

# **A *Vibrio cholerae* Core Genome Multilocus Sequence Typing Scheme to Facilitate the Epidemiological Study of Cholera**

## **SUPPLEMENTAL MATERIAL**

Kevin Y. H. Liang,<sup>a</sup> Fabini D. Orata,<sup>a</sup> Mohammad Tarequl Islam,<sup>a</sup> Tania Nasreen,<sup>a</sup>  
Munirul Alam,<sup>b</sup> Cheryl L. Tarr,<sup>c</sup> Yann F. Boucher<sup>a#</sup>

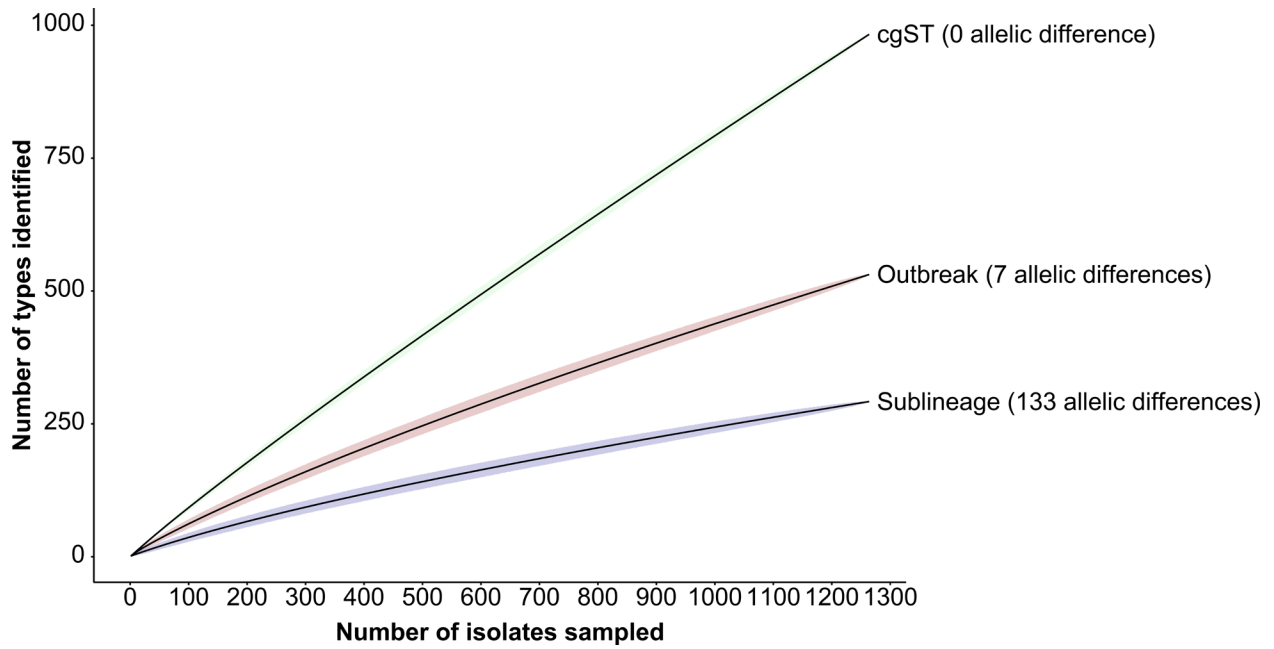
<sup>a</sup>Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

