Limitations of Ribotyping as Genotyping Method for *Corynebacterium ulcerans*

Tsuyoshi Sekizuka, Chihiro Katsukawa,[1] Makoto Kuroda, Keigo Shibayama, Ken Otsuji, Mitsumasa Saito, Akihiko Yamamoto, Masaaki Iwaki

We conducted molecular typing of a *Corynebacterium ulcerans* isolate from a woman who died in Japan in 2016. Genomic DNA modification might have affected the isolate's ribotyping profile. Multilocus sequence typing results (sequence type 337) were more accurate. Whole-genome sequencing had greater ability to discriminate lineages at high resolution.

Corynebacterium ulcerans is a zoonotic pathogen that causes an illness categorized in World Health Organization documents as diphtheria (*1*). Genotyping methods such as ribotyping, multilocus sequence typing (MLST), and whole-genome sequencing are used to classify isolates. During the 1990s and early 2000s, the standard molecular typing method of *Corynebacterium diphtheriae* was conventional ribotyping (*2*,*3*). Ribotyping is also used to classify *C. ulcerans* (*4*) and compare isolates (*5*–*9*). Today, the standard method is MLST because of its objectivity and reproducibility (*8*,*10*). We sequenced 3 isolates of *C. ulcerans* from patients in Japan to analyze the accuracy of conventional ribotyping, MLST, and whole-genome sequencing.

The Study

In 2016, a 66-year-old woman in Fukuoka, Japan, died of a diphtheria-like disease. Otsuji et al. isolated toxigenic *C. ulcerans* from the patient's tracheal pseudomembrane and blood (*6*). We analyzed the isolate (FH2016-1) from the pseudomembrane alongside the first (*11*) and second (*5*) *C. ulcerans* isolates taken from patients in Japan; the first isolate (0102) was taken in 2001 and the second isolate (0211) in 2002.

Author affiliations: National Institute of Infectious Diseases, Tokyo, Japan (T. Sekizuka, M. Kuroda, K. Shibyama, A. Yamamoto, M. Iwaki); Osaka Prefectural Institute of Public Health, Osaka, Japan (C. Katsukawa); University of Occupational and Environmental Health, Kitakyushu, Japan (K. Otsuji, M. Saito)

DOI: https://doi.org/10.3201/eid2610.200086

Strains 0102 and 0211 (named for the initial isolates taken in 2001 and 2002) are the 2 major ribotypes of *C. ulcerans* in Japan. Our conventional ribotyping of the isolates found the pattern obtained from FH2016-1 was indistinguishable from that of 0102, indicating that FH2016-1 belongs to strain 0102 (Figure 1, panel A).

We also whole-genome sequenced strains FH2016-1 and 0211 using the NextSeq500 Illumina (for strain FH2016-1 [Illumina, https://www.illumina.com]), Illumina GAII (for strain 0211 [Illumina]), ABI 3730xl (Thermo Fisher, https://www.thermofisher.com), and PacBio Sequel (Pacific Biosciences of California, Inc., https://www.pacb.com) sequencers, followed by de novo assembly. We deposited complete sequences and assembly methods in GenBank under accession nos. AP019663 (strain FH2016-1) and AP019662 (strain 0211). Using these sequences and the previously published genome sequence (*12*) of strain 0102 (GenBank accession no. AP012284), we conducted in silico ribotyping of *Bst*EII-digested fragments that hybridized with OligoMix5 probes, producing a predicted pattern for each sequence (*13*). The predicted patterns of all 3 strains matched the conventional ribotype pattern of strain 0211. However, the conventional ribotyping patterns of strains FH2016-1 and 0102 did not match the in silico–predicted ribotype pattern (Figure 1, panel A).

The discrepancy between the conventional and in silico–predicted patterns is caused by impaired restriction digestion at specific *Bst*EII sites. In these strains, the conventional (modified) ribotype pattern differed from the in silico–predicted (unmodified) ribotype pattern by a shift of 4 fragments (Appendix Figure 1, panel A, https://wwwnc.cdc.gov/EID/article/26/10/20-0086-App1.pdf). For example, in silico typing predicted that 3 *Bst*EII sites would be digested at nt 770,000 of strain FH2016-1. PacBio modification analysis revealed that 1 of these sites might have been modified

[1]Current affiliation: Osaka Prefecture University, Osaka, Japan.

