Check for updates

RESEARCH ARTICLE

Whole-genome sequence analysis of *Vibrio cholerae* from three outbreaks in Uganda, 2014 - 2016 [version 1; peer review: 2 not approved]

Dickson Aruhomukama ¹, Ivan Sserwadda ², Gerald Mboowa ^{1,2}

¹Department of Medical Microbiology, College of Health Sciences, School of Biomedical Sciences, Makerere University, Kampala, P.O Box 7072, Uganda

²Department of Immunology and Molecular Biology, School of Biomedical Sciences, College of Health Sciences, School of Biomedical Sciences, Makerere University, Kampala, P.O Box 7072, Uganda

V1 First published: 02 Aug 2019, 8:1340
<https://doi.org/10.12688/f1000research.20048.1>

Latest published: 02 Aug 2019, 8:1340
<https://doi.org/10.12688/f1000research.20048.1>

Abstract

Background: Cholera remains a serious public health problem in Uganda and Africa. The aim of this study was to provide the complete array of antimicrobial resistance genes, integrative and conjugative elements, virulence genes, pathogenicity islands, plasmids, and insertion sequences in the strains. In addition, this study also aimed to provide a single nucleotide polymorphism (SNP) based phylogenetic analysis of the strains.

Methods: In the analysis, both Linux and web-based bioinformatics approaches were used to analyze the study sequences. Databases used included; FastQC, MultiQC, Snippy, PANTHER, PATRIC, Unicycler, ISFinder, Center for Genomic Epidemiology pipelines (i.e. MLST, PlasmidFinder, MyDbFinder, and ResFinder), MashTree and IcyTree.

Results: The 10 sequenced strains of *Vibrio cholerae* were found to carry virulence-associated genes including *MakA*, *ctxA*, *ctxB*, *carA*, *carB*, *trpB*, *clpB*, *ace*, *toxR*, *zot*, *rtxA*, *ompW*, *ompR*, *gmhA*, *fur*, *hlyA*, and *rstR*. Also identified were: genes of the Type VI secretion system including *vasA-L*, *vgrG-2*, *vgrG-3*, *vipA/mglA*, and *vipB/mglB*; *alsD* (VC1589), involved in the synthesis of 2,3-butanediol; *alsR*, involved in the acetate-responsive LysR-type regulation; *makA*, the flagella-mediated cytotoxin gene; Type VI pilus genes including *tcpA-F*, *tcpH-J*, *tcpN*, *tcpP-T*, and *icmF/vask*; adherence genes *acfA-D* and *IlpA*; and quorum sensing system genes *luxS* and *cqsA*. Pathogenicity islands identified comprised of VSP-1 and VSP-2, as well as VPI-1 and VPI-2. In addition, *strA* and *B*, *APH(3'')-I*, *APH(3'')-Ib*, *APH(6)-Id*, *APH(6)-Ic*, *murA*, *pare*, *dfrA1*, *floR*, *catB*, and *catB9* were among the antimicrobial resistance genes found in the sequences. Analysis for SNPs shared among the sequences showed that the sequenced strains shared 218 SNPs and of these, 98 SNPs were missense. Gene enrichment analysis of these SNPs showed enrichment in genes that mediate transmembrane-

Open Peer Review

Approval Status

	1	2
version 1
02 Aug 2019	view	view
1. Jason Miller , Shepherd University, Shepherdstown, USA		
2. Samit Watve , Tufts University, Medford, USA		

Any reports and responses or comments on the article can be found at the end of the article.

signaling receptor activity, peptidyl-prolyl cis-trans isomerase activity, and phosphor-relay response regulator activity.

Conclusions: This study applied bioinformatics approaches to provide comprehensive genomic analysis of *V. cholerae* genomes obtained from Uganda.

Keywords

Vibrio cholerae, Whole-genome sequencing, Bioinformatics, Genomics,

This article is included in the **Pathogens** gateway.

This article is included in the **Antimicrobial Resistance** collection.

Corresponding authors: Dickson Aruhomukama (dickson.aruhomukama@chs.mak.ac.ug), Gerald Mboowa (gerald.mboowa@chs.mak.ac.ug)

Author roles: **Aruhomukama D:** Conceptualization, Data Curation, Formal Analysis, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Sserwadda I:** Conceptualization, Data Curation, Formal Analysis, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Mboowa G:** Conceptualization, Data Curation, Formal Analysis, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Gerald Mboowa is supported through the DELTAS Africa Initiative [DEL15011] to THRIVE-2 (the Training Health Researchers into Vocational Excellence in East Africa). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences' (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107742] and the UK Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Aruhomukama D *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Aruhomukama D, Sserwadda I and Mboowa G. **Whole-genome sequence analysis of *Vibrio cholerae* from three outbreaks in Uganda, 2014 - 2016 [version 1; peer review: 2 not approved]** F1000Research 2019, 8:1340 <https://doi.org/10.12688/f1000research.20048.1>

