

Comprehensive genomic analysis of *Vibrio cholerae* from three outbreaks in Uganda, 2014 - 2016

Ivan Sserwadda

Makerere University College of Health Sciences

Dickson Aruhomukama

Makerere University College of Health Sciences

Gerald Mboowa (✉ gmbwoowa@gmail.com)

Makerere University College of Health Sciences <https://orcid.org/0000-0001-8445-9414>

Research note

Keywords: *V. cholerae*, Whole-genome sequencing, Bioinformatics, Genomics, Uganda

Posted Date: July 2nd, 2019

DOI: <https://doi.org/10.21203/rs.2.10810/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective: This study aimed at providing a comprehensive genomic analysis of whole-genome sequence data obtained from *V. cholerae* isolates from different outbreaks in Uganda using bioinformatics approaches. Results: The ten sequenced strains of *V. cholerae* were found to carry virulence-associated genes MakA, ctxA, ctxB, carA, carB, trpB, clpB, ace, toxR, zot, rtxA, ompW, ompR, gmhA, fur, hlyA, rstR, Type IV secretion system genes, T6SS genes vasA-L, vgrG-2, vgrG-3, vipA/mglA, and vipB/mglB; alsD; alsR; the flagella-mediated cytotoxin gene, makA; Type IV pilus genes tcpA-F, tcpH-J, tcpN, tcpP-T, and icmF/vasK; adherence genes acfA-D, IlpA, and quorum sensing system genes luxS and cqsA. Pathogenicity islands identified were VSP-1, VSP-2, VPI-1, and VPI-2. Antimicrobial resistance genes identified in the strains included strA and B, APH(3")-I, APH(3")-Ib, APH(6)-Id, APH(6)-Ic, murA, pare, dfrA1, floR, catB, and catB9. The strains also contained genes of the VC1786 integrative and conjugative element. SNP-based phylogenetic analysis revealed 218 shared SNPs in the strains, of these, 98 were missense. Gene enrichment analysis of the 98 SNPs revealed enrichment in genes that mediate transmembrane-signaling activity, peptidyl-prolyl cis-trans isomerase activity, and phosphorelay response regulator activity.

Introduction

Cholera remains a serious public health problem in Uganda and Africa as a whole (1,2). In Uganda, the detection, monitoring, and surveillance of cholera rely upon the isolation of *V. cholerae* using culture-based methods (2). The use of these methods has been associated with several challenges (1,3). Unlike these methods, high-throughput sequencing culture-independent methods have been documented to provide an unprecedented view of pathogen biology as well as delivery of 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 comprehensive genomic data on *V. cholerae* remains unavailable. Therefore, this study aimed at providing a comprehensive genomic analysis of whole-genome sequence data obtained from *V. cholerae* isolates from different outbreaks in Uganda using bioinformatics approaches.

Methods

This was a cross-sectional study that analyzed 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) at the SRA database of the National Center for Biotechnology Information (NCBI) under a study titled, "Molecular characterization of *Vibrio cholerae* responsible for cholera epidemics in Uganda by PCR, MLVA and WGS" and Project Accession Number: PRJNA439310.

Sequence data was downloaded using the NCBI's SRA toolkit fastq-dump v2.9.3. Following this, a comprehensive bioinformatics analysis was done.

An overview of the bioinformatics workflow adopted in this study has been provided (Figure 1).

Figure 1. Overview of the bioinformatics workflow. Each box represents a component corresponding to a series of tasks that provide a certain well-defined output. Major bioinformatics utilities employed in each module are also mentioned.

In the analysis, untrimmed sequence data quality reports were generated with FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v1.7 (<https://multiqc.info/>) using default settings.

Bacterial SNP calling was done using Snippy 3.2-dev. *V. cholerae* genome assembly accession: GCF_002892855.1 obtained from NCBI's nucleotide archive was used as a reference during variant calling. BCFtools v1.9 was used to extract all SNPs that were shared by the *V. cholerae* strain. Custom bash scripts were used to extract only missense SNPs. Gene ontology enrichment analysis (<http://geneontology.org/>) was performed using PANTHER.

V. cholerae genomic reads were assembled using Unicycler v0.4.8-beta (8) to generate contigs. The Pathosystems Resource Integration Center - PATRIC v3.5.39 was used to annotate the assembled genomes.

