



# Searching whole genome sequences for biochemical identification features of emerging and reemerging pathogenic *Corynebacterium* species

André S. Santos<sup>1,2</sup> · Rommel T. Ramos<sup>3</sup> · Artur Silva<sup>3</sup> · Raphael Hirata Jr<sup>4</sup> · Ana L. Mattos-Guaraldi<sup>4</sup> · Roberto Meyer<sup>2</sup> · Vasco Azevedo<sup>1</sup> · Liza Felicori<sup>1</sup> · Luis G. C. Pacheco<sup>2</sup>

Received: 16 November 2017 / Revised: 28 March 2018 / Accepted: 16 April 2018 / Published online: 11 May 2018

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Biochemical tests are traditionally used for bacterial identification at the species level in clinical microbiology laboratories. While biochemical profiles are generally efficient for the identification of the most important corynebacterial pathogen *Corynebacterium diphtheriae*, their ability to differentiate between *biovars* of this bacterium is still controversial. Besides, the unambiguous identification of emerging human pathogenic species of the genus *Corynebacterium* may be hampered by highly variable biochemical profiles commonly reported for these species, including *Corynebacterium striatum*, *Corynebacterium amycolatum*, *Corynebacterium minutissimum*, and *Corynebacterium xerosis*. In order to identify the genomic basis contributing for the biochemical variabilities observed in phenotypic identification methods of these bacteria, we combined a comprehensive literature review with a bioinformatics approach based on reconstruction of six specific biochemical reactions/pathways in 33 recently released whole genome sequences. We used data retrieved from curated databases (MetaCyc, PathoSystems Resource Integration Center (PATRIC), The SEED, TransportDB, UniProtKB) associated with homology searches by BLAST and profile Hidden Markov Models (HMMs) to detect enzymes participating in the various pathways and performed ab initio protein structure modeling and molecular docking to confirm specific results. We found a differential distribution among the various strains of genes that code for some important enzymes, such as beta-phosphoglucosyltransferase and fructokinase, and also for individual components of carbohydrate transport systems, including the fructose-specific phosphoenolpyruvate-dependent sugar phosphotransferase (PTS) and the ribose-specific ATP-binding cassette (ABC) transporter. Horizontal gene transfer plays a role in the biochemical variability of the isolates, as some genes needed for sucrose fermentation were seen to be present in genomic islands. Noteworthy, using profile HMMs, we identified an enzyme with putative alpha-1,6-glycosidase activity only in some specific strains of *C. diphtheriae* and this may aid to understanding of the differential abilities to utilize glycogen and starch between the *biovars*.

**Keywords** *Corynebacterium* spp. · Bacterial identification · Biochemical tests · Comparative genomics

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10142-018-0610-3>) contains supplementary material, which is available to authorized users.

✉ Luis G. C. Pacheco  
luis.pacheco@ufba.br

<sup>1</sup> Bioinformatics Post-Graduate Program, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte, MG, Brazil

<sup>2</sup> Institute of Health Sciences, Federal University of Bahia (UFBA), Salvador, BA, Brazil

<sup>3</sup> Institute of Biological Sciences, Federal University of Pará (UFPA), Belém, PA, Brazil

<sup>4</sup> Faculty of Medical Sciences, Rio de Janeiro State University (UERJ), Rio de Janeiro, RJ, Brazil

## Introduction

The genus *Corynebacterium* includes a number of pathogenic bacterial species of importance for human and animal health (Bernard 2012). The most well-studied pathogen in this genus, *Corynebacterium diphtheriae*, is the main causative agent of diphtheria, which is a potentially fatal infection of humans. Though incidence of classical, toxin-mediated, diphtheria has considerably dropped due to widespread vaccination, there is a need for continued surveillance as novel cases remain being reported in various regions (Zakikhany and Efstratiou 2012; Forbes 2017). Also, there are several recent reports of outbreaks of cutaneous diphtheria, a highly contagious skin manifestation

of infection, including in developed countries (Meinel et al. 2016; Reynolds et al. 2016; Belchior et al. 2017; Kolios et al. 2017). Moreover, the emergences of toxigenic strains of *Corynebacterium ulcerans* in cases of diphtheria-like illness and of non-toxigenic strains of *C. diphtheriae* as causative agents of severe invasive infections have been matters of concern worldwide (Edwards et al. 2011; Viguetti et al. 2012; Zasada 2013; Hacker et al. 2016).

Similarly, novel cases of severe infections caused by non-diphtherial *Corynebacterium* spp. are increasingly being registered, involving the emerging pathogenic species *Corynebacterium striatum*, *Corynebacterium amycolatum*, *Corynebacterium minutissimum*, and, less frequently, *Corynebacterium xerosis* (Martins et al. 2009; Hahn et al. 2016; Chandran et al. 2016). These so-called diphtheroids can be found as normal constituents of the human microbiota, but their clinical significance has been demonstrated by various studies (Martins et al. 2009; Chandran et al. 2016; Hahn et al. 2016; Leal et al. 2016). Particularly, rapid acquisition of a multidrug resistant phenotype has been shown for the species *C. striatum*, and this justifies the need for rapid and accurate identification of diphtheroids to the species level (Hahn et al. 2016; Ajmal et al. 2017).

Traditional phenotypic identification methods based on batteries of biochemical reactions, such as the API Coryne system (bioMérieux) or the RapID CB Plus Kit (Thermo Fisher), are the most widely employed to identify corynebacterial isolates to the species level (Bernard 2012; Bernard and Funke 2012). These tests are considered as efficient for identification of the species *C. diphtheriae* (Zakikhany and Efstratiou 2012), though some biochemical variability can be observed between isolates of different *biovars*, especially in the tests for nitrate reduction and utilizations of sucrose, glycogen, maltose, and starch (Hall et al. 2010; Bernard and Funke 2012; Sangal et al. 2014). Noteworthy, recent genomic studies have already challenged the usefulness of classifying *C. diphtheriae* isolates into *biovars* with basis on biochemical tests, due to extensive genomic diversity within the species (Sangal and Hoskisson 2016). Additionally, the effectiveness of biochemical tests for unambiguous identification of some diphtheroids may vary considerably (reported range = 41.7–90.5% success) and might require additional tests (Funke et al. 1997; Renaud et al. 1998; Almuzara et al. 2006; Adderson et al. 2008; Alibi et al. 2015). Confounding biochemical identification profiles are commonly reported for the species *C. striatum*, *C. amycolatum*, *C. minutissimum*, and *C. xerosis* (Funke et al. 1996; Renaud et al. 1998; Adderson et al. 2008; Palacios et al. 2010). Palacios et al. (2010) reported the existence of five different biochemical profiles for the *C. amycolatum* isolates studied and reported difficulties for identification of *C. xerosis* due to variabilities in the tests for nitrate reduction and alpha-glucosidase (Palacios et al. 2010). Biochemical variabilities of isolates of these four diphtheroids also frequently include the reactions of

sucrose, maltose, ribose, trehalose, and galactose utilizations (Funke et al. 1996; Renaud et al. 1998; Palacios et al. 2010; Bernard and Funke 2012).

In order to identify the genomic basis contributing to the biochemical variabilities observed in phenotypic identification methods of isolates of these emerging and reemerging species of the genus *Corynebacterium*, we combined a comprehensive literature review with a bioinformatics approach based on reconstruction of specific biochemical reactions/pathways in 33 recently released whole genome sequences. We used data retrieved from curated databases (MetaCyc, PathoSystems Resource Integration Center (PATRIC), The SEED, TransportDB, UniProtKB) associated with homology searches by BLAST and profile Hidden Markov Models (HMMs) to detect enzymes participating in the various pathways and performed ab initio protein structure modeling and molecular docking in order to confirm specific results. Our data demonstrate the usefulness of an integrated approach (literature review, homology searches, structural analysis) for successful predictions of phenotypes from whole genomic sequences. Noteworthy, the level of genome assembly (draft or complete) and the quality of annotation significantly impacted phenotypic predictions.

## Materials and methods

### Whole genome sequences and genome re-annotation

Table 1 lists all genomic sequences used in this study. In total, 33 whole genome sequences were retrieved from NCBI's Genome Database: 14 *C. diphtheriae* strains, 2 *C. ulcerans* strains, 10 *Corynebacterium striatum* strains, 2 *C. xerosis* strains, 3 *C. minutissimum* strains, and 2 *C. amycolatum* strains. To standardize genomic annotations, all genome sequences were re-annotated using RASTtk (Brettin et al. 2015), through the fully automated genome annotation functionality of the PATRIC (Wattam et al. 2017). When required, existing divergences between annotations were detected through comparative analysis using Artemis software (Rutherford et al. 2000) and NCBI's Sequence Viewer application (<http://www.ncbi.nlm.nih.gov/projects/sviewer/>). Annotations compared were from the International Nucleotide Sequence Database (INSDC), which are generated by authors, and from NCBI Reference Sequence Database (RefSeq), automatically generated by NCBI's Prokaryotic Genome Annotation Pipeline, besides newly re-annotated genome sequences using RASTtk.

### Literature review and identification of variable biochemical pathways

The target metabolic pathways were identified through literature review of studies aimed at evaluating the reliability of

**Table 1** Genomic sequences used in this study

Species/strains	Accession no.	Total size (Mbp)	Contigs	Genes	Pseudogenes	References
<i>Corynebacterium diphtheriae</i> (n = 14)						
<i>C. diphtheriae</i> 241	NC_016782.1	2.43	1	2252	35	Trost et al. (2012)
<i>C. diphtheriae</i> HC01	NC_016786.1	2.43	1	2255	25	Trost et al. (2012)
<i>C. diphtheriae</i> HC02	NC_016802.1	2.47	1	2285	38	Trost et al. (2012)
<i>C. diphtheriae</i> HC03	NC_016787.1	2.48	1	2280	29	Trost et al. (2012)
<i>C. diphtheriae</i> HC04	NC_016788.1	2.48	1	2299	44	Trost et al. (2012)
<i>C. diphtheriae</i> HC07	JRUZ00000000.1	2.49	70	2332	51	Encinas et al. (2015)
<i>C. diphtheriae</i> 31A	NC_016799.1	2.53	1	2383	60	Trost et al. (2012)
<i>C. diphtheriae</i> BH8	NC_016800.1	2.48	1	2359	68	Trost et al. (2012)
<i>C. diphtheriae</i> C7(beta)	NC_016801.1	2.50	1	2357	57	Trost et al. (2012)
<i>C. diphtheriae</i> NCTC 13129	NC_002935.2	2.49	1	2291	38	Cerdeño-Tárraga et al. (2003)
<i>C. diphtheriae</i> VA01	NC_016790.1	2.40	1	2196		Trost et al. (2012)
<i>C. diphtheriae</i> PW8	NC_016789.1	2.53	1	2355	89	Trost et al. (2012)
<i>C. diphtheriae</i> CDCE 8392	NC_016785.1	2.43	1	2270	67	Trost et al. (2012)
<i>C. diphtheriae</i> INCA 402	NC_016783.1	2.45	1	2279	50	Trost et al. (2012)
<i>Corynebacterium ulcerans</i> (n = 2)						
<i>C. ulcerans</i> 809	NC_017317	2.50	1	2275	32	Trost et al. (2011)
<i>C. ulcerans</i> BR-AD22	NC_015683	2.60	1	2407	49	Trost et al. (2011)
<i>Corynebacterium xerosis</i> (n = 2)						
<i>C. xerosis</i> ATCC 373	NZ_LAYS00000000.1	2.75	152	2497	448	Pacheco et al. (2015)
<i>C. xerosis</i> NBRC 16721	NZ_BCRD00000000.1	2.69	14	2375	109	NCBI BioProject #: PRJDB439
<i>Corynebacterium striatum</i> (n = 10)						
<i>C. striatum</i> 1961	NZ_ACGE00000000.1	2.72	176	2574	142	Mattos-Guaraldi et al. (2015)
<i>C. striatum</i> 1329 CAUR	NZ_JVTL00000000.1	2.90	94	2705	85	Roach et al. (2015)
<i>C. striatum</i> 1327 CAUR	NZ_JVTN00000000.1	2.91	94	2718	85	Roach et al. (2015)
<i>C. striatum</i> 963 CAUR	NZ_JULV00000000.1	2.83	151	2628	113	Roach et al. (2015)
<i>C. striatum</i> 962 CAUR	NZ_JULW00000000.1	2.83	141	2624	109	Roach et al. (2015)
<i>C. striatum</i> 797 CAUR	NZ_JUSN00000000.1	2.74	116	2518	106	Roach et al. (2015)
<i>C. striatum</i> 587 CAUR	NZ_JVBA00000000.1	2.78	184	2558	129	Roach et al. (2015)
<i>C. striatum</i> 542 CAUR	NZ_JVCW00000000.1	2.83	91	2592	97	Roach et al. (2015)
<i>C. striatum</i> 2230	LTBG00000000.1	2.89	37	2721	158	NCBI BioProject #: PRJNA313271 <sup>a</sup>
<i>C. striatum</i> 2245	LTBG00000000.1	2.92	14	2700	365	NCBI BioProject #: PRJNA313271 <sup>a</sup>
<i>Corynebacterium minutissimum</i> (n = 3)						
<i>C. minutissimum</i> 1941	NZ_JSEF00000000.1	2.66	38	2450	61	Pacheco et al. (2015)
<i>C. minutissimum</i> NBRC15371	NZ_BCRC00000000	2.66	42	2456	62	NCBI BioProject #: PRJDB439
<i>C. minutissimum</i> ATCC 23348	NZ_JSEF00000000.1	2.66	38	2450	61	NCBI BioProject #: PRJNA264738
<i>Corynebacterium amycolatum</i> (n = 2)						
<i>C. amycolatum</i> Sk46	NZ_ABZU00000000.1	2.51	48	2167	26	NCBI BioProject #: PRJNA31007
<i>C. amycolatum</i> ICIS53	NZ_MIFV00000000.1	2.46	41	2172	34	Gladysheva et al. (2016)