First published: 02 Aug 2019, 8:1340 <https://doi.org/10.12688/f1000research.20048.1>

Introduction

Cholera remains a serious public health problem in Uganda and Africa as a whole^{1,2}. It is characterized by a large disease burden, recurrent outbreaks, high case fatality rates, as well as tenacious endemicity^{1,2}. Over the last four decades, Uganda has experienced several cholera outbreaks². The detection, monitoring, and surveillance of cholera in Uganda rely upon the isolation of *Vibrio cholerae* using culture-based methods in microbiology laboratories^{2,3}. However, these methods are faced with several challenges, including: associated long turn-around times (24–48 hours); limited microbiology laboratory capacity; lack of laboratory supplies; poor laboratory infrastructure, particularly electricity necessary to operate laboratory equipment; as well as limited reliable and rapid diagnostic tests². Unlike culture-based methods, high-throughput sequencing, a culture-independent method, has been documented to be less affected by most of the challenges facing culture-based methods and at the same time provides an unprecedented view of pathogen biology and delivers high-resolution genomic epidemiology via rapid and cheap whole-genome sequencing^{4–7}. Despite this knowledge, sequencing remains a less desirable option for most scientists in Uganda and hence, data on the genomic characteristics of *V. cholerae* remains scarce, likely attributable to the underdeveloped bioinformatics capacity and lack of expertise necessary for analyzing whole-genome sequencing data⁷. This study set out to use bioinformatics approaches to analyze whole-genome sequence data obtained from *V. cholerae* isolates from different outbreaks in Uganda, with the aim of providing the complete array of virulence genes, pathogenicity islands, antimicrobial resistance genes, integrative and conjugative elements, and antimicrobial resistance genes associated with these elements, plasmids, and insertion sequences. In addition, this study also provided a single nucleotide polymorphism (SNP) based phylogenetic analysis of the strains.

Methods

Study design

This was a cross-sectional study that analyzed 10 whole-genome sequences of *V. cholerae* isolates. These isolates were collected during three different cholera outbreaks in Uganda between 2014 and 2016 and sequenced by a group from the University of Maryland (Bwire *et al.*, 2018). The whole-genome sequencing data was deposited in the NCBI's Sequence Read Archive (SRA) with the accession number [SRP136117](#).

Sample collection and whole-genome sequencing

Procedures and considerations in sample collection and whole-genome sequencing are described by Bwire *et al.*, 2018. Briefly, whole-genome sequencing was done using three or four representative samples that have been obtained from each of the three Multiple-Locus Variable Number Tandem-Repeat Analysis clonal complexes that had been identified during the period 2014–2016. Steps in whole-genome sequencing were: library preparation from fragmented DNA, this was achieved using an appropriate library preparation kit (KAPA High Throughput Library Preparation Kit, Millipore-Sigma, St. Louis MO); following this, enrichment and barcoding were done, and subsequently, libraries were sequenced using a 100bp paired-end run on an Illumina HiSeq2500 (Illumina, San Diego, CA).

Bioinformatics workflow

Whole-genome sequencing data for the 10 *V. cholerae* isolates were downloaded from NCBI's SRA using the toolkit fastq-dump v2.9.3. An overview of the bioinformatics workflow adopted in this study has been provided ([Figure 1](#)).

Quality control of untrimmed sequence data

Untrimmed sequence data quality reports were generated with [FastQC](#) v0.11.8 and [MultiQC](#) v1.7 using default settings.

Figure 1. Outline of the bioinformatics workflow. IS, insertion sequences; MLST, multilocus sequence typing; AMR, antimicrobial resistance; SNP, single nucleotide polymorphism; PATRIC, Pathosystems Resource Integration Center.

Bacterial variant calling and gene ontology enrichment analysis

Bacterial SNP calling was done using Snippy 3.2-dev, a tool for rapid and core genome alignments. *V. cholerae* genome assembly (accession number [GCF_002892855.1](#)) was obtained from NCBI's nucleotide archive and used as a reference during variant calling. We then used BCFtools v1.9 to extract all SNPs that were shared by the 10 *V. cholerae* isolates. Custom bash scripts were used to extract only missense SNPs (nonsynonymous), available on [GitHub](#) (see *Software availability*)⁸. Gene ontology enrichment analysis was performed using PANTHER Overrepresentation Test (released 2019-06-06), annotation version PANTHER version 14.1 (Released 2019-03-12) and a reference list of *V. cholerae*. Biological process and molecular function enrichment analyses were also carried out using the same database.

