

Genomic epidemiology of nontoxigenic *Corynebacterium diphtheriae* from King County, Washington State, USA between July 2018 and May 2019

Lingzi Xiaoli¹, Eileen Benoliel², Yanhui Peng³, Janessa Aneke¹, Pamela K. Cassidy³, Meagan Kay², Shelly McKeirnan², Jeffery S. Duchin², Vance Kawakami², Scott Lindquist⁴, Anna M. Acosta³, Chas DeBolt⁴, Maria Lucia Tondella³ and Michael R. Weigand^{3,*}

Abstract

Between July 2018 and May 2019, *Corynebacterium diphtheriae* was isolated from eight patients with non-respiratory infections, seven of whom experienced homelessness and had stayed at shelters in King County, WA, USA. All isolates were microbiologically identified as nontoxigenic *C. diphtheriae* biovar mitis. Whole-genome sequencing confirmed that all case isolates were genetically related, associated with sequence type 445 and differing by fewer than 24 single-nucleotide polymorphisms (SNPs). Compared to publicly available *C. diphtheriae* genomic data, these WA isolates formed a discrete cluster with SNP variation consistent with previously reported outbreaks. Virulence-related gene content variation within the highly related WA cluster isolates was also observed. These results indicated that genome characterization can readily support epidemiology of nontoxigenic *C. diphtheriae*.

DATA SUMMARY

Genome sequence data have been deposited in GenBank, organized under BioProject accession number PRJNA541849. The complete genome assembly for PC0646 is available under accession number CP040557.

INTRODUCTION

Diphtheria is characterized as toxin-mediated respiratory or cutaneous infections caused by toxigenic strains of *Corynebacterium diphtheriae* [1]. Following widespread application of effective vaccines against the diphtheria toxin, respiratory diphtheria cases have decreased significantly worldwide [2]. However, increased infections by nontoxigenic *C. diphtheriae* strains, which are not preventable by current vaccines targeting diphtheria toxin, have been reported

[3–6]. Nontoxigenic *C. diphtheriae* infections often occur at cutaneous lesions but can progress to severe diseases such as bacteraemia, septic arthritis and endocarditis [3–6]. Population groups particularly at risk of *C. diphtheriae* infections include refugees [7], foreign travellers [8], intravenous drug users, or persons experiencing homelessness in metropolitan areas [3–6].

Molecular approaches developed for the surveillance and outbreak clustering of diphtheria can also inform epidemiology of *C. diphtheriae*. A multilocus sequence typing (MLST) scheme targeting housekeeping genes has been widely used for *C. diphtheriae* subtyping and currently includes more than 664 reported sequence types (STs) [9]. However, MLST provides limited resolution to study bacterial genetic relatedness, particularly within outbreaks that include case isolates of a single ST. MLST is also incapable of detecting gene content

Received 13 August 2020; Accepted 16 October 2020; Published 04 December 2020

Author affiliations: ¹IHRC, Inc., Atlanta, Georgia, USA; ²Public Health Seattle and King County, Seattle, WA, USA; ³Division of Bacterial Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA; ⁴Washington State Department of Health, Shoreline, WA, USA.

*Correspondence: Michael R. Weigand, yrh8@cdc.gov

Keywords: *C. diphtheriae*; genomics; nontoxigenic; SNP; virulence factor.

Abbreviations: MIC, minimum inhibitory concentration; MLST, multilocus sequence typing; SNP, single nucleotide polymorphism; ST, sequence type; VFDB, Virulence Factors Database; WGS, whole-genome sequencing.

Genome sequence data have been deposited under BioProject accession number PRJNA541849 in GenBank. The complete genome assembly for PC0646 is available under accession number CP040557. The raw reads are under accession numbers SRR12270031–SRR12270040.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Six supplementary tables and one supplementary figure are available with the online version of this article.

000467



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

variations such as acquisition of certain antibiotic resistance genes and virulence factors. Alternatively, whole-genome sequencing (WGS) captures sequence variations across the full genome and thus can simultaneously be used to elucidate genetic links between cases and identify antibiotic resistance markers and virulence factors among isolates. Previous studies have successfully used whole-genome single nucleotide polymorphisms (SNPs) to confirm the clustering of isolates from diphtheria cases with known epidemiological or geographical associations [7, 10, 11]. Together, these reports suggest that isolates recovered from outbreak clusters of diphtheria frequently differ by fewer than 150 SNPs, whereas unlinked sporadic isolates differ by an average of 30000 SNPs [7].

