Evolution, epidemiology and diversity of *Corynebacterium diphtheriae*: new perspectives on an old foe

Vartul Sangal[1] and Paul A. Hoskisson[2*]

[1]Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne - NE1 8ST, UK

[2]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow. UK G4 0RE, UK. Email: Paul.hoskisson@strath.ac.uk Tel +44 141 548 2819

*Correspondence:

Keywords: *Corynebacterium diphtheriae;* biovar; evolution; pathogenesis; MLST; secreted proteins.

ABSTRACT

Diphtheria is a debilitating disease caused by toxigenic *Corynebacterium diphtheriae* strains and has been effectively controlled by the toxoid vaccine, yet several recent outbreaks have been reported across the globe. Moreover, non-toxigenic *C. diphtheriae* strains are emerging as a major global health concern by causing severe pharyngitis and tonsillitis, endocarditis, septic arthritis and osteomyelitis. Molecular epidemiological investigations suggest the existence of outbreak-associated clones with multiple genotypes circulating around the world. Evolution and pathogenesis appears to be driven by recombination as major virulence factors, including the *tox* gene and pilus gene clusters, are found within genomic islands that appear to be mobile between strains. The number of pilus gene clusters and variation introduced by gain or loss of gene function correlate with the variable adhesive and invasive properties of *C. diphtheriae* strains. Genomic variation does not support the separation of *C. diphtheriae* strains into biovars which correlates well with findings of studies based on multilocus sequence typing. Genomic analyses of a relatively small number of strains also revealed a recombination driven diversification of strains within a sequence type and indicate a wider diversity among *C. diphtheriae* strains than previously appreciated. This suggests that there is a need for increased effort from the scientific community to study *C. diphtheriae* to help understand the genomic diversity and pathogenicity within the population of this important human pathogen.

1. Introduction

Toxigenic *Corynebacterium diphtheriae* are responsible for diphtheria in humans, a toxin-mediated disease of the upper respiratory tract which is generally characterized by the presence of an inflammatory pseudomembrane on the tonsils, oropharynx and pharynx causing sore throat, high temperature and potentially death (Hadfield et al., 2000). The toxin

is encoded by the *tox* gene within the lysogenised β-corynephage (Sangal and Hoskisson, 2014a) and can be effectively controlled by the diphtheria toxoid vaccine (Baxter, 2007). The cases of diphtheria were significantly reduced following the global immunization initiative (Galazka, 2000). Yet in the 1990s, the Newly Independent States (largely Former Soviet Union) observed the largest outbreaks of Diphtheria since the introduction of mass vaccination (Vitek & Wharton, 1998). In addition, there is still considerable morbidity and mortality around the world caused by this organism (www.WHO.int) and we need to remain vigilant.

Non-toxigenic *C. diphtheriae* strains (those that lack the *tox* gene) are now emerging as the cause of significant disease, especially invasive infections such as endocarditis, septic arthritis and osteomyelitis (Barakett et al., 1993; Belko et al., 2000; Edwards et al., 2011; Farfour et al., 2012; Patey et al., 1997; Poilane et al., 1995; Romney et al., 2006; Tiley et al., 1993). There is also the potential for *C. diphtheriae* to cause skin infections which result in cutaneous diphtheria across the globe in patients with varying vaccination status and travel histories (Gordon et al., 2011; Romney et al., 2006; Huhulescu et al., 2014; Cassir et al., 2015; Nelson et al., 2016). These infections are often associated with travel to *C. diphtheriae* prevalent endemic areas (FitzGerald et al., 2015; Lindhusen-Lindhe et al., 2012; May et al., 2014). More recently, non-toxigenic *tox* gene-bearing strains (NTTB) have also been reported from Europe (Zakikhany et al., 2014). These NTTB strains possess the *tox* gene, however mutation (a nucleotide deletion or disruption by an insertion sequence) in the A-subunit of the gene prevents expression (Zakikhany et al., 2014). These strains pose a potential threat to public through genetic reversion resulting in toxin production. Moreover, carriage of non-toxigenic strains in healthy individuals, as part of the normal upper respiratory tract flora is poorly understood, but has the potential to act as a reservoir of bacteria that can undergo phage-conversion and dissemination.

C. diphtheriae strains have historically been subdivided into the four biovars - gravis, intermedius, mitis and belfanti (Funke et al., 1997; Goodfellow et al., 2012). However, this biochemical differentiation appears to be dependent on technical capabilities of the laboratory and is unsupported by genomic analysis (Sangal et al., 2014a). This view is also supported by the quality assurance (Elek) tests for diphtheria diagnostics by the European diphtheria surveillance network (EDSN) where several participating laboratories could not correctly identify these biovars, particularly biovars intermedius and belfanti (Both et al., 2014; Neal and Efstratiou, 2009).

Related pathogenic corynebacteria including *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis* generally cause zoonotic infection in humans (Peel et al., 1997; Taylor et al., 2010; Wagner et al., 2011; Sangal et al., 2014b) whereas *C. diphtheriae* appears to be largely human specific. Recent reports highlight potential host jump of *C. diphtheriae* to and from domesticated and wild animals (Sing et al., 2015; Zakikhany et al., 2014). This is particularly important as the *tox* gene carrying β-corynephage is able to lysogenize all three species – *C. diphtheriae, C. ulcerans* and *C. pseudotuberculosis* and the promiscuous nature of the corynephage may result in human outbreaks of diphtheria and diphtheria-like diseases caused by non-*C. diphtheriae* strains.

Here we aim to provide an overview of global epidemiology and evolutionary dynamics of *C. diphtheriae* in the light of recent work in the field, with particular emphasis on the impact of whole genome sequencing in understanding the evolution and pathogenicity of different *C. diphtheriae* strains.

2. *C. diphtheriae* is genetically diverse

Despite an estimated 86% global coverage of the vaccine, 7,321 cases of diphtheria were reported in 2014, mainly from the developing countries (www.WHO.int). A diphtheria

epidemic in the former Soviet Union in the 1990s resulted in >157,000 cases claiming ~5000 lives (Dittmann et al., 2000). Yet, this pathogen is not under control, and the have been multiple outbreaks in different countries since 2000 including Colombia (Landazabal et al., 2001), India (Parande et al., 2014; Saikia et al., 2010), Norway (Rasmussen et al., 2011), Nigeria (Besa et al., 2014), Thailand (Wanlapakorn et al., 2014), and more recently in Brazil (Santos et al., 2015), Laos (Nanthavong et al., 2015) and Indonesia (Hughes et al., 2015).