<sup>a</sup> Whole genome sequencing, assembly, and annotations were made available by Universidade Federal do Pará (UFPA), in November 2017. Strains 2230 and 2245 of *C. striatum* were isolated in the Faculty of Medical Sciences, Rio de Janeiro State University (UERJ), Rio de Janeiro, Brazil

various biochemical reactions for differential identification of the six *Corynebacterium* species studied: *C. diphtheriae*, *C. ulcerans*, *C. striatum*, *C. amycolatum*, *C. minutissimum*, and *C. xerosis* (Martinez-Martinez et al. 1995; Funke et al. 1996, 1997; Renaud et al. 1998, 2001; Wauters et al. 1998; Almuzara et al. 2006; Letek et al. 2006; Adderson et al. 2008; Funke and Frodl 2008; Palacios et al. 2010; Bernard 2012; Bernard and Funke 2012; Torres et al. 2013). Reactions that

are commonly used in phenotypic identification methods and that presented conflicting results between the various studies were selected: (i) nitrate reduction and utilization of (ii) sucrose, (iii) maltose, (iv) glycogen, (v) galactose, and (vi) ribose (Table 2). These biochemical reactions are part of the most traditionally used commercial identification kits, such as the API Coryne system (bioMérieux) (Table 2).

## Retrieval of enzyme commission numbers for each pathway

The general workflow of the bioinformatics analysis is shown in Fig. 1. The various enzymatic reactions that compose each of the six variable biochemical pathways were initially retrieved from the MetaCyc database ([www.metacyc.org](http://www.metacyc.org)). The following pathways were selected: sucrose degradation I

**Table 2** Common biochemical profiles of isolates of *C. diphtheriae*, *C. ulcerans*, and *C. xerosis*/*C. striatum*/*C. minutissimum*/*C. amycolatum* (XSMA), according to literature review

Biochemical reaction <sup>a</sup>	<i>Cd</i>	<i>Cu</i>	<i>Cx</i>	<i>Cs</i>	<i>Cm</i>	<i>Ca</i>
<b>NIT</b>	+/( <i>belfanti</i> )	–	+/-	+/- <sup>c</sup>	–	+/-
<b>PYZ</b>	–	–	+	+	+	+/-
<b>PyrA</b>	–	–	–	–	–	–
<b>PAL</b>	–	+	+	+	+	+
<b>βGUR</b>	–	–	–	–	–	–
<b>βGAL</b>	–	–	–	–	–	–
<b>αGLU</b>	+	+	+/-	–	–	+/-
<b>BNAG</b>	–	–	–	–	–	–
<b>ESC</b>	–	–	–	–	–	–
<b>URE</b>	–	+	–	–	–	+/-
<b>GEL</b>	–	+	–	–	–	–
<b>O</b>	–	–	–	–	–	–
<b>GLU</b>	+	+	+	+/-	+	+/-
<b>RIB</b>	+	+	+	+/-	+/-	+/-
<b>XYL</b>	–	–	–	–	–	–
<b>MAN</b>	–	+/-	–	–	+/-	–
<b>MAL</b>	+/- <sup>b</sup>	+	+	–	+	+/-
<b>LAC</b>	–	–	–	+/-	–	+/-
<b>SAC</b>	+/- <sup>b</sup>	+/-	+	+/-	+/-	+/-
<b>GLYG</b>	+( <i>gravis</i> )/-	+/-	–	–	–	–
<b>CAT</b>	+	+	+	+	+	+

*Cx* *C. xerosis*, *Cs* *C. striatum*, *Cm* *C. minutissimum*, *Ca* *C. amycolatum*, *Cd* *C. diphtheriae*, *Cu* *C. ulcerans*, **NIT** nitrate reduction, **PYZ** pyrazinamidase, **PyrA** pyrrolidonyl arylamidase, **PAL** alkaline phosphatase, **βGUR** β-glucuronidase, **βGAL** β-galactosidase, **αGLU** α-glucosidase, **BNAG** N-acetyl-β glucosaminidase, **ESC** esculin, **URE** urease, **GEL** gelatin hydrolysis, **O** fermentation control, **GLU** glucose fermentation, **RIB** ribose fermentation, **XYL** xylose fermentation, **MAN** mannitol fermentation, **MAL** maltose fermentation, **LAC** lactose fermentation, **SAC** sucrose fermentation, **GLYG** glycogen fermentation, **CAT** catalase, + positive, – negative, +/- variable

<sup>a</sup> Biochemical reactions that compose the API Coryne system (bioMérieux). The five most variable reactions among the studied species are marked bold

<sup>b</sup> A negative reaction for maltose fermentation and a positive reaction for sucrose fermentation are atypical findings for the species *C. diphtheriae* (Hall et al. 2010; Mattos-Guaraldi and Formiga 1998)

<sup>c</sup> A negative result for nitrate reductase activity is a variable finding for *C. striatum* (Martinez-Martinez et al. 1995; Baio et al. 2013)

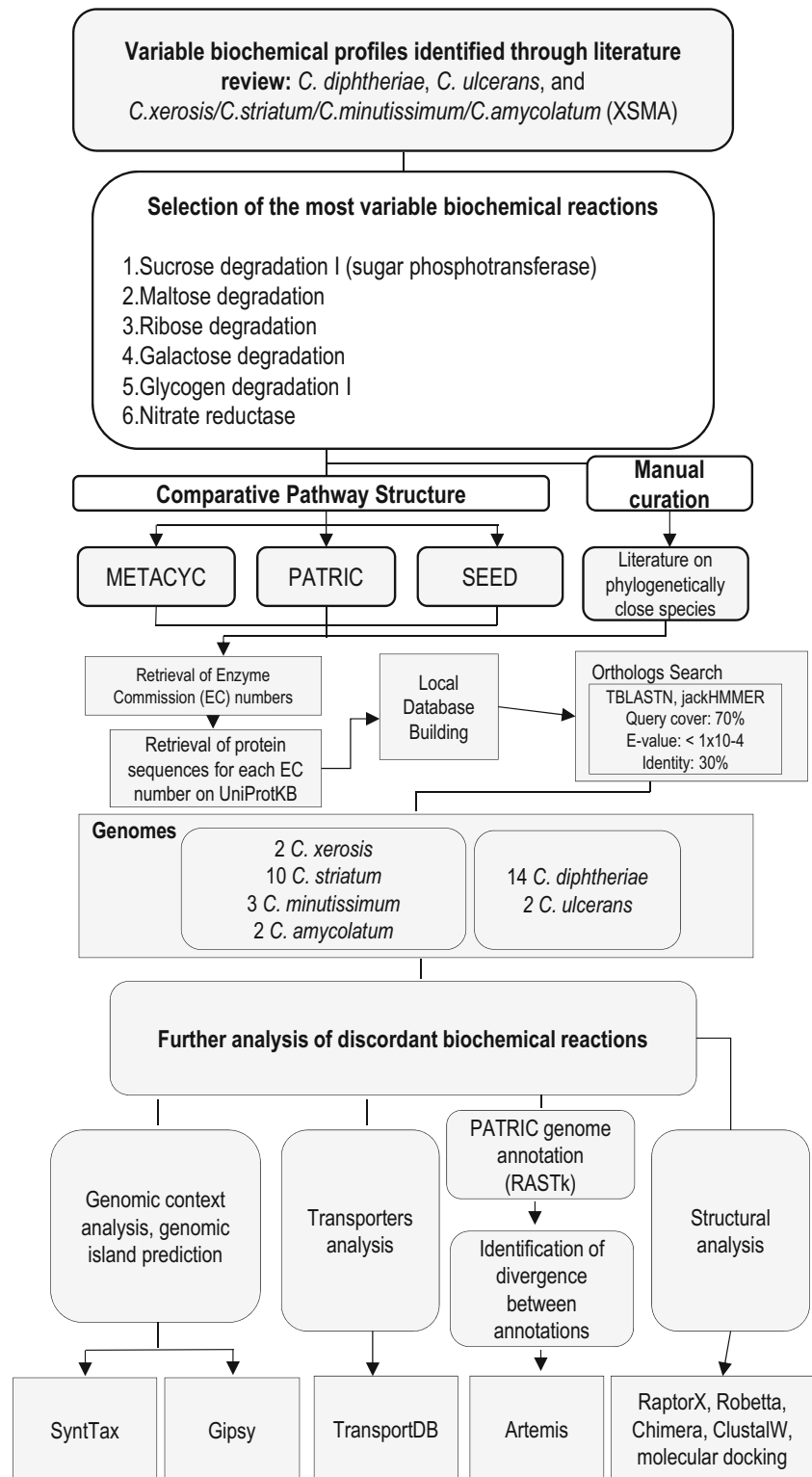
(sucrose phosphotransferase), maltose degradation, ribose degradation, D-galactose degradation I, glycogen degradation I, and nitrate reductase. The enzymatic reactions composing each curated pathway were retrieved through the associated enzyme commission (EC) numbers, using the reference bacterial species available in the database. For comparison with MetaCyc, the manually curated subsystems in the SEED (Overbeek et al. 2005) that correspond to each of the biochemical pathways were also retrieved, this time choosing phylogenetically close bacterial species. Additionally, automatic metabolic reconstructions for each of the studied genomes were retrieved from PATRIC v.3.3.10 ([www.patricbrc.org/](http://www.patricbrc.org/)), using the Comparative Pathway tool.

## Ortholog search

The EC numbers that compose each of the studied biochemical pathways were used to retrieve protein sequences from UniProtKb ([www.uniprot.org](http://www.uniprot.org)); sequences from phylogenetically close organisms were prioritized. A local nucleotide database was created with the 33 whole genome sequences, using the makeblastdb tool, available at the NCBI's BLAST+ suite. Similarity searches for orthologs were done with the tBLASTn stand-alone version 2.6.0 (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>). Cutoff values were as follows: *E* value ≤ 10E-4, query cover ≥ 70%, and identity ≥ 30%. When required, ortholog searches were performed using profile Hidden Markov Models, with jackHMMER (Finn et al. 2015).

## Literature-based curation of carbohydrate utilization pathways and mapping of transporters

To support automatic predictions of metabolic pathways, a literature review of experimentally confirmed carbohydrate utilization pathways in the species *Corynebacterium glutamicum*, a biotechnologically important species, was performed to identify genes demonstrated to participate in the metabolism of hexoses (glucose and fructose), pentoses (D-ribose, L-arabinose, D-xylose), disaccharides (sucrose and maltose), and other relevant carbohydrates. Information was also retrieved on carbohydrate transport in *C. glutamicum* using phosphoenolpyruvate-dependent sugar phosphotransferase (PTS) and non-PTS, ATP-binding cassette (ABC)-type, systems (Blombach and Seibold 2010; Ikeda 2012). Additionally, sugar transport systems were predicted in the whole genome sequences using TransportDB 2.0 relational database (Elbourne et al. 2017).

