RESEARCH ARTICLE

Whole genome sequence of *Vibrio cholerae* directly from dried spotted filter paper

Angèle H. M. Bénard[1], **Etienne Guenou**[2,3], **Maria Fookes**[1], **Jerome Ateudjieu**[2,4,5], **Watipaso Kasambara**[6], **Matthew Siever**[7], **Stanislas Rebaudet**[8,9], **Jacques Boncy**[10], **Paul Adrien**[11], **Renaud Piarroux**[12], **David A. Sack**[7], **Nicholas Thomson**[1,13], **Amanda K. Debes**[7]*

1 Wellcome Trust Sanger Institute, Genome campus, Hinxton United Kingdom, 2 M.A. SANTE (Meilleur Accès aux Soins de Santé), Yaoundé, Cameroon, 3 Department of Microbiology and Parasitology, Faculty of Science, University of Buea, Buea, Cameroon, 4 Department of Public Health, Faculty of Medicine and Pharmaceutical Sciences, University of Dschang, Cameroon Dschang Cameroon, 5 Clinical Research Unit, Division of Health Operations Research, Ministry of Public Health, N˚8, quartier du Lac (Yaoundé III), Cameroon, 6 Ministry of Health, Lilongwe, Malawi, 7 John Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, 8 Assistance Publique–Hôpitaux de Marseille (APHM), Marseille, France, 9 Hôpital Européen, Marseille, France, 10 National Laboratory of Public Health in Haiti (LNSP), Ministry of Public Health and Population, Haiti, 11 Directorate for Epidemiology, Laboratory and Research, Ministry of Public Health and Population, Haiti, 12 Sorbonne Université, INSERM, Institut Pierre-Louis d'Epidémiologie et de Santé Publique, APHP, Hôpital Pitié-Salpêtrière, Paris, France, 13 London School of Hygiene and Tropical Medicine, Keppel St, Bloomsbury, London WC1E 7HT, United Kingdom

* adebes1@jhu.edu

Abstract

Background

Global estimates for cholera annually approximate 4 million cases worldwide with 95,000 deaths. Recent outbreaks, including Haiti and Yemen, are reminders that cholera is still a global health concern. Cholera outbreaks can rapidly induce high death tolls by overwhelming the capacity of health facilities, especially in remote areas or areas of civil unrest. Recent studies demonstrated that stool specimens preserved on filter paper facilitate molecular analysis of *Vibrio cholerae* in resource limited settings. Specimens preserved in a rapid, low-cost, safe and sustainable manner for sequencing provides previously unavailable data about circulating cholera strains. This may ultimately contribute new information to shape public policy response on cholera control and elimination.

Methodology/Principal findings

Whole genome sequencing (WGS) recovered close to a complete sequence of the *V. cholerae* O1 genome with satisfactory genome coverage from stool specimens enriched in alkaline peptone water (APW) and *V. cholerae* culture isolates, both spotted on filter paper. The minimum concentration of *V. cholerae* DNA sufficient to produce quality genomic information was 0.02 ng/μL. The genomic data confirmed the presence or absence of genes of epidemiological interest, including cholera toxin and pilus loci. WGS identified a variety of diarrheal pathogens from APW-enriched specimen spotted filter paper, highlighting the potential for this technique to explore the gut microbiome, potentially identifying co-

infections, which may impact the severity of disease. WGS demonstrated that these specimens fit within the current global cholera phylogenetic tree, identifying the strains as the 7th pandemic El Tor.

Conclusions

WGS results allowed for mapping of short reads from APW-enriched specimen and culture isolate spotted filter papers. This provided valuable molecular epidemiological sequence information on *V. cholerae* strains from remote, low-resource settings. These results identified the presence of co-infecting pathogens while providing rare insight into the specific *V. cholerae* strains causing outbreaks in cholera-endemic areas.