The molecular epidemiology and diversity of *C. diphtheriae* has been investigated using a number of genotyping approaches including ribotyping, amplified fragment length polymorphism (AFLP), pulse-field gel electrophoresis (PFGE), random amplified polymorphic DNA (RAPD), clustered regularly interspaced short palindromic repeat (CRISPR) based spoligotyping and multilocus sequence typing (MLST) (Bolt et al., 2010; Damian et al., 2002; De Zoysa et al., 2008; Grimont et al., 2004; Kolodkina et al., 2006; Mokrousov et al., 2007; Mokrousov et al., 2005; Mokrousov et al., 2009; Titov et al., 2003). Most of the typing approaches exhibited some degree of correspondence (Damian et al., 2002; De Zoysa et al., 2008; Kolodkina et al., 2006; Titov et al., 2003). Ribotyping was found to be more discriminatory than PFGE and AFLP (De Zoysa et al., 2008) and was the gold standard for genotyping *C. diphtheriae* prior to the introduction of a robust MLST approach (Bolt et al., 2010; Grimont et al., 2004). The main Ribotyping scheme adhered to is that of Grimont et al., (2004) with each ribotype being allocated a geographical name based on the location of isolation; however, some previous studies followed an arbitrary nomenclature to represent different ribotypes. Ribotyping identified 34 ribotypes among 167 *C. diphtheriae* strains from Romania, the Russian Federation and the Republic of Moldova (Damian et al., 2002). The strains belonging to two ribotypes, C1 and C5 were predominant in Russia and Moldova whereas ribotypes C3 and C7 were isolated more frequently in Romania (Damian et al., 2002). The majority of *C. diphtheriae* strains were found to belong

to ribotypes D1 and D4 in Belarus (Titov et al., 2003). Remarkably, the distribution of ribotypes was found to alter between 1996 and 2005 (Kolodkina et al., 2006). Interestingly, this may be the result of increased vaccination in these areas following the outbreaks, perhaps indicating some level of vaccine-driven population selection in *C. diphtheriae.* Overall, all these studies identified prevalent clones associated with different outbreaks, but also found that multiple genotypes were circulating within different continents, suggesting great diversity of *C. diphtheriae* strains within the human population (Damian et al., 2002; De Zoysa et al., 2008; Kolodkina et al., 2006; von Hunolstein et al., 2003).

CRISPR based spoligotyping offered additional resolution within these ribotypes and was successfully used to characterize outbreak-associated strains from countries of former Soviet Union (Mokrousov, 2013; Mokrousov et al., 2005; Mokrousov et al., 2009). The epidemic strains from Russia that belonged to two ribotypes (Sankt-Peterburg and Rossija) were subdivided into 45 spoligotypes (Mokrousov, 2013; Mokrousov et al., 2007; Mokrousov et al., 2005). Due to the higher diversity within ribotype Sankt-Peterburg, it was proposed to have evolved prior to the emergence ribotype Rossija, indicating that new strains are emerging regularly within this species (Mokrousov, 2013).

While most genotypic approaches are focused on outbreak characterization and high resolution strain discrimination, MLST is more appropriate to investigate long-term evolutionary dynamics and has been applied to a number of microorganisms prior to the emergence of cost effective genome sequencing (Maiden, 2006). A robust MLST scheme was developed for *C. diphtheriae* in 2010 and sequence types (STs) were shown to be consistent with the previously determined *C. diphtheriae* ribotypes and offered higher resolution in most cases (Bolt et al., 2010). One important feature of the MLST studies was that they revealed a lack of correlation between the STs and the widely used biovar system and also showed no correlation with the severity of the disease caused by different strains (Bolt et al., 2010;

Farfour et al., 2012). While some eBURST groups, the so called clonal complexes, were found to be associated with certain countries, others were reported from multiple continents, indicating wide dissemination of strains (Bolt et al., 2010). MLST diversity has grown since 2010 and the data for 384 reference STs is available from the MLST website (http://pubmlst.org/cdiphtheriae/; accessed in November 2015). A total of 115 of these STs formed 11 major eBURST groups where the predicted founder had three or more single locus variants (Fig. 1). However, some of these data belong to *C. ulcerans* strains and may also contain some erroneous submissions to the database by the public.

More recently, whole genome sequences of 20 *C. diphtheriae* strains have been analysed (Cerdeno-Tarraga et al., 2003; Sangal et al., 2015; Sangal et al., 2014; Sangal et al., 2012a, b; Trost et al., 2012), revealing the genetic diversity amongst and within the major STs. Approximately 60% of the genome appears to be functionally conserved within *C. diphtheriae* strains with 1,625 genes belonging to the core genome (Sangal et al., 2015). However, enough diversity has accumulated within the core genes to allow discrimination of most *C. diphtheriae* strains from each other. Strains within STs appear to show close relationships indicating the robust nature of the MLST approach (Fig. 2; Bolt et al., 2010; Sangal et al., 2015). Similar groupings were also obtained from the genome-wide single nucleotide polymorphism analysis (SNPs; Sangal et al., 2014). The accessory genome varied greatly among *C. diphtheriae* strains (Sangal et al., 2015) even when a relatively small number of genomes was considered (14 known STs; Fig. 1). This indicates that most of the *C. diphtheriae* diversity remains to be discovered and will be crucial in our understanding of the molecular epidemiology, global transmission and carriage of this pathogen.

3. Evolutionary dynamics

Despite the global emergence of non-toxigenic strains and multiple recent outbreaks caused by *C. diphtheriae*, little is known about the evolutionary dynamics of this pathogen and most of the current understanding comes from the genomic analyses. MLST analyses indicated that there is significant recombination within *C. diphtheriae* populations (Bolt et al., 2010). Recombination plays an important role in bacterial evolution and is often linked to the increased virulence in some strains (Joseph et al., 2011; Suarez et al., 2004; Wirth et al., 2006). Indeed, the primary niche of *C. diphtheriae* in humans is the upper respiratory tract which is a hot-bed of horizontal gene transfer between bacterial strains (Marks et al., 2012).