A few investigations of nontoxigenic *C. diphtheriae* have included WGS analyses, but the retrospective designs of those studies have been complicated by large temporal or geographical distributions among studied case isolates as well as sparse epidemiological data [12, 13]. Nontoxigenic *C. diphtheriae* is of public health concern due to the lack of protection provided by current diphtheria vaccines and the potential for such circulating strains to readily become toxigenic through lysogenization by toxin-encoding bacteriophages [14]. In the present study, we employed WGS to compare nontoxigenic *C. diphtheriae* isolates primarily recovered from skin lesions of epidemiologically linked cases in King County, WA, USA and quantify SNP accumulation during outbreak transmission.

METHODS

Outbreak information

Between July 2018 and May 2019, 10 *C. diphtheriae* isolates were recovered from eight patients, seven of whom experienced homelessness and had stayed at shelters in King County, WA, USA. Patients were predominantly male (87.5%), had a median age of 51.5 (range, 35–67) and had a documented

Outcome

In this study, we investigated an outbreak cluster of nontoxigenic *C. diphtheriae* emerging in King County, WA, USA using whole-genome sequencing. Our genomic analyses confirmed clustering of these cases and revealed that *C. diphtheriae* can exhibit gene content variation during outbreak transmission. Thus, increased awareness of the risks for persistent nontoxigenic *C. diphtheriae* cutaneous infections among at-risk populations is needed.

history of intravenous drug (patients 2, 5 and 8) or methamphetamine abuse (patients 1, 4 and 7). These individuals registered stays at multiple homeless shelters in King County, including three in downtown Seattle (shelters A, B and C) separated by 0.3 miles and another (shelter D) approximately 2 miles away. The first *C. diphtheriae* case isolate (PC0646) was collected in July 2018, followed by seven cases reported between Jan 2019 and May 2019. Patients 1, 4, 5 and 6 registered overlapping visits at shelter A, ranging from 1 to 166 nights between October 2018 and April 2019, and all four patients were present in this shelter on one night. During investigation of the outbreak, a second report of *C. diphtheriae* infection was identified for two patients (patients 1 and 2), resulting in two isolates each from these individuals, while all other patients contributed one isolate to the study. Basic epidemiological case information is summarized in Table 1.

Microbiological identification

Isolates were grown on trypticase soy agar with 5% sheep blood (BBL, Sparks, MD, USA) and biochemically identified using API Coryne strips (bioMérieux, Durham,

Table 1. Epidemiological and microbiological description of the WA *C. diphtheriae* isolates

Patient no.	Isolate	Collection date	Homeless	Location	Isolate source	MLST	Penicillin MIC
1	PC0647	Jan–2019	Yes	Shelter A	Leg wound	445	0.25
1	PC0650	Feb–2019	Yes	Shelter A	Leg wound	445	0.25
2	PC0646	Jul–2018	Yes	Street	Arm wound	445	0.19
2	PC0652	Mar–2019	Yes	Shelter D	Abdomen abscess	NT*	0.25
3	PC0648	Jan–2019	Yes	Shelter A	Blood	445	0.38
4	PC0649	Feb–2019	Yes	Shelters A, B	Leg wound	445	0.25
5	PC0651	Mar–2019	Yes	Shelter A	Buttock wound	445	0.25
6	PC0653	Mar–2019	Yes	Shelters A, C	Neck wound	445	0.38
7	PC0654	Apr–2019	No	NA	Foot ulcer	445	0.19
8	PC0655	May–2019	Yes	Shelter B	Hand wound	445	0.38

*Single-locus variant of ST445 with synonymous SNP in *fusA*.
MIC, minimum inhibitory concentration (mg l⁻¹); NT, new type; NA, not applicable.