Author summary

Cholera affects more than 4 million people globally every year; people predominantly living in poverty or in resource-constrained conditions including political crises or natural disasters. Cholera's typical presentation is characterized by rapid onset of acute watery diarrhea and vomiting which can progress from watery stool to shock in as little as four hours. Laboratory conditions needed for culture confirmation and strain preservation are rarely to never present in these affected areas. In fact, many cholera endemic areas in Sub-Saharan African are so remote that even treatment response alone is often challenging. Here we present the genomic analysis of DNA extracted from dried filter paper, which is a low-cost, low-tech and sustainable method. Previously this method has facilitated cholera confirmation by PCR, but we demonstrate that this method is also suitable for whole genome sequencing and subsequent strain characterization by presenting the analysis of samples from an outbreak in a remote area of Cameroon. This method will facilitate the understanding of the molecular epidemiology in cholera-prone areas, which were previously too challenging to attempt. It also introduces a method that can be used on a broader scale for diarrheal disease surveillance, including providing a window into co-infection and microbiome analyses.

Introduction

Global estimates for moderate to severe diarrhea are estimated to account for 1.6 million deaths annually worldwide and total a burden close to 75 million disability-adjusted life years (DALY) with costs approximating 3.11 billion USD in 2010 [1]. Recent studies have demonstrated the value of whole genome sequencing (WGS) in understanding the molecular evolution and transmission of the etiologic agents that cause moderate to severe diarrhea, including *Vibrio cholerae*. An analysis of > 700 *V. cholerae* isolate genomes originating from Asia, Africa, Latin America and the Caribbean spanning a period of more than half a century demonstrated that epidemics of cholera in Africa and the Americas stem from the introduction of a single pandemic lineage from Africa and South Asia [2]. This understanding of cholera was only possible because WGS data provided phylogenetically robust measures of relatedness. This analysis revealed that the currently circulating seventh pandemic of cholera can largely be attributed to a closely related genetic sublineage known as the seventh pandemic El Tor lineage (7PET) that forms a single branch within the diverse *V. cholerae* species phylogeny.

Over the past 20 years, WGS has improved our understanding of the molecular evolution and transmission of the etiologic agents that cause moderate to severe diarrhea, including *V. cholerae* [3]. Molecular epidemiology has become critical to determine the areas at highest risk for infections, to improve intervention measures such as vaccination campaigns but also to inform the development of new vaccines targeting the appropriate species and genotypes [4–6]. In addition, molecular epidemiological data also provides the information needed to combat the growing threats posed by antimicrobial resistance (AMR) spread [7]. Molecular epidemiology has played an important role in recent years for cholera. The OCV stockpile was created in 2013 and the need to target the doses available to those at greatest risk is important due to limits in vaccine availability[8–10]. Cholera often strikes in remote or resource-poor settings where laboratory capacity is limited, and therefore, it is not always possible to culture specimens from stool. Moreover, the time necessary from specimen collection to isolate preservation can take up to five days, if culture capability is available on site, and up to several weeks if specimens have to be stored on Cary Blair and subsequently transported to a central facility. In addition, specimen storage facilities are not available in every laboratory, and biosafety shipping of infectious strains to specialized genotyping laboratories may be very challenging from cholera-affected countries. Hence, the need for alternative approaches is critical to facilitate access to genomic information from samples collected in all cholera outbreak areas, particularly remote and resource-poor areas.

The primary analysis in this paper is based upon samples collected during a cholera surveillance and response program in the remote area of the Far North Region in Cameroon (FNC), which established and validated the use of dried filter paper for stool specimen preservation [11]. During an outbreak in the fall of 2014, cholera cases were primarily reported from an island in Lake Chad where specimen preservation is generally not possible due to lack of laboratory capacity and lack of cold chain transport. After prior stool enrichment in alkaline peptone water (APW), the enriched specimens were either directly spotted onto filter paper (APW-enriched specimen spotted filter paper) or cultured overnight, from which a single colony was picked and spotted onto filter paper (culture isolate spotted filter paper). The dried spot methodology, with or without prior culture step, was instrumental in providing the ability to preserve genomic material in order to characterize the outbreak. The study demonstrated that dried specimen spotted filter papers can be stored for up to 2 years at room temperature prior to DNA extraction and PCR amplification [12]. Once extracted, the DNA was analysed using multi-locus variable-number tandem repeat analysis (MLVA) and demonstrated that advanced molecular DNA methods could be used on dried specimen spotted filter papers preserved samples [12]. A similar study was performed in 2012 during a cholera epidemic in Sierra Leone, where watery diarrhea was directly sampled on filter paper without prior enrichment in APW, and stored for nearly three years at room temperature before successful MLVA genotyping [13]. Given the limitations of MLVA in providing detailed high-resolution molecular epidemiological features, we sought to determine if the same spotted filter paper material could also be used to perform WGS. This is a proof of principle study demonstrating that DNA extracted from simple, low-cost, APW-enriched and culture isolate spotted filter paper can generate high-quality accurate sequence data that has the potential to inform public health decisions by providing essential information on cholera genotype and co-infection burden.