<sup>b</sup>Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

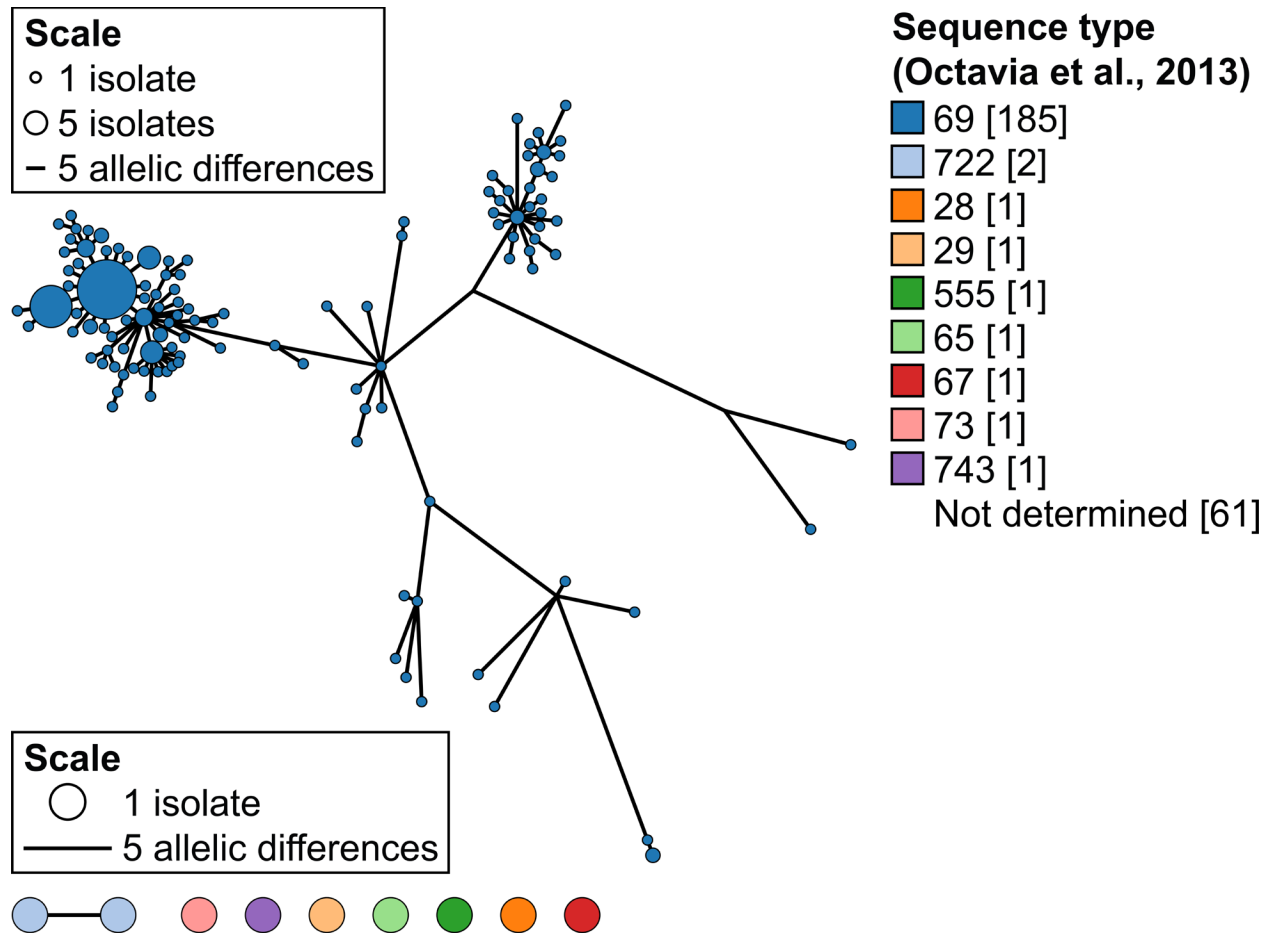
<sup>c</sup>Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

#Address correspondence to Yann F. Boucher, [yboucher@ualberta.ca](mailto:yboucher@ualberta.ca)