NC, USA). Toxin production was confirmed by Elek test [15] and antibiotic resistance to 11 different antibiotics (amoxicillin, azithromycin, clarithromycin, clindamycin, daptomycin, erythromycin, levofloxacin, meropenem, penicillin, rifampin and vancomycin) was determined by Etest (bioMérieux, Durham, NC, USA). Antibiotic resistance interpretative categories followed the Clinical and Laboratory Standard Institute (CLSI) 2015 guidelines, which for penicillin included 'sensitive' ($\text{MIC} \leq 0.12 \text{ mg l}^{-1}$), 'intermediate' ($\text{MIC} > 0.25 - 2.0 \text{ mg l}^{-1}$), or resistant ($\text{MIC} \geq 4.0 \text{ mg l}^{-1}$). As clinical breakpoints for *Corynebacterium* species are not defined for amoxicillin, azithromycin, clarithromycin, or levofloxacin, these were not assigned an interpretive category.

WGS

All *C. diphtheriae* isolates were grown on trypticase soy agar with 5% sheep blood at 37°C for 24 h. Genomic DNA was extracted using the Maxwell RSC Whole Blood DNA kit (Promega, San Luis Obispo, CA, USA). Genomic DNA for PacBio sequencing was further cleaned by salt/chloroform washing [16]. Genomic DNA concentrations were determined using the Qubit dsDNA broad range quantification kit (Thermo Fisher Scientific, Waltham, MA, USA). Paired-end libraries (2×250 bp) were prepared with the NEBNext Ultra DNA Library Prep kit (New England Biolabs, Ipswich, MA, USA) and sequenced on a Miseq using reagent v2, 500-cycle kit (Illumina, San Diego, CA, USA). The Illumina raw reads were checked for quality using FastQC v0.11.5 [17] and trimmed by Cutadapt [18]. Trimmed reads were *de novo* assembled using SPAdes v3.9 [19] and evaluated by QUAST v4.5 [20].

An additional PacBio library was prepared from PC0646 using the SMRTbell Template Prep kit 1.0 and Polymerase Binding kit P4. It was sequenced on the PacBio RSII sequencer (Pacific Biosciences, Menlo Park, CA, USA). *De novo* assembly was performed using the Hierarchical Genome Assembly Process (HGAP v3) (Pacific Biosciences, Menlo Park, CA, USA) [21]. The resulting consensus sequence was manually checked for circularity using Gepard v1.30 [22] and further polished by mapping Illumina reads in CLC Genomics Workbench v12 (CLC Bio, Boston, MA, USA). The complete genome assembly of PC0646 was annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP).

MLST and diphtheria toxin gene detection

MLST allele profiles were determined from the assembled contigs according to pubMLST [9], and further confirmed with trimmed sequencing reads using SRST2 [23]. The presence of diphtheria toxin homologues was determined by TBLASTN [24] query of the assembled genomes using select references: corynephage beta A and B subunits (NCBI accession number P00588), corynephage omega beta diphtheria toxin (accession number P00587), and corynephage beta diphtheria toxin homologue (accession number P00589).

Additionally, trimmed sequencing reads from all *C. diphtheriae* isolates were mapped to the reference genome of NCTC13129 (accession number NC_002935.2) using Snippy v4.3.8 with default settings [25] and coverage across the diphtheria toxin gene was calculated with SAMtools [26]. The detected presence or absence of the diphtheria toxin gene in each *C. diphtheriae* isolate is reported in Table S1 (available in the online version of this article).

Whole-genome SNP phylogeny and characterization

Publicly available *C. diphtheriae* genomic data for 305 isolates was retrieved from the NCBI, including 284 raw Illumina sequencing read sets from SRA (downloaded on 22 November 2019) and 21 complete genomes (Table S2). Raw reads were trimmed and filtered as described above. SNPs were determined by mapping trimmed reads to the reference genome of NCTC13129 using Snippy v4.3.8 with default settings [25]. The resulting core SNP alignment was used to estimate the phylogeny using maximum-likelihood with RaxML v8.2.9 [27]. The final tree was visualized with iTOL v4 [28]. Construction of additional subtrees, including for each of the clusters reported previously, followed a similar procedure using a reference genome selected from within each cluster. SNPs among the WA isolates were determined relative to the complete genome assembly of PC0646 (accession number CP040557) using Snippy and further characterized using SnpEff v4.3 [29]. Pairwise SNP distances were calculated from the subtree core alignments using SNP-Dist v4.0 [30].