**Fig. 1** General workflow of the bioinformatics analysis

### Prediction of genomic islands, genomic context analysis, and pan-genome comparisons

Metabolic genomic islands were predicted in the studied genomes using GIPSY software (Soares et al. 2016). The genomic context of the genes belonging to the various biochemical

pathways was analyzed with the SyntTax tool (<http://archaea.u-psud.fr/SyntTax/>), using default parameters (Oberto 2013). A pan-genome analysis was performed with re-annotated genome sequences using the Bacterial Pan-Genome Analysis (BPGA) pipeline (Chaudhari et al. 2016), to identify shared proteins and genome-specific proteins.

## Protein structural analysis

A structural approach was used to aid prediction of functionally of the mapped PTS-type carbohydrate transporters and of strain-restricted proteins. Ab initio 3D protein modeling was performed with the RaptorX (Källberg et al. 2012) and Robetta (Song et al. 2013) servers. Template-based modeling was performed with the software Modeller v9.18. Protein structure refinement was done with i3Drefine (Bhattacharya and Cheng 2013). Protein structure evaluation and validation was done with Ramachandran plots generated with RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage2.php>), and Z score values and energy distribution plots obtained with ProSA-web (Wiederstein and Sippl 2007). Structural alignments were done with the RaptorX structure alignment server (Wang et al. 2011, 2013) and visualized using UCSF Chimera (Pettersen et al. 2004). Functional domains were predicted with Conserved Domains Database (CDD) (Marchler-Bauer et al. 2014) and CATH structure database (Dawson et al. 2017). Multiple protein sequence alignments were performed through PRABI server ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_server.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html)). Molecular docking was performed using the software autodocktools v1.5.6 (Huey and Morris 2003).

## Results and discussion

### Variable biochemical reactions

The target biochemical reactions were identified through literature review of conflicting results obtained in phenotypic identification studies of six emerging and reemerging pathogenic *Corynebacterium* species: *C. diphtheriae* (including biovar *gravis*, *mitis*, *belfanti*, and *intermedius*) and *C. ulcerans*; *C. xerosis*, *C. striatum*, *C. minutissimum*, and *C. amycolatum* (hereafter collectively referred to as X SMA group). Six biochemical reactions were identified as the most variable among the various isolates, of which five reactions are component of the most traditionally used identification test, the API Coryne system (bioMérieux): nitrate reduction and fermentations of ribose, maltose, sucrose, glycogen, and galactose (the latter not included in API Coryne) (Table 2). Bacterial identification by the API Coryne system is based on a numerical code generated by the APIweb application; this code combines results of 21 different reactions, and biochemical variabilities in the aforementioned reactions may render unreliable identifications.

The nitrate reduction (NIT) test normally renders positive results for the species *C. striatum*, though some studies have reported on nitrate-negative isolates (Table 2). It is also positive for the biovars *gravis*, *mitis*, and *intermedius* of *C. diphtheriae*, while the biovar *belfanti* is negative. The

species *C. ulcerans* and *C. minutissimum* are also expected to yield negative results, whereas variable results are reported for the species *C. xerosis* and *C. amycolatum* (Table 2).

Ribose fermentation (RIB) is positive for isolates of *C. diphtheriae*, *C. ulcerans*, and *C. xerosis* but variable for the species *C. striatum*, *C. minutissimum*, and *C. amycolatum* (Table 2). Maltose fermentation (MAL) presents variable results for *C. amycolatum* and is negative for *C. striatum*. This test is generally positive for *C. xerosis*, *C. ulcerans*, and *C. diphtheriae*, but there are reports of atypical *C. diphtheriae* isolates presenting negative results (Hall et al. 2010). Similarly, the sucrose fermentation (SAC) test, which is mostly negative for the species *C. diphtheriae*, might present some atypical positive results for some isolates (Mattos-Guaraldi and Formiga 1998). This reaction is expected to be positive for *C. xerosis* but is variable for all the other species (Table 2). Glycogen utilization (GLYG) presents variability within the species *C. diphtheriae*, with isolates of the biovar *gravis* normally being expected to render a positive result; this test is also expected to render positive results for *C. ulcerans* but is negative for the remaining four species studied (Table 2).

### Enzymatic reactions composing each biochemical pathway

A comparative analysis was performed using data from different databases and literature-based curation, in order to identify key enzymatic reactions and potential variations in the enzymatic activities that compose the six target metabolic pathways (Fig. 1). Annotated EC numbers for the various enzymatic reactions were retrieved from the MetaCyc database for the following bacterial species: *Bacillus subtilis*, *Escherichia coli*, *Lactobacillus brevis*, *Lactococcus lactis*, and *Neisseria meningitidis*. From The SEED database, we retrieved annotated enzymatic reactions for the species *C. diphtheriae*. Finally, a literature review of the various reactions was performed for the species *C. glutamicum*. This latter species was chosen for manual curation of pathways.

We identified variations in the annotated carbohydrate utilization pathways in the different sources (Supplementary Figs. S1 and S2). The maltose utilization pathway in *C. glutamicum* requires internalization by the ABC-type transporter MuseFGK2I (EC 3.6.3.19) (Henrich et al. 2013); then, it involves the actions of a 4- $\alpha$ -glucanotransferase (encoded by *malQ* gene) (EC 2.4.1.25) and of maltose phosphorylase (*malP* gene) (EC 2.4.1.8), glucokinase (*glk* gene) (EC 2.7.1.2), and phosphoglucomutase (*pgm* gene) (EC 5.4.2.2), to generate maltodextrins, glucose-1-phosphate, and glucose-6-phosphate (Blombach and Seibold 2010; Clermont et al. 2015). The sucrose degradation pathway in this species involves the transport of this carbohydrate by a specific PTS-type transporter (EC 2.7.1.191), followed by cleavage of

sucrose-6-phosphate by a sucrose hydrolase (*scrB* gene) (EC 3.2.1.-), to generate glucose-6-phosphate and fructose; unlike annotations for various bacteria in MetaCyc and the SEED, *C. glutamicum* does not present fructokinase (*scrK* gene) activity (EC 2.7.1.4) (Supplementary Figs. S1 and S2). Alternatively, it is believed to export fructose by an as-yet-unidentified efflux mechanism and then this sugar is re-imported through a fructose-specific PTS to yield fructose-1-phosphate; it can then be phosphorylated by the enzyme FruK 1-phosphofructokinase (*pfkB* gene) to generate the glycolytic intermediate fructose-1,6-bisphosphate (Moon et al. 2005; Engels et al. 2008; Ikeda 2012) (Supplementary Figs. S1 and S2). Glycogen utilization involves the following reactions: cleavage of  $\alpha$ -1-4-glycosidic linkages by glycogen phosphorylase (*glgP* gene) (EC 2.4.1.1); generation of maltodextrins by the glycogen debranching enzyme GlgX (*glgX* gene) (EC 3.2.1.196); and generation of glucose and glucose-6-phosphate involves the activities of the enzymes maltodextrin phosphorylase (*malP* gene) (EC 2.4.1.1), 4- $\alpha$ -glucanotransferase (*malQ* gene) (EC 2.4.1.125), maltodextrin glucosidase (*malZ* gene) (EC 3.2.1.20), glucokinase (*glk* gene) (EC 2.7.1.2), and alpha-phosphoglucosyltransferase (*pgm* gene) (EC 5.4.2.2) (Seibold et al. 2009; Von Zaluskowski 2015) (Supplementary Figs. S1 and S2). D-ribose utilization involves the activity of the RbsACBD transporter (EC 3.6.3.17) and generation of D-ribose-5-phosphate by RbsK1 and RbsK2 ribokinases (*rbsK1* and *rbsK2* genes) (EC 2.7.1.15) (Nentwich et al. 2009). Since *C. glutamicum* is not able to utilize galactose, we retrieved EC numbers from the annotated pathway in *E. coli*; the pathway involves conversion of  $\beta$ -D-galactose in  $\alpha$ -D-galactose by galactose-1-epimerase (*galM* gene) (EC 5.1.3.3), phosphorylation by galactokinase (*galK* gene) (EC 2.7.1.6), transfer of UMP from UDP-glucose to  $\alpha$ -D-galactose-1-phosphate by galactose-1-phosphate uridylyl transferase (*galT* gene) (EC 2.7.7.12), and conversion of UDP-galactose in UDP-glucose by UDP-glucose-4-epimerase (*galE* gene) (EC 5.1.3.2); phosphoglucosyltransferase (*pgm* gene) (EC 5.4.2.2) is then required to convert glucose-1-phosphate in glucose-6-phosphate (Holden et al. 2003) (Supplementary Figs. S1 and S2).

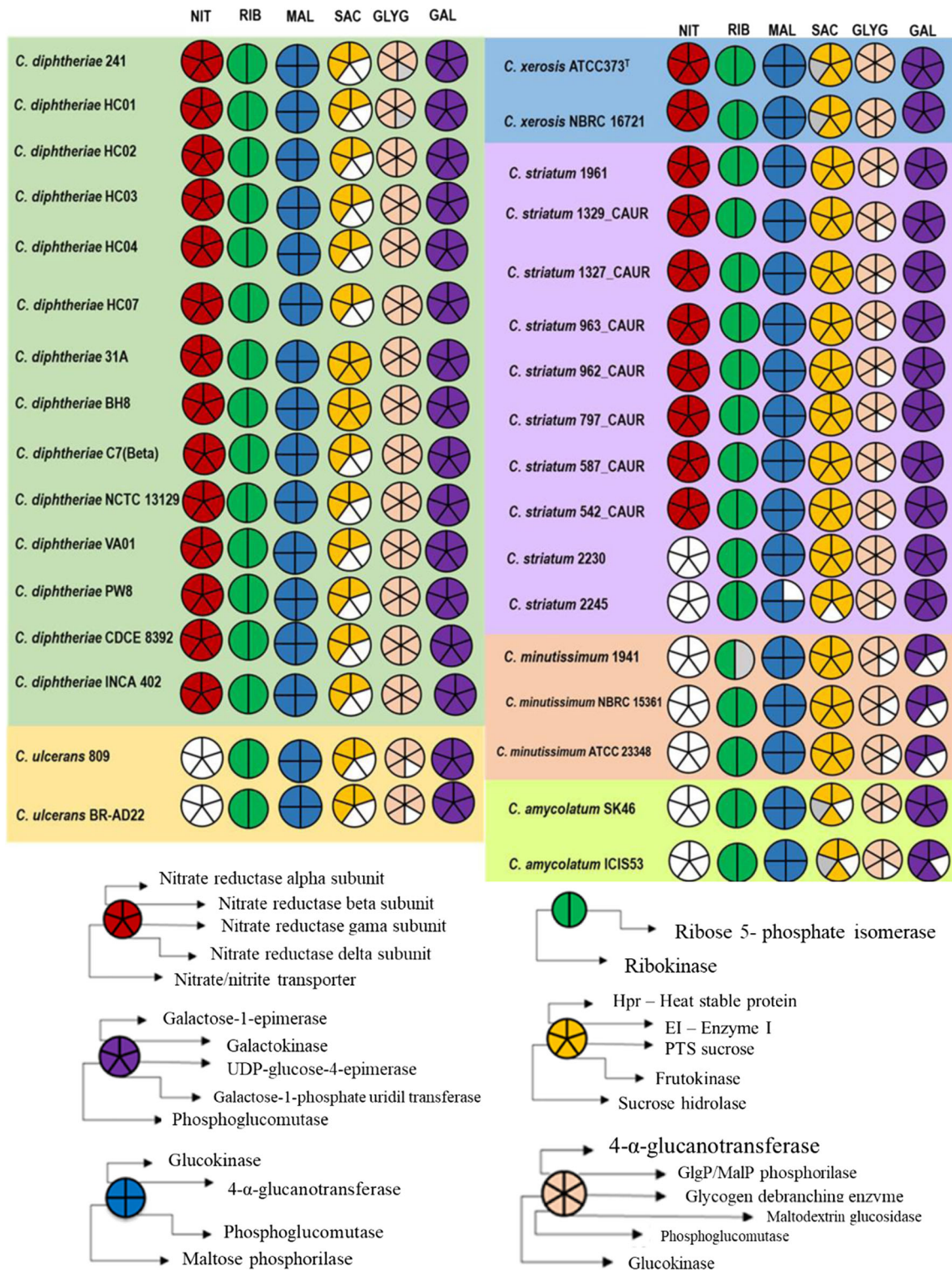
Finally, the nitrate reductase activity (EC 1.7.5.1) in *C. glutamicum* involves the *narKGHJI* gene cluster (Nishimura et al. 2007), with high similarity to the system annotated for *E. coli* and other bacteria in MetaCyc. It involves the transport of nitrate ions by the NarK transporter and nitrate reduction by the alpha subunit of NarG. This subunit contains the guanylyl molybdenum cofactor and there is an iron-sulfur cluster [4Fe-4S]. The beta subunit NarH, which contains three [4Fe-4S] and one [3Fe-4S] iron-sulfur clusters, is anchored to the membrane, together with NarG, by the NarI gamma subunit, that contains two heme groups for electron transfer; delta subunit NarJ is needed for assembly of the complex (Rothery et al. 1998; Almeida et al. 2017).