Bacterial genome assembly and annotation

V. cholerae genomic reads were assembled using Unicycler v0.4.8-beta⁹ to generate contigs. The [Pathosystems Resource Integration Center \(PATRIC\)](#) v3.5.39 was used to annotate the assembled genomes.

Identification of antimicrobial resistance genes, virulence genes, insertion sequences, integrative and conjugative elements, pathogenicity islands, and plasmids

[PATRIC](#) v3.5.39 was used to generate genome assembly metrics, identify antimicrobial resistance genes, and virulence factors. We used [ISfinder](#), a dedicated database for bacterial insertion sequences, to screen for the presence of insertion sequences in our assembled bacterial genomes. In addition, we performed a number of analyses using the different pipelines at the [Center for Genomic Epidemiology](#) to analyze the assembled bacterial genomes. These analyses were multilocus sequence typing (MLST) using MLST v2.0, plasmids searches using PlasmidFinder v2.0, phenotyping using BLAST-based on the *V. cholerae* database using MyDbFinder v1.2, and identification

of acquired antibiotic resistance genes using ResFinder. For all the above pipelines, default settings were used.

SNP-based phylogenetic analysis

Using the Mashree command-line based tool, we generated a Newick file. This file was then uploaded to [IcyTree](#), a browser-based phylogenetic tree viewer.

Results

Genomic characterization of the *V. cholerae* strains

The 10 sequenced strains of *V. cholerae* belonging to Inaba and Ogawa serotypes were characterized through analysis of the whole-genome sequencing data. Except for one strain, which was a non-O1, all the other strains belonged to the serogroup O1 due to the presence of the *rfbV*-O1 gene. All 10 sequenced strains had biotype-specific genes *ctxB*, *rstR*, and *tcpA*; hence were all atypical EI Tor biotype variants of *V. cholerae* and these also belonged to the third wave of the seventh pandemic. *In silico* MLST revealed that the sequenced strains belonged to two different sequence types (STs); ST69 and ST515. [Table 1](#) shows the genomic characteristics of the *V. cholerae* strains.

In addition, all the 10 sequenced strains were found to carry virulence-associated genes. These included *MakA*, *ctxA*, *ctxB*, *carA*, *carB*, *trpB*, *clpB*, *ace*, *toxR*, *zot*, *rtxA*, *ompW*, *ompR*, *gmhA*, *fur*, *hlyA*, and *rstR*. The 10 sequenced strains also carried genes for the Type VI secretion system (T6SS), including *vasA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, and *L*, *vgrG*-2, *vgrG*-3, *vipA/mglA* and *vipB/mglB*. The strains were also found to have the following genes: *alsD* (VC1589), involved in the synthesis of 2,3-butanediol; *alsR*, involved in the acetate-responsive LysR-type regulation; *makA*, the flagella-mediated cytotoxin gene; Type IV pilus genes, including *tcpA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*, *J*, *N*, *P*, *Q*, *R*, *S*, and *T*, as well as *icmF/vasK*; adherence genes *acfA*, *B*, *C*, *D*, and *IlpA*; and quorum sensing system genes *luxS*, and *cqsA*. Pathogenicity islands were also present in all the 10 sequenced

Table 1. Biosample data, serotype and genomic sequence data of the *V. cholerae* strains.

Isolate ID	Serotype	Year of collection	Month of collection	District of origin	Coverage	Genome size (bp)	No. of contigs	No. of coding sequences
SRR6871251	Inaba	2014	Apr	Arua	155	4025190	88	3733
SRR6871254	Inaba	2014	May	Moyo	240	4025367	85	3743
SRR6871247	Ogawa	2015	Apr	Kasese	200	4012560	88	3697
SRR6871245	Inaba	2015	Apr	Kasese	260	4030572	96	3734
SRR6871253	Inaba	2015	Jul	Arua	200	4024003	89	3743
SRR6871248	Inaba	2015	May	Kasese	260	4034890	102	3745
SRR6871246	Inaba	2015	Sep	Hoima	250	4032599	90	3734
SRR6871252	Inaba	2015	Sep	Hoima	200	4025376	87	3731
SRR6871249	Ogawa	2016	Jan	Mbale	300	3998859	82	3688
SRR6871250	Ogawa	2016	Jan	Mbale	250	4011573	89	3700

strains; namely, the *Vibrio* seventh pandemic islands VSP-1 and VSP-2 as well as VPI-1 and VPI-2.

Genotypic antimicrobial resistance and mobile genetic elements

The 10 sequenced strains all showed genotypic resistance to streptomycin, aminoglycosides, fosfomycin, fluoroquinolones, sulphonamides, trimethoprim, chloramphenicol/florfenicol, and tetracyclines, as illustrated in [Table 2](#).