Virulence factor and resistance gene detection

The VF analyser from the Virulence Factors DataBase (VFDB) [31] was used to predict the virulence factor profiles of WA isolates using their genome assemblies. The predicted nucleotide sequences for *srtB*, *srtC* and *spaD* were extracted from genome assemblies of PC0653 and PC0654 for BLASTP [24] query against the NCBI non-redundant (nr) protein database. The unique region harbouring the above three genes was extracted from PC0653 and PC0654 and translated with GeneMark [32]. Additional protein-coding genes predicted by GeneMark were further BLASTP [24] queried in the NCBI nr protein database. In a similar manner, the presence of the novel penicillin binding protein encoding gene *pbp2m* [33] was also evaluated by querying the assembled contigs using BLASTN [24].

RESULTS

Molecular characterization and antimicrobial susceptibility

Diagnostic microbiological assays identified all case isolates as nontoxigenic *C. diphtheriae* biovar mitis with intermediate resistance to penicillin (Table 1). The isolates were susceptible to 10 additional antibiotics and the *pbp2m* beta-lactam resistance gene was not detected in any assembled genomes. MLST determined from either assembled genomes or trimmed reads indicated that nine isolates shared ST445 (*aptA*, 3; *dnaE*, 2;

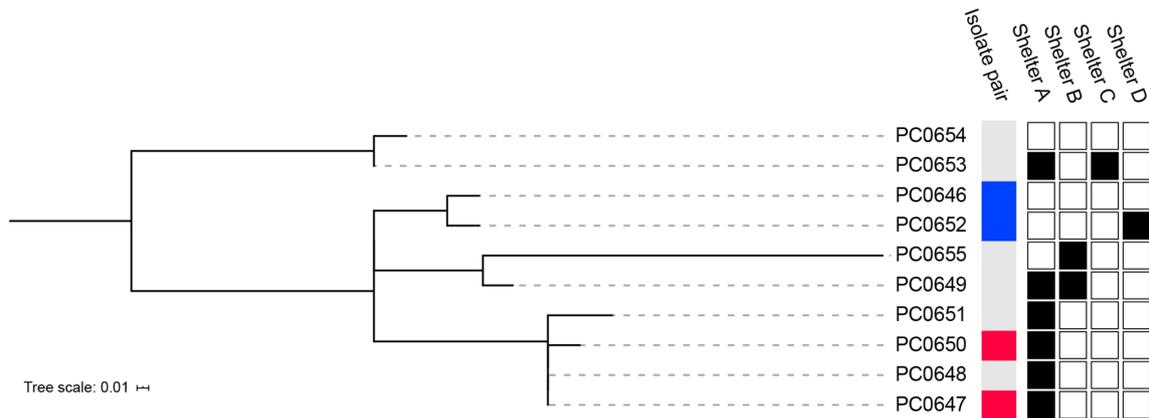


Fig. 1. Phylogenetic reconstruction of the WA *C. diphtheriae* isolates from 37 core variable sites using maximum likelihood. Isolate pairs recovered from the same patient and visits to specific shelter locations are indicated next to the tree.

dnaK, 60; *fusA*, 30; *leuA*, 3; *odhA*, 3; *rpoB*, 2), while PC0652 exhibited a single locus variant due to a synonymous SNP in *fusA*, whose sequence has been submitted to pubMLST and assigned as *fusA* 83.

Core SNP phylogeny

The phylogeny of WA *C. diphtheriae* isolates was reconstructed from 37 core SNPs to investigate their genetic relatedness using the complete genome of PC0646 as a reference (Fig. 1). The average pairwise distance between isolates was 11.6 SNPs (range: 0–24) (Table S3), consistent with other reported outbreaks of *C. diphtheriae* (Table S4). Isolates from three patients (PC0647, PC0648, PC0650, PC0651), all of whom attended shelter A, formed a tight cluster and differed from each other by an average of 1.5 SNPs (range: 0–3) (Fig. 1). Pairs of isolates recovered from individual patients (PC0646 and PC0652, PC0647 and PC0650) clustered together. In total, 51 unique mutations were detected among the WA isolates, 35 of which appeared in protein-coding genes (12 synonymous, 17 nonsynonymous and 6 indels) of various predicted functions that did not include known virulence factors (Table S5). Only a single SNP was observed between PC0647 and PC0650, a

missense mutation in a gene encoding a predicted cell surface protein with high sequence similarity to substrate-binding components of ABC transporters. PC0646 and PC0652 differed by 2 SNPs, a synonymous substitution in *fusA* and an intergenic substitution upstream of an AraC family transcriptional regulator (Table 2).