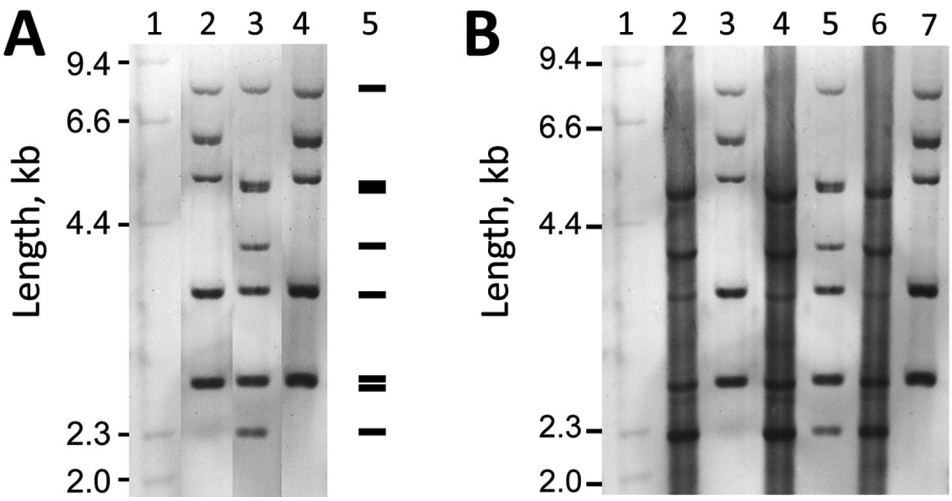


Figure 1. Alteration of ribotyping patterns by genomic DNA modification of *Corynebacterium ulcerans* strains 0102, 0211, and FH2016–1, Japan, 2001–2016. Ribotyping was performed as described previously (*4*,*11*). *Hin*dIII-digested, digoxigenin-labeled λ phage DNA segments were used as length markers. A) Conventional ribotyping patterns of strains 0102, 0211, and FH2016-1. 1, λ*Hin*dIII; 2, 0102; 3, 0211; 4, FH2016-1; 5, Pattern predicted by in silico typing. B) Ribotyping patterns of genomic DNA and whole-genome amplified DNA as substrates. 1, λ*Hin*dIII; 2, 0102 WGA; 3, 0102 native; 4, 0211 WGA; 5, 0211 native; 6, FH2016-1 WGA; 7, FH2016-1 native. The label "WGA" indicates whole-genome amplified DNA as a substrate; "native" indicates genomic DNA. WGA (unmodified) DNA of the 3 strains show identical patterns. The pattern matches that of native 0211 (unmodified genomic DNA). In contrast, native FH2016-1 and 0102 are modified and show different patterns from their WGA counterparts.

(Appendix Figure 1, panel B). *Bst*EII is sensitive to methylation and other types of DNA modification (*14*). Thus, the difference in restriction fragment patterns was closely related to the nucleotide modifications within *Bst*EII recognition sites (Appendix Figure 1, panel B). Other *Bst*EII sites also might have been modified, resulting in the 4-fragment shift. Accordingly, we did not observe this shift in ribotypes of unmodified DNA substrate prepared by whole-genome amplification of the 3 strains (*15*) (Figure 1, panel B). The patterns of unmodified DNA matched the pattern of strain 0211 (Figure 1, panel B) and the in silico–predicted pattern (Figure 1, panel A). The ≥6.1-kb bands seen in "native" lanes were not visible in whole-genome amplification lanes, potentially because of the failure of whole-genome amplification to generate large fragments. These results indicate that ribotyping patterns might be substantially affected by DNA modification.

The sequences of strains FH2016-1, 0102, and 0211 were highly homologous. For example, they shared complete sequence identity (data not shown) for a structural gene (locus tag CULCFH20161_03390) encoding a DNA methylase. However, we observed small differences in their genomes (Table, https://wwwnc.cdc.gov/eid/article/26/10/20-0086-t1; Figure 2; Appendix Table 1). We expected factors contributing to genomic DNA modification to be common between strains FH2016-1 and 0102, but not 0211. Scanning the genomes of the 3 strains for such factors resulted in 15 candidate open reading frames (ORFs) (Table). None of these ORFs contained motifs related to DNA methylation; however, these ORFs might still contribute to DNA modification of other gene products. The nature of the modification(s) remains unknown.

Conventional ribotyping (Figure 1, panel A) showed that strains FH2016-1 and 0102 were closely related. However, comparison of 30 genome sequences of strains from around the world (Appendix Table 2, Figure 2) revealed that all 3 strains from Japan belong to a single phylogenetic cluster and sequence type (ST) 337. Whether the 3 isolates represent the entire population of *C. ulcerans* in Japan is unclear. However, more than half the isolates we have analyzed (≈20) are ST337 (M. Iwaki and A. Yamamoto, unpub. data), suggesting a small amount of genetic diversity among the *C. ulcerans* population in Japan.

Close-up view of the phylogenetic tree showed that these strains from Japan divided into 2 different lineages. At most, 117 single nucleotide variations and 59 insertions/deletions existed between any 2 strains (Figure 2). Although this result indicated low variability

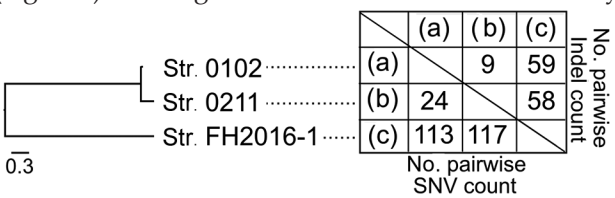


Figure 2. Genetic similarity among 3 selected strains of *Corynebacterium ulcerans*, Japan, 2001–2016. Strain 0102 is represented by (a), strain 0211 by (b), and strain FH2016–1 by (c). Numbers of SNVs and indels between strains are shown. A phylogenetic tree generated by SNV data are shown on the left. Indel, insertion/deletion; SNV, single-nucleotide variation.

among the 3 strains, it also showed that strain FH2016-1 was genetically distinct from 0102 and 0211 (Figure 2). Thus, the genome sequence analysis indicated that conventional ribotyping did not reflect lineage accurately and resulted in a misleading classification of these specimens. In contrast, MLST, which is now the preferred method of molecular typing (*8*,*10*), provided more accurate results. We queried the genomic sequences of the 3 strains on the PubMLST website (https://pubmlst.org) and analyzed them at 7 alleles (*atpA*, *dnaE*, *dnaK*, *fusA*, *leuA*, *odhA*, and *rpoB*). The same sequence type (ST337) was assigned to all 3 strains, reflecting the low genetic variability among these strains.