Furthermore, all the 10 sequenced strains contained the VC1786 integrative and conjugative elements (VC1786ICE genes). Antimicrobial resistance genes associated with resistance to chloramphenicol, streptomycin, sulfamethoxazole, and trimethoprim usually found on the VC1786ICE, such as *strA* and *strB*, *floR* as well as *sul2*, were found present in the strains according to MyDbFinder 1.2. The sequenced strains were also found to have a genomic organization of the integrative and conjugative element similar to that of the *V. cholerae* ICEVchHail reference strain¹⁰.

In addition, all the sequenced strains had no plasmids according to PlasmidFinder 2.0, particularly the IncA/C plasmids, and

cryptic plasmids pSDH1-2 were also absent in all the strains according to MyDbFinder 1.2 and in BLAST atlas.

Insertion sequences were, however, present in all the sequenced strains; these included TS200/IS605, IS630, IS66, IS3, and IS4 ([Table 3](#)).

Phylogenetic comparison of the *V. cholerae* strains

SNP-based phylogenetic analysis showed an overall SNP difference of 120 among the 10 sequenced *V. cholerae* strains. Close relatedness was observed among strains SRR6871252, SRR6871253, and SRR6871254 (only seven SNP differences); strains SRR6871247, SRR6871249, and SRR6871250 (only four SNP differences), and among strains SRR6871245, SRR6871246, and SRR6871248 (only six SNP differences) ([Figure 2](#)).

Furthermore, analysis for shared SNPs among the sequences showed that the sequenced strains shared 218 SNPs. Of these, 98 SNPs were missense (non-synonymous). Gene enrichment analysis of the SNPs using the PANTHER GO Ontology database showed enrichment in genes that mediate transmembrane-signaling receptor activity, peptidyl-prolyl cis-trans isomerase activity, and phosphor-relay response regulator activity.

Table 2. Antimicrobial resistance genes in the *V. cholerae* strains.

Antibiotic category	Genes associated with resistance	Resistance phenotype
Tetracycline	<i>Tet (35)</i>	Tetracycline resistance
Sulphonamides	<i>sul2</i>	Sulphonamide resistance
Trimethoprim	<i>dfrA1</i>	Trimethoprim resistance
Phenicol	<i>catB9</i> , <i>catB</i> , <i>floR</i>	Chloramphenicol resistance
Fosfomycin	<i>MurA</i>	Fosfomycin resistance
Fluoroquinolone	<i>ParE</i>	Fluoroquinolone resistance
Aminoglycoside	<i>APH(3'')-I</i> , <i>APH(3'')-Ib</i> , <i>APH(6)-Id</i> , <i>APH(6)-Ic</i>	Aminoglycoside resistance

Table 3. Insertion sequences in the *V. cholerae* strains.

Isolate ID	TS200/IS605	IS630	IS66	IS3	IS4
SRR6871251	+	+	+	-	-
SRR6871254	+	+	+	-	-
SRR6871247	+	+	+	-	-
SRR6871245	+	+	+	-	-
SRR6871253	+	+	+	-	-
SRR6871248	+	+	+	-	-
SRR6871246	-	-	+	+	+
SRR6871252	+	+	+	-	-
SRR6871249	+	+	+	-	-
SRR6871250	+	+	+	+	-

Key: + / - Presence or absence

Figure 2. Single nucleotide polymorphism-tree showing the phylogenetic relationship among the *V. cholerae* strains.

Discussion

This was a cross-sectional study that aimed at providing a comprehensive genomic analysis of 10 whole-genome sequences of *V. cholerae* isolates collected during three different cholera outbreaks in Uganda between 2014 and 2016, submitted by the University of Maryland (Baltimore, MD, United States) to the NCBI SRA database under a study titled, “Molecular characterization of *Vibrio cholerae* responsible for cholera epidemics in Uganda by PCR, MLVA and WGS”.

In the genomic analysis, this study confirmed the identity of the isolates, provided the complete array of virulence genes, pathogenicity islands, antimicrobial resistance genes, integrative and conjugative elements, and antimicrobial resistance genes associated with these elements, plasmids, and insertion sequences. In addition, this study also provided a SNP-based phylogenetic analysis of the strains.

The identity of the isolates was in tandem with what was reported by Bwire *et al.*, 2018. In addition, the finding of this study in regards to most of the isolates belonging to the O1 serotype are consistent with other studies elsewhere; these studies attributed this to the presence of the *rfbV*-O1 gene in isolates classified as O1⁵. Unlike a similar study done in the East African region⁵ in which MLST revealed that their isolates

belonged to a single ST (ST69), this study revealed that the isolates belonged to two STs; namely, ST69 and ST515. Strains of *V. cholerae* belonging to the ST515 have been reported elsewhere¹¹.

The virulence genes reported in this study are similar to those reported in studies elsewhere^{5,12–14}. These genes included, among others, those belonging to the Type IV secretion system, those involved in adherence, Type IV pilus genes, and those involved in quorum sensing.

