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

Dickson Aruhomukama, Ivan Sserwadda, Gerald Mboowa

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

**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-signaling receptor activity, peptidyl-prolyl cis-trans isomerase activity, and phospho-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,

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 Licence, 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: 1 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. It is characterized by a large disease burden, recurrent outbreaks, high case fatality rates, as well as tenacious endemicity. Over the last four decades, Uganda has experienced several cholera outbreaks. The detection, monitoring, and surveillance of cholera in Uganda rely upon the isolation of *Vibrio cholerae* using culture-based methods in microbiology laboratories. 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 tests. 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. 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 data. 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.](image-url)
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 (non-synonymous), available on GitHub (see Software availability). 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-beta2\(^3\) 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 Mashtree 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 El 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, tcpB, clpB, ace, toxR, zot, ctxA, ompW, ompR, gmhA, fur, hlyA, and rstR. The 10 sequenced strains also carried genes for the Type V1 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-butenediol; 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.

<table>
<thead>
<tr>
<th>Isolate ID</th>
<th>Serotype</th>
<th>Year of collection</th>
<th>Month of collection</th>
<th>District of origin</th>
<th>Coverage</th>
<th>Genome size (bp)</th>
<th>No. of contigs</th>
<th>No. of coding sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRR6871251</td>
<td>Inaba</td>
<td>2014</td>
<td>Apr</td>
<td>Arua</td>
<td>155</td>
<td>4025190</td>
<td>88</td>
<td>3733</td>
</tr>
<tr>
<td>SRR6871254</td>
<td>Inaba</td>
<td>2014</td>
<td>May</td>
<td>Moyo</td>
<td>240</td>
<td>4025367</td>
<td>85</td>
<td>3743</td>
</tr>
<tr>
<td>SRR6871247</td>
<td>Ogawa</td>
<td>2015</td>
<td>Apr</td>
<td>Kasese</td>
<td>200</td>
<td>4012560</td>
<td>88</td>
<td>3697</td>
</tr>
<tr>
<td>SRR6871245</td>
<td>Inaba</td>
<td>2015</td>
<td>Apr</td>
<td>Kasese</td>
<td>260</td>
<td>4030572</td>
<td>96</td>
<td>3734</td>
</tr>
<tr>
<td>SRR6871253</td>
<td>Inaba</td>
<td>2015</td>
<td>Jul</td>
<td>Arua</td>
<td>200</td>
<td>4024003</td>
<td>89</td>
<td>3743</td>
</tr>
<tr>
<td>SRR6871248</td>
<td>Inaba</td>
<td>2015</td>
<td>May</td>
<td>Kasese</td>
<td>260</td>
<td>4034890</td>
<td>102</td>
<td>3745</td>
</tr>
<tr>
<td>SRR6871246</td>
<td>Inaba</td>
<td>2015</td>
<td>Sep</td>
<td>Hoima</td>
<td>250</td>
<td>4032599</td>
<td>90</td>
<td>3734</td>
</tr>
<tr>
<td>SRR6871252</td>
<td>Inaba</td>
<td>2015</td>
<td>Sep</td>
<td>Hoima</td>
<td>200</td>
<td>4025376</td>
<td>87</td>
<td>3731</td>
</tr>
<tr>
<td>SRR6871249</td>
<td>Ogawa</td>
<td>2016</td>
<td>Jan</td>
<td>Mbale</td>
<td>300</td>
<td>3998859</td>
<td>82</td>
<td>3688</td>
</tr>
<tr>
<td>SRR6871250</td>
<td>Ogawa</td>
<td>2016</td>
<td>Jan</td>
<td>Mbale</td>
<td>250</td>
<td>4011573</td>
<td>89</td>
<td>3700</td>
</tr>
</tbody>
</table>
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/llorfenicol, 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* ICEVchHai1 reference strain95.

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.

<table>
<thead>
<tr>
<th>Antibiotic category</th>
<th>Genes associated with resistance</th>
<th>Resistance phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetracycline</td>
<td>Tet (35)</td>
<td>Tetracycline resistance</td>
</tr>
<tr>
<td>Sulphonamides</td>
<td>sul2</td>
<td>Sulphonamide resistance</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>dfrA1</td>
<td>Trimethoprim resistance</td>
</tr>
<tr>
<td>Phenicols</td>
<td>catB9, catB, floR</td>
<td>Chloramphenicol resistance</td>
</tr>
<tr>
<td>Fosfomycin</td>
<td>MurA</td>
<td>Fosfomycin resistance</td>
</tr>
<tr>
<td>Fluoroquinolone</td>
<td>ParE</td>
<td>Fluoroquinolone resistance</td>
</tr>
<tr>
<td>Aminoglycoside</td>
<td>APH(3&quot;)-I, APH(3&quot;)-Ib, APH(6)-Id, APH(6)-Ic</td>
<td>Aminoglycoside resistance</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Isolate ID</th>
<th>TS200/IS605</th>
<th>IS630</th>
<th>IS66</th>
<th>IS3</th>
<th>IS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRR6871251</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871254</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871247</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871245</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871253</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871248</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871246</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SRR6871252</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871249</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRR6871250</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Key: + / - Presence or absence
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 O1. Unlike a similar study done in the East African region 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 elsewhere.

The virulence genes reported in this study are similar to those reported in studies elsewhere. 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, VPI-1, VSP-2 as well as VPI-1 and VPI-2. 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.

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 

The presence of integrative and conjugative elements (VC1786I{CE}), containing resistance genes associated with sulfamethoxazole and trimethoprim, chloramphenicol, and streptomycin resistance, are also reported in this study. These are genomically similar to \textit{V. cholerae} ICEVchHai1 

This study found no plasmids or \textit{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 findings 

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 

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 cells 

The results obtained from the SNP-based phylogenetic analysis show the relatedness of the Ugandan \textit{V. cholerae} strains and are in agreement with the results obtained by Bwire \textit{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 \textit{V. cholerae}, a process of cell-cell communication that allows these bacteria to share information about cell density and adjust gene expression accordingly 

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, \textit{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 \textit{Vibrio cholera} isolates from Sequence Read Archive, Accession number SRR136117: https://identifiers.org/insdc.sra/SRP136117

\textit{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
- 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


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