## Identification of the various enzyme genes in the studied genomes

Homology searches using the tBLASTn tool were initially performed against a local database, using protein sequences retrieved from UniProtKb that represent each of the enzymatic activities involved in the variable biochemical reactions. The distribution of the various enzyme genes in the 33 studied genomes is shown in Fig. 2. The differential occurrence of the genes is also color-coded in Supplementary Figs. S1 and S2. Complete results of the BLAST searches for each genome are presented in supplementary material (Supplementary Table S1).

In accordance with expected phenotypic results, all genes required for the nitrate reduction pathway were found in most strains of *C. striatum*, *C. xerosis*, and *C. diphtheriae* (Fig. 2 and Table 2). Two strains of *C. striatum* (2230 and 2245) and all studied strains of *C. ulcerans*, *C. minutissimum*, and *C. amycolatum* did not present any of the genes required for respiratory nitrate reductase activity (Fig. 2). As for the biovar *belfanti* of *C. diphtheriae*, we identified through BLAST searches that the nitrate reductase pathway might be non-functional in strain INCA 402, due to the absence of the *narG* gene. Although this is in accordance with the expected phenotypic profile for strains of the biovar *belfanti*, we could later relate this finding to incorrect assembly of the genomic region that contained the *narGHJ* operon in this strain, in the genome release used. Nevertheless, Sangal et al. (2014) have previously reported that this *belfanti* strain indeed possesses a mutated *narJ* gene that might explain the nitrate-negative phenotype. We further investigated this in four newly released draft genomic sequences of isolates assigned to the biovar *belfanti* (strains 631, 1734, 1137, and 2937—GenBank accession numbers GCA\_002202425.1, GCA\_002202575.1, GCA\_002202505.1, GCA\_002202655.1). Although the *narKGHJ* gene cluster was complete in all strains, the gene coding for the molybdopterin biosynthesis enzyme MoaB, which is a cofactor for nitrate reductase activity, is annotated as pseudogene due to frameshifts in three *belfanti* strains (data not shown).

The main enzyme gene participating in the ribose utilization pathway, *rbsK*, which codes for ribokinase, was found in all genomic sequences, along with *rpiA*, coding for ribose-5-phosphate isomerase (Fig. 2). Unlike *C. glutamicum* that has two ribokinases, the six species studied only presented one ribokinase gene. Given the low complexity of the main reactions in the pathway, the observed variability could potentially originate from the functionality of the ribose-specific transporter. According to Ikeda (2012), ribose uptake in *C. glutamicum* occurs through the RbsACBD transporter (Nentwich et al. 2009). However, only *C. minutissimum* 1941 presented all four components of the transporter. The other strains showed only the RbsABC components (*C. diphtheriae*, *C. ulcerans*,



**Fig. 2** Graphical representation of tBLASTn-based homology searches on the local database containing genomic sequences of the 33 strains. Pie charts represent the different metabolic pathways and each fraction corresponds to an essential protein/subunit required for that pathway. Red:

nitrate reductase reaction. Green: ribose utilization. Blue: maltose utilization. Yellow: sucrose utilization. Pink: glycogen utilization. Purple: galactose utilization. Gray fractions indicate genes annotated as pseudogenes and white fractions indicate the absence of specific genes

*C. striatum*, and *C. minutissimum*) or RbsAC (*C. xerosis* and *C. amycolatum*) (Tables 3, 4, 5, and 6). According to Clifton et al. (2015), the ribose-specific ABC transporter in *E. coli*

differs from other ABC transporters, being composed only by RbsABC2 subunits; the two ATPase components are fused to the RbsA subunit. Van Der Heide and Poolman (2002)



**Table 3** Genomic status of transporter systems in *C. diphtheriae* strains

Transporter system	Protein subunits involved	Genetic support on chromosome
PTS sucrose	EI, EIIA <sup>suc</sup> , EIIB <sup>suc</sup> , EIIC <sup>suc</sup> , and Hpr	The strains 31A and BH8 were the only <i>C. diphtheriae</i> strains with complete coding sequences
PTS fructose	EI, EIIA <sup>fru</sup> , EIIB <sup>fru</sup> , EIIC <sup>fru</sup> , and Hpr	Encoding genes are present but annotated as pseudogenes and assigned as nonfunctional due to frameshifts
PTS glucose	EI, EIIA <sup>glu</sup> , EIIB <sup>glu</sup> , EIIC <sup>glu</sup> , and Hpr	All coding sequences are annotated as functional. There is high similarity with PTS sucrose, generating ambiguous annotations
Galactose ATP-binding cassette (ABC) transporter	MglA, MglB, and MglC	All genes are present; however, there is some ambiguity in annotations due to other carbohydrate transport systems
Ribose ATP-binding cassette (ABC) transporter	RbsA, RbsB, RbsC, and RbsD	All strains had <i>rbsABC</i> but lacked <i>rbsD</i> ; the transport system might require only the subunits RbsABC in this species
Maltose ATP-binding cassette (ABC) transporter	MalK, MalE, MalF, and MalG	All strains presented coding sequences for the four components
Nitrate transporter	NarK	All strains had the complete coding sequence

identified the existence of chimeric ABC proteins having the substrate-binding subunit (SBP) fused to the translocator, in which case the RbsC translocating subunit could have the substrate-binding site incorporated into the sequence. Also, the existence of low-affinity ribose transporters cannot be ruled out; as shown by Kim et al. (1997) for *E. coli*, the D-alose ABC transporter is also capable of transporting ribose. Considering the absence of the *rbsD* gene in the majority of strains in this study, we hypothesize that the ribose transport in these Corynebacteria may follow a pathway very similar to the one in *E. coli*.

In the maltose utilization pathway, all studied genomes also presented the four main enzymes predicted according to the annotated pathway in *C. glutamicum* (Bloombach and Seibold 2010) (Fig. 2); nevertheless, the presence of the *pgmB* gene (coding for beta-phosphoglucomutase) in some species (*C. xerosis* and *C. amycolatum*) is an indication that the degradation pathway might also follow an alternative pathway

that involves the genes *malP*, *pgmB*, and *glk*, as described by Andersson and Rådström (2002). The essentiality of beta-phosphoglucomutase activity for maltose metabolism has also been demonstrated for *L. lactis* and *Streptococcus mutans* (Buckley et al. 2014).

*C. striatum* is phenotypically negative for maltose utilization but possessed all the genes for maltose degradation according to the pathway annotated in *C. glutamicum*. By using a refined search using jackHMMER, we could also find in this species all the components of the maltose-specific transporter MusKEFG (Henrich et al. 2013) (Tables 4 and 6); thus, it is feasible to propose that this species might need beta-phosphoglucomutase for efficient maltose utilization. *C. minutissimum*, in turn, did not present the *pgmB* gene but is phenotypically positive for the maltose fermentation test. This species presented all genes for the MusKEFG transporter components, so the pathway used for maltose utilization seems to be closer to the functional variant of *C. glutamicum*

**Table 4** Genomic status of transporter systems in strains of the XSMA group

Transporter system	Protein subunits involved	Genetic support on chromosome
PTS sucrose	EI, EIIA <sup>suc</sup> , EIIB <sup>suc</sup> , EIIC <sup>suc</sup> , and Hpr	All the strains of the <i>C. xerosis</i> and <i>C. amycolatum</i> lacked the transporter. <i>C. striatum</i> and <i>C. minutissimum</i> strains had all subunits
PTS fructose	EI, EIIA <sup>fru</sup> , EIIB <sup>fru</sup> , EIIC <sup>fru</sup> , and Hpr	All the strains had all the components of the transporter
PTS glucose	EI, EIIA <sup>glu</sup> , EIIB <sup>glu</sup> , EIIC <sup>glu</sup> , and Hpr	All the strains had all the components of the transporter
Galactose ATP-binding cassette (ABC) transporter	MglA, MglB, and MglC	All <i>C. striatum</i> and <i>C. minutissimum</i> strains had alignments to all components; however, the annotation indicates ambiguous or different function
Ribose ATP-binding cassette (ABC) transporter	RbsA, RbsB, RbsC, and RbsD	Only <i>C. minutissimum</i> 1941 had the <i>rbsD</i> gene; the remaining strains presented a transporter similar to the one in <i>E. coli</i> , consisting only of RbsABC
Maltose ATP-binding cassette (ABC) transporter	MalK, MalE, MalF, and MalG	All <i>C. minutissimum</i> strains and <i>C. amycolatum</i> ICIS53 had all components
Nitrate transporter	NarK	All the <i>C. xerosis</i> and <i>C. striatum</i> strains had the complete coding sequence, although annotated as pseudogene in strain ATCC 373 of <i>C. xerosis</i>

**Table 5** Genomic status of the variable reactions in 14 *C. diphtheriae* strains

Biochemical reaction	Predicted phenotypes	Genes involved	Genetic support on chromosome
Maltose utilization	Potentially utilized by 14 strains	<i>pgm</i> , <i>glk</i> , <i>malQ</i> , <i>malP</i> , <i>malE</i> , <i>malF</i> , <i>malG</i> , and <i>malK</i>	All 14 genomes showed the presence of the basic genetic components for maltose utilization. The ATP-binding cassette (ABC)-type transporter was complete in all strains
Nitrate reduction	Potentially performed by 14 strains	<i>narG</i> , <i>narH</i> , <i>narI</i> , <i>narJ</i> , and <i>narK</i>	The strain INCA 402 (biovar <i>belfanti</i> ) is the only negative strain. As demonstrated previously (Sangal et al. 2014), an insertion at the 3'-end in <i>narJ</i> inactivates this gene in this strain
Sucrose utilization	Potentially utilized by 2 strains	<i>scrA</i> , <i>ptsF</i> , <i>scrB</i> , <i>scrK</i> , EI, and Hpr	The genes <i>scrA</i> and <i>scrK</i> were identified inside a predicted genomic island on strains 31A and BH8. All other strains lack <i>scrK</i> (fructokinase), similarly to <i>C. glutamicum</i> ; the transporter <i>ptsF</i> is annotated as non-functional due to frameshift in all strains
Glycogen utilization	Potentially utilized by 14 strains	<i>glgP</i> , <i>glgX</i> , <i>malZ</i> , <i>pgm</i> , <i>malP</i> , and <i>glk</i>	All 14 genomes had all genetic components. <i>glgP</i> is annotated as a pseudogene due to frameshift in strains 241 and HC01. However, only strains of the biovar <i>gravis</i> and one strain of the biovar <i>mitis</i> presented an ortholog of <i>spuA</i>
Galactose utilization	Potentially utilized by 14 strains	<i>mglA</i> , <i>mglB</i> , <i>mglC</i> , <i>galE</i> , <i>galM</i> , <i>galK</i> , <i>galT</i> , and <i>pgm</i>	All 13 genomes had all genetic components; orthologs of the galactose specific ABC transporter MglABC were detected
Ribose utilization	Potentially utilized by 14 strains	<i>rbsA</i> , <i>rbsB</i> , <i>rbsC</i> , <i>rbsD</i> , <i>rbsK1</i> , and <i>rbsK2</i>	All 13 strains lacked the subunit RbsD of the ribose-specific transporter RbsABCD; however, uptake may be functional with only the RbsABC components

