M, Platzer M, Szafranski K. 2017.** PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Research* **45**:1–11 DOI 10.1093/nar/gkx179.

**Sangal V, Burkovski A, Hunt AC, Edwards B, Blom J, Hoskisson PA. 2014.** A lack of genetic basis for biovar differentiation in clinically important Corynebacterium diphtheriae from whole genome sequencing. *Infection, Genetics and Evolution* **21**:54–57 DOI 10.1016/j.meegid.2013.10.019.

**Schneider K, Dimroth P, Bott M. 2000.** Biosynthesis of the prosthetic group of citrate lyase. *Biochemistry* **39**:9438–9450 DOI 10.1021/bi000401r.

**Schneider A, Souvorov A, Sabath N, Landan G, Gonnet GH, Graur D. 2009.** Estimates of positive darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biology and Evolution* **1**:114–118 DOI 10.1093/gbe/evp012.

**Scolnick E, Tompkins R, Caskey T, Nirenberg M. 1968.** Release factors differing in specificity for terminator codons. *Proceedings of the National Academy of Sciences of the United States of America* **61**:768–774 DOI 10.1073/pnas.61.2.768.

**Selim SA, Mohamed FH, Hessain AM, Moussa IM. 2015.** Immunological characterization of diphtheria toxin recovered from Corynebacterium pseudotuberculosis. *Saudi Journal of Biological Sciences* **23**:282–287 DOI 10.1016/j.sjbs.2015.11.004.

**Sharma NC, Efstratiou A, Mokrousov I, Mutreja A, Das B, Ramamurthy T. 2019.** Diphtheria. *Nature Reviews Disease Primers* **5**:81 DOI 10.1038/s41572-019-0131-y.

**Shende G, Haldankar H, Barai RS, Bharmal MH, Shetty V, Idicula-Thomas S. 2016.** PBIT: pipeline builder for identification of drug targets for infectious diseases. *Bioinformatics* **33(6)**:929–931 DOI 10.1093/bioinformatics/btw760.

**Smith JM. 1992.** Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**:126–129 DOI 10.1007/BF00182389.

**Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV, Baumbach J, Figueiredo HCP, Miyoshi A, Tauch A, Silva A, Azevedo V. 2016.** GIPSy: genomic island prediction software. *Journal of Biotechnology* **232**:2–11 DOI 10.1016/j.jbiotec.2015.09.008.

**Storey JD, Tibshirani R. 2003.** Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**:9440–9445 DOI 10.1073/pnas.1530509100.

**Timson DJ. 2016.** Metabolic enzymes of helminth parasites: potential as drug targets. *Current Protein & Peptide Science* **17**:280–295 DOI 10.2174/1389203717999160226180733.

**Trost E, Ott L, Schneider J, Schröder J, Jaenicke S, Goesmann A, Husemann P, Stoye J, Dorella FA, Rocha FS, Soares SDC, D'Afonseca V, Miyoshi A, Ruiz J, Silva A, Azevedo V, Burkovski A, Guiso N, Join-Lambert OF, 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 DOI 10.1186/1471-2164-11-728.

**Truelove SA, Keegan LT, Moss WJ, Chaisson LH, Macher E, Azman AS, Lessler J. 2020.** Clinical and epidemiological aspects of diphtheria: a systematic review and pooled analysis. *Clinical Infectious Diseases* **71**:89–97 DOI 10.1093/cid/ciz808.

**Volkamer A, Kuhn D, Rippmann F, Rarey M. 2012.** DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**:2074–2075 DOI 10.1093/bioinformatics/bts310.

**Wu C-C, Lin JLJ, Yuan HS. 2020.** Structures, mechanisms, and functions of his-me finger nucleases. *Trends in Biochemical Sciences* **45**:935–946 DOI 10.1016/j.tibs.2020.07.002.

**Yang Z. 2005.** Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**:1107–1118 DOI 10.1093/molbev/msi097.

**Yang Z, Dos Reis M. 2011.** Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution* **28**:1217–1228 DOI 10.1093/molbev/msq303.

**Yang Z, Nielsen R. 2002.** Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution* **19**:908–917 DOI 10.1093/oxfordjournals.molbev.a004148.

**Zasada AA. 2013.** Nontoxigenic highly pathogenic clone of Corynebacterium diphtheriae, Poland, 2004–2012. *Emerging Infectious Diseases* **19**:1870–1872 DOI 10.3201/eid1911.130297.

**Zasada AA. 2014.** Antimicrobial susceptibility and treatment. In: *Corynebacterium Diphtheriae and Related Toxigenic Species*. Dordrecht: Springer Netherlands, 239–246.

**Zhang R. 2004.** DEG: a database of essential genes. *Nucleic Acids Research* **32**:271D–272D DOI 10.1093/nar/gkh024.

**Zhang J. 2005.** Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution* **22**:2472–2479 DOI 10.1093/molbev/msi237.

**Zhang Y, Zhang H, Zhou T, Zhong Y, Jin Q. 2011.** Genes under positive selection in Mycobacterium tuberculosis. *Computational Biology and Chemistry* **35**:319–322 DOI 10.1016/j.compbiolchem.2011.08.001.