A total of 57 genomic islands have been reported in *C. diphtheriae* and the distribution was found to vary significantly between strains (Trost et al., 2012). The genomic islands can be horizontally acquired from other bacteria, suggesting that recombination is shaping the current genetic diversity in *C. diphtheriae*. Some of the genomic islands carried phage associated genes while others harboured the genes that encode proteins for different cellular activities including siderophore biosynthesis and transport, degradation of polysaccharides and hydrocarbon derivatives such as 3-hydroxyphenylpropionic acid, antibiotic and heavy metal resistance (Trost et al., 2012). The major virulence factor of *C. diphtheriae*, the *tox* gene, is carried on a bacteriophage that can also move between strains, resulting in phage conversion (Barksdale and Pappenheimer, 1954; Freeman, 1951; Sangal and Hoskisson, 2014). Genomic islands carrying different *spa* operons introduced the variation in the ability of *C. diphtheriae* strains to form pili and interact with the host. These *spa* operons harbour genes encoding subunits of different types of pili and the gain or loss of the function of these genes correlate to the number and expression of pili on the cell surface (Ott et al., 2010; Chang et al., 2011; Trost et al., 2012).

Approximately one-third of the *C. diphtheriae* genome encodes accessory genes that vary widely between strains (Sangal et al., 2015). The strains within individual STs differed

199 from each other by the presence or absence of up to 290 genes, many of which are present on
200 the genomic islands (Sangal et al., 2015). These observations indicate likely differences in
201 recombination frequencies between *C. diphtheriae* strains. The frequencies of recombination
202 may vary widely between different strains within a species (Sangal et al., 2010), and may
203 reflect the difference in strain propensities for acquiring foreign DNA, which may result in
204 variation in pathogenicity of strains. Restriction-modification systems, bacteriophage defence
205 systems and CRISPR-Cas systems are major barriers to recombination that have been
206 reported in the genomes of *C. diphtheriae* strains (Hoskisson & Smith, 2007; Sangal et al.,
207 2013).

208 Genomic analyses of *C. diphtheriae* strains revealed the presence of two types of
209 CRISPR-Cas systems in three different configurations (Sangal et al., 2013). These systems
210 are comprised of CRISPR-associated proteins (Cas proteins encoded by *cas* genes) and
211 CRISPR arrays of short spacer sequences acquired from invading bacteriophages or plasmids
212 that are separated by repeat sequences. These arrays are transcribed into crRNA that
213 recognizes the invasion by the same nucleic acids and activate their cleavage by Cas
214 ribonucleoprotein complex (Marraffini, 2015). The acquisition of each spacer sequence
215 represents a unique evolutionary event, an encounter of the bacterial cell with the
216 bacteriophage or plasmid that may be unique to particular environment.

217 The majority of *C. diphtheriae* strains carried a type II-C CRISPR-Cas system,
218 however this was replaced by a type I-E-a in some strains or *vice versa* (Sangal et al., 2013).
219 A few strains with a type II-C system possessed an additional CRISPR-Cas system, type I-E-
220 b, at a different location in the genome. The variation in the G+C content and the
221 phylogenetic analyses of *cas1* gene, along with the direct repeat sequences in the CRISPR
222 arrays suggest three independent horizontal acquisitions of these CRISPR-Cas systems by *C.*
223 *diphtheriae*. Most of the spacer sequences are unique to CRISPR arrays in different strains,

suggesting that these strains evolved in different environments and encountered a range of different bacteriophages or plasmids (Sangal et al., 2013). Some strains were found to share spacer sequences at the distal end of the array, which may represent common strain ancestry or abundance of a particular foreign DNA type (bacteriophages/plasmids). The type of CRISPR-Cas systems and most of the spacer sequences in the arrays were shared between individuals of the same ST, which is consistent with their evolution from a recent common ancestor. These results also support CRISPR loci as useful molecular markers for strain identification and epidemiological studies (Mokrousov, 2013; Mokrousov et al., 2007).

Overall, the genomic and spacer diversities found in *C. diphtheriae* strains indicate unique evolutionary trajectories for different *C. diphtheriae* strains after they separated from their last common ancestor. However, no clear geographic or temporal association of *C. diphtheriae* strains has been reported. Interestingly, this may simply reflect a sampling bias, as available genomes reflect <10% of the current *C. diphtheriae* diversity observed from MLST analysis (Fig. 1). These data highlight the need to expand the genome sequencing effort for this species to fully understand the evolutionary dynamics of this pathogen.

4. Genetic basis of biochemical differentiation

The biochemical differentiation of *C. diphtheriae* strains into biovars is complex and unreliable, however for historical reasons it is still routinely followed by reference labortories (Both et al., 2014; Neal and Efstratiou, 2009; Sangal et al., 2014). The key characteristics include lipophilism of biovar intermedius strains - the need lipids for optimal growth and the formation of small gray or translucent colonies on agar plates (Funke et al., 1997). The strains of other biovars generally form large white or opaque colonies. The strains of biovar belfanti can not reduce nitrate and only biovar gravis strains seem to definitely utilize glycogen and

starch as carbon sources (Efstratiou et al., 2000; Efstratiou and George, 1999; Goodfellow et al., 2012).

Comparative genomic analyses identified that four genes involved in carbohydrate metabolism are absent or are pseudogenes in the intermedius strain (Sangal et al., 2014), potentially suggesting that this biovar may have compromised abilities to effectively use carbohydrates as the energy source and require alternate carbon source such as lipids, for optimal growth in the host. We have previously highlighted an insertion at the 3' end of *narJ* gene in the only sequenced belfanti genome, that results in an extended coding sequence in comparison to its homolog DIP0498 in NCTC 13129 (Sangal et al., 2014). However, the annotation of strain NCTC 13129 has recently been revised (GenBank accession number: NC_002935.2; new locus tag for DIP0498: DIP_RS13825) and the protein sequence of *narJ* is of the same length as observed in belfanti. Therefore, genetic basis of the belfanti strains not being able to reduce nitrate remains unclear. The phylogenomic analyses of core genome, accessory genome and genome-wide SNPs revealed an absence of a biovar specific grouping. Therefore, the biochemical seperation of *C. diphtheriae* into the traditional biovars is not supported by genomic diversityand is unsuitable for modern epidemiological studies (Sangal et al., 2015; Sangal et al., 2014; Trost et al., 2012). Genome sequencing results are consistent with the MLST phylogeny where the major *C. diphtheriae* lineage included strains from all four biovars (Bolt et al., 2010). However, a smaller second befanti-specific lineage can be observed from the MLST analyses which is not detected in the genomic study, potentially because the genome sequence of only one strain for each of the biovars belfanti and intermedius is available that highlights a clear need for more strains of these biovars to be sequenced.