PATRIC v3.5.39 was used to generate genome assembly metrics, identify antimicrobial resistance genes, and virulence factors. We used the ISfinder 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 included Multilocus sequence typing using MLST v2.0, plasmids searches with PlasmidFinder v2.0, MyDbFinder v1.2 to phenotype using blast based on *V. cholerae* database, and ResFinder for identification of acquired antibiotic resistance genes. For all the above pipelines, we used default settings.

Results

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 the sequenced strains had biotype-specific genes *ctxB*, *rstR*, and *tcpA* hence were all atypical El Tor biotype variants of *V. cholerae*, this implies that they belonged to the third wave of the seventh pandemic. The *in silico* MLST revealed that the sequenced strains belonged to two different sequence types, the ST69, and ST515. Table 1 shows the genomic characteristics of the *V. cholerae* strains.

Table 1: Biosample data, Serotype and Genomic sequence data of the *V. cholerae* strains

In addition, all the strains were found to carry virulence-associated genes *MakA*, *ctxA*, *ctxB*, *carA*, *carB*, *trpB*, *clpB*, *ace*, *toxR*, *zot*, *rtxA*, *ompW*, *ompR*, *gmhA*, *fur*, *hlyA*, *rstR*, genes of the Type IV secretion system, T6SS genes, these included *vasA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, and *L*, *vgrG-2*, *vgrG-3*, *vipA/mglA*, *vipB/mglB*;

alsD involved in the synthesis of 2,3-butanediol; *alsR* involved in the acetate-responsive LysR-type regulation; the flagella-mediated cytotoxin gene, *makA*; Type IV pilus genes *tcpA, B, C, D, E, F, H, I, J, N, P, Q, R, S,* and *T*, and *icmF/vask*; adherence genes *acfA, B, C, D, lfpA* as well as quorum sensing system genes *luxS*, and *cqsA*.

Pathogenicity islands were also present in all the strains, these were VSP-1, VSP-2, VPI-1, and VPI-2.

The sequenced strains all showed genotypic resistance to streptomycin, aminoglycosides, fosfomycin, fluoroquinolones, sulphonamides, trimethoprim, chloramphenicol/florfenicol, and tetracyclines (Table 2).

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

Furthermore, all the sequenced strains contained the VC1786 integrative and conjugative element genes. Also, antimicrobial resistance genes associated with sulfamethoxazole and trimethoprim, chloramphenicol and streptomycin resistance normally found on the VC1786ICE such as *sul2, floR, strA,* and *strB* were found present in the strains according to MyDbFinder 1.2.

In addition, all the sequenced strains had no plasmids according to PlasmidFinder 2.0, MyDbFinder 1.2 and Blast atlas.

SNP-based phylogenetic analysis showed an overall SNP difference of 120 among the 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).

Analysis for shared SNPs among the sequences showed that the sequenced strains shared 218 SNPs, of these, 98 SNPs were missense. Gene enrichment analysis of the 98 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.

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.

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. 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; also, the finding of this study in regards to most of the isolates except one 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 (5). Unlike a similar study done in the East African region (5) in which MLST revealed that their isolates belonged to a single sequence type (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 (9).

The virulence genes reported in this study are similar to those reported in similar studies elsewhere (5,10–12), these genes included among others those belonging to the Type IV secretion system, those involved in adherence, Type IV pilus genes as well as 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, these were the VSP-1, VSP-2 as well as VPI-1 and VPI-2 (13,14). These have not only been documented to encode virulence-associated genes in *V. cholerae* but have been reported to facilitate the better understanding of the evolutionary events that lead to the emergence of pathogenic *V. cholerae* clones (13,14).

Despite the World Health Organization (WHO) recommendations (https://www.who.int/topics/cholera/publications/WHO_CDD_SER_91_15/en/) in regards to the management of cholera with oral rehydration salts in addition to antibiotics namely streptomycin, aminoglycosides, trimethoprim, fosfomycin, fluoroquinolones, sulphonamides, chloramphenicol/florfenicol, and tetracyclines, this study reports genotypic resistance in the isolates to the same antibiotics. Similar resistance has been reported in similar studies done in the East African region and elsewhere (5,15–17).

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 (15,17).

This study found no plasmids present in the isolates, this could have been 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 (5).

The presence of Type IV secretion system, T6SS genes in the study isolates could explain their genotypic antimicrobial resistance gene profile. T6SS-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; 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 (18).

The results obtained from the SNP-based phylogenetic analysis show the closeness of the Ugandan *V. cholerae* strains, these are in agreement with the results 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 (19–21). Quorum sensing has been documented to regulate the expression of virulence factors in *V. cholerae* (19–21).