**Table 6** Genomic status of variable reactions in 17 strains of the XSMA group

Biochemical reaction	Predicted phenotypes	Genes involved	Genetic support on chromosome
Maltose utilization	Potentially utilized by 4 strains	<i>pgm</i> , <i>glk</i> , <i>malQ</i> , <i>malP</i> , <i>malE</i> , <i>malF</i> , <i>malG</i> , and <i>malK</i>	The presence of the gene <i>pgmB</i> (beta-phosphoglucomutase) in <i>C. xerosis</i> and <i>C. amycolatum</i> strains suggest that maltose degradation in these species could be different from the pathway in <i>C. glutamicum</i>
Nitrate reduction	Potentially performed by 9 strains	<i>narG</i> , <i>narH</i> , <i>narI</i> , <i>narJ</i> , and <i>narK</i>	All <i>C. xerosis</i> and most <i>C. striatum</i> strains had all necessary genes. All genes were in the same operon. The <i>narK</i> gene was annotated as pseudogene in <i>C. xerosis</i> ATCC 373. Two <i>C. striatum</i> strains (2230 and 2245) did not present any of the genes
Sucrose utilization	Potentially utilized by 12 strains	<i>scrA</i> , <i>ptsF</i> , <i>scrB</i> , <i>scrK</i> , EI, and Hpr	All genes were identified for most <i>C. striatum</i> and <i>C. minutissimum</i> strains suggesting that sucrose pathway is different from <i>C. glutamicum</i> due to the presence of <i>scrK</i> ; the <i>C. xerosis</i> and <i>C. amycolatum</i> strains lacked <i>scrB</i> and <i>scrA</i>
Glycogen utilization	Potentially utilized by 14 strains	<i>glgP</i> , <i>glgX</i> , <i>malZ</i> , <i>pgm</i> , <i>malP</i> , and <i>glk</i>	Most strains lack the <i>malZ</i> gene, but the essential <i>glgX</i> and <i>glgP</i> genes are present in 14 strains. <i>glgX</i> is annotated as pseudogene in one strain of <i>C. xerosis</i> . The intracellular potential of glycogen utilization in glycogen-negative species indicates that differential transport may be a key component
Galactose utilization	Potentially utilized by 10 strains	<i>mglA</i> , <i>mglB</i> , <i>mglC</i> , <i>galE</i> , <i>galM</i> , <i>galK</i> , <i>galT</i> , and <i>pgm</i>	All <i>C. striatum</i> strains had all genes; both <i>C. xerosis</i> and <i>C. amycolatum</i> lacked the <i>galK</i> gene. Orthologs of the galactose specific ABC transporter MglABC were detected
Ribose utilization	Potentially utilized by 14 strains	<i>rbsA</i> , <i>rbsB</i> , <i>rbsC</i> , <i>rbsD</i> , <i>rbsK1</i> , and <i>rbsK2</i>	Similarly to <i>C. diphtheriae</i> , all the species (except <i>C. minutissimum</i> ) lacked the subunit RbsD of the ribose-specific transporter RbsABCD; however, uptake may be functional with only the RbsABC components

(Table 4). The distribution of homologs of the *malZ* gene in the strains is controversial, as only strains of *C. xerosis* and *C. diphtheriae* presented homologs with identities above 60%. However, the predicted proteins that are encoded by these genes, when evaluated in databases of protein domain prediction, indicate an enzyme from the same protein family but with slightly different catalytic site. Whereas the maltodextrin glycosidase MalZ has catalytic activity at alpha-1,4-glycosidic linkages, the alignment results in CATHdb indicate an oligo-1,6 glycosidase. Seibold et al. (2009) also failed to identify the *malZ* gene in *C. glutamicum*, though they have demonstrated experimentally the expected enzymatic activity.

In the sucrose utilization pathway, the main enzyme for differentiation between functional alternative pathways is fructokinase (encoded by the *scrK* gene); the ability to phosphorylate fructose in the cytoplasm is important to determine if the pathway follows the classical model, using the enzymes ScrB and ScrK and the transporter PTS(Suc), or the functional variant, in which fructose is exported to the extracellular environment and then imported through a fructose-specific PTS. Our data indicate that *C. diphtheriae* probably assimilates sucrose following a classic model that requires the three proteins, given the absence of PTS(Suc) in most strains. Additional evidence would be the annotation of the fructose-specific transporter PTS(Fru) in all strains of this species as a pseudogene, due to frameshifts. The absence of the ability to phosphorylate fructose through PTS(Fru) impairs sucrose utilization, as *C. diphtheriae* strains also do not possess a functional fructokinase. Nonetheless, several previous studies have reported on the identification of atypical sucrose-fermenting strains of *C. diphtheriae* (Pennie et al. 1996; Efstratiou and George 1996; Mattos-Guaraldi and Formiga 1998; Pimenta et al. 2008; Viguetti et al. 2012). We could identify the presence of the *ptsS*, *scrB* and *scrK* genes in *C. diphtheriae* strains 31A and BH8, confirming the strain-specific presence of the reaction. However, as reported by Viguetti et al. (2012), other strains used in this study also present a positive sucrose fermentation reaction (strains HC01, HC02, HC03, INCA 402, and 241); we only identified in these strains the presence of the *scrB* gene. This is controversial as the absence of the sucrose-specific PTS transporter PTS(Suc) would be sufficient to abolish the sucrose fermentation capacity (Moon et al. 2005).

In the strains of the XSMA group, the presence of the *scrK* gene in 14 genomic sequences indicates that the sucrose fermentation in these species may follow the classical pathway. The species *C. xerosis* and *C. amycolatum* presented the gene coding for the hydrolase ScrB, however annotated as pseudogene due to frameshifts. These species are phenotypically positive for sucrose fermentation according to the literature; however, the annotation indicated the presence of frameshifts in the sequence of *scrB* in both strains of *C. xerosis* and *C. amycolatum*, what would make it impossible to cleave the

sucrose molecules. Most strains of *C. minutissimum* and *C. striatum* showed all the necessary genes, with exception of strain 2245, which lacks a detectable *scrK* gene. Noteworthy, this strain is also phenotypically negative for fructose utilization (Ramos 2014). The high number of annotated pseudogenes in the genomic sequences of the *C. xerosis* strains raises the possibility that these have arisen due to assembly errors. Interestingly, the absence of PTS(Suc) and ScrK also occurs in the diphtheric species *C. pseudotuberculosis* and *C. ulcerans*, which have positive and variable phenotypes for sucrose fermentation, respectively. *C. pseudotuberculosis* is phylogenetically close to *C. diphtheriae* but does not present all the genes required for utilization of sucrose; no orthologs of the PTS(Suc) transporter could be identified in genomic sequences of *C. pseudotuberculosis* strains (data not shown). This points to the existence of an alternative sucrose utilization pathway operating in these species, but this needs to be explored further. We did investigate the possibility of alternative transport of sucrose in these species and found an *N*-acetylglucosamine-specific transporter with great structure similarity with the *C. glutamicum* transporter PTS(Suc) and with the ones found in the strains 31A and BH8 of *C. diphtheriae* (data not shown).

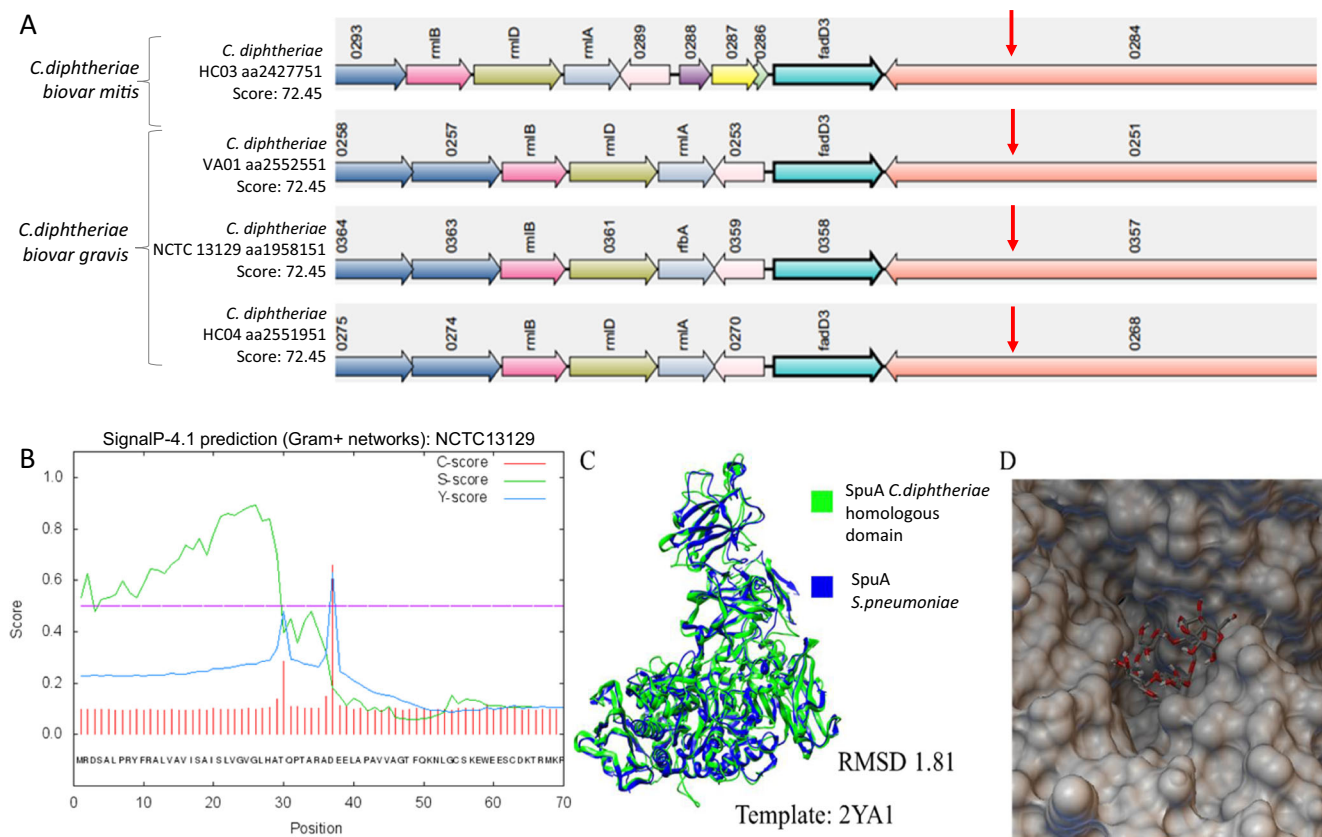
The glycogen degradation pathway relates to the maltose utilization pathway, as glycogen is formed by glucose chains of 8 to 12 branched molecules and connected by 1,4 and 1,6-glycosidic linkages between the branches (Berg et al. 2012), and maltose is formed by two glucose molecules connected by 1,4-glycosidic linkages; therefore, sharing of enzymatic reactions between the two pathways is expected. Among the shared enzymes are Glk, MalP, MalZ, MalQ, and  $\alpha$ -Pgm (Seibold et al. 2009).

According to literature review, all species of the XSMA group are negative for glycogen utilization, but all had homologs to the main genes of the metabolic pathway. Most of the strains only presented one  $\alpha$ -glucan phosphorylase (*malP* or *glgP* genes), indicating a possible difference with the metabolic pathway annotated for *E. coli* (Seibold et al. 2009). As the enzymes MalP and GlgP catalyze the same reaction, the presence of one enzyme is sufficient for functionality of the pathway. MalP and GlgX are the main enzymes for glycogen degradation (Von Zaluskowski 2015); annotations for *C. xerosis* ATCC 373 and *C. striatum* 1961 indicated a pseudogene in *glgX*, suggesting inability to metabolize glycogen. The *malQ* gene is also annotated as a pseudogene in the strain 1941 of *C. minutissimum*. Even though this gene is not essential, it could alter the dynamics of production of free glucose and decrease the rate of maltose absorption. The other genomic sequences presented all the genes with identities above 50%, except for *malZ* which was the one with the lowest identity among all strains.