5. Variation in pathogenicity and invasive strains

C. diphtheriae is considered a paradigm of mucosal pathogenicity, with much of the research focused on toxin production and pseudomembrane formation, almost to the neglect of studying other virulence mechanisms, such that the discovery of invasive strains of *C. diphtheriae* was a surprise to researchers. The *tox* gene, encoding the diphtheria toxin, is harboured on the genome of the β-corynephage, which integrates into *C. diphtheriae* genome between duplicated arginine tRNA genes (Sangal and Hoskisson, 2014; Trost et al., 2012). Only one prophage is present in most toxigenic strains, with the exception of strain PW8 where two copies of corynephage ω^{tox+} is found (Sangal and Hoskisson, 2014; Trost et al., 2012). While the nucleotide sequence of different corynephages show high levels of diversity, the sequence of the *tox* gene is highly conserved and also reflects the efficacy of the toxoid vaccine. The transcription of *tox* gene is controlled by the DtxR regulon, which is a key determinant for iron homeostasis (De Zoysa et al., 2005; Fourel et al., 1989). Iron is involved in a number of cellular activities and the induction of toxin in low iron availability might help pathogens to compete with the host for iron (Ganz and Nemeth, 2015; Trost et al., 2012) or liberate iron through killing of host cells. The gene composition of DtxR regulons in different *C. diphtheriae* strains may vary due to gain or loss of the genes that may affect the iron supply to the bacterial cell and hence, the expression of the *tox* gene (Litwin and Calderwood, 1993; Trost et al., 2012).

Non-toxigenic *C. diphtheriae* strains by definition do not contain the *tox* carrying β-corynephage, but do vary in their abilities to adhere to host cells, intracellular viability and their ability to stimulate cytokine production by the host immune system which may influence the severity of the disease due to infection (Bertuccini et al., 2004; Hirata et al., 2002; Peixoto et al., 2014; Puliti et al., 2006). These strains differ from each other in the presence and organisation of different pilus gene clusters, *spaA*, *spaD* and *spaH* (Sangal et al., 2015; Trost et al., 2012). Two pilus gene clusters, *spaD and spaH*, were present in four *C.*

298 *diphtheriae* strains that exhibited different adhesive and invasive properties. Interestingly, the
299 *spaA* operon was only present in the two strains with higher adhesion to pharyngeal D562
300 cell lines (Ott et al., 2010; Sangal et al., 2015). SpaA pili have been shown to interact with
301 the pharyngeal epithelial cells and SpaD and SpaH with the laryngeal and lung epithelial cell
302 types (Mandlik et al., 2007; Reardon-Robinson and Ton-That, 2014) suggesting niche
303 specialised roles for specific pilus types. However, some genes were found to be pseudogenes
304 in these clusters (Sangal et al., 2015), for example, *srtB* gene that encodes sortase for
305 incorporation of SpaE into the SpaD subunit of SpaD-type pili, *spaG* encoding a subunit of
306 SpaH-type pili and *spaB* encoding pilus base subunit of SpaA-type pili were pseudogenes in
307 strains ISS 4060, ISS 3319 and ISS 4746, respectively (Reardon-Robinson and Ton-That,
308 2014; Sangal et al., 2015). In addition, a gene *spaF* that encodes surface anchored fimbrial
309 subunit of *spaD*-type pili was pseudogenitised both in ISS 4746 and ISS 4749. Strain ISS
310 4749 with two intact gene clusters (SpaA and SpaH) exhibited highest number of pili at the
311 cell surface and highest adhesion to the cell lines when compared to ISS 3319 (SpaD gene
312 cluster) and ISS 4746 (SpaH gene cluster) with only one intact gene cluster (Bertuccini et al.,
313 2004; Ott et al., 2010; Sangal et al., 2015). Although SpaH gene cluster appears to be fully
314 functional in ISS 4060 strain, no surface pili were observed, suggesting there may be
315 variation in the levels of gene expression. However, adhesive properties of this strain were
316 comparable to ISS 3319 (Bertuccini et al., 2004; Ott et al., 2010; Sangal et al., 2015).
317 Therefore, the macromolecular surface structure and cell adhesion properties generally
318 correlate to the presence of pilus gene clusters in *C. diphtheriae* and expression of these
319 genes may be subject to unknown gene regulation mechanisms.
320 ISS 4746 and ISS 4749 were also shown to induce higher cytokine (IL-1 and IL-6)
321 production and caused higher incidences and severity of arthritis in mice in comparison to
322 ISS 3319 (Puliti et al., 2006). In addition to the membrane associated proteins, comparative

genomic analyses revealed a variation in predicted secreted proteins including lipoproteins and non-classical secreted proteins among these strains, which may be associated with the variation in the degree of pathogenesis (Sangal et al., 2015). Most of these proteins are hypothetical and a molecular characterization of these proteins might further improve understanding of the mechanisms of adhesion, invasion and immune induction in *C. diphtheriae*.

6. Conclusions

C. diphtheriae is still a major human pathogen, with multiple contemporary outbreaks around the world. Moreover, non-toxigenic strains are beginning to cause significant invasive disease in patients. Genomic analyses not only identified potential genes involved in adhesive, invasive and virulence characteristics of *C. diphtheriae* strains but also highlighted the impact of horizontal gene transfer in acquisition of these genes. These analyses also raise concerns about the use of biochemical separation of *C. diphtheriae* strains into biovars in clinics as a biovar encompasses genetically distinct strains. The evolutionary dynamics and the global diversity in *C. diphtheriae* are poorly characterized, clearly emphasizing the need of a community-based genome sequencing program that will improve the understanding of global transmission and local adaptation and will facilitate the development of effective surveillance policies and preventive strategies, amid multiple ongoing outbreaks. It will also inform on future vaccine development, perhaps to augment existing toxoid-based vaccines with universal surface proteins from *C. diphtheriae* which may be more effective in reducing carriage and the invasive diseases caused by non-toxigenic strains.

Acknowledgements

VS is supported by Anniversary Research Fellowship of the Northumbria University, Newcastle upon Tyne. *C. diphtheriae* work in the PAH laboratory is supported by Medical Research Scotland (Grant 422 FRG), the University of Strathclyde and the Microbiology Society.

References

Barakett, V., Morel, G., Lesage, D., Petit, J.C., 1993. Septic arthritis due to a nontoxigenic strain of *Corynebacterium diphtheriae* subspecies *mitis*. Clin Infect Dis 17, 520-521.

Barksdale, W.L., Pappenheimer, A.M., Jr., 1954. Phage-host relationships in nontoxigenic and toxigenic diphtheria bacilli. J Bacteriol 67, 220-232.

Baxter, D., 2007. Active and passive immunity, vaccine types, excipients and licensing. Occ. Med. 57, 552-556.

Belko, J., Wessel, D.L., Malley, R., 2000. Endocarditis caused by *Corynebacterium diphtheriae*: case report and review of the literature. Pediatr Infect Dis J 19, 159-163.