Conclusions

Our study of 3 strains of *C. ulcerans* showed that conventional ribotyping is less accurate than other methods of phylogenetic analysis. In comparison, MLST is less erroneous, and whole-genome sequencing produces results with greater resolution than those of conventional ribotyping. MLST produced results with lower resolution than whole-genome sequencing while maintaining a high level of accuracy. MLST and whole-genome sequencing improve the accuracy and efficiency of phylogenetic analysis of *C. ulcerans*.

Acknowledgments

We are grateful to Kaoru Umeda for her comments on the manuscript.

This research was supported by the Japan Agency for Medical Research and Development under grant nos. JP18fk0108017, JP19fk0108097, and JP19fk0108103.

About the Author

Dr. Sekizuka is chief at Laboratory of Bacterial Genomics, Pathogen Genomics Center, National Institute of Infectious Diseases (Shinjuku-ku, Tokyo). His primary research interests include genomic molecular epidemiology with genomics and bioinformatics, and pathogen identification with metagenomics.

References

1. World Health Organization. Diphtheria. In: Vaccine preventable diseases surveillance standards. Geneva: The Organization; 2018.
2. De Zoysa A, Hawkey P, Charlett A, Efstratiou A. Comparison of four molecular typing methods for characterization of *Corynebacterium diphtheriae* and determination of transcontinental spread of *C. diphtheriae* based on *Bst*EII rRNA gene profiles. J Clin Microbiol. 2008;46:3626–35. https://doi.org/10.1128/JCM.00300-08
3. Grimont PAD, Grimont F, Efstratiou A, De Zoysa A, Mazurova I, Ruckly C, et al.; European Laboratory Working Group on Diphtheria. International nomenclature for *Corynebacterium diphtheriae* ribotypes. Res Microbiol. 2004;155:162–6. https://doi.org/10.1016/j.resmic.2003.12.005
4. De Zoysa A, Hawkey PM, Engler K, George R, Mann G, Reilly W, et al. Characterization of toxigenic *Corynebacterium ulcerans* strains isolated from humans and domestic cats in the United Kingdom. J Clin Microbiol. 2005;43:4377–81. https://doi.org/10.1128/JCM.43.9.4377-4381.2005
5. Komiya T, Seto Y, De Zoysa A, Iwaki M, Hatanaka A, Tsunoda A, et al. Two Japanese *Corynebacterium ulcerans* isolates from the same hospital: ribotype, toxigenicity and serum antitoxin titre. J Med Microbiol. 2010;59:1497–504. https://doi.org/10.1099/jmm.0.022491-0
6. Otsuji K, Fukuda K, Endo T, Shimizu S, Harayama N, Ogawa M, et al. The first fatal case of *Corynebacterium ulcerans* infection in Japan. JMM Case Rep. 2017;4:e005106. https://doi.org/10.1099/jmmcr.0.005106
7. Yasuda I, Matsuyama H, Ishifuji T, Yamashita Y, Takaki M, Morimoto K, et al. Severe pneumonia caused by toxigenic *Corynebacterium ulcerans* infection, Japan. Emerg Infect Dis. 2018;24:588–91. https://doi.org/10.3201/eid2403.171837
8. Katsukawa C, Komiya T, Umeda K, Goto M, Yanai T, Takahashi M, et al. Toxigenic *Corynebacterium ulcerans* isolated from a hunting dog and its diphtheria toxin antibody titer. Microbiol Immunol. 2016;60:177–86. https://doi.org/10.1111/1348-0421.12364
9. Katsukawa C, Umeda K, Inamori I, Kosono Y, Tanigawa T, Komiya T, et al. Toxigenic *Corynebacterium ulcerans* isolated from a wild bird (ural owl) and its feed (shrew-moles): comparison of molecular types with human isolates. BMC Res Notes. 2016;9:181. https://doi.org/10.1186/s13104-016-1979-5
10. König C, Meinel DM, Margos G, Konrad R, Sing A. Multilocus sequence typing of *Corynebacterium ulcerans* provides evidence for zoonotic transmission and for increased prevalence of certain sequence types among toxigenic strains. J Clin Microbiol. 2014;52:4318–24. https://doi.org/10.1128/JCM.02291-14
11. Hatanaka A, Tsunoda A, Okamoto M, Ooe K, Nakamura A, Miyakoshi M, et al. *Corynebacterium ulcerans* diphtheria in Japan. Emerg Infect Dis. 2003;9:752–3. https://doi.org/10.3201/eid0906.020645
12. Sekizuka T, Yamamoto A, Komiya T, Kenri T, Takeuchi F, Shibayama K, et al. *Corynebacterium ulcerans* 0102 carries the gene encoding diphtheria toxin on a prophage different from the *C. diphtheriae* NCTC 13129 prophage. BMC Microbiol. 2012;12:72. https://doi.org/10.1186/1471-2180-12-72
13. Regnault B, Grimont F, Grimont PAD. Universal ribotyping method using a chemically labelled oligonucleotide probe mixture. Res Microbiol. 1997;148:649–59. https://doi.org/10.1016/S0923-2508(99)80064-3
14. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015;43(D1):D298–9. https://doi.org/10.1093/nar/gku1046
15. Nelson JR. Random-primed, Phi29 DNA polymerase-based whole genome amplification. Curr Protoc Mol Biol. 2014;105:15.13.1.