Similarly to the findings for the XSMA group, all the strains of *C. diphtheriae* presented genes for glycogen

degradation. However, according to the literature, only strains of the biovar *gravis*, and eventually biovar *intermedius*, can utilize glycogen. The strains 241 and HC01, both of biovar *mitis*, presented frameshifts in the *glgP* gene, despite also possessing all genes. Although the glycogen utilization in *C. glutamicum* describes the main reactions needed for maintenance of the pathway intracellularly, they do not warrant the degradation of glycogen in the external environment. Due to its size, glycogen could not normally be transported through the bacterial membrane into the cytoplasm, raising doubts about the mechanism by which it is fragmented and transported in strains of the biovar *gravis*. Abbott et al. (2010) emphasize the importance of *spuA* and *malX* genes in the metabolism of extracellular glycogen in *Streptococcus pneumoniae*. The pneumococcal SpuA acts on glycogen depolymerization through the hydrolysis of  $\alpha$ -1,6 glycosidic linkages and formation of malto-oligosaccharides of different sizes, while MalX interacts with high affinity with the different malto-oligosaccharides produced by SpuA activity. GlgX also acts on cleavage of  $\alpha$ -1,6 glycosidic linkages but has no signal peptide for exportation from the intracellular environment; thus, SpuA is essential for assimilation of external glycogen. A profile HMM-based

search for an ortholog of the streptococcal extracellular enzyme SpuA (pullulanase) rendered positive results only for some genomic sequences from *C. diphtheriae*, particularly of the biovar *gravis* ( $E$  value =  $7.4e-189$ ) (Fig. 3a). This SpuA ortholog identified in *C. diphtheriae* has a secretory signal peptide (Fig. 3b) and is putatively able to bind to glycogen, as shown by molecular docking (Fig. 3c, d). Out of 200 genomic sequences (draft and complete) of *C. diphtheriae* available in public databases, we could identify the putative pullulanase in 72 genomes: 53 of the biovar *gravis*; 2 of the biovar *mitis*; and 17 without biovar assignment. All *C. diphtheriae* strains presented the gene coding for MalX that binds to malto-oligosaccharides greater than 11 glucose units that are unable to be transported normally (Abbott et al. 2010), reinforcing the need for the SpuA ortholog. The classification of *C. diphtheriae* strains into biovars based on biochemical reactions is still complex and unreliable (Sangal et al. 2014). Sangal et al. (2014) reported on the difficulties for finding a genetic basis that supports biovar differentiation. Our finding that strains of the biovar *gravis* present a SpuA ortholog may aid to this knowledge as this might help to explain the differential ability to utilize glycogen in this biovar. The presence of this gene in three strains of the biovar *mitis* could also help to



**Fig. 3** **a** *C. diphtheriae* biovar *gravis* enzyme with putative alpha-1,6-glycosidase activity, similar to streptococcal SpuA. **a** Genomic context analysis demonstrating the strain-specific presence of the gene coding for the *C. diphtheriae* alpha-1,6-glycosidase. **b** Prediction of signal peptide

for protein extracellular localization. **c** Structural alignment of the *C. diphtheriae* alpha-1,6-glycosidase with SpuA from *Streptococcus pneumoniae*. **d** Molecular docking of *C. diphtheriae* VA01 putative alpha-1,6-glycosidase and the glycogen molecule

explain the rare ability to utilize starch. In the XSMA group, no homologs were found for *spuA* and *malX*, indicating that these strains are not capable of utilizing external glycogen, though they may still be able to degrade it intracellularly.

The galactose degradation pathway is originally not present in the most traditionally used biochemical identification tests of Corynebacteria, such as the API Coryne system. However, this reaction is widely used in the identification of microorganisms in clinical microbiology laboratories and its high variability in the studied species may aid understanding of the genetic events contributing to biochemical variability. The presence of the *galK*, *galE*, and *galT* genes is essential for the maintenance of the galactose utilization pathway. *C. xerosis* presented homologs for all genes in the pathway; however, the strain ATCC 373 possesses a frameshift in the *galK* and *galT* genes, indicating an inability to phosphorylate galactose and impairment of the subsequent reactions. The *C. minutissimum* genotype corresponds to the expected negative phenotype, as *galM* and *galT* genes were not identified in any of the strains. There were divergences between *C. amycolatum* strains, as strain SK46 presented all the essential genes for galactose utilization, whereas strain ICIS 53 did not present *galT*, possibly explaining variable phenotype observed for this species. All *C. diphtheriae* strains presented essential genes for degradation of galactose but prediction of transporter components rendered ambiguous results with protein domains for transport of other substrates, such as ribose and xylose. A similar result was found for prediction of ABC transport systems in the other species (Tables 3 and 4).

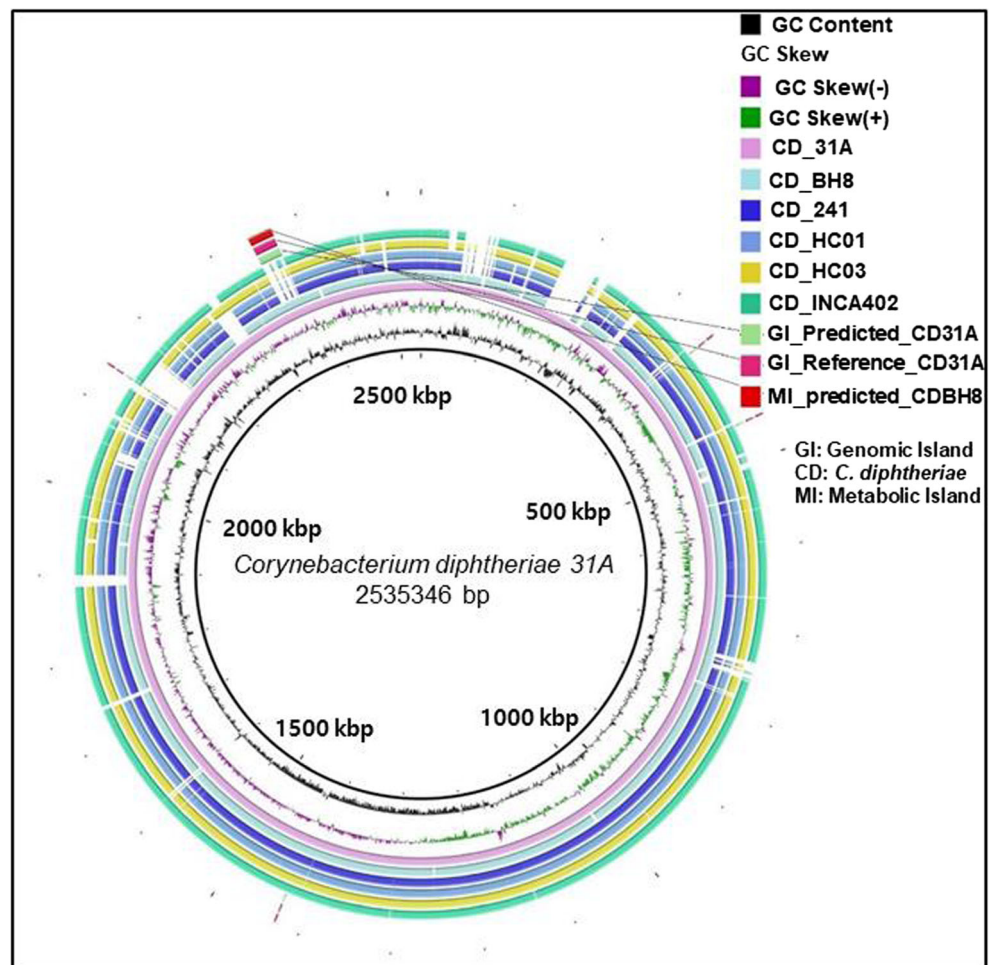
### Genomic context analysis and prediction of genomic islands

To explore the contributions of events of gene gain and loss for the differential distribution of the various enzyme genes, we performed a genomic context analysis of the main genes that are essential for the six variable biochemical pathways and evaluated their occurrences in horizontally acquired genomic islands. We found that the main genes in the sucrose utilization pathway—*scrA*, *scrB*, and *scrK*—are in close proximity in the genome, whereas the most conserved components of the PTS system, Hpr and EI, encoded by the *ptsH* and *ptsI* genes, respectively, are found in distinct regions of the genomes. A prediction of genomic islands (Fig. 4) indicated the existence of a metabolic island that contains the main genes for sucrose utilization, only in the strains 31A and BH8 of *C. diphtheriae*. Although these genes were also found in other species, prediction of genomic islands in strains of the XSMA group was hampered due to the high fragmentation of the genome assemblies into various contigs (Table 1).

### Mapping of carbohydrate transporters

We mapped the carbohydrate transporter systems in the various genomic sequences in order to evaluate whether differential occurrence of transport-related protein components might contribute to biochemical variabilities of the strains. In addition to evaluating transport of the target carbohydrates, we also observed the presence of transporters for monosaccharides, fructose and glucose, which are some of the basic components of the target carbohydrates. The genes coding for the common Hpr and EI components of the PTS system were present in all strains (Tables 3 and 4). The glucose- and fructose-specific PTS transporters were identified in all strains; however, all *C. diphtheriae* strains presented frameshifts in the fructose-specific PTS (*ptsF*) gene, and it is annotated as pseudogene. The *ptsF* gene is a key component of sucrose metabolism in *C. glutamicum*, due to the lack of fructokinase activity; the only pathway for phosphorylation of fructose derived from sucrose is the export and subsequent uptake of fructose through PTS(Fru). All strains of the XSMA group presented homologous genes of the non-PTS *iol1* and *iol2* glucose transporters (Ikeda et al. 2011), while no strains of *C. diphtheriae* showed homologs for these transporters (Tables 3 and 4). An automatic analysis in the curated database TransportDB identified the presence of the EIIAB components of fructose in all strains of *C. diphtheriae*, with exception of the strain INCA 402. A sucrose-specific PTS was found in all strains of *C. striatum* and *C. minutissimum* (Tables 3 and 4); however, only two strains of *C. diphtheriae* (31A and BH8) and one strain each of *C. xerosis* (NBRC16721) and *C. amycolatum* (SK46) presented this sucrose-specific transporter. Maltose transport is carried out by a specific ABC-type transporter; homologs with high similarities to the four *C. glutamicum* transporter components were identified in all strains of *C. diphtheriae*, *C. minutissimum*, and in *C. amycolatum* strain ICIS53 (Tables 3 and 4). Ribose transport is also performed by an ABC-type transporter; however, homologs of the four component proteins were only found in the *C. minutissimum* 1941 strain, an unexpected result given the observed positive and variable profiles of the studied species (Tables 3 and 4). The transport of galactose in its free form is also carried out by an ABC carrier, according to the Transport Classification Database (TCDB). The transport of galactose in *E. coli* is carried out by three components, encoded by the genes *mgIA*, *mgIB*, and *mgIC* and by the symporter permease encoded by *galP*. Protein orthologs sharing more than 30% identity were found in all *C. diphtheriae* strains; however, protein annotation and domain database searching indicated these components as part of the ribose-specific ABC transporter, given the similarities between the families of ABC transporters of ribose and galactose. *C. striatum* and *C. minutissimum* also rendered results with more than 30% identities for *mgIA* and *mgIC*, but these are potentially false-positive results as the prediction of

**Fig. 4** Genomic island alignments in *C. diphtheriae* strains. Graphical representation of the alignment of the genomes of *C. diphtheriae* strains with potential for sucrose fermentation. The genomic islands were aligned with the five genomes but were not found in some of the strains described as positive for sucrose utilization. The genome of *C. diphtheriae* 31A was used as a reference. GI: Genomic Island; MI: Metabolic Island; CD: *C. diphtheriae*



domains and annotations in different databases also indicated the transport of other substrates. *C. xerosis* and *C. amycolatum* only showed homologs for *mglA* (Tables 3 and 4). BLAST-based homology searches did not detect orthologs of the gene encoding the galactose-binding periplasmic protein MglB; nevertheless, by using profile HMMs through jackHMMER (Finn et al. 2015), we could find statistically relevant hits for the MglB substrate-ligand protein in *C. diphtheriae*, *C. striatum*, *C. minutissimum*, and *C. amycolatum* (Tables 3 and 4). Noteworthy, some ABC transporter components present ambiguous annotations due to high sequence conservation among various protein family members.

## Concluding remarks

The results obtained solely by BLAST-based homology searches of the main enzyme genes needed for each of the biochemical reactions (Fig. 2) clearly do not explain the phenotypic profiles of the various bacterial isolates (Table 2). Besides, discordant phenotypic predictions from genotypes are obtained if only these results are considered, without curation of the biochemical pathway reconstructions.

Additionally, the differential occurrence of carbohydrate transporters may impact the phenotypic predictions (Tables 5 and 6).

For instance, prediction of maltose utilization using the pathway structure annotated for the species *C. glutamicum* will render positive results for all genomic sequences analyzed (Fig. 2), even though this reaction is expected to be negative for the species *C. striatum*. Automatic reconstruction of this pathway using the PATRIC platform would yield similar results (data not shown). However, when considering the structure of the pathway annotated in MetaCyc for other bacterial species, that requires the enzyme beta-phosphoglucomutase (*pgmB* gene) (EC 5.4.2.6), then *C. striatum* would render a negative result, while the species *C. xerosis* and *C. amycolatum* would be considered positive (Fig. 5). Therefore, the occurrence of the *pgmB* gene permits a better prediction of maltose utilization ability in species of the XSMA group.

