Infection, Genetics and Evolution 43 (2016) 364–370

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Review

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

Vartul Sangal [a], Paul A. Hoskisson [b,*]

[a] Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
[b] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow, G4 0RE, UK

ARTICLE INFO

Article history:
Received 4 April 2016
Received in revised form 7 June 2016
Accepted 8 June 2016
Available online 9 June 2016

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.

Contents

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, 2014) 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 and Wharton, 1998). In addition, there is still considerable morbidity and

* Corresponding author.
E-mail address: Paul.hoskisson@strath.ac.uk (P.A. Hoskisson).

http://dx.doi.org/10.1016/j.meegid.2016.06.024
1567-1348/

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, 7321 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, *C*1 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,