Address for correspondence: Masaaki Iwaki, Department of Bacteriology II and Management Department of Biosafety and Laboratory Animal, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashimurayama-shi, Tokyo 208-0011, Japan; email: miwaki@nih.go.jp

Article DOI: https://doi.org/10.3201/eid2610.200086

Limitations of Ribotyping as Genotyping Method for *Corynebacterium ulcerans*

Appendix

Appendix Table 1. Mutation sites in 3 strains of *Corynebacterium ulcerans*, Japan, 2001–2016*

Chromosome 0102	Position, nt	Mutation type	Strain		
			102	211	FH2016-1
AP012284.1	5468	SNV	T	T	C
AP012284.1	13074	SNV	G	G	A
AP012284.1	13243	Ins	T	TTC	T
AP012284.1	21848	SNV	G	G	A
AP012284.1	23960	SNV	C	C	T
AP012284.1	27867	Del	AGGAACTACCTCCGCGTACGCGGAGAAAAGCGTGTCTCCGGATCGAGGGAAGAACTTCTGCG	AGGAACTACCTCCGCGTACGCGGAGAAAAGCGTGTCTCCGGATCGAGGGAAGAACTTCTGCG	A
AP012284.1	28234	Del	GGGAACTACCTCCGCGTACGCGGAGAAAAGGCGTTCAGGTTTCCGCCGGCCTCAGCCAGCTT	GGGAACTACCTCCGCGTACGCGGAGAAAAGGCGTTCAGGTTTCCGCCGGCCTCAGCCAGCTT	G
AP012284.1	31398	SNV	G	G	A
AP012284.1	38660	SNV	G	G	A
AP012284.1	43805	SNV	T	T	C
AP012284.1	66246	SNV	C	C	T
AP012284.1	72966	SNV	G	G	A
AP012284.1	105558	Ins	A	A	AG
AP012284.1	109678	SNV	G	G	A
AP012284.1	147936	SNV	T	C	T
AP012284.1	148229	SNV	C	C	A
AP012284.1	157816	SNV	A	A	G
AP012284.1	219707	SNV	T	T	C
AP012284.1	270551	Del	AT	A	AT
AP012284.1	306538	Ins	C	C	CG
AP012284.1	309164	SNV	G	G	A
AP012284.1	347372	SNV	A	A	G
AP012284.1	352293	SNV	T	T	C
AP012284.1	356626	SNV	T	T	C
AP012284.1	364122	SNV	T	T	C
AP012284.1	389417	SNV	G	G	A
AP012284.1	399537	SNV	C	C	A
AP012284.1	399986	SNV	G	G	A
AP012284.1	408576	SNV	C	T	C

Chromosome 0102	Position, nt	Mutation type	Strain		
			102	211	FH2016-1
AP012284.1	448658	SNV	T	T	C
AP012284.1	471264	SNV	G	G	A
AP012284.1	471701	SNV	T	T	A
AP012284.1	484237	SNV	C	C	T
AP012284.1	502123	SNV	G	G	A
AP012284.1	520521	SNV	T	T	C
AP012284.1	528107	SNV	T	C	C
AP012284.1	531308	Ins	C	C	CCGGGTTAAAGCCGAGG
AP012284.1	531313	Ins	T	T	TTG
AP012284.1	531315	Ins	A	A	ACT
AP012284.1	531316	Ins	A	A	ACTCGTGTGGTCGCATACGGTGG
AP012284.1	531318	Ins	A	A	CCTCGCAGTGCGTGAGGTCATGG AGCCGCCAGCGAGGCTGGCAGC CCGGTTGTAG
AP012284.1	532752	SNV	T	T	C
AP012284.1	561350	Ins	A	AT	A
AP012284.1	562902	SNV	G	A	A
AP012284.1	577038	Ins	T	T	TG
AP012284.1	584561	SNV	T	T	C
AP012284.1	584754	Ins	G	G	GT
AP012284.1	599536	SNV	C	C	T
AP012284.1	601617	SNV	A	G	G
AP012284.1	604180	SNV	G	G	A
AP012284.1	611818	Del	CA	C	C
AP012284.1	619349	SNV	G	G	A
AP012284.1	626851	SNV	G	G	A
AP012284.1	634142	SNV	T	C	T
AP012284.1	648703	SNV	A	A	G
AP012284.1	692823	Del	AG	A	A
AP012284.1	734318	SNV	C	C	T
AP012284.1	739758	Ins	T	T	TTGTGCCAATAAGCGCGGTGGCG TTGGGGTAGTGATGTACGGTCTC GCGGA
AP012284.1	741335	SNV	A	A	C
AP012284.1	742304	Del	AG	AG	A
AP012284.1	770130	Ins	C	C	CGA
AP012284.1	770138	Ins	A	A	AG
AP012284.1	770142	Del	CTT	CTT	C
AP012284.1	779707	SNV	G	G	T
AP012284.1	787007	SNV	C	C	T
AP012284.1	803736	SNV	A	A	G
AP012284.1	820419	SNV	G	T	G
AP012284.1	859288	SNV	A	A	G
AP012284.1	891192	SNV	A	A	C
AP012284.1	976642	SNV	G	G	A
AP012284.1	981747	SNV	C	T	C
AP012284.1	987876	SNV	T	T	C
AP012284.1	989653	SNV	C	T	C