Methods

Ethics statement

APW-enriched specimen and culture isolate spotted filter papers included in the study were isolated from participants enrolled in the "Sustainable Cholera Surveillance for Cameroon"

project. The Johns Hopkins Bloomberg School of Public Health Institutional Review Board reviewed and approved this study, IRB No. IRB00003981. Written informed consent was obtained from each participant or their caretaker prior to initiation of study activities.

Epidemiology and site description

The surveillance methodology, specimen collection, laboratory testing and findings have been previously reported [11,12]. Sixty-five isolates from two distinct outbreaks in the Far North of Cameroon were collected during this time.

APW-enriched and culture isolate spotted filter paper samples

Cameroonian stool specimens from 65 patients tested positive for *V. cholerae* by Crystal VC[TM] dipstick kit (Arkray Healthcare Pvt Ltd.,Surat, India) and subsequently culture-confirmed. Of these specimens, only 16 were processed according to two different protocols, called hereafter APW-enriched specimen spotted filter paper and culture isolate spotted filter paper.

The specimens referred to as APW-enriched specimen spotted filter papers are derived from stool samples enriched for 6-hours in 1X alkaline peptone water (APW) solution at room temperature. Two drops of the APW-enriched stool specimen were aliquoted onto two circles of a Whatman 903 filter paper card for preservation [12].

The specimens referred to as culture isolate spotted filter paper refers to the same 16 Cameroonian patients' specimens from which an isolate was able to be cultured. Briefly, the APW-enriched stool specimen was transferred via Cary Blair media to the main health facility for microbiological culture. The specimen was streaked onto TCBS medium overnight at 37˚C. From each TCBS culture, a single colony was selected, diluted in 50µL of phosphate-buffered saline (PBS) and aliquoted onto filter paper.

Viability of *V. cholerae* from spotted filter papers

To evaluate spotted filter paper as a new sample preservation method, the viability of *V. cholerae* after drying on filter paper blots was tested. *V. cholerae* serogroup O1 was grown in liquid culture to a confluence of $1x10^8$CFU/mL; 50µL of bacterial suspension was aliquoted onto Whatman filter paper and allowed to air dry overnight at room temperature (17h). Simultaneously, bacterial suspensions were aliquoted into 4 tubes for *Vibrio* viability experiments: heat-killing, ethanol-killing, bleach-killing and UV-light irradiation were evaluated for their sterilisation potential. After timed incubations with each potential killing agent, the bacterial suspension was spotted onto filter papers and allowed to air-dry overnight. A dried spot was excised from each filter paper and subsequently incubated in APW for 6 hours at 37˚C. Following enrichment, specimens were tested via Crystal VC dipstick to assess for any *V. cholerae* growth. Concurrently, each specimen was streaked on both TCBS and Luria Agar (LA) plates for overnight culture.

DNA preparation

For APW-enriched specimen and culture isolate spotted filter papers, a single spot of filter paper was excised at Johns Hopkins facilities, using scissors and inserted into a micro-centrifuge tube, washed twice with 1ml 1X PBS, and boiled with 200µL 1.5% Chelex solution for eight minutes. After a 1-minute centrifugation, the supernatant was transferred to a sterile micro-centrifuge tube. The presence of *V. cholerae* was confirmed via multiplex PCRs, first targeting an outer membrane protein, OmpW, in combination with primers targeting cholera toxin A, ctxA. A second PCR confirmed the presence of the rfb gene specific for the O1

serogroup following previously described methods [14,15]. To optimize for WGS following Chelex extraction, the DNA extracted from culture isolate spotted filter papers was subsequently purified by ethanol precipitation as described by Sambrook et al. [16]. The DNA was resuspended with ddH2O and sent for quantification by Qubit 2.0 Fluorometer (Thermofisher) and qPCR (StepOnePlus Real-Time PCR System) before WGS. Only samples with greater than a 0.001 ng/μLμl concentration of *V. cholerae* DNA were submitted for WGS.