In the glycogen utilization pathway, the major variation between all genomic sequences was the occurrence of the *malZ* gene (that codes for maltodextrin glucosidase) (Fig. 2). Biochemical pathway reconstructions in PATRIC indicate differential occurrence of the *malP* gene (not shown). All strains

		NIT	MAL	SAC	RIB	GLYG	GAL
<b><i>C. diphtheriae</i></b>							
HC01 ( <i>mitis</i> )	Phenot.	+	+	+	+	-	NA
	Bioinf.	+	+	-	+	-	+
HC07 ( <i>gravis</i> )	Phenot.	+	+	-	+	+	-
	Bioinf.	+	+	-	+	+	+
VA01 ( <i>gravis</i> )	Phenot.	+	+	-	+	+	NA
	Bioinf.	+	+	-	+	+	+
<b><i>C. ulcerans</i></b>							
809	Phenot.	-	+	NA	+	+	NA
	Bioinf.	-	+	-	+	+	+
BR-AD22	Phenot.	-	+	NA	+	+	NA
	Bioinf.	-	+	-	+	+	+
<b><i>C. striatum</i></b>							
1961	Phenot.	+	-	+	-	-	+
	Bioinf.	+	-	+	+	-	+
2230	Phenot.	-	-	+	-	NA	+
	Bioinf.	-	-	+	+	-	+
2245	Phenot.	-	-	+	-	NA	+
	Bioinf.	-	-	-	+	-	+

**Fig. 5** Comparisons between true phenotypic profiles and bioinformatic predictions of phenotypes from genomic data. Phenotypic profiles were obtained from literature review for specific strains of *C. diphtheriae*, *C. ulcerans*, and *C. striatum* (Trost et al. 2011; Baio et al. 2013; Ramos 2014; Encinas et al. 2015). Curated bioinformatic predictions considered

presented the essential genes *glgX* and *glgP* for intracellular degradation of glycogen; considering that mostly strains of the biovar *gravis* of *C. diphtheriae* were expected to be positive for glycogen fermentation, these findings pointed to a possible variability in the process of import of glycogen derivatives from the extracellular environment. In fact, as mentioned earlier, a profile HMM-based search for an ortholog of the streptococcal extracellular enzyme SpuA (pullulanase) rendered positive results only for genomic sequences from *C. ulcerans* and *C. diphtheriae*, particularly for strains of the biovar *gravis* (Fig. 3). Therefore, we propose that a positive phenotype prediction for glycogen utilization in strains of *C. diphtheriae* should include the detection of a SpuA ortholog in the genomic sequence (Fig. 5).

In the ribose utilization pathway, there was a variability in the occurrence of the components of the RbsACBD transporter between the studied genomes (Tables 5 and 6). With exception of the pathway annotated in The SEED database, the other variants of this pathway required the component D-ribose-pyranase (*rbsD* gene); only one strain of *C. minutissimum* (strain 1941) presented this component (Tables 5 and 6), even though the ribose utilization test renders positive results for the species *C. diphtheriae* and *C. xerosis* and is potentially positive for the remaining species. PATRIC reconstructions indicate this biochemical pathway as positive for all genomics sequences (not shown). Intriguingly, false-

**Improvements of bioinformatic predictions:**

Detection of a putative secreted alpha-1,6-glycosidase (*spuA* homolog) aided better prediction of glycogen utilization (GLYG) phenotypes.

The maltose utilization (MAL) phenotype was correctly predicted for *C. striatum* when we considered the need for a beta-phosphoglucomutase (*pgmB* gene).

**Limitations:**

Though presence of a fructokinase encoding gene (*scrK* gene) was a good predictor of sucrose utilization (SAC) ability, the strains HC01 of *C. diphtheriae* and 2245 of *C. striatum* did not present this gene. Of note, this latter strain is phenotypically negative for fructose utilization.

Strain HC07 showed all genes for galactose utilization (GAL) but is phenotypically negative. This may be due to difficulties in reliable prediction of specific ABC-transport components.

All *C. striatum* strains are negative for ribose utilization (RIB), but presented all required genes. This may be also due to unreliable prediction of specific ABC-transport components.

results obtained by tBLASTn- and profile HMM-based orthologs search, prediction of carbohydrate transport systems, and literature-based curation of results. Improvements over bioinformatic predictions based solely on BLAST searches are indicated. Limitations of the strategy are also highlighted

positive predictions of phenotype from genotype were obtained for the species *C. striatum* when we assumed that the lack of *rbsD* would not compromise the ability to utilize ribose (Fig. 5). On the other hand, the same assumption could be made for the species *C. diphtheriae* (Fig. 5).

Similarly, in the galactose utilization pathway, all required genes were found in genomic sequences of strains of *C. diphtheriae*, *C. ulcerans*, and *C. striatum* (Fig. 2); however, strains of *C. diphtheriae* which are phenotypically negative for galactose utilization also show the presence of all genes in the pathway (Fig. 5). As mentioned earlier, reliable prediction of specific ABC transport systems might be hampered by high similarities between the different systems and this will need further studies.

In the sucrose utilization pathway, there was significant variation in the distribution of the major component genes *scrA* (sucrose-specific PTS permease) and *scrB* (sucrose phosphate hydrolase) between the genomes. Interestingly, strains of *C. striatum* and *C. minutissimum* presented a gene coding for ScrK (fructokinase), which is absent in most *C. diphtheriae* strains and in *C. glutamicum*. This latter species uses a *ptsF* for uptake and phosphorylation of fructose exported during sucrose utilization. Although some *C. diphtheriae* strains used in this study are reported to present the atypical sucrose-fermenting phenotype (Viguetti et al. 2012), the *ptsF* gene is annotated as non-functional due to frameshifts in all

strains. Importantly, an automated pathway reconstruction using PATRIC would indicate the presence of all components for sucrose utilization in most *C. diphtheriae* strains, including a sucrose-specific PTS (not shown). Conversely, manual curation indicates the horizontal acquisition of the entire gene set only in two *C. diphtheriae* strains (31A and BH8). Therefore, prediction of sucrose utilization ability by identification of a *scrK* homolog gene in the genomic sequences renders reliable results; however, the absence of this gene and of a detectable sucrose-specific PTS is not sufficient to rule out this ability. These results suggest the existence of an alternative pathway for sucrose utilization.

**Acknowledgements** This work was supported by grants from FAPESB (JCB0031/2013) and CAPES (PROCAD 071/2013).

## References

- Abbott DW, Higgins MA, Hyrnuik S, Pluvinae B, Lammerts van Bueren A, Boraston AB (2010) The molecular basis of glycogen breakdown and transport in *Streptococcus pneumoniae*. *Mol Microbiol* 77:183–199
- Adderson EE, Boudreaux JW, Cummings JR, Pounds S, Wilson DA, Procop GW, Hayden RT (2008) Identification of clinical coryneform bacterial isolates: comparison of biochemical methods and sequence analysis of 16S rRNA and *rpoB* genes. *J Clin Microbiol* 46:921–927
- Ajmal S, Saleh OA, Beam E (2017) Development of high-grade daptomycin resistance in a patient being treated for *Corynebacterium striatum* infection. *Antimicrob Agents Chemother* 61:e00705–e00717. <https://doi.org/10.1128/AAC.00705-17>
- Alibi S, Ferjani A, Gaillot O, Marzouk M, Courcol R, Boukadida J (2015) Identification of clinically relevant *Corynebacterium* strains by Api Coryne, MALDI-TOF-mass spectrometry and molecular approaches. *Pathol Biol* 63:153–157
- Almeida S, Sousa C, Abreu V, Diniz C, Dorneles EMS, Lage AP, Barh D, Azevedo V (2017) Exploration of nitrate reductase metabolic pathway in *Corynebacterium pseudotuberculosis*. *Int J Genomics* 2017: 1–12
- Almuzara MN, De Mier C, Rodríguez CR et al (2006) Evaluación del sistema API Coryne, versión 2.0, para la identificación de bacilos gram-positivos difteroides de importancia clínica. *Rev Argent Microbiol* 38(4):197–201
- Andersson U, Rådström P (2002)  $\beta$ -glucose 1-phosphate-interconverting enzymes in maltose- and trehalose-fermenting lactic acid bacteria. *Environ Microbiol* 4:81–88
- Baio PV, Mota HF, Freitas AD, Gomes DL, Ramos JN, Sant'Anna LO, Souza MC, Camello TC, Hirata Junior R, Vieira VV, Mattos-Guaraldi AL (2013) Clonal multidrug-resistant *Corynebacterium striatum* within a nosocomial environment, Rio de Janeiro, Brazil. *Mem Inst Oswaldo Cruz* 108(1):23–29
- Belchior E, Henry S, Badell E, Collet L, Benoit-Cattin T, de Montera AM, Guiso N, Patey O, Levy-Bruhl D, Filleul L, Chieze F, Olivier S (2017) Diphtheria in Mayotte, 2007–2015. *Emerg Infect Dis* 23: 1218–1220
- Berg JM, Tymoczko JL, Stryer L (2012) *Biochemistry* (7th, International ed). WH Freeman and company 7:338
- Bernard K (2012) The genus *Corynebacterium* and other medically relevant coryneform-like bacteria. *J Clin Microbiol* 50:3152–3158
- Bernard KA, Funke G (2012) Genus *Corynebacterium*. In: *Bergey's manual of systematic bacteriology*, vol 5, pp 245–289
- Bhattacharya D, Cheng J (2013) i3Drefine software for protein 3D structure refinement and its assessment in CASP10. *PLoS One* 8:e69648
- Blombach B, Seibold GM (2010) Carbohydrate metabolism in *Corynebacterium glutamicum* and applications for the metabolic engineering of L-lysine production strains. *Appl Microbiol Biotechnol* 86:1313–1322
- Brettin T, Davis JJ, Disz T et al (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365
- Buckley AA, Faustoferri RC, Quivey RG Jr (2014)  $\beta$ -Phosphoglucomutase contributes to aciduricity in *Streptococcus mutans*. *Microbiology* 160:818–827
- Cerdeño-Tárraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, De Zoysa A, Chillingworth T, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Quail MA, Rabinowitsch E, Rutherford KM, Thomson NR, Unwin L, Whitehead S, Barrell BG, Parkhill J (2003) The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res* 31(22):6516–6523
- Chandran R, Puthukkichal DR, Suman E, Mangalore SK (2016) Diphtheroids—important nosocomial pathogens. *J Clin Diagn Res* 10:28–31
- Chaudhari NM, Gupta VK, Dutta C (2016) BPGA—an ultra-fast pan-genome analysis pipeline. *Sci Rep* 6:24373. <https://doi.org/10.1038/srep24373>
- Clermont L, Macha A, Müller L et al (2015) The  $\alpha$ -glucan phosphorylase MalP of *Corynebacterium glutamicum* is subject to transcriptional regulation and competitive inhibition by ADP-glucose. *J Bacteriol* 197:1394–1407
- Clifton MC, Simon MJ, Erramilli S et al (2015) In vitro reassembly of the ribose ATP-binding cassette transporter reveals a distinct set of transport complexes. *J Biol Chem* 290:5555–5565
- Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I (2017) CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* 45: D289–D295
- Edwards B, Hunt AC, Hoskisson PA (2011) Recent cases of non-toxicogenic *Corynebacterium diphtheriae* in Scotland: justification for continued surveillance. *J Med Microbiol* 60:561–562
- Efstratiou A, George RC (1996) Screening tests for the presumptive identification of *Corynebacterium diphtheriae* in a diagnostic laboratory. *J Clin Microbiol* 34:3251
- Elbourne LD, Tetu SG, Hassan KA, Paulsen IT (2017) TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res* 45(D1):D320–D324
- Encinas F, Marin MA, Ramos JN, Vieira VV, Mattos-Guaraldi AL, Vicente AC (2015) Genomic analysis of a nontoxicogenic, invasive *Corynebacterium diphtheriae* strain from Brazil. *Mem Inst Oswaldo Cruz* 110(6):817–819
- Engels V, Georgi T, Wendisch VF (2008) ScrB (Cg2927) is a sucrose-6-phosphate hydrolase essential for sucrose utilization by *Corynebacterium glutamicum*. *FEMS Microbiol Lett* 289(1):80–89
- Finn RD, Clements J, Arndt W et al. (2015) HMMER web server: 2015 update. *Nucleic Acids Res* 43(W1):W30–W38. <https://doi.org/10.1093/nar/gkv397>
- Forbes BA (2017) Did I hear you correctly? The organism identified was *Corynebacterium diphtheriae*? *Clin Microbiol News* 139:35–41
- Funke G, Frodl R (2008) Comprehensive study of *Corynebacterium freneyi* strains and extended and emended description of *Corynebacterium freneyi* Renaud, Aabel, Riegel, Meugnier, and Bollet 2001. *J Clin Microbiol* 46(2):638–643
- Funke G, Lawson PA, Bernard KA, Collins MD (1996) Most *Corynebacterium xerosis* strains identified in the routine clinical