Phylogenetic placement of the WA isolates was further investigated within the context of all publicly available *C. diphtheriae* genomic data from 305 isolates, including reported outbreak clusters from Switzerland and other geographically defined clusters (Fig. S1, Table 1). Among the publicly available genomic data, FRC0157 (ST367) was the most closely related to the WA isolates but still differed by more than 12000 SNPs.

Virulence factor profiles

All WA isolates were nontoxicogenic as determined by the Elek test. The absence of the toxin-encoding corynebacteriophage was also confirmed by the genomic data. Genome sequence queries with a collection of known virulence factors detected the presence of 23 putative matches, most of which were

Table 2. Single-nucleotide polymorphism annotations for the WA *C. diphtheriae* isolates from the same patient

Position*	Ref	Alt	Type	Gene	Amino acids	Function
PC0652 vs PC0646						
370633	C	T	synonymous_variant	<i>fusA</i>	p.Ser314Ser	elongation factor G
808375	T	C	upstream_gene_variant	FGA20_04195		AraC family transcriptional regulator
PC0647 vs PC0650						
2323370	G	T	missense_variant	FGA20_11260	p.Leu508Met	Cell surface protein

*Position in complete assembly of PC0646 (accession number CP040557).

Table 3. Predicated virulence factor profiles for the WA isolates and reference genomes

Virulence factors	Genes	VFDB ID	NCTC13129* (Locus ID)	CD10†	PC0653 PC0654	PC0646‡
SpaA-type pili	<i>spaA</i>	VFG002201	DIP_RS21255	Y	Y	Y
	<i>spaB</i>	VFG002200	DIP_RS21245	Y	Y	Y
	<i>spaC</i>	VFG002199	DIP_RS21240	Y	Y	Y
	<i>srtA</i>	VFG013668	DIP_RS21250	Y	Y	Y
SpaD-type pili	<i>spaD</i>	VFG002202	DIP_RS12565	-	Y	-
	<i>spaE</i>	VFG002203	DIP_RS12575	-	-	-
	<i>spaF</i>	VFG002204	DIP_RS12580	-	-	-
	<i>srtB</i>	VFG013672	DIP_RS12555	Y	Y	-
	<i>srtC</i>	VFG013676	DIP_RS12570	Y	Y	-
SpaH-type pili	<i>spaG</i>	VFG002207	DIP_RS22255	-	-	-
	<i>spaH</i>	VFG002206	DIP_RS22250	-	-	-
	<i>spaI</i>	VFG002205	DIP_RS22235	-	-	-
	<i>srtD</i>	VFG013685	DIP_RS22245	-	-	-
	<i>srtE</i>	VFG013682	DIP_RS22240	-	-	-
Surface-anchored pilus proteins	<i>sapA</i>	VFG013691	DIP_RS21475	-	-	-
	<i>sapD</i>	VFG013693	DIP_RS13555	Y	Y	Y
	<i>sapE</i>	VFG013695	-	-	-	-
ABC transporter	<i>fagA</i>	VFG013742	DIP_RS16495	Y	Y	Y
	<i>fagB</i>	VFG013736	DIP_RS16490	Y	Y	Y
	<i>fagC</i>	VFG013730	DIP_RS16485	Y	Y	Y
	<i>fagD</i>	VFG013748	DIP_RS16500	Y	Y	Y
ABC-type haem transporter	<i>hmuT</i>	VFG013704	DIP_RS14430	Y	Y	Y
	<i>hmuU</i>	VFG013709	DIP_RS14435	Y	Y	Y
	<i>hmuV</i>	VFG013715	DIP_RS14440	Y	Y	Y
Siderophore-dependent iron uptake system	<i>irp6A</i>	VFG013698	DIP_RS11960	Y	Y	Y
	<i>irp6B</i>	VFG013700	DIP_RS11965	Y	Y	Y
	<i>irp6C</i>	VFG013702	DIP_RS11970	Y	Y	Y
ciu iron uptake and siderophore biosynthesis system	<i>ciuA</i>	VFG013721	DIP_RS14220	Y	Y	Y
	<i>ciuB</i>	VFG013723	DIP_RS14225	Y	Y	Y
	<i>ciuC</i>	VFG013725	DIP_RS14230	Y	Y	Y
	<i>ciuD</i>	VFG013727	DIP_RS14235	Y	Y	Y
	<i>ciuE</i>	VFG013729	DIP_RS14240	Y	Y	Y
Diphtheria toxin repressor DtxR	<i>dtxR</i>	VFG013754	DIP_RS18250	Y	Y	Y
Diphtheria toxin (DT)	<i>tox</i>	VFG002198	DIP_RS12515	-	-	-

*The NCBI accession number for NCTC13129 is NC002935.2.