Chromosome 0102	Position, nt	Mutation type	Strain		
			102	211	FH2016-1
AP012284.1	991460	SNV	A	G	G
AP012284.1	1015997	Ins	C	C	CA
AP012284.1	1058136	SNV	G	A	G
AP012284.1	1072030	SNV	G	G	T
AP012284.1	1086565	SNV	G	G	A
AP012284.1	1094809	SNV	T	C	T
AP012284.1	1116335	SNV	C	T	T
AP012284.1	1140849	SNV	A	A	G
AP012284.1	1150120	SNV	G	G	T
AP012284.1	1153661	SNV	C	C	T
AP012284.1	1167108	Del	TG	TG	T
AP012284.1	1177278	Ins	A	A	AGG
AP012284.1	1187307	SNV	G	G	T
AP012284.1	1193209	SNV	G	G	A
AP012284.1	1202772	SNV	T	T	A
AP012284.1	1211380	SNV	G	G	A
AP012284.1	1221468	SNV	C	C	T
AP012284.1	1237855	SNV	T	T	C
AP012284.1	1239122	SNV	G	G	A
AP012284.1	1249696	SNV	C	C	T
AP012284.1	1256330	Ins	G	G	GT
AP012284.1	1265029	SNV	C	C	A
AP012284.1	1271793	SNV	A	A	G
AP012284.1	1275272	SNV	C	C	T
AP012284.1	1276489	SNV	T	T	A
AP012284.1	1359233	Del	ACTAACCAGAATTTTGACGGAGCCCTAAACTTCACCGGCTGCAGTACTGGGAGCTTGTCAGGAAAATTTGAGCAGCGCGAGGGTGGACTTTTCCACGATG	ACTAACCAGAATTTTGACGGAGCCCTAAACTTCACCGGCTGCAGTACTGGGAGCTTGTCAGGAAAATTTGAGCAGCGCGAGGGTGGACTTTTCCACGATG	A
AP012284.1	1359874	Del	CA	CA	C
AP012284.1	1360594	Ins	C	C	CTCATCGTCTACGACTACGCAGACCACAGAGCCAACGTCCTCATCGTCTACGACTACGCAGACCACGGAACCAACGTCGTCATCGTCTACGACTACGCAGACCACGGAACCAACGTCG
AP012284.1	1365000	Ins	T	T	TTTTTA
AP012284.1	1365003	Del	CTAAAG	CTAAAG	C
AP012284.1	1394742	Del	GCAC	GCAC	G
AP012284.1	1394745	Ins	C	C	CGT
AP012284.1	1394748	Del	TC	TC	T
AP012284.1	1394782	Del	CT	CT	C
AP012284.1	1462352	SNV	A	C	A
AP012284.1	1501754	SNV	G	A	A
AP012284.1	1511297	SNV	C	C	A
AP012284.1	1525192	SNV	G	G	A
AP012284.1	1527488	SNV	A	A	G
AP012284.1	1530546	Del	GC	G	G