Accessory genetic elements, particularly pathogenicity islands previously reported to commonly occur in *V. cholerae* were also reported in this study; namely, VSP-1, VSP-2 as well as VPI-1 and VPI-2^{15,16}. These have not only been documented to encode virulence-associated genes in *V. cholerae*, but have also been reported to facilitate a better understanding of the evolutionary events that lead to the emergence of pathogenic *V. cholerae* clones^{15,16}.

Despite the **World Health Organization (WHO) recommendations** in regards to the management of cholera with oral rehydration salts in addition to antibiotics such as streptomycin, aminoglycosides, trimethoprim, fosfomycin, fluoroquinolones, sulphonamides, chloramphenicol/florfenicol, and tetracyclines, this study reports

genotypic resistance in the isolates to these same antibiotics. Similar resistance has been reported in similar studies from the East African region and elsewhere^{5,17–19}.

The presence of integrative and conjugative elements (VC1786ICE), containing resistance genes associated with sulfamethoxazole and trimethoprim, chloramphenicol, and streptomycin resistance, are also reported in this study. These are genomically similar to *V. cholerae* ICEVchHai1^{17,19}.

This study found no plasmids or *intI* genes. This could be attributed to the presence of integrative and conjugative elements, a factor that made them insignificant in regards to the encoding of antimicrobial resistance. Studies similar to this have reported similar findings⁵.

This study also reported the presence of insertion sequences IS605, IS66, IS3, and IS4. Insertion sequences have been described in various studies to be drivers of genetic variability. These studies have also alluded to them being fixed by natural selection each time that a mutation induced by these elements is selected^{20,21}.

The presence of T6SS genes in the study isolates could explain their antimicrobial resistance gene profile. TS66-dependent killing of other bacteria is mostly directed to neighboring cells. These consequently release their DNA, which is ultimately taken up by the killer cells and in the process, these can integrate valuable genes including those that encode antimicrobial resistance. These may consequently evolve, leading to antimicrobial resistance in the killer cells²².

The results obtained from the SNP-based phylogenetic analysis show the relatedness of the Ugandan *V. cholerae* strains and are in agreement with the results obtained by Bwire *et al.*, 2018.

The analysis for shared SNPs among the sequences and, consequently, gene enrichment revealed the enrichment in genes that mediate transmembrane-signaling receptor activity, peptidyl-prolyl cis-trans isomerase activity, and phosphor relay response regulator activity. These play a fundamental role in quorum sensing in *V. cholerae*, a process of cell-cell communication that allows these bacteria to share information about cell density and adjust gene expression accordingly^{23–25}. Quorum sensing has been documented to regulate the expression of virulence factors in *V. cholerae*^{23–25}.

Conclusions

Despite the fact that bioinformatics capacity remains underdeveloped in Uganda and Africa as a whole, this study demonstrated the ability to apply bioinformatics approaches to zoom into genomes (in this case, *V. cholerae* genomes obtained from Uganda) to provide a comprehensive genomic analysis. This study sets a stage that encourages more sequencing work with potential public health consequences to be done in African settings. Furthermore, it also encourages the need to build bioinformatics capacity in African settings to enable analysis of whole-genome sequence data generated from the continent.

Data availability

Underlying data

Whole-genome sequences of the ten *Vibrio cholera* isolates from Sequence Read Archive, Accession number SRP136117: <https://identifiers.org/insdc.sra/SRP136117>

Vibrio cholera reference genome assembly from NCBI Assembly, Accession number GCF_002892855.1: https://www.ncbi.nlm.nih.gov/assembly/GCF_002892855.1

Software availability

- Source code available from: <https://github.com/gmboowa/extractonlymissenseSNPs->
- Archived source code at the time of publication: <https://doi.org/10.5281/zenodo.3354469>⁸
- License: GPL-3.0

Grant information

Gerald Mboowa is supported through the DELTAS Africa Initiative [DEL15011] to THRiVE-2 (the Training Health Researchers into Vocational Excellence in East Africa). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences' (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107742] and the UK Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Mengel MA, Delrieu I, Heyerdahl L, *et al.*: **Cholera outbreaks in Africa**. In: *Cholera outbreaks*. Springer; *Curr Top Microbiol Immunol*. 2014; **379**: 117–44. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Bwire G, Orach CG, Abdallah D, *et al.*: **Alkaline peptone water enrichment with a dipstick test to quickly detect and monitor cholera outbreaks**. *BMC Infect Dis*. 2017; **17**(1): 726. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Kasolo F, Yoti Z, Bakayita N, *et al.*: **IDSR as a platform for implementing IHR in African countries**. *Biosecur Bioterror*. 2013; **11**(3): 163–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Pallen MJ, Loman NJ, Penn CW: **High-throughput sequencing and clinical microbiology: progress, opportunities and challenges**. *Curr Opin Microbiol*. 2010; **13**(5): 625–31. [PubMed Abstract](#) | [Publisher Full Text](#)