†Core virulence factor profile for nontoxigenic *C. diphtheriae* ST8 circulating in Canada [13].

‡Seven WA isolates (PC0647, PC0648, PC0649, PC0650, PC0651, PC0652 and PC0655) share the same virulence factor profiles with PC0646.

conserved among all the WA isolates as well as nontoxigenic Canadian isolate CD10 (Table 3). Eight isolates (PC0646, PC0647, PC0648, PC0649, PC0650, PC0651, PC0652 and PC0655) shared an identical complement of virulence genes. Compared to VFDB reference NCTC13129, the WA isolate genomes encoded the same complement of iron and haem acquisition systems, but only one intact pilus locus (*spaABC*) was identified (Table 3). However, additional virulence genes (*srtB*, *srtC* and *spaD*) were detected in PC0653 and PC0654. These three genes were encoded within an 8.7 kb island that was absent from the complete genome assembly of PC0646. Compared to corresponding orthologues in NCTC13129, the relative sequence identity was 58% for SrtB, 63% SrtC and 41% for SpaD. This 8.7 kb region was observed in the middle of a long contig (>25 kb) that had a high coverage (>100×), indicating that it was not an assembly artefact, and exhibited very high similarity (99% identity) to *C. diphtheriae* isolate FRC0435 (ST411). Further BLASTP query of additional proteins encoded within the 8.7 kb region to the NCBI nr database revealed two possible matches to cell surface/membrane proteins found in *C. diphtheriae*, which had no homology to SpaE and SpaF (Table S6).

DISCUSSION

To better understand the molecular diversity of nontoxigenic *C. diphtheriae* cutaneous infections, we performed WGS with 10 *C. diphtheriae* case isolates recovered from patients in King County, WA, USA to study their phylogeny and virulence factor profiles. The phylogenetic reconstruction showed that all WA isolates were closely related, but quite different from other sequenced isolates of *C. diphtheriae*. This indicated that these cutaneous infections were caused by a single cluster of nontoxigenic *C. diphtheriae*, which had accumulated enough mutations to resolve linkages between cases with shared epidemiology. We observed very few SNPs among pairs of isolates when recovered from the same patient, including those from different body sites, suggesting both patients likely suffered from a single, persistent infection rather than repeated infections.

Although all WA isolates are phylogenetically related, the genomes of two isolates (PC0653 and PC0654) possessed an additional region encoding more virulence factors (SrtB, SrtC and SpaD). Such virulence factor content variation among isolates of the same ST has been reported previously [13]. SrtB, SrtC and SpaD belong to the SpaD-like pilus gene cluster (*srtB-spaD-srtC-spaE-spaF*) and are essential for sortase-mediated pilus assembly, which mediates bacterial attachment and colonization of host tissues [34]. Similar pilus gene clusters are often reported on horizontally acquired genomic islands, and the number and organization of SpaD-like pilus gene clusters vary [35]. The gene content and organization of the region detected in PC0653 and PC0654 differed from previously reported SpaD-like pilus gene clusters [35, 36], suggesting that it may encode either a novel SpaD-like pilus or nonfunctional relic. Varied pilus gene cluster content is common among *C. diphtheriae* isolates and while it appears

to alter macromolecular surface composition and influence adherence to specific host cell types, correlation with pathogenicity or invasion remains unclear [35–38]. The WA isolate genomes otherwise encoded the full complement of virulence genes in NCTC13129, specifically iron and haem acquisition systems essential for colonizing the low-iron host environment [39–42], corroborating their capacity for disease and concern of serious respiratory infection should they gain toxin-encoding bacteriophage. Regardless, these data illustrate that the *C. diphtheriae* genome is dynamic, varying in gene content even during transmission within a well-defined outbreak where pairwise SNP distance remains very low.