Limitations

This study did not provide a complete draft genome of V. cholerae, this is because the available data constituted of only short paired end sequence reads. A combination of short and long read sequence data (PacBio and Oxford Nanopore) would have enable the construction of a draft complete genome.

List Of Abbreviations

List of abbreviations

IS- Insertion sequence

SRA- Sequence Read Archive

CDS- Coding DNA sequences

WGS- Whole-genome sequencing

WHO- World Health Organization

MLST- Multilocus sequence typing

SNP- Single nucleotide polymorphism

ICE- Integrative and Conjugative element

NCBI- National Centre for Biotechnology Information

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publish

Not applicable

Availability of data and material

The genomic sequence data analyzed here is freely available and was deposited by the University of Maryland (Baltimore, MD, United States) at the SRA database of the NCBI under a study titled, “Molecular characterization of *Vibrio cholerae* responsible for cholera epidemics in Uganda by PCR, MLVA, and WGS”.

Competing interests

The authors declare that they have no competing interests.

Funding

There was no funding for this research

Authors' contributions

IS, DA, and GM contributed equally in conceptualization, data curation, formal analysis, methodology, validation, visualization, writing of the original draft, and review as well as editing the final manuscript.

Acknowledgments

Gerald Mboowa is supported through the DELTAS Africa Initiative (Grant no. 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 (Grant no. 107742/Z/15/Z) and the UK Government

References

1. Bwire G, Orach CG, Abdallah D, Debes AK, Kagirita A, Ram M, et al. Alkaline peptone water enrichment with a dipstick test to quickly detect and monitor cholera outbreaks. *BMC Infect Dis.* 2017;17(1):726.
2. Mengel MA, Delrieu I, Heyerdahl L, Gessner BD. Cholera outbreaks in Africa. In: *Cholera outbreaks.* Springer; 2014. p. 117–44.
3. Kasolo F, Yoti Z, Bakyaita N, Gaturuku P, Katz R, Fischer JE, et al. IDSR as a platform for implementing IHR in African countries. *Biosecurity bioterrorism biodefense Strateg Pract Sci.* 2013;11(3):163–9.
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.
5. Hounmanou YMG, Leekitcharoenphon P, Hendriksen RS, Dougnon T V, Mdegela RH, Olsen J, 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. p. 52.
6. Bwire G, Sack DA, Almeida M, Li S, Voeglein JB, Debes AK, 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.
 7. Aruhomukama D SI and MG. Investigating colistin drug resistance: The role of high-throughput sequencing and bioinformatics. *F1000Research*. 2019;8(150).
 8. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017;13(6):e1005595.
 9. Fu S, Hao J, Jin S, Wu K, Wang Y, Ye 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.
 10. Ceccarelli D, Garriss G, Choi SY, Hasan NA, Stepanauskas R, Pop M, et al. Characterization of Two Cryptic Plasmids Isolated in Haiti from Clinical *Vibrio cholerae* Non-O1/Non-O139. *Front Microbiol*. 2017;8:2283.
 11. Aliabad NH, Bakhshi B, Pourshafie MR, Sharifnia A, Ghorbani M. Molecular diversity of CTX prophage in *Vibrio cholerae*. *Lett Appl Microbiol*. 2012;55(1):27–32.
 12. Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL, Mekalanos JJ. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc Natl Acad Sci*. 2002;99(5):3129–34.
 13. Faruque SM, Mekalanos JJ. Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol*. 2003;11(11):505–10.
 14. 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.
 15. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, 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.
 16. Hounmanou YMG, Mdegela RH, Dougnon T V, Mhongole OJ, Mayila ES, Malakalinga J, 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.
 17. Kaas RS, Ngandjio A, Nzouankeu A, Siriphap A, Fonkoua M-C, Aarestrup FM, 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.

18. Borgeaud S, Metzger LC, Scignari T, Blokesch M. The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. *Science* (80-). 2015;347(6217):63–7.
19. Rutherford ST, Bassler BL. Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb Perspect Med*. 2012;2(11):a012427.
20. Parker CT, Sperandio V. Cell-to-cell signalling during pathogenesis. *Cell Microbiol*. 2009;11(3):363–9.
21. 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.