Chromosome 0102	Position, nt	Mutation type	Strain		
			102	211	FH2016-1
AP012284.1	1549974	Del	CA	CA	C
AP012284.1	1615719	Ins	A	A	ACAGCAGGTTTTGGAGCGGCTGGCTTCGCGGAAGCCGGCTTGGGTGCTGCAGGCTTTGCGG
AP012284.1	1681351	SNV	A	A	G
AP012284.1	1683276	SNV	T	T	G
AP012284.1	1709596	SNV	G	T	G
AP012284.1	1725453	SNV	A	G	G
AP012284.1	1741037	Del	GA	GA	G
AP012284.1	1756915	SNV	G	G	A
AP012284.1	1768592	SNV	A	A	G
AP012284.1	1780351	SNV	C	C	T
AP012284.1	1787891	SNV	A	A	G
AP012284.1	1789178	SNV	T	T	A
AP012284.1	1817372	SNV	A	A	G
AP012284.1	1819250	SNV	C	C	T
AP012284.1	1835048	SNV	T	T	A
AP012284.1	1847653	Ins	T	T	TG
AP012284.1	1886000	SNV	C	T	C
AP012284.1	1886801	SNV	T	T	C
AP012284.1	1909561	Ins	C	C	CG
AP012284.1	1931545	SNV	C	C	T
AP012284.1	1936049	SNV	C	C	T
AP012284.1	1943886	SNV	G	G	A
AP012284.1	1963329	SNV	C	C	T
AP012284.1	1967703	SNV	T	T	C
AP012284.1	1977322	SNV	G	G	T
AP012284.1	2007494	SNV	T	C	C
AP012284.1	2025107	Del	GATA	GATA	G
AP012284.1	2034743	SNV	G	G	A
AP012284.1	2038449	Del	GGCGCGGCTTTTCTCCGCGTATGCGGAGGTAGTTCCCACCGCCGGGCAGCACTCCCTCAAAGCACAAGCTTTTCTCCGCGTATGCGGAGGTAGTTCCGGATCGGGCGCGGGCTGCGATTTCGA	GGCGCGGCTTTTCTCCGCGTATGCGGAGGTAGTTCCCACCGCCGGGCAGCACTCCCTCAAAGCACAAGCTTTTCTCCGCGTATGCGGAGGTAGTTCCGGATCGGGCGCGGGCTGCGATTTCGA	G
AP012284.1	2038758	Del	GGCCTTTTCTCCGCGTATGCGGAGGTAGTTCCCATCATTGGGGAATCCCCCCGAGCATGGGTGACTTTTCTCCGCGTATGCGGAGGTAGTTCCCATCCGGCGTTAGTGACCGCTGCGGTCGCA	G	GGCCTTTTCTCCGCGTATGCGGAGGTAGTTCCCATCATTGGGGAATCCCCCCGAGCATGGGTGACTTTTCTCCGCGTATGCGGAGGTAGTTCCCATCCGGCGTTAGTGACCGCTGCGGTCGCA
AP012284.1	2039309	Del	GGCTTTTCTCCGCGTATGCGGAGGTAGTTCCGCAAATCTTCAAGAACATCTTTGTCCAAGGTCCTTTTCTCCGCGTATGCGGAGGTAGTTCTGAGCTGGGTTTTGTGGCGTCTCTTTCTCGTC	GGCTTTTCTCCGCGTATGCGGAGGTAGTTCCGCAAATCTTCAAGAACATCTTTGTCCAAGGTCCTTTTCTCCGCGTATGCGGAGGTAGTTCTGAGCTGGGTTTTGTGGCGTCTCTTTCTCGTC	G
AP012284.1	2040619	Del	TCCAC	TCCAC	T
AP012284.1	2040630	Del	CG	CG	C

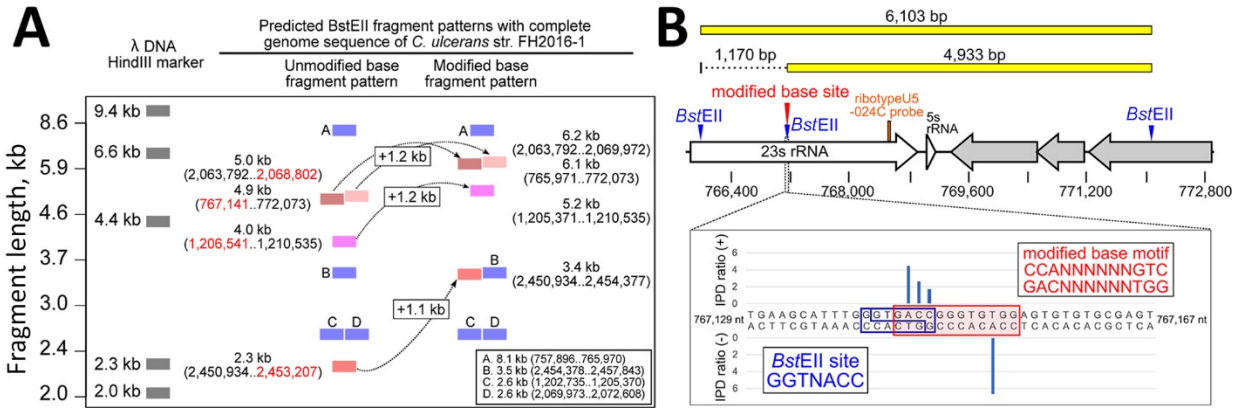
Chromosome 0102	Position, nt	Mutation type	Strain		
			102	211	FH2016-1
AP012284.1	2040634	Del	GGC	GGC	G
AP012284.1	2040639	Ins	A	A	AGCTGGTC
AP012284.1	2040645	Del	GCCCC	GCCCC	G
AP012284.1	2040650	Ins	G	G	GA
AP012284.1	2040653	Ins	G	G	GGAA
AP012284.1	2054399	Ins	T	T	TA
AP012284.1	2055051	Ins	A	A	AC
AP012284.1	2057376	SNV	G	G	T
AP012284.1	2059553	SNV	C	C	T
AP012284.1	2066689	Del	GAA	GAA	G
AP012284.1	2066698	Ins	T	T	TC
AP012284.1	2066705	Ins	C	C	CCT
AP012284.1	2084128	SNV	G	G	A
AP012284.1	2104276	SNV	G	A	G
AP012284.1	2126945	SNV	C	T	C
AP012284.1	2131950	SNV	G	G	A
AP012284.1	2147458	SNV	G	G	A
AP012284.1	2157633	SNV	G	G	T
AP012284.1	2159106	SNV	G	G	A
AP012284.1	2247082	Del	CA	CA	C
AP012284.1	2258634	SNV	C	C	T
AP012284.1	2261240	SNV	T	T	C
AP012284.1	2276928	Del	TC	T	T
AP012284.1	2276937	Del	AC	AC	A
AP012284.1	2283843	Ins	A	A	AG
AP012284.1	2284180	SNV	T	T	G
AP012284.1	2284351	SNV	G	G	A
AP012284.1	2295862	SNV	C	C	T
AP012284.1	2296083	SNV	T	T	C
AP012284.1	2299003	SNV	C	C	T
AP012284.1	2324071	SNV	T	T	C
AP012284.1	2395978	SNV	G	G	A
AP012284.1	2400888	SNV	A	C	C
AP012284.1	2401441	SNV	T	C	T
AP012284.1	2416052	SNV	C	C	T
AP012284.1	2417504	SNV	G	G	T
AP012284.1	2421828	Ins	C	C	CT
AP012284.1	2450685	SNV	C	C	G
AP012284.1	2468623	Ins	A	ACAGCGTGTCAG	ACAGCGTGTCAG
AP012284.1	2487864	SNV	C	T	T
AP012284.1	2496366	Ins	C	C	CA
AP012284.1	2515718	SNV	A	A	G
AP012284.1	2549156	SNV	T	T	G
AP012284.1	2560560	SNV	T	T	C