Bertuccini, L., Baldassarri, L., von Hunolstein, C., 2004. Internalization of non-toxigenic *Corynebacterium diphtheriae* by cultured human respiratory epithelial cells. Microb Pathog 37, 111-118.

Besa, N.C., Coldiron, M.E., Bakri, A., Raji, A., Nsuami, M.J., Rousseau, C., Hurtado, N., Porten, K., 2014. Diphtheria outbreak with high mortality in northeastern Nigeria. Epidemiol Infect 142, 797-802.

Bolt, F., Cassiday, P., Tondella, M.L., Dezoysa, A., Efstratiou, A., Sing, A., Zasada, A., Bernard, K., Guiso, N., Badell, E., Rosso, M.L., Baldwin, A., Dowson, C., 2010. Multilocus sequence typing identifies evidence for recombination and two distinct lineages of *Corynebacterium diphtheriae*. J Clin Microbiol 48, 4177-4185.

Both, L., Neal, S., De Zoysa, A., Mann, G., Czumbel, I., Efstratiou, A., Members of the European Diphtheria Surveillance, N., 2014. External quality assessments for microbiologic diagnosis of Diphtheria in Europe. J Clin Microbiol 52, 4381-4384.

Cassir, N., Bagnères, D., Fourneir, P. E., Broqui, P., Rossi, P. M. 2015. Cutaneous diphtheria: easy to be overlooked. Int. J. Infect. Dis. 33, 104–105

Chang, C., Mandlik, A., Das, A., Ton-That, H., 2011. Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. Mol. Microbiol. 79, 1236–1247.

Cerdeno-Tarraga, A.M., Efstratiou, A., Dover, L.G., Holden, M.T., Pallen, M., Bentley, S.D., Besra, G.S., Churcher, C., James, K.D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M.A., Rabbinowitsch, E., Rutherford, K.M., Thomson, N.R., Unwin, L., Whitehead, S., Barrell, B.G., Parkhill, J., 2003. The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. Nucleic Acids Res 31, 6516-6523.

Damian, M., Grimont, F., Narvskaya, O., Straut, M., Surdeanu, M., Cojocaru, R., Mokrousov, I., Diaconescu, A., Andronescu, C., Melnic, A., Mutoi, L., Grimont, P.A., 2002. Study of *Corynebacterium diphtheriae* strains isolated in Romania, northwestern Russia and the Republic of Moldova. Res Microbiol 153, 99-106.

De Zoysa, A., Efstratiou, A., Hawkey, P.M., 2005. Molecular characterization of diphtheria toxin repressor (*dtxR*) genes present in non-toxigenic *Corynebacterium diphtheriae* strains isolated in the United Kingdom. J Clin Microbiol 43, 223-228.

De Zoysa, A., Hawkey, P., Charlett, A., Efstratiou, A., 2008. Comparison of four molecular typing methods for characterization of *Corynebacterium diphtheriae* and determination of transcontinental spread of *C. diphtheriae* based on BstEII rRNA gene profiles. J Clin Microbiol 46, 3626-3635.

Dittmann, S., Wharton, M., Vitek, C., Ciotti, M., Galazka, A., Guichard, S., Hardy, I., Kartoglu, U., Koyama, S., Kreysler, J., Martin, B., Mercer, D., Ronne, T., Roure, C., Steinglass, R., Strebel, P., Sutter, R., Trostle, M., 2000. Successful control of epidemic diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. J Infect Dis 181 Suppl 1, S10-22.

Edwards, B., Hunt, A.C., Hoskisson, P.A., 2011. Recent cases of non-toxigenic *Corynebacterium diphtheriae* in Scotland: justification for continued surveillance. J Med Microbiol 60, 561-562.

Efstratiou, A., Engler, K.H., Mazurova, I.K., Glushkevich, T., Vuopio-Varkila, J., Popovic, T., 2000. Current approaches to the laboratory diagnosis of diphtheria. J Infect Dis 181 Suppl 1, S138-145.

Efstratiou, A., George, R.C., 1999. Laboratory guidelines for the diagnosis of infections caused by *Corynebacterium diphtheriae* and *C. ulcerans*. Comm. Dis. Publ. Health 2, 250-257.

Farfour, E., Badell, E., Zasada, A., Hotzel, H., Tomaso, H., Guillot, S., Guiso, N., 2012. Characterization and comparison of invasive *Corynebacterium diphtheriae* isolates from France and Poland. J Clin Microbiol 50, 173-175.

FitzGerald, R.P., Rosser, A.J., Perera, D.N., 2015. Non-toxigenic penicillin-resistant cutaneous *C. diphtheriae* infection: a case report and review of the literature. Journal of infection and public health 8, 98-100.

Fourel, G., Phalipon, A., Kaczorek, M., 1989. Evidence for direct regulation of diphtheria toxin gene transcription by an Fe2+-dependent DNA-binding repressor, DtoxR, in *Corynebacterium diphtheriae*. Infect Immun 57, 3221-3225.

Freeman, V.J., 1951. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. J Bacteriol 61, 675-688.

Funke, G., von Graevenitz, A., Clarridge, J.E., 3rd, Bernard, K.A., 1997. Clinical microbiology of coryneform bacteria. Clin Microbiol Rev 10, 125-159.

Galazka, A., 2000. The changing epidemiology of diphtheria in the vaccine era. J Infect Dis 181 Suppl 1, S2-9.

Ganz, T., Nemeth, E., 2015. Iron homeostasis in host defence and inflammation. Nat. Rev. Immunol. 15, 500-510.

Goodfellow, M., Kaempfer, P., Busse, H.-J., Trujillo, M.E., Suzuki, K.-i., Ludwig, W., Whitman, W.B., 2012. Bergey's Manual of Systematic Bacteriology, in: Whitman, W.B. (Ed.), The *Actinobacteria*, Part A, 2 ed. Springer, London, p. 1034.

Gordon, C.L., Fagan, P., Hennessy, J., Baird, R., 2011. Characterization of *Corynebacterium diphtheriae* isolates from infected skin lesions in the Northern Territory of Australia. J Clin Microbiol 49, 3960-3962.

Grimont, P.A., Grimont, F., Efstratiou, A., De Zoysa, A., Mazurova, I., Ruckly, C., Lejay-Collin, M., Martin-Delautre, S., Regnault, B., European Laboratory Working Group on, D., 2004. International nomenclature for *Corynebacterium diphtheriae* ribotypes. Res Microbiol 155, 162-166.