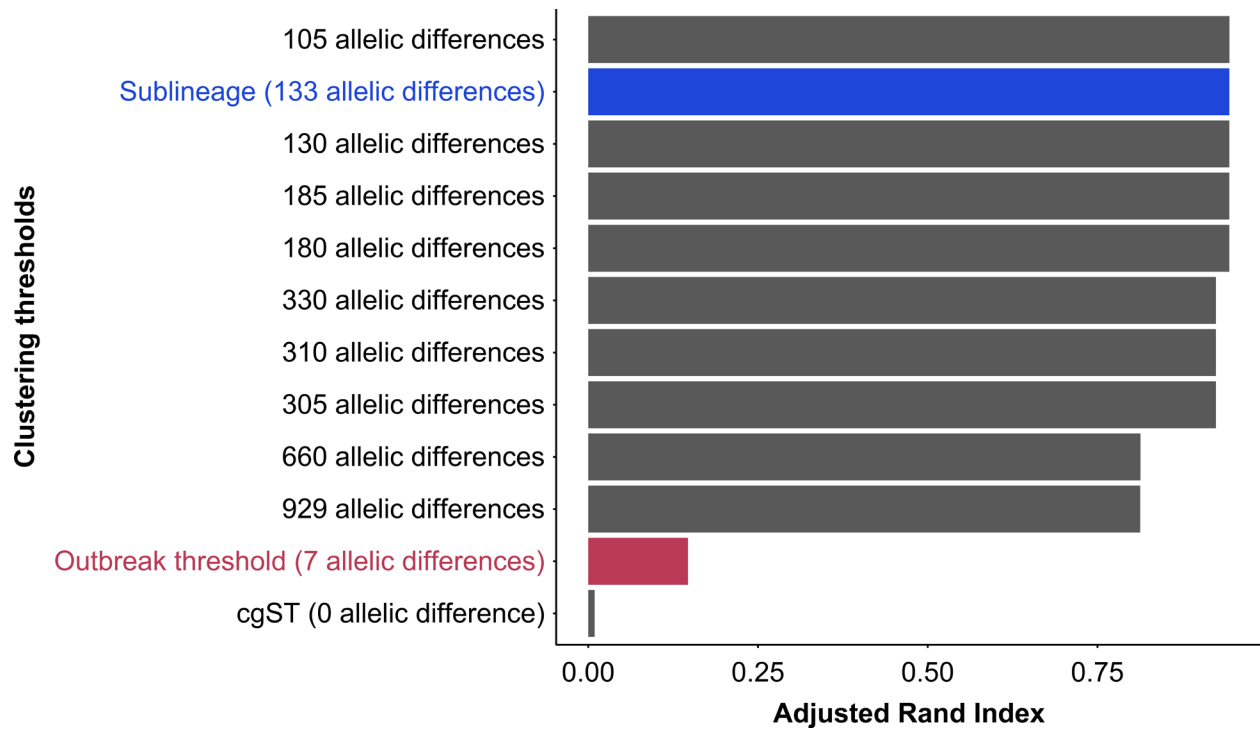
## SUPPLEMENTAL FIGURES



**FIG S1** Rarefaction curves computed using mothur with default parameters (1) for cgST (i.e., 0 allelic difference), the outbreak threshold (i.e., 7 allelic differences), and the sublineage threshold (i.e., 133 allelic differences).



**FIG S2** Evaluation of network similarities between the cgMLST sublineage threshold and MLST ST. Networks of all sublineages identified using only *V. cholerae* isolates from Bangladesh ( $n = 255$ ). Each cluster represents a sublineage and includes isolates with less than or equal to 133 allelic differences with each other. Each node represents a cgST and is colored by ST based on the 2013 MLST scheme (2). Size of the nodes are proportional to the number of isolates. The length of the connecting lines within a cluster is proportional to the number of allelic differences.



**FIG S3** Adjusted Rand Index for individual pairwise comparisons between predefined clustering thresholds (FIG 2) and the 2013 MLST scheme (2). The sublineage clustering threshold (i.e., 133 allelic differences) and outbreak threshold (i.e., 7 allelic differences) are indicated in blue and red bars, respectively.