5. Hounmanou YMG, Leekitcharoenphon P, Hendriksen RS, *et al.*: **Environmental surveillance and genomic characterization of *Vibrio cholerae* O1 from fish, phytoplankton and water in Lake Victoria.** In: *13th DWF Water Research Conference: Danish Water Forum.* University of Copenhagen; 2019; 52.
[Reference Source](#)
6. Bwire G, Sack DA, Almeida M, *et al.*: **Molecular characterization of *Vibrio cholerae* responsible for cholera epidemics in Uganda by PCR, MLVA and WGS.** *PLoS Negl Trop Dis.* 2018; 12(6): e0006492.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Aruhomukama D, Sserwadda I, Mboowa G: **Investigating colistin drug resistance: The role of high-throughput sequencing and bioinformatics [version 2; peer review: 2 approved, 1 approved with reservations].** *F1000Res.* 2019; 8: 150.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Mboowa G: **gmboowa/extractonlymissenseSNPs: Extracts Genomic Variant Class and Counts (Version v1.0.0).** *Zenodo.* 2019.
<http://www.doi.org/10.5281/zenodo.3354469>
9. Wick RR, Judd LM, Gorrie CL, *et al.*: **Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.** *PLoS Comput Biol.* 2017; 13(6): e1005595.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Reimer AR, Van Domselaar G, Stroika S, *et al.*: **Comparative genomics of *Vibrio cholerae* from Haiti, Asia, and Africa.** *Emerg Infect Dis.* 2011; 17(11): 2113–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Fu S, Hao J, Jin S, *et al.*: **A Human Intestinal Infection Caused by a Novel Non-O1/O139 *Vibrio cholerae* Genotype and Its Dissemination Along the River.** *Front Public Heal.* 2019; 7: 100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Ceccarelli D, Garriss G, Choi SY, *et al.*: **Characterization of Two Cryptic Plasmids Isolated in Haiti from Clinical *Vibrio cholerae* Non-O1/Non-O139.** *Front Microbiol.* 2017; 8: 2283.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Aliabad NH, Bakhshi B, Pourshafie MR, *et al.*: **Molecular diversity of CTX prophage in *Vibrio cholerae*.** *Lett Appl Microbiol.* 2012; 55(1): 27–32.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Zhu J, Miller MB, Vance RE, *et al.*: **Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*.** *Proc Natl Acad Sci U S A.* 2002; 99(5): 3129–34.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Faruque SM, Mekalanos JJ: **Pathogenicity islands and phages in *Vibrio cholerae* evolution.** *Trends Microbiol.* 2003; 11(11): 505–10.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Murphy RA, Boyd EF: **Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates.** *J Bacteriol.* 2008; 190(2): 636–47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Hendriksen RS, Price LB, Schupp JM, *et al.*: **Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak.** *mBio.* 2011; 2(4): e00157–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Hounmanou YM, Mdegela RH, Dougnon TV, *et al.*: **Toxigenic *Vibrio cholerae* O1 in vegetables and fish raised in wastewater irrigated fields and stabilization ponds during a non-cholera outbreak period in Morogoro, Tanzania: an environmental health study.** *BMC Res Notes.* 2016; 9(1): 466.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Kaas RS, Ngandjio A, Nzouankeu A, *et al.*: **The Lake Chad Basin, an Isolated and Persistent Reservoir of *Vibrio cholerae* O1: A Genomic Insight into the Outbreak in Cameroon, 2010.** *PLoS One.* 2016; 11(5): e0155691.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Syvanen M: **Insertion sequences and their evolutionary role.** In: *Bacterial Genomes.* Springer; 1998; 213–20.
[Publisher Full Text](#)
21. Siguier P, Goubeyre E, Chandler M: **Bacterial insertion sequences: their genomic impact and diversity.** *FEMS Microbiol Rev.* 2014; 38(5): 865–91.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Borgeaud S, Metzger LC, Scignari T, *et al.*: **The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer.** *Science.* 2015; 347(6217): 63–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Rutherford ST, Bassler BL: **Bacterial quorum sensing: its role in virulence and possibilities for its control.** *Cold Spring Harb Perspect Med.* 2012; 2(11): pii: a012427.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Parker CT, Sperandio V: **Cell-to-cell signalling during pathogenesis.** *Cell Microbiol.* 2009; 11(3): 363–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Waters CM, Bassler BL: **The *Vibrio harveyi* quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers.** *Genes Dev.* 2006; 20(19): 2754–67.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 18 September 2019