*Del, deletion; Ins, insertion; SNV, single-nucleotide variation. Dashes indicate not applicable.

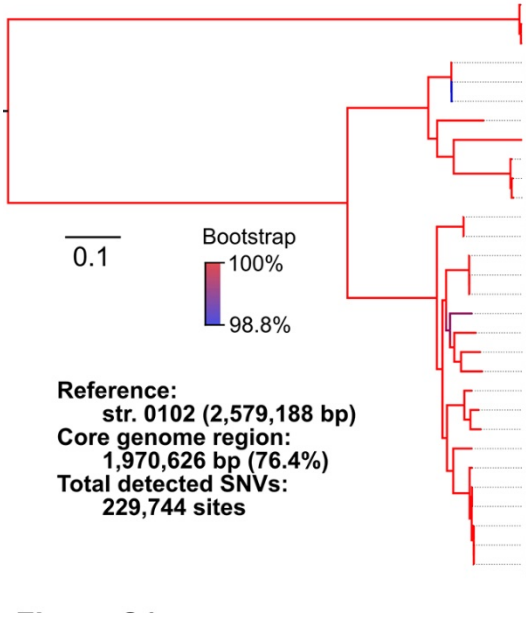
Appendix Table 2. Strain information used in core genome phylogenetic analysis of *Corynebacterium ulcerans**

Strain	Country	Sequence type	tox (%)¶	BioSample ID	Assembly ID	Sequence completion status
PO100/5	Portugal	New ST	94.77	SAMN07109093	GCA_002162115.1	Complete
KL1196	Germany	ST578	94.66	SAMN10477295	GCA_004382105.1	Contig
W25	Germany	ST578	94.66	SAMN12027930	GCA_006370535.1	Contig
FRC11	France	ST335	–	SAMN03099770	GCA_000767685.1	Complete
NCTC8666	NA	ST335	95.19	SAMEA3923577	GCA_900447245.1	Contig
NCTC12077	NA	ST335	–	SAMN02470561	GCA_000498915.1	Contig
03-8664	France	ST341	99.23	SAMN03938368	GCA_001298285.2	Contig
131002	Brazil	New ST	–	SAMN03418694	GCA_000968945.1	Complete
04-7514	France	ST344	99.7	SAMN04041014	GCA_001302345.1	Contig
LSPQ-04227	Canada	ST344	98.57	SAMN03387456	GCA_000988285.1	Contig
LSPQ-04228	Canada	ST344	98.57	SAMN03387457	GCA_000988295.1	Contig
210932	Brazil	ST329	99.47	SAMN03025288	GCA_000767415.1	Complete
131001	Brazil	ST329	99.47	SAMN03316967	GCA_001281445.1	Complete
NCTC7908	NA	ST552	–	SAMEA3893446	GCA_900475635.1	Complete
NCTC8639	UK	ST345	–	SAMEA4017706	GCA_900475775.1	Complete
NCTC7910	UK	ST345	–	SAMEA4504038	GCA_900187135.1	Complete
2590	Brazil	New ST	–	SAMN06029387	GCA_002866525.1	Contig
4940	Belarus	ST349	–	SAMN04497012	GCA_002872395.1	Contig
809	Brazil	ST338	–	SAMN02603079	GCA_000215645.1	Complete
BR-AD 2649	Brazil	New ST	–	SAMN06029388	GCA_002866485.1	Contig
FRC58	France	ST328	99.7	SAMN02471303	GCA_000499805.2	Complete
210931	Brazil	ST332	99.7	SAMN03025287	GCA_000767645.1	Complete
05146	Brazil	ST331	99.64	SAMN03104569	GCA_000769635.1	Complete
BR-AD22	Brazil	ST339	–	SAMN02603080	GCA_000215665.1	Complete
MRi49	UK	ST543	95.19	SAMN13521090	GCA_009789155.1	Complete
04-3911	France	New ST	100.00	SAMN03938367	GCA_001298275.1	Contig
KZN-2016-48390	South Africa	ST325	100.00	SAMN05729438	GCA_001876325.1	Contig
FH2016-1	Japan	ST337	100.00	SAMD00168434	GCA_009002285.1	Complete
0102	Japan	ST337	100.00	SAMD00012381	GCA_000306825.1	Complete
0211	Japan	ST337	100.00	SAMD00168433	GCA_008995395.1	Complete

*NA, not available; ST, sequence type. Dashes indicate strains without the toxin gene.
¶ nucleotide identity % against diphtheria-like toxin of str. 0102.