- laboratory correspond to *Corynebacterium amycolatum*. J Clin Microbiol 34:1124–1128
- Funke G, Renaud FN, Freney J, Riegel P (1997) Multicenter evaluation of the updated and extended API (RAPID) Coryne database 2.0. J Clin Microbiol 35(12):3122–3126
- Gladysheva IV, Cherkasov SV, Khlopko YA, Plotnikov AO, Gogoleva NE (2016) Draft Genome Sequence of Strain ICIS 53 Isolated from a Female Urogenital Tract. Genome Announc 4(6):e01267–16.
- Hacker E, Antunes CA, Mattos-Guaraldi AL, Burkovski A, Tauch A (2016) *Corynebacterium ulcerans*, an emerging human pathogen. Future Microbiol 11:1191–1208
- Hahn WO, Werth BJ, Butler-Wu SM, Rakita RM (2016) Multidrug-resistant *Corynebacterium striatum* associated with increased use of parenteral antimicrobial drugs. Emerg Infect Dis 22:1908–1914
- Hall AH, Cassiday PK, Bernard KA et al (2010) Novel *Corynebacterium diphtheriae* in domestic cats. Emerg Infect Dis 16:688–691
- Henrich A, Kuhlmann N, Eck A et al (2013) Maltose uptake by the novel ABC transport system MusEFGK2I causes increased expression of ptsG in *Corynebacterium glutamicum*. J Bacteriol 195:2573–2584
- Holden HM, Rayment I, Thoden JB (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. J Biol Chem 278:43885–43888
- Huey R, Morris GM (2003) AutoDock tools. The Scripps Research Institute, La Jolla
- Ikeda M (2012) Sugar transport systems in *Corynebacterium glutamicum*: features and applications to strain development. Appl Microbiol Biotechnol 96:1191–1200
- Ikeda M, Mizuno Y, Awane S, Hayashi M, Mitsuhashi S, Takeno S (2011) Identification and application of a different glucose uptake system that functions as an alternative to the phosphotransferase system in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 90:1443–1451
- Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7:1511–1522
- Kim C, Song S, Park C (1997) The D-allose operon of *Escherichia coli* K-12. J Bacteriol 179:7631–7637
- Kolios AGA, Cozzio A, Zinkernagel AS, French LE, Kündig TM (2017) Cutaneous *Corynebacterium* infection presenting with disseminated skin nodules and ulceration. Case Rep Dermatol 9:8–12
- Leal SM Jr, Jones M, Gilligan PH (2016) Clinical significance of commensal gram-positive rods routinely isolated from patient samples. J Clin Microbiol 54:2928–2936
- Letek M, Ordonez E, Fernández-Natal I et al (2006) Identification of the emerging skin pathogen *Corynebacterium amycolatum* using PCR-amplification of the essential divIVA gene as a target. FEMS Microbiol Lett 265(2):256–263
- Marchler-Bauer A, Derbyshire MK, Gonzales NR et al (2014) CDD: NCBI's conserved domain database. Nucleic Acids Res 43(Database issue):D222–D226. <https://doi.org/10.1093/nar/gku1221>
- Martinez-Martinez L, Suárez AI, Winstanley J et al (1995) Phenotypic characteristics of 31 strains of *Corynebacterium striatum* isolated from clinical samples. J Clin Microbiol 33(9):2458–2461
- Martins C, Faria L, Souza M (2009) Microbiological and host features associated with corynebacteriosis in cancer patients: a five-year study. Mem Inst Oswaldo Cruz 104:905–913
- Mattos-Guaraldi AL, Formiga LCD (1998) Bacteriological properties of a sucrose-fermenting *Corynebacterium diphtheriae* strain isolated from a case of endocarditis. Curr Microbiol 37:156–158
- Mattos-Guaraldi AL, Guimarães LC, Santos CS, Veras AA, Carneiro AR, Soares SC, Ramos JN, Souza C, Vieira VV, Hirata R, Azevedo V, Pacheco LG, Silva A, Ramos RT (2015) Draft genome sequence of 1961 BR-RJ/09, a multidrug-susceptible strain isolated from the urine of a hospitalized 37-year-old female patient. Genome Announc 3(4):e00869–15. <https://doi.org/10.1128/genomeA.00869-15>
- Meinel DM, Kuehl R, Zbinden R, Boskova V, Garzoni C, Fadini D, Dolina M, Blümel B, Weibel T, Tschudin-Sutter S, Widmer AF, Bielicki JA, Dierig A, Heininger U, Konrad R, Berger A, Hinic V, Goldenberger D, Blaich A, Stadler T, Battagay M, Sing A, Egli A (2016) Outbreak investigation for toxigenic *Corynebacterium diphtheriae* wound infections in refugees from Northeast Africa and Syria in Switzerland and Germany by whole genome sequencing. Clin Microbiol Infect 22:1003.e1–1003.e8
- Moon M-W, Kim HJ, Oh TK et al (2005) Analyses of enzyme II gene mutants for sugar transport and heterologous expression of fructokinase gene in *Corynebacterium glutamicum* ATCC 13032. FEMS Microbiol Lett 244:259–266
- Nentwich SS, Brinkrolf K, Gaigal L, Huser AT, Rey DA, Mohrbach T, Marin K, Puhler A, Tauch A, Kalinowski J (2009) Characterization of the LacI-type transcriptional repressor RbsR controlling ribose transport in *Corynebacterium glutamicum* ATCC 13032. Microbiology 155:150–164
- Nishimura T, Vertès AA, Shinoda Y, Inui M, Yukawa H (2007) Anaerobic growth of *Corynebacterium glutamicum* using nitrate as a terminal electron acceptor. Appl Microbiol Biotechnol 75:889–897
- Oberto J (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. BMC Bioinformatics 14:4
- Overbeek R, Begley T, Butler R, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goessmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy A, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33:5691–5702
- Pacheco LG, Mattos-Guaraldi AL, Santos CS, Veras AA, Guimarães LC, Abreu V, Pereira FL, Soares SC, Dorella FA, Carvalho AF, Leal CG, Figueiredo HC, Ramos JN, Vieira VV, Farfour E, Guiso N, Hirata R Jr, Azevedo V, Silva A, Ramos RT (2015) Draft genome sequences of two species of “difficult-to-identify” human-pathogenic *Corynebacteria*: implications for better identification tests. J Genomics 3:82–84. <https://doi.org/10.7150/jgen.12886>
- Palacios L, Vela AI, Molin K, Fernandez A, Latre MV, Chacon G, Falsen E, Fernandez-Garayzabal JF (2010) Characterization of some bacterial strains isolated from animal clinical materials and identified as *Corynebacterium xerosis* by molecular biological techniques. J Clin Microbiol 48:3138–3145
- Pennie RA, Malik AS, Wilcox L (1996) Misidentification of toxigenic *Corynebacterium diphtheriae* as a *Corynebacterium* species with low virulence in a child with endocarditis. J Clin Microbiol 34:1275–1276
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
- Pimenta FP, Hirata R Jr, Rosa AC et al (2008) A multiplex PCR assay for simultaneous detection of *Corynebacterium diphtheriae* and differentiation between non-toxicogenic and toxicogenic isolates. J Med Microbiol 57:1438–1439
- Ramos JN (2014) Caracterização de estirpes sugestivas de corinebactérias isoladas de sítios intravenosos. Dissertation, Instituto Nacional de Controle de Qualidade em Saúde, Fundação Oswaldo Cruz. Rio de Janeiro
- Renaud FN, Dautaur M, Daoud S, Aubel D, Riegel P, Monget D, Freney J (1998) Differentiation of *Corynebacterium amycolatum*, *C. minutissimum*, and *C. striatum* by carbon substrate assimilation tests. J Clin Microbiol 36(12):3698–3702

- Renaud FN, Aubel D, Riegel P, Meugnier H, Bollet C (2001) *Corynebacterium freneyi* sp. nov., alpha-glucosidase-positive strains related to *Corynebacterium xerosis*. *Int J Syst Evol Microbiol* 51(5): 1723–1728
- Reynolds GE, Saunders H, Matson A, O’Kane F, Roberts SA, Singh SK, Voss LM, Kiedrzyński T (2016) Public health action following an outbreak of toxigenic cutaneous diphtheria in an Auckland refugee resettlement centre. *Commun Dis Intell Q Rep* 40:E475–E481
- Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM, Cookson BT, Shendure J, Salipante SJ (2015) A Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota. *PLoS Genet* 11(7):e1005413. <https://doi.org/10.1371/journal.pgen.1005413>
- Rothery RA, Magalon A, Giordano G, Guigliarelli B, Blasco F, Weiner JH (1998) The molybdenum cofactor of Escherichia coli nitrate reductase A (NarGHI). Effect of a mobAB mutation and interactions with [Fe-S] clusters. *J Biol Chem* 273:7462–7469
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945
- Sangal V, Hoskisson PA (2016) Evolution, epidemiology and diversity of *Corynebacterium diphtheriae*: new perspectives on an old foe. *Infect Genet Evol* 43:364–370
- Sangal V, Burkovski A, Hunt AC et al (2014) A lack of genetic basis for biovar differentiation in clinically important *Corynebacterium diphtheriae* from whole genome sequencing. *Infect Genet Evol* 21: 54–57
- Seibold GM, Wurst M, Eikmanns BJ (2009) Roles of maltodextrin and glycogen phosphorylases in maltose utilization and glycogen metabolism in *Corynebacterium glutamicum*. *Microbiology* 155:347–358
- Soares SC, Geyik H, Ramos R et al (2016) GIPSy: genomic island prediction software. *J Biotechnol* 232:2–11
- Song Y, DiMaio F, Wang RY et al (2013) High-resolution comparative modeling with RosettaCM. *Structure* 21:1735–1742
- Torres Lde F, Ribeiro D, Hirata R Jr, Pacheco LG, Souza MC, dos Santos LS, dos Santos CS, Salah M, Costa MM, Ribeiro MG, Selim SA, Azevedo VA, Mattos-Guaraldi AL (2013) Multiplex polymerase chain reaction to identify and determine the toxigenicity of *Corynebacterium* spp with zoonotic potential and an overview of human and animal infections. *Mem Inst Oswaldo Cruz* 108(3):272–279
- Trost E, Al-Dilaimi A, Papavasiliou P, Schneider J, Viehovever P, Burkovski A, Soares SC, Almeida SS, Dorella FA, Miyoshi A, Azevedo V, Schneider MP, Silva A, Santos CS, Santos LS, Sabbadini P, Dias AA, Jr HR, Mattos-Guaraldi AL, Tauch A (2011) Comparative analysis of two complete *Corynebacterium ulcerans* genomes and detection of candidate virulence factors. *BMC Genomics* 12:383
- Trost E, Blom J, de Castro Soares S, Huang IH, al-Dilaimi A, Schroder J, Jaenicke S, Dorella FA, Rocha FS, Miyoshi A, Azevedo V, Schneider MP, Silva A, Camello TC, Sabbadini PS, Santos CS, Santos LS, Hirata R, Mattos-Guaraldi AL, Efstratiou A, Schmitt MP, Ton-That H, Tauch A (2012) Pangenomic study of *Corynebacterium diphtheriae* that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. *J Bacteriol* 194:3199–3215
- Van Der Heide T, Poolman B (2002) ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep* 3:938–943
- Viguetti SZ, Pacheco LG, Santos LS (2012) Multilocus sequence types of invasive *Corynebacterium diphtheriae* isolated in the Rio de Janeiro urban area, Brazil. *Epidemiol Infect* 140:617–620
- Von Zalusowski PRA (2015) Glycogen metabolism in *Corynebacterium glutamicum*: effects of environmental factors and of metabolic disturbances. PhD thesis. Universität Ulm
- Wang S, Peng J, Xu J (2011) Alignment of distantly related protein structures: algorithm, bound and implications to homology modeling. *Bioinformatics* 27:2537–2545
- Wang S, Ma J, Peng J, Xu J (2013) Protein structure alignment beyond spatial proximity. *Sci Rep* 3:1448
- Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL (2017) Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* 45:D535–D542
- Wauters G, Van Bosterhaut B, Janssens M, Verhaegen J (1998) Identification of *Corynebacterium amycolatum* and other nonlipophilic fermentative corynebacteria of human origin. *J Clin Microbiol* 36(5):1430–1432
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410
- Zakikhany K, Efstratiou A (2012) Diphtheria in Europe: current problems and new challenges. *Future Microbiol* 7:595–607
- Zasada AA (2013) Nontoxigenic highly pathogenic clone of *Corynebacterium diphtheriae*, Poland, 2004–2012. *Emerg Infect Dis* 19:1870–1872