Tables

Table 1: Biosample data, Serotype and Genomic sequence data of the *V. cholerae* strains

<i>e ID</i>	<i>VC serotype</i>	<i>Year of collection</i>	<i>Month of collection</i>	<i>District of origin</i>	<i>Coverage</i>	<i>Genome size (bp)</i>	<i>No. of contigs</i>	<i>No. of CDS</i>
1251	<i>Inaba</i>	2014	<i>Apr</i>	<i>Arua</i>	155	4025190	88	3733
1254	<i>Inaba</i>	2014	<i>May</i>	<i>Moyo</i>	240	4025367	85	3743
1247	<i>Ogawa</i>	2015	<i>Apr</i>	<i>Kasese</i>	200	4012560	88	3697
1245	<i>Inaba</i>	2015	<i>Apr</i>	<i>Kasese</i>	260	4030572	96	3734
1253	<i>Inaba</i>	2015	<i>Jul</i>	<i>Arua</i>	200	4024003	89	3743
1248	<i>Inaba</i>	2015	<i>May</i>	<i>Kasese</i>	260	4034890	102	3745
1246	<i>Inaba</i>	2015	<i>Sep</i>	<i>Hoima</i>	250	4032599	90	3734
1252	<i>Inaba</i>	2015	<i>Sep</i>	<i>Hoima</i>	200	4025376	87	3731
1249	<i>Ogawa</i>	2016	<i>Jan</i>	<i>Mbale</i>	300	3998859	82	3688
1250	<i>Ogawa</i>	2016	<i>Jan</i>	<i>Mbale</i>	250	4011573	89	3700

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

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

Figures

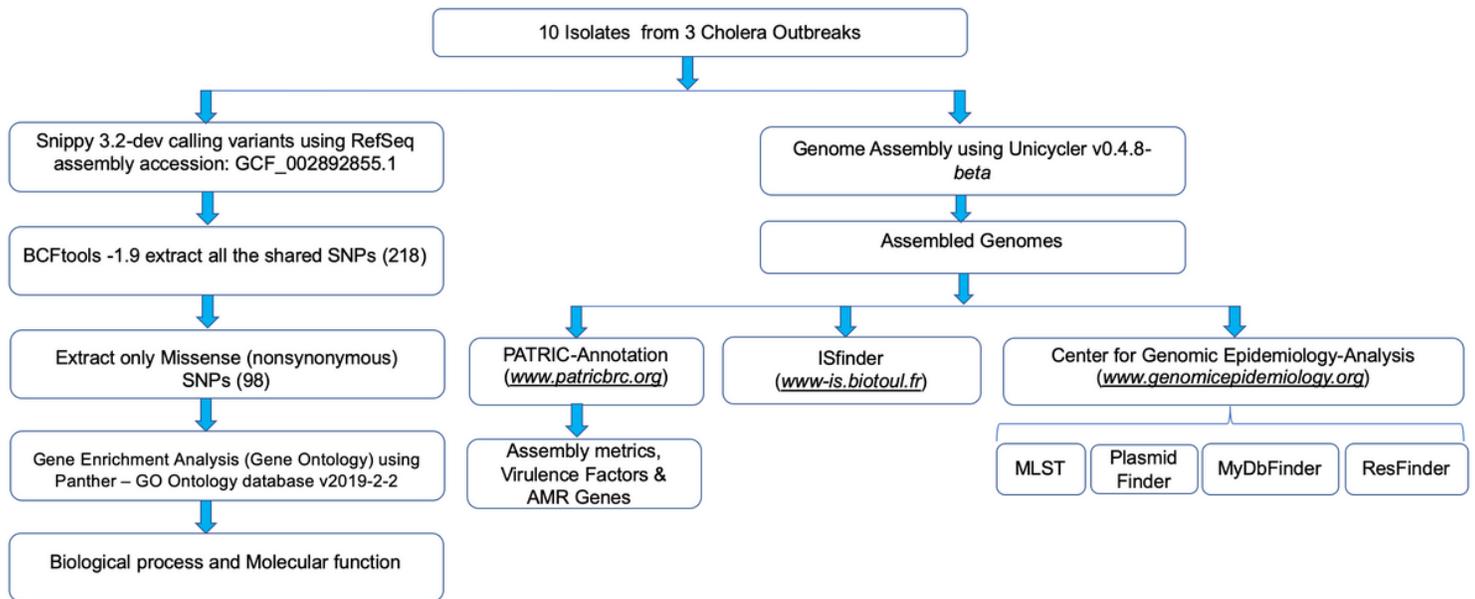


Figure 1

Overview of the bioinformatics workflow. Each box represents a component corresponding to a series of tasks that provide a certain well-defined output. Major bioinformatics utilities employed in each module are also mentioned. In the analysis, untrimmed sequence data quality reports were generated with FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v1.7 (<https://multiqc.info/>) using default settings.