<https://doi.org/10.5256/f1000research.22012.r53477>

© 2019 Watve S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Samit Watve

Sackler School of Graduate Biomedical Sciences, Tufts University, Medford, MA, USA

General comments:

This study is tough to review because there is no clear scientific question being addressed and also lacks particularly novel or interesting findings. The authors claim, "The aim of this study was to provide the complete array of antimicrobial resistance genes, integrative and conjugative elements, virulence genes, pathogenicity islands, plasmids, and insertion sequences in the strains. In addition, this study also aimed to provide a single nucleotide polymorphism (SNP) based phylogenetic analysis of the strains." What is unclear is what value this data array provides over the currently available genomic data. The study exclusively uses genomes that have been previously sequenced, assembled, analyzed and reported by Bwire *et al.*, and are publicly available in the NCBI repository under accession number [SRP136117](#). One could argue that having a pre-compiled dataset describing the presence/absence of antimicrobial resistance genes, integrative and conjugative elements, virulence genes, pathogenicity islands, plasmids, and insertion sequences enables faster access to important information for future researchers. If so, then why limit such reference dataset to a handful of genomes? A comprehensive database covering either all or a large proportion of the ~1300 or so publicly available *V. cholerae* genomes would be more appropriate for this kind of study. The original Bwire *et al.*, study used genome sequencing as a tool to track disease spread across geography and time. This study uses the same information to report the presence of virulence, type VI secretion, type IV secretion and pathogenicity islands. These features are present almost universally in the seventh pandemic El Tor strains of *Vibrio cholerae* including the one in Uganda and this study does not add significantly to what's already known about *Vibrio cholerae* biology. At best, it is a re-analysis of the Bwire *et al.* study with specific gene feature annotations extracted and tabulated. Since this study doesn't represent an important advance to the state of knowledge in the field, in my view, this study does not merit publication.

Specific comments:

1. The section on sample collection and sequencing should be removed. Since the biological sample collection and sequencing was not performed in this study, it is inappropriate to

include those methods here.

2. Based on the workflow outlined in fig 1, it seems that for one aspect of the study (SNP variation analysis) the authors used the original RefSeq assemblies from Bwire *et al.*, while for other aspects, such as looking for antibiotic resistance genes, finding plasmids etc. they generated new assemblies. The authors should explain why they used two different sets of assemblies and whether using different assemblies would generate significantly different results.
3. Fig 2 doesn't have a scale bar for genetic distance or a distantly related outgroup to compare against. Therefore, the distance between two strains is impossible to determine.
4. Some claims don't make sense. For example, in the Discussion the authors claim, "In the genomic analysis, this study confirmed the identity of the isolates, ..." or "The identity of the isolates was in tandem with what was reported by Bwire *et al.*, 2018." Since the genomes were from the strains that Bwire *et al.* sequenced, why was this ever in doubt?
5. The authors also claim, "The presence of T6SS genes in the study isolates could explain their antimicrobial resistance gene profile. T6SS-dependent killing of other bacteria is mostly directed to neighbouring cells. These consequently release their DNA, which is ultimately taken up by the killer cells and in the process, these can integrate valuable genes including those that encode antimicrobial resistance. These may consequently evolve, leading to antimicrobial resistance in the killer cells."

Several studies have demonstrated that an active T6SS can lyse neighbouring cells, release their contents and allow for horizontal gene transfer. Even though this is a potential mechanism for horizontal transfer, there is no evidence that the presence of active T6SS in these strains has led to acquisition of antibiotic resistance genes. Indeed, *Vibrio cholerae* strains lacking an active T6SS can also have high efficiency of natural transformation and T6SS is by no means a requirement for horizontal gene transfer in *Vibrio cholerae*. Alternatively, antibiotic resistance genes can also be picked up through conjugation or transduction.

6. "The analysis for shared SNPs among the sequences and, consequently, gene enrichment revealed the enrichment in genes that mediate transmembrane-signaling receptor activity, peptidyl-prolyl cis-trans isomerase activity, and phosphorelay response regulator activity. These play a fundamental role in quorum sensing in *V. cholerae*, a process of cell-cell communication that allows these bacteria to share information about cell density and adjust gene expression accordingly²³⁻²⁵. Quorum sensing has been documented to regulate the expression of virulence factors in *V. cholerae*²³⁻²⁵."

While it is true that the quorum sensing receptors LuxQ and CqsS have transmembrane-signaling and receptor activity as well as phosphorelay activity, the authors do not specify

whether the *luxQ* and *cqsS* genes themselves carry any SNP's. Many other two component signaling system proteins also have transmembrane-signaling and phosphorelay activity.

7. Please avoid vague or overly general sentences like: "Similar resistance has been reported in similar studies from the East African region and elsewhere^{5,17-19}." or "This study found no plasmids or *intI* genes... Studies similar to this have reported similar findings⁵."

