Appendix

Figures and Tables

Appendix Table 1. *V. cholerae* isolates selected for sequencing.

Appendix Table 2. Sequencing metrics for *V. cholerae* O1 genomes.

NR, not recorded. Percent genome, fraction of the N16961 reference genome covered by at least 100x read depth.

Appendix Table 3. Quinolone susceptibility and mutations in the quinolone-resistance determining regions of DNA gyrase and topoisomerase IV in *Vibrio cholerae* O1.

S: Serine; I: Isoleucine; L: Leucine; A: Alanine; E: Glutamate; --: Not tested; ND: Not Determined. No mutations in *parE* genes were observed compared to the reference peptide sequence (susceptible). The two non-O1 genomes (NGA_148_2019 and NGA_252_2019) were not included in this analysis.

Appendix Table 4. Phenotypic antibiotic resistance profile versus genotypic profile.

Pr: Gene present; Abs: Gene absent; R: Resistant isolate; S: Susceptible isolate; I: Intermediate; --: Not tested; AMC: amoxicillin-clavulanic acid; AMP: ampicillin; NAL: nalidixic acid; CHL: chloramphenicol; PMB: polymyxin B; CST: colistin. All samples had *tet(34)* present and were susceptible to tetracycline; all samples had *farA* present and were not tested for antibacterial free fatty acids. The following samples were not tested and thus not included in this analysis: NGA_236_2019, NGA_263_2018, NGA_148_2019 and NGA_252_2019.

Data availability

Raw data for all sequenced isolates is available under NCBI BioProject accession: PRJNA616029.

Sample collection and cholera confirmation

V. cholerae isolates were collected from clinically confirmed cholera cases in Cameroon, Niger and Nigeria between 2011 and 2019 (**Appendix Table 1**). In all three countries, suspected cholera stools were cultured on TCBS medium. Phenotypic identification of *V. cholerae* colonies was based on morphology, motility, and biochemical characteristics (positive oxidase, saccharose, indole and gelatinase). *V. cholerae* isolates were confirmed with agglutination tests with anti-O1 or anti-O139 serum (WHO antisera & Denka Seiken Agglutinating Sera Vibrio cholerae Antisera Set).

DNA extraction and quantification

DNA extraction was performed at Centre Pasteur du Cameroun (Yaounde, Cameroun), Centre de Recherche Médicale et Sanitaire (Niamey, Niger), and at National Reference Laboratory (NRL; Nigeria Centre for Disease Control, Abuja, Nigeria). Confirmed *V. cholerae* isolates were cultured on selective TCBS, MH and/or HCK agar plates and incubated at 37°C overnight. In all countries, *V. cholerae* DNA was extracted following the standard protocol from the Qiagen QIAamp DNA Mini Kit with a final elution volume of 200µL. Extracted DNA from *V. cholerae* isolates were stored in 1.5 mL Eppendorf tubes at 4^oC or at room temperature until quantification at NRL. DNA concentrations were measured with the Qubit Fluorometer 4.0 (Thermo Fisher) using the dsDNA HS assay standard protocol. Input concentrations for each isolate are recorded in **Appendix Table 2**.

Oxford Nanopore library construction and sequencing

Library preparation and sequencing was performed at the NRL or at Johns Hopkins University (Baltimore, Maryland, United States). In both locations, quantified, extracted DNA was diluted to 1000ng in 48µL input material, as required by the SQK-LSK109 library preparation kit from Oxford Nanopore. For samples with a concentration below 20ng/µL, 48µL of extracted DNA was used, regardless of the final amount (<1000ng).

DNA repair and end-prep was performed with 2µL NEBNext FFPE DNA Repair Mix, 3.5µL NEBNext FFPE DNA Repair Buffer, 3µL Ultra II End-Prep Reaction Mix, and 3.5µL Ultra II End-Prep Enzyme Buffer (New England Biolabs). The reaction was incubated at 20°C for 20 minutes and then 65°C for 10 minutes, and then purified with 1x AMPure XP beads (Beckman Coulter) and eluted in 25µL Elution Buffer (Promega). A unique 2.5µL Native Barcode (Oxford Nanopore EXP-NBD104) and 25µL Blunt/TA Ligase Master Mix (New England Biolabs) were added to dA-tailed DNA and the reaction mixture was incubated for 30 minutes at room temperature. As before, the reaction was purified with 1x AMPure XP beads. The purified DNA was eluted in 10µL Elution Buffer. Equimolar amounts of 6-9 isolates were pooled to an approximate total DNA amount of 1000ng, and the pool was diluted to 65µL in Elution Buffer. The barcoded pool was added to 5µL Adapter Mix II (Oxford Nanopore EXP-NBD104), 20µL NEBNext Quick Ligation Reaction Buffer (New England Biolabs), and 10µL Quick T4 DNA Ligase (New England Biolabs). The reaction was incubated at room temperature for 30 minutes and purified with 0.5x AMPure XP beads, using Long Fragment Buffer (Oxford Nanopore SQK-LSK109) instead of ethanol in the wash steps. The final library was eluted in 15µL Elution Buffer and quantified using the Qubit Fluorometer 4.0 to confirm the process had been successful.