Hadfield, T.L., McEvoy, P., Polotsky, Y., Tzinserling, V.A., Yakovlev, A.A., 2000. The pathology of diphtheria. J Infect Dis 181 Suppl 1, S116-120.

Hirata, R., Napoleao, F., Monteiro-Leal, L.H., Andrade, A.F., Nagao, P.E., Formiga, L.C., Fonseca, L.S., Mattos-Guaraldi, A.L., 2002. Intracellular viability of toxigenic *Corynebacterium diphtheriae* strains in HEp-2 cells. FEMS Microbiol Lett 215, 115-119.

Hoskisson, P. A. & Smith, M. C. M., 2007. Hypervariation and phase variation in the bacteriophage 'resistome'. Curr. Opin. Microbiol. 10, 396–400.

Hughes, G. J., Mikhail, A. F. W., Husada, D., Irawan, E., Kafatos, G., Bracebridge, S., Pebody, R., Efstratiou, A., 2015. Seroprevalence and Determinants of Immunity to Diphtheria for Children Living in Two Districts of Contrasting Incidence During an Outbreak in East Java, Indonesia. Pediatr. Infect. Dis. J. 34, 1152–1156.

Huhulescu, S., Hirk, S., Zeinzinger, V., Hasenberger, P., Skvara, P., Müllegger, R., Allerberger, F Indra., A.2014. Letter to the editor: cutaneous diphtheria in a migrant from an endemic country in east Africa, Austria May 2014. Euro Surveill. 19.

Joseph, B., Schwarz, R.F., Linke, B., Blom, J., Becker, A., Claus, H., Goesmann, A., Frosch, M., Muller, T., Vogel, U., Schoen, C., 2011. Virulence evolution of the human pathogen *Neisseria meningitidis* by recombination in the core and accessory genome. PLoS One 6, e18441.

Kolodkina, V., Titov, L., Sharapa, T., Grimont, F., Grimont, P.A., Efstratiou, A., 2006. Molecular epidemiology of *C. diphtheriae* strains during different phases of the diphtheria epidemic in Belarus. BMC Infect Dis 6, 129.

Landazabal, G.N., Burgos Rodriguez, M.M., Pastor, D., 2001. Diphtheria outbreak in Cali, Colombia, August-October 2000. Epidemiological bulletin 22, 13-15.

Lindhusen-Lindhe, E., Dotevall, L., Berglund, M., 2012. Imported laryngeal and cutaneous diphtheria in tourists returning from western Africa to Sweden, March 2012. Euro Surveill 17.

Litwin, C.M., Calderwood, S.B., 1993. Role of iron in regulation of virulence genes. Clin Microbiol Rev 6, 137-149.

Maiden, M.C., 2006. Multilocus sequence typing of bacteria. Annu Rev Microbiol 60, 561-588.

Mandlik, A., Swierczynski, A., Das, A., Ton-That, H., 2007. *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. Mol Microbiol 64, 111-124.

Marks, L.R., Reddinger, R.M., Hakansson, A.P., 2012. High Levels of Genetic Recombination during Nasopharyngeal Carriage and Biofilm Formation in *Streptococcus pneumoniae*. MBio 3, e00200-00212.

Marraffini, L.A., 2015. CRISPR-Cas immunity in prokaryotes. Nature 526, 55-61.

May, M.L., McDougall, R.J., Robson, J.M., 2014. *Corynebacterium diphtheriae* and the returned tropical traveler. J. Travel Med. 21, 39-44.

Mokrousov, I., 2013. *Corynebacterium diphtheriae*, in: de Filippis, I., McKee, M.L. (Eds.), Molecular Typing in Bacterial Infections. Humana Press, New York, USA, pp. 283-300.

Mokrousov, I., Limeschenko, E., Vyazovaya, A., Narvskaya, O., 2007. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. Biotechnol J 2, 901-906.

Mokrousov, I., Narvskaya, O., Limeschenko, E., Vyazovaya, A., 2005. Efficient discrimination within a *Corynebacterium diphtheriae* epidemic clonal group by a novel macroarray-based method. J Clin Microbiol 43, 1662-1668.

Mokrousov, I., Vyazovaya, A., Kolodkina, V., Limeschenko, E., Titov, L., Narvskaya, O., 2009. Novel macroarray-based method of *Corynebacterium diphtheriae* genotyping: evaluation in a field study in Belarus. Eur J Clin Microbiol Infect Dis 28, 701-703.

Nanthavong, N., Black, A.P., Nouanthong, P., Souvannaso, C., Vilivong, K., Muller, C.P., Goossens, S., Quet, F., Buisson, Y., 2015. Diphtheria in Lao PDR: Insufficient Coverage or Ineffective Vaccine? PLoS One 10, e0121749.

Neal, S.E., Efstratiou, A., 2009. International external quality assurance for laboratory diagnosis of diphtheria. J Clin Microbiol 47, 4037-4042.

Nelson, T. G., Mitchell, C. D., Sega-Hall, G. M. & Porter, R. J. 2016. Cutaneous ulcers in a returning traveller: a rare case of imported diphtheria in the UK. Clin. Exp. Dermatol. 41, 57–59

Ott, L., Holler, M., Rheinlaender, J., Schaffer, T.E., Hensel, M., Burkovski, A., 2010. Strain-specific differences in pili formation and the interaction of *Corynebacterium diphtheriae* with host cells. BMC Microbiol 10, 257.

Parande, M.V., Parande, A.M., Lakkannavar, S.L., Kholkute, S.D., Roy, S., 2014. Diphtheria outbreak in rural North Karnataka, India. JMM Case Rep 1, 1-3.

Patey, O., Bimet, F., Riegel, P., Halioua, B., Emond, J.P., Estrangin, E., Dellion, S., Alonso, J.M., Kiredjian, M., Dublanchet, A., Lafaix, C., 1997. Clinical and molecular study of

Corynebacterium diphtheriae systemic infections in France. Coryne Study Group. J Clin Microbiol 35, 441-445.

Peel, M.M., Palmer, G.G., Stacpoole, A.M., Kerr, T.G., 1997. Human lymphadenitis due to *Corynebacterium pseudotuberculosis*: report of ten cases from Australia and review. Clin Infect Dis 24, 185-191.

Peixoto, R.S., Pereira, G.A., Sanches dos Santos, L., Rocha-de-Souza, C.M., Gomes, D.L., Silva Dos Santos, C., Werneck, L.M., Dias, A.A., Hirata, R., Jr., Nagao, P.E., Mattos-Guaraldi, A.L., 2014. Invasion of endothelial cells and arthritogenic potential of endocarditis-associated *Corynebacterium diphtheriae*. Microbiology 160, 537-546.