Appendix Figure 1. Alteration of ribotyping patterns by genomic DNA modification of *Corynebacterium ulcerans* strains 0102, 0211, and FH2016–1, Japan, 2001–2016. A) Fragments affected by DNA modification indicated by the pink, brown, magenta, and orange horizontal bars. Unaffected fragments (A, B, C, and D) indicated by blue horizontal bars; their sizes are displayed in the square at the lower right corner of the panel. B) Inhibition of *Bst*EII digestion. The genomic region around nt 770,000 is shown. The inset indicates the *Bst*EII site affected by genomic DNA modification that contributes to the shift in fragment size and mobility. The staggered blue frame represents the *Bst*EII site. The base motif (red rectangle) and target nucleotides for modification (blue bars), were identified by the PacBio modification analysis (Pacific Biosciences of California, Inc., https://www.pacb.com) using a motif finding algorithm (*1*). Longer bars represent higher relative probability of modification.



Strain	Country	ST	*tox* (%)¶	BioSample ID	Assembly ID
PO100/5	Portugal	New ST	94.77	SAMN07109093	GCA_002162115.1
KL1196	Germany	ST578	94.66	SAMN10477295	GCA_004382105.1
W25	Germany	ST578	94.66	SAMN12027930	GCA_006370535.1
FRC11	France	ST335	N.D.	SAMN03099770	GCA_000767685.1
NCTC8666	N/A	ST335	95.19	SAMEA3923577	GCA_900447245.1
NCTC12077	N/A	ST335	N.D.	SAMN02470561	GCA_000498915.1
03-8664	France	ST341	99.23	SAMN03938368	GCA_001298285.2
131002	Brazil	New ST	N.D.	SAMN03418694	GCA_000968945.1
04-7514	France	ST344	99.7	SAMN04041014	GCA_001302345.1
LSPQ-04227	Canada	ST344	98.57	SAMN03387456	GCA_000988285.1
LSPQ-04228	Canada	ST344	98.57	SAMN03387457	GCA_000988295.1
210932	Brazil	ST329	99.47	SAMN03025288	GCA_000767415.1
131001	Brazil	ST329	99.47	SAMN03316967	GCA_001281445.1
NCTC7908	N/A	ST552	N.D.	SAMEA3893446	GCA_900475635.1
NCTC8639	UK	ST345	N.D.	SAMEA4017706	GCA_900475775.1
NCTC7910	UK	ST345	N.D.	SAMEA4504038	GCA_900187135.1
2590	Brazil	New ST	N.D.	SAMN06029387	GCA_002866525.1
4940	Belarus	ST349	N.D.	SAMN04497012	GCA_002872395.1
809	Brazil	ST338	N.D.	SAMN02603079	GCA_000215645.1
BR-AD 2649	Brazil	New ST	N.D.	SAMN06029388	GCA_002866485.1
FRC58	France	ST328	99.7	SAMN02471303	GCA_000499805.2
210931	Brazil	ST332	99.7	SAMN03025287	GCA_000767645.1
05146	Brazil	ST331	99.64	SAMN03104569	GCA_000769635.1
BR-AD22	Brazil	ST339	N.D.	SAMN02603080	GCA_000215665.1
MRi49	UK	ST543	95.19	SAMN13521090	GCA_009789155.1
04-3911	France	New ST	100	SAMN03938367	GCA_001298275.1
KZN-2016-48390	South Africa	ST325	100	SAMN05729438	GCA_001876325.1
FH2016-1	Japan	ST337	100	SAMD00168434	GCA_009002285.1
0102	Japan	ST337	100	SAMD00012381	GCA_000306825.1
0211	Japan	ST337	100	SAMD00168433	GCA_008995395.1

¶ nucleotide identity % against *tox* of str. 0102

Appendix Figure 2. Phylogenetic analysis of 30 strains of *Corynebacterium ulcerans*, including 3 sequences (FH2016-1, 0102 and 0211) from Japan. Phylogenetic tree constructed with maximum-likelihood analysis. ¶nucleotide identity % against diphtheria-like toxin of str. 0102. NA, not available; ND, no data (indicates strains without toxin gene); SNV, single-nucleotide variation; ST, sequence type; str, strain.

Reference

1. Kelleher P, Murphy J, Mahony J, van Sinderen D. Identification of DNA base modifications by means of Pacific Biosciences RS sequencing technology. Methods Mol Biol. 2018;1681:127–37. PubMed https://doi.org/10.1007/978-1-4939-7343-9_10