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetics of *Vibrio cholerae* with a focus on Type six secretion, natural transformation, quorum sensing as well as expertise in genome sequencing and data analysis.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 19 August 2019

<https://doi.org/10.5256/f1000research.22012.r52002>

© 2019 Miller J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jason Miller

College of Natural Sciences and Mathematics, Shepherd University, Shepherdstown, WV, USA

Summary of the Manuscript: Genome assemblies of ten strains of *Vibrio cholerae* were generated from public sequence data. The assembled contigs were analyzed for gene content, insertion sequence content, SNPs content, etc.

Summary of the Review: This reviewer is uncertain whether this manuscript contains publishable information. The experiment seems to be a repeat of a previously published experiment, using the same data exclusively. This experiment may have been designed as a test of the reproducibility of that previous study, but it is not presented as such. The Results section lists several ways that this study confirms findings in several previous publications but it leaves unclear what is novel. The Conclusions section of this study focuses on a different topic that was hardly addressed by the study: whether bioinformatics can be applied to genomic sequence, particularly in Uganda.

Major Revision:

The conclusions as stated are not supported by the data. Please refer to the conclusions section of the Abstract as well as the entire Conclusions section of the manuscript. These sections present a finding that researchers, particularly researchers in Uganda, can use bioinformatics to study cholera. If that truly was the research question, then the manuscript needs a new title and the report should be presented as a case study, N=1. In fact, the focus of the work seems to be cholera. If so, then the conclusions must pertain to cholera. If this was regionally the first study of its kind, then a sentence to that effect could be warranted in the Discussion section.

This manuscript needs to clarify whether this was a re-analysis of a previously published whole-genome sequencing project. The relationship to that previous study is easy to miss. Several sections are misleading in this regard. The Title refers to "Whole-genome sequence analysis" which usually includes sequencing. The Abstract refers to "The 10 sequenced strains" rather than "Ten previously-sequenced strains." The Introduction says, "sequencing remains a less desirable option for most scientists in Uganda", without pointing out that no sequencing was performed for this study. The Methods section says, "the whole-genome sequencing data was deposited in NCBI's SRA with accession SRP136117" but that was the data of Bwire 2018; as a previously published fact, the deposition is not appropriate for this manuscript to report.

The manuscript must distinguish its novel findings from confirmations of previously published findings. This issue is addressed broadly in the Discussion section but not specifically in the Results. For example, the Results section says, "... all the 10 sequenced strains were found to carry virulence-associated genes. These included *MakA*, *ctxA*, *ctxB*, *carA*, *carB*, *trpB*, *clpB*, *ace*, *toxR*, *zot*, *rtxA*, *ompW*, *ompR*, *gmhA*, *fur*, *hlyA*, and *rstR*." No prior work is cited there. The text fails to point out that the Bwire 2018 paper already reported that "All 63 isolates tested positive for *ompW*, *toxR*, and *ctxA* ..." based on similar analysis of the very same sequencing reads.

The manuscript needs to explain why the re-analysis is worthy of publication. The Bwire 2018 paper already reported that the ten samples were Illumina-sequenced and assembled with the SPAdes software. The authors of that study posted ten assemblies at NCBI under BioProject PRJNA439310 and analyzed them in their publication. This manuscript presents ten assemblies of the same data using Unicycler (which uses SPAdes). Are the new assemblies different? Did the re-analysis highlight any flaws or shortcomings of the previous study? Did the re-analysis reveal

anything new?

Minor Revision:

In the Discussion, this study is self-described as a “cross-sectional study” which implies a look across samples from one timepoint. However, this study looked at three outbreaks over two years. Thus there was opportunity to include longitudinal and geo-spacial population analysis. This opportunity should be addressed in discussion, if possible.

I presume the reads from the ten samples were assembled separately, but this is not stated in the Methods and not illustrated in Figure1 .

The phrase in the Abstract, “... in the strains”, should be “... in 10 strains isolated in Uganda” or similar. Basically, you cannot refer to “the” strains before saying which strains.

Within the Abstract, the background says one goal was to describe phylogeny by SNP-analysis but the abstract has no further mention of this. Either omit this from the background or include something about it in the results of the abstract.

The inline citations use notation like “Bwire 2018” but the references are numbered and listed in order of citation. This makes it hard to find the reference for a citation. Could the authors adopt one convention or the other?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, expression analysis.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for

reasons outlined above.

Comments on this article

Version 1

Reader Comment 06 Aug 2019

Gildas Fils de Roger, University of Copenhagen, Copenhagen, Denmark

For your reference number 5, cite the original paper at <https://doi.org/10.3389/fmicb.2019.00901> rather than the conference proceeding.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research