Previous reports from Canada, Poland and Germany have also investigated the circulation of nontoxigenic *C. diphtheriae* in urban impoverished populations, each involving a unique predominant ST primarily colonizing skin ulcerations [3, 6, 12, 13]. In addition, Lowe *et al.* observed that certain patients' skin ulcerations developed into severe non-respiratory diseases including bacteraemia and endocarditis [3]. Poor hygiene conditions, homelessness, drug use and alcoholism were also identified as risk factors. Similarly, 7/8 (90%) of the WA patients in this study experienced homelessness and most were documented drug users. Reduced susceptibility to first-line antimicrobials for treatment could further complicate persistent infections in this population, but a recent study suggests that breakpoints defining the intermediate penicillin resistance observed in isolates here may lack clinical importance [43]. As skin lesions may facilitate efficient transmission, particularly when coupled with poor hygiene, there is a need for increased awareness of the risk for persistent cutaneous infections of nontoxigenic *C. diphtheriae* in homeless populations.

Funding information

This work was supported by internal funds at the US Centers for Disease Control and Prevention.

Acknowledgements

We thank Amy Poel (WA State Department of Health) for assisting with gathering epidemiological information and Matthew Cole (IHRC, Inc.) for isolate data management. The findings and conclusions in this report are those of the authors, and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade names and commercial sources is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, or the US Department of Health and Human Services.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

Samples were obtained by Public Health Seattle and King County with patient consent and submitted to CDC for diagnostic characterization and patient information was anonymized for surveillance activities. This research was reviewed by a CDC Human Subject Contact and determined to not involve human subjects.

References

1. CDC. 2019. Diphtheria (*Corynebacterium diphtheriae*) 2019 case definition. <https://www.cdc.gov/nndss/conditions/diphtheria/case-definition/2019/>
2. Galazka A. The changing epidemiology of diphtheria in the vaccine era. *J Infect Dis* 2000;181 Suppl 1:S2–S9.