Poilane, I., Fawaz, F., Nathanson, M., Cruaud, P., Martin, T., Collignon, A., Gaudelus, J., 1995. *Corynebacterium diphtheriae* osteomyelitis in an immunocompetent child: a case report. Euro. J.Pediatr. 154, 381-383.

Puliti, M., von Hunolstein, C., Marangi, M., Bistoni, F., Tissi, L., 2006. Experimental model of infection with non-toxigenic strains of *Corynebacterium diphtheriae* and development of septic arthritis. J Med Microbiol 55, 229-235.

Rasmussen, I., Wallace, S., Mengshoel, A.T., Hoiby, E.A., Brandtzaeg, P., 2011. Diphtheria outbreak in Norway: lessons learned. Scand. J. Infect. Dis. 43, 986-989.

Reardon-Robinson, M.E., Ton-That, H., 2014. Assembly and function of *Corynebacterium diphtheriae* pili, in: Burkovski, A. (Ed.), *Corynebacterium diphtheriae* and related toxigenic species. Springer, Heidelberg, pp. 123-141.

Romney, M.G., Roscoe, D.L., Bernard, K., Lai, S., Efstratiou, A., Clarke, A.M., 2006. Emergence of an invasive clone of nontoxigenic *Corynebacterium diphtheriae* in the urban poor population of Vancouver, Canada. J Clin Microbiol 44, 1625-1629.

Saikia, L., Nath, R., Saikia, N.J., Choudhury, G., Sarkar, M., 2010. A diphtheria outbreak in Assam, India. Southeast Asian J. Trop. Med. Public Health 41, 647-652.

Sangal, V., Blom, J., Sutcliffe, I.C., von Hunolstein, C., Burkovski, A., Hoskisson, P.A., 2015. Adherence and invasive properties of *Corynebacterium diphtheriae* strains correlates with the predicted membrane-associated and secreted proteome. BMC Genomics 16, 765.

Sangal, V., Burkovski, A., Hunt, A.C., Edwards, B., Blom, J., Hoskisson, P.A., 2014. A lack of genetic basis for biovar differentiation in clinically important *Corynebacterium diphtheriae* from whole genome sequencing. Infect Genet Evol 21, 54-57.

Sangal, V., Fineran, P.C., Hoskisson, P.A., 2013. Novel configurations of type I and II CRISPR-Cas systems in *Corynebacterium diphtheriae*. Microbiology 159, 2118-2126.

Sangal, V., Harbottle, H., Mazzoni, C.J., Helmuth, R., Guerra, B., Didelot, X., Paglietti, B., Rabsch, W., Brisse, S., Weill, F.X., Roumagnac, P., Achtman, M., 2010. Evolution and Population Structure of *Salmonella enterica* Serovar Newport. J Bacteriol 192, 6465-6476.

Sangal, V., Hoskisson, P.A., 2014a. Corynephages: infections of the infectors, in: Burkovski, A. (Ed.), *Corynebacterium diphtheriae* and related toxigenic species. Springer, Heidelberg, pp. 67-82.

Sangal, V., Nieminen, L., Weinhardt, B., Raeside, J., Tucker, N. P., Florea, C-D., Pollock, K. G., Hoskisson, P. A., 2014b Genome sequencing of a toxigenic *Corynebacterium ulcerans* strain causing a diphtheria-like disease. Emerg. Infect. Dis. 20, 1257-1258.

Sangal, V., Tucker, N.P., Burkovski, A., Hoskisson, P.A., 2012a. Draft genome sequence of *Corynebacterium diphtheriae* biovar intermedius NCTC 5011. J Bacteriol 194, 4738.

Sangal, V., Tucker, N.P., Burkovski, A., Hoskisson, P.A., 2012b. The draft genome sequence of *Corynebacterium diphtheriae* bv. mitis NCTC 3529 reveals significant diversity between the primary disease-causing biovars. J Bacteriol 194, 3269.

Santos, L.S., Sant'anna, L.O., Ramos, J.N., Ladeira, E.M., Stavracakis-Peixoto, R., Borges, L.L., Santos, C.S., Napoleao, F., Camello, T.C., Pereira, G.A., Hirata, R., Vieira, V.V., Cosme, L.M., Sabbadini, P.S., Mattos-Guaraldi, A.L., 2015. Diphtheria outbreak in Maranhao, Brazil: microbiological, clinical and epidemiological aspects. Epidemiol Infect 143, 791-798.

Sing, A., Konrad, R., Meinel, D.M., Mauder, N., Schwabe, I., Sting, R., 2015. *Corynebacterium diphtheriae* in a free-roaming red fox: case report and historical review on diphtheria in animals. Infection.pp 1-5.

Suarez, D.L., Senne, D.A., Banks, J., Brown, I.H., Essen, S.C., Lee, C.W., Manvell, R.J., Mathieu-Benson, C., Moreno, V., Pedersen, J.C., Panigrahy, B., Rojas, H., Spackman, E., Alexander, D.J., 2004. Recombination resulting in virulence shift in avian influenza outbreak, Chile. Emerg Infect Dis 10, 693-699.

Taylor, J., Saveedra-Campos, M., Harwood, D., Pritchard, G., Raphaely, N., Kapadia, S., Efstratiou, A., White, J., Balasegaram, S., 2010. Toxigenic *Corynebacterium ulcerans* infection in a veterinary student in London, United Kingdom, May 2010. Euro Surveill 15.

Tiley, S.M., Kociuba, K.R., Heron, L.G., Munro, R., 1993. Infective endocarditis due to nontoxigenic *Corynebacterium diphtheriae*: report of seven cases and review. Clin Infect Dis 16, 271-275.

Titov, L., Kolodkina, V., Dronina, A., Grimont, F., Grimont, P.A., Lejay-Collin, M., de Zoysa, A., Andronescu, C., Diaconescu, A., Marin, B., Efstratiou, A., 2003. Genotypic and phenotypic characteristics of *Corynebacterium diphtheriae* strains isolated from patients in belarus during an epidemic period. J Clin Microbiol 41, 1285-1288.

Trost, E., Blom, J., Soares Sde, C., Huang, I.H., Al-Dilaimi, A., Schroder, J., Jaenicke, S., Dorella, F.A., Rocha, F.S., Miyoshi, A., Azevedo, V., Schneider, M.P., Silva, A., Camello, T.C., Sabbadini, P.S., Santos, C.S., Santos, L.S., Hirata, R., Jr., Mattos-Guaraldi, A.L., Efstratiou, A., Schmitt, M.P., 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.