For sequencing, 190ng pooled final library was diluted to 12µL in Elution Buffer and combined with 37.5µL Sequencing Buffer and 25.5µL Loading Beads (Oxford Nanopore SQK-LSK109) before loading onto a primed R9.4.1 Oxford Nanopore MinION flowcell. For each sequencing run, the MinION was run for 48 hours using an offline-capable version of MinKNOW made available by Oxford Nanopore Technologies specifically for this project.

All sequencing files were basecalled using Guppy version 3.0.3 with the flip-flop model (dna_r9.4.1_450bps_fast.cfg). Adapters were removed from reads and reads were demultiplexed into individual samples with Porechop (16). Filtlong (https://github.com/rrwick/Filtlong) was used with the following options to reduce file size and filter out low quality sequencing reads: '--keep_percent 90 - target bases 800000000.' The resulting filtered FASTQ files were used as input for all subsequent analyses, including assembly.

Reference-based genome assembly

Reference-based assembly was performed with nanopolish version 0.11.1 (17) using the seventh pandemic strain N16961 as a reference (accession: AE003852/AE003853). Briefly, reads were indexed with nanopolish, then mapped to the reference with minimap2 version 2.17 (18), then sorted and indexed with samtools version 1.10 (19). Variants to the reference genome were then called with nanopolish, and bcftools version 1.9 (20) was used to filter out variants with less than 75% support and sites with less than 100x coverage. A custom Python3 script was used to create consensus genomes from the VCF files, keeping only sites with at least 100x coverage. The full reference-based genome assembly pipeline for *V. cholerae* is publicly available at: https://github.com/HopkinsIDD/minion-vc.

Maximum likelihood estimation

A maximum likelihood tree was generated on a multiple sequence alignment of the genomes from Cameroon, Niger, and Nigeria as well as 1280 previously-published *V. cholerae* whole genome sequences (**Appendix Table 5**) (1,2,11). Recombinant sites were masked as previously described (1) (see https://figshare.com/s/d6c1c6f02eac0c9c871e for masking sites) and using Gubbins version 2.3.4 (21). The maximum likelihood tree was generated on the resulting masked SNP alignment using IQ-TREE version 1.6.10 (22) with a GTR substitution model and 1000 bootstrap iterations. The tree was rooted on A6 (accession: ERR025382). Trees were visualized in FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Vibrio cholerae typing

Reads from each isolate were mapped to reference sequences *ctxA* (accession: AF463401), *wbeO1* (accession: KC152957), and *wbfO139* (accession: AB012956), as described in Greig et al (13). Reads were mapped using minimap2 version 2.17 (18) and then sorted with samtools version 1.10 (19). The number of reads mapping at each reference gene site was counted using samtools.

Antibiotic susceptibility assays

Antibiotic susceptibility testing of *V. cholerae* isolates were done by disc diffusion on MH agar plates at Centre Pasteur du Cameroun, Centre de Recherche Médicale et Sanitaire, and NRL. The following antibiotics were tested for resistance: amoxicillin/clavulanic acid, nalidixic acid, polymyxin, tetracycline, ciprofloxacin, and chloramphenicol. The interpretation of resistance/susceptibility profiles was done according to the CASFM 2007 guidelines for Enterobacteriaceae.

Antibiotic resistance gene detection

Genes associated with antibiotic resistance were detected in two ways: first, by using the software abricate (https://github.com/tseemann/abricate) with the NCBI (23) and CARD (24) databases. And second, by translating the nucleotide sequences on the EMBOSS (25) website and aligning the resulting translations to target genes downloaded from https://pubmlst.org/vcholerae/ and https://www.ncbi.nlm.nih.gov/ with the CLUSTALW (26) website. Resistance to quinolones was determined by comparing the assemblies to reference strains, as described in previous studies (27–29).

References Cited in Appendix

16. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017 Oct;3(10):e000132.

17. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015 Aug;12(8):733–5.

18. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100.

19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

20. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov 1;27(21):2987–93.

21. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015 Feb 18;43(3):e15.

22. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.

23. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. Antimicrob Agents Chemother [Internet]. 2019 Nov;63(11). Available from: http://dx.doi.org/10.1128/AAC.00483-19

24. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and modelcentric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017 Jan 4;45(D1):D566– 73.

25. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000 Jun;16(6):276–7.

26. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994 Nov 11;22(22):4673–80.

27. Chien J-Y, Chiu W-Y, Chien S-T, Chiang C-J, Yu C-J, Hsueh P-R. Mutations in gyrA and gyrB among Fluoroquinolone- and Multidrug-Resistant Mycobacterium tuberculosis Isolates. Antimicrob Agents Chemother. 2016 Apr;60(4):2090–6.

28. Zhou Y, Yu L, Li J, Zhang L, Tong Y, Kan B. Accumulation of mutations in DNA gyrase and topoisomerase IV genes contributes to fluoroquinolone resistance in Vibrio cholerae O139 strains. Int J Antimicrob Agents. 2013 Jul;42(1):72–5.

29. Divya MP, Sivakumar KC, Sarada Devi KL, Remadevi S, Thomas S. Novel multiple mutations in the topoisomerase gene of Haitian variant Vibrio cholerae O1. Antimicrob Agents Chemother. 2014 Aug;58(8):4982–3.