**SUPPLEMENTAL TABLES\***

**TABLE S1** Meta-information for all 1,262 isolates used in this study. ST designations are based on the 2016 MLST scheme by Kirchberger and colleagues (3) and the 2013 MLST scheme by Octavia and colleagues (2).

**TABLE S2** Allelic profiles for the cgMLST scheme (as defined in this study), the 2016 MLST scheme by Kirchberger and colleagues (3), and the 2013 MLST scheme by Octavia and colleagues (2). The cgSTs and their corresponding STs and PubMLST IDs are indicated. All missing genes are indicated as NA. The most likely cgSTs are indicated in parentheses where applicable.

**TABLE S3** *V. cholerae* isolates from the Yemen cholera outbreak and neighbouring countries, as well as other isolates from different lineages.

**TABLE S4** Genome completeness for the cgMLST scheme (using 2,443 core genes). All genomes with less than 90% completeness were subsequently removed.

**TABLE S5** Genome completeness information for the final set of 679 genomes. Completeness for the cgMLST scheme is represented as the percentage of the of the 2,443 core genes present in each genome.

**TABLE S6** All NCBI accession numbers for isolates, PubMLST IDs, and PubMLST links to online storage of genomes.

*\*See separate Excel sheets for complete supplemental tables.*

## REFERENCES

1. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–41. <https://doi.org/10.1128/AEM.01541-09>.
2. Octavia S, Salim A, Kurniawan J, Lam C, Leung Q, Ahsan S, Reeves PR, Nair GB, Lan R. 2013. Population structure and evolution of non-O1/non-O139 *Vibrio cholerae* by multilocus sequence typing. *PLoS One* 8:e65342. <https://doi.org/10.1371/journal.pone.0065342>.
3. Kirchberger PC, Orata FD, Barlow EJ, Kauffman KM, Case RJ, Polz MF, Boucher Y. 2016. A small number of phylogenetically distinct clonal complexes dominate a coastal *Vibrio cholerae* population. *Appl Environ Microbiol* 82:5576–5586. <https://doi.org/10.1128/AEM.01177-16>.
4. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon YS, Kim DW, Lee JH, Brettin TS, Bruce DC, Challacombe JF, Detter JC, Han CS, Munk AC, Chertkov O, Meincke L, Saunders E, Walters RA, Huq A, Nair GB, Colwell RR. 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 106:15442–15447. <https://doi.org/10.1073/pnas.0907787106>.
5. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406:477–483. <https://doi.org/10.1038/35020000>.
6. Katz LS, Petkau A, Beaulaurier J, Tyler S, Antonova ES, Turnsek MA, Guo Y, Wang S, Paxinos EE, Orata F, Gladney LM, Stroika S, Folster JP, Rowe L, Freeman MM, Knox N, Frace M, Boncy J, Graham M, Hammer BK, Boucher Y, Bashir A, Hanage WP, Van Domselaar G, Tarr CL. 2013. Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. *MBio* 4:e00398–13. <https://doi.org/10.1128/mBio.00398-13>.
7. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M, Niyogi SK, Kim EJ, Ramamurthy T, Chun J, Wood JL, Clemens JD, Czerkinsky C, Nair GB, Holmgren J, Parkhill J, Dougan G. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 477:462–465. <https://doi.org/10.1038/nature10392>.

8. Weill FX, Domman D, Njamkepo E, Almesbahi AA, Naji M, Nasher SS, Rakesh A, Assiri AM, Sharma NC, Kariuki S, Pourshafie MR, Rauzier J, Abubakar A, Carter JY, Wamala JF, Seguin C, Bouchier C, Malliavin T, Bakhshi B, Abulmaali HHN, Kumar D, Njoroge SM, Malik MR, Kiiru J, Luquero FJ, Azman AS, Ramamurthy T, Thomson NR, Quilici ML. 2019. Genomic insights into the 2016-2017 cholera epidemic in Yemen. *Nature* 565:230–233. <https://doi.org/10.1038/s41586-018-0818-3>.