3. Lowe CF, Bernard KA, Romney MG. *Cutaneous diphtheria* in the urban poor population of Vancouver, British Columbia, Canada: a 10-year review. *J Clin Microbiol* 2011;49:2664–2666.
4. Romney MG, Roscoe DL, Bernard K, Lai S, Efstratiou A et al. Emergence of an invasive clone of nontoxigenic *Corynebacterium diphtheriae* in the urban poor population of Vancouver, Canada. *J Clin Microbiol* 2006;44:1625–1629.
5. Tiley SM, Kociuba KR, Heron LG, Munro R. Infective endocarditis due to nontoxigenic *Corynebacterium diphtheriae*: report of seven cases and review. *Clin Infect Dis* 1993;16:271–275.
6. Zasada AA. Nontoxigenic highly pathogenic clone of *Corynebacterium diphtheriae*, Poland, 2004–2012. *Emerg Infect Dis* 2013;19:1870–1872.
7. Meinel DM, Kuehl R, Zbinden R, Boskova V, Garzoni C et al. 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* 2016;22:1003.e1–101003.
8. FitzGerald RP, Rosser AJ, Perera DN. Non-toxigenic penicillin-resistant cutaneous *C. diphtheriae* infection: a case report and review of the literature. *J Infect Public Health* 2015;8:98–100.
9. pubMLST. 2010. *Corynebacterium diphtheriae* MLST database. <https://pubmlst.org/cdiphtheriae/>
10. du Plessis M, Wolter N, Allam M, de Gouveia L, Moosa F et al. Molecular characterization of *Corynebacterium diphtheriae* outbreak isolates, South Africa, March–June 2015. *Emerg Infect Dis* 2017;23:1308–1315.
11. Mohd Khalid MKN, Ahmad N, Hii SYF, Abd Wahab MA, Hashim R et al. Molecular characterization of *Corynebacterium diphtheriae* isolates in Malaysia between 1981 and 2016. *J Med Microbiol* 2019;68:105–110.
12. Dangel A, Berger A, Konrad R, Bischoff H, Sing A et al. Geographically diverse clusters of nontoxigenic *Corynebacterium diphtheriae* infection, Germany, 2016–2017. *Emerg Infect Dis* 2018;24:1239–1245.
13. Chorlton SD, Ritchie G, Lawson T, Romney MG, Lowe CF et al. Whole-Genome sequencing of *Corynebacterium diphtheriae* isolates recovered from an inner-city population demonstrates the predominance of a single molecular strain. *J Clin Microbiol* 2020;58
14. Freeman VJ. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol* 1951;61:675–688.
15. Engler KH, Glushkevich T, Mazurova IK, George RC, Efstratiou A et al. A modified Elek test for detection of toxigenic corynebacteria in the diagnostic laboratory. *J Clin Microbiol* 1997;35:495–498.
16. PacBio. 2014. Guidelines for using a salt:chloroform wash to clean up gDNA. <https://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Guidelines-for-Using-a-Salt-Chloroform-Wash-to-Clean-Up-gDNA.pdf>
17. Andrews S. 2010. FastQC: a quality control tool for high throughput sequencing data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
18. Martin M. 2011. Cutadapt removes adapter sequences from high throughput sequencing reads. <http://journal.embnet.org/index.php/embnetjournal/article/view/200>
19. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
20. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
21. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013;10:563–569.
22. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007;23:1026–1028.
23. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ et al. SRST2: rapid genomic surveillance for public health and hospital microbiology Labs. *Genome Med* 2014;6:90.
24. Madden T. 2002. The BLAST sequence analysis tool. <https://www.ncbi.nlm.nih.gov/books/NBK21097/>
25. Seemann T. 2015. snippy: fast bacterial variant calling from NGS reads. <https://github.com/tseemann/snippy>
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
27. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
28. Letunic I, Bork P. Interactive tree of life (iTOL) V4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
29. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 2012;6:80–92.
30. Seemann T. 2018. snp-dists: pairwise SNP distances from FASTA sequence alignments. <https://github.com/tseemann/snp-dists>
31. Liu B, Zheng D, Jin Q, Chen L, Yang J et al. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 2019;47:D687–D692.
32. Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 2005;33:W451–W454.
33. Hennart M, Panunzi LG, Rodrigues C, Gaday Q, Baines SL et al. Population genomics and antimicrobial resistance in *Corynebacterium diphtheriae*. *bioRxiv* 2020.
34. Gaspar AH, Ton-That H. Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. *J Bacteriol* 2006;188:1526–1533.
35. Trost E, Blom J, Soares SdeC, Huang I-H, Al-Dilaimi A et al. 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* 2012;194:3199–3215.
36. Timms VJ, Nguyen T, Crighton T, Yuen M, Sintchenko V et al. Genome-wide comparison of *Corynebacterium diphtheriae* isolates from Australia identifies differences in the Pan-genomes between respiratory and cutaneous strains. *BMC Genomics* 2018;19:869.
37. Ott L, Höller M, Rheinlaender J, Schäffer TE, Hensel M et al. Strain-specific differences in pili formation and the interaction of *Corynebacterium diphtheriae* with host cells. *BMC Microbiol* 2010;10:257.
38. Sangal V, Hoskisson PA. Evolution, epidemiology and diversity of *Corynebacterium diphtheriae*: New perspectives on an old foe. *Infection, Genetics and Evolution* 2016;43:364–370.
39. Billington SJ, Esmay PA, Songer JG, Jost BH. Identification and role in virulence of putative iron acquisition genes from *Corynebacterium pseudotuberculosis*. *FEMS Microbiol Lett* 2002;208:41–45.
40. Drazek ES, Hammack CA, Schmitt MP. *Corynebacterium diphtheriae* genes required for acquisition of iron from haemin and haemoglobin are homologous to ABC haemin transporters. *Mol Microbiol* 2000;36:68–84.
41. Kunkle CA, Schmitt MP. Analysis of a DtxR-regulated iron transport and siderophore biosynthesis gene cluster in *Corynebacterium diphtheriae*. *J Bacteriol* 2005;187:422–433.
42. Qian Y, Lee JH, Holmes RK. Identification of a DtxR-regulated operon that is essential for siderophore-dependent iron uptake in *Corynebacterium diphtheriae*. *J Bacteriol* 2002;184:4846–4856.
43. Zou J, Chorlton SD, Romney MG, Payne M, Lawson T et al. Phenotypic and genotypic correlates of penicillin susceptibility in nontoxigenic *Corynebacterium diphtheriae*, British Columbia, Canada, 2015–2018. *Emerg Infect Dis* 2020;26:97–103.