Vitek, C. R. & Wharton, M., 1998. Diphtheria in the former Soviet Union: reemergence of a pandemic disease. Emerg Infect Dis 4, 539–550.

von Hunolstein, C., Alfarone, G., Scopetti, F., Pataracchia, M., La Valle, R., Franchi, F., Pacciani, L., Manera, A., Giammanco, A., Farinelli, S., Engler, K., De Zoysa, A., Efstratiou,

A., 2003. Molecular epidemiology and characteristics of *Corynebacterium diphtheriae* and *Corynebacterium ulcerans* strains isolated in Italy during the 1990s. J Med Microbiol 52, 181-188.

Wagner, K.S., White, J.M., Neal, S., Crowcroft, N.S., Kupreviciene, N., Paberza, R., Lucenko, I., Joks, U., Akbas, E., Alexandrou-Athanassoulis, H., Detcheva, A., Vuopio, J., von Hunolstein, C., Murphy, P.G., Andrews, N., Efstratiou, A., 2011. Screening for *Corynebacterium diphtheriae* and *Corynebacterium ulcerans* in patients with upper respiratory tract infections 2007-2008: a multicentre European study. Clin Microbiol Infect 17, 519-525.

Wanlapakorn, N., Yoocharoen, P., Tharmaphornpilas, P., Theamboonlers, A., Poovorawan, Y., 2014. Diphtheria outbreak in Thailand, 2012; seroprevalence of diphtheria antibodies among Thai adults and its implications for immunization programs. Southeast Asian J. Trop. Med. Public Health 45, 1132-1141.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C., Ochman, H., Achtman, M., 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60, 1136-1151.

Zakikhany, K., Neal, S., Efstratiou, A., 2014. Emergence and molecular characterisation of non-toxigenic tox gene-bearing *Corynebacterium diphtheriae* biovar mitis in the United Kingdom, 2003-2012. Euro Surveill 19.

604 **Figure Legends**

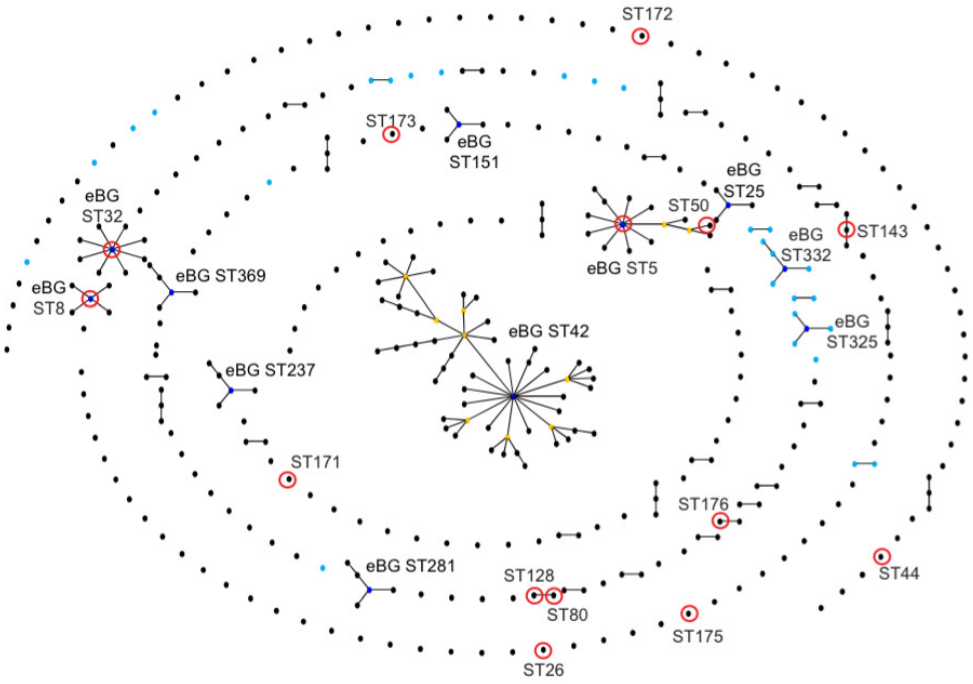
605 **Fig. 1.** An eBURST diagram from the MLST profiles of reference STs from the MLST
606 website (http://pubmlst.org/cdiphtheriae/). The predicted founder STs are shown in blue and
607 co-founder STs are shown in yellow. Single locus variants (SLVs) are connected to each
608 other and major groups where predicted founder has three or more SLVs are labelled. The
609 known STs for *C. ulcerans* are shown in cyan. ST with some genome sequenced strains are
610 encircled in red.

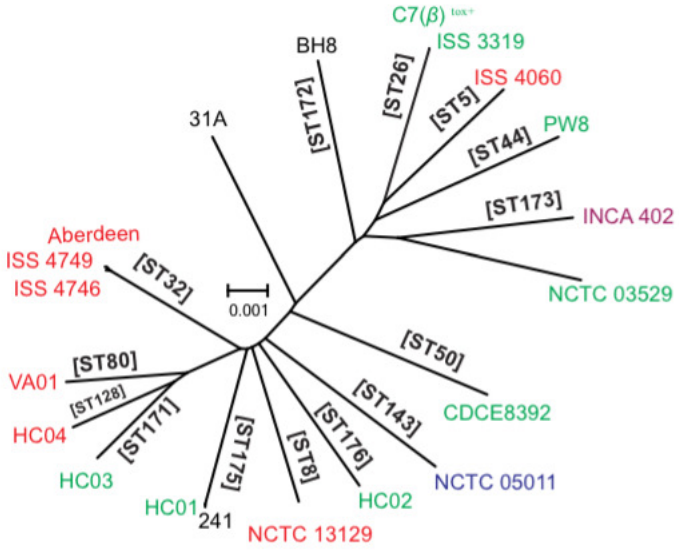
611

612 **Fig. 2.** A phylogenetic tree from the core genome of *C. diphtheriae* (adapted from Sangal et
613 al., 2015). ST designations are mapped on the tree in parentheses, if known. The strains
614 biovars gravis, mitis, belfanti and intermedius are labelled in red, green, purple and blue,
615 respectively.

616

Fig 1.


ST172
ST173
eBG ST151
eBG ST25
ST50
eBG ST32
eBG ST5
ST143
eBG ST332
eBG ST369
eBG ST8
eBG ST42
eBG ST325
eBG ST237
ST171
ST176
eBG ST281
ST128
ST44
ST80
ST175
ST26



630 Fig. 2.

631

632