WGS generation and analysis

WGS was performed at the Wellcome Trust Sanger Institute on an Illumina HiSeq 2500 platform to generate 100 bp paired-end reads. Short read data are available in the European Nucleotide Archive (ENA) database (S3 Table).

Sequence reads were mapped against reference genome *V. cholerae* O1 El Tor reference N16961 (accession numbers LT907989/LT907990) using SMALT v0.7.4 [17]. SMALT was used to index the reference using a k-mer size of 13 and a step size of 6 (-k 13 -s 6) and the reads were aligned using default parameters but with the maximum insert size (i) set as 3 times the mean fragment size of the sequencing library. PCR duplicate reads were identified using Picard v1.92 [18] and flagged as duplicates. A reference-based alignment was obtained by mapping paired-end Illumina reads to DNA sequenced from APW-enriched specimen spotted filter papers and isolate spotted filter papers to the *V. cholerae* O1 El Tor reference N16961. Automated annotation was performed using PROKKA v1.11 [19] and a genus-specific database from RefSeq [20].

Variation detection was performed using SamtoolsMpileup v0.1.19 [21] with parameters "-d 1000 -DSugBf" and bcftools v0.1.19 [22] to produce a BCF file of all variant sites. All bases were filtered to remove those with uncertainty in the base call. The bcftools variant quality score was required to be greater than 50 and the mapping quality to be greater than 30. If all reads did not give the same base call, the allele frequency, as calculated by bcftools, was required to be either 0 for bases called the same as the reference, or 1 for bases called as a single nucleotide polymorphism (SNP) (af1 < 0.95). The majority base call was required to be present in at least 75% of reads mapping at the base, (ratio < 0.75), and the minimum mapping depth required was 4 reads, at least two of which had to map to each strand (depth < 4, depth_strand < 2). Finally, strand bias was required to be less than 0.001, map bias less than 0.001 and tail bias less than 0.001. If any of these filters were not met, the base was called as uncertain. A pseudo-genome was constructed by substituting the base call at each site (variant and non-variant) in the BCF file into the reference genome and any site called as uncertain was substituted with an N. Insertions with respect to the reference genome were ignored and deletions with respect to the reference genome were filled with N's in the pseudo-genome to keep it aligned and the same length as the reference genome used for read mapping. Mapping was visualised with Artemis [23] and ACT [24].

Assemblies

Short reads from the Cameroonian samples were assembled *de novo* using SPAdes v3.10.0 [25], reordered against the reference sequence with ABACAS [26], and then a metaSPAdes [27] was performed with parameters—meta -t 8 -m 15. Statistics from assemblies were extracted with metaQUAST with parameters—fast -no-check [28]. Genome completeness estimates and checks for contamination were performed using CheckM lineage wf with the following parameters -t 8 -x fa—reduced tree [29]. Kraken v0.10.6 [30] was used to assign taxonomic labels using default parameters and the database Refseq release 72 (27/08/2015). Annotation was performed using the RAST server [31].

Phylogenetic analysis

A genome distance matrix was obtained by using MASH on SPAdes assemblies as previously described [32][5]. To accurately place our samples into a phylogenetic context, we supplemented our analyses with previously published genomes taken from Weill et al [2]. A Neighbor-Joining tree was generated based on the distance matrix using MASHtree [33]. An outgroup composed of M66, CNRVC960188 and CNRVC961190 was used to reroot the tree using Figtree. The resulting phylogenetic tree and corresponding metadata were visualized using Microreact [34].

Results

Stool samples collection, DNA extraction and sequence analysis of *V. cholerae* genome from spotted filter papers

Viability tests demonstrated that there was no viable *V. cholerae* after drying spotted filter paper overnight (17h), Viability was negative on all specimens evaluated via overnight culture on both TCBS and Luria Agar. This demonstrated that the specimens were no longer infectious for *V. cholerae* allowing safe shipping regarding biological risks.

Sixteen *V. cholerae* O1 positive specimen pairs were included in this study, comparing APW-enriched specimen spotted filter paper to culture isolate spotted filter paper, both derived from the same original stool specimen. As measured by RT PCR, total DNA concentration was on average 3 times higher when recovered from APW-enriched specimen spotted filter paper than from culture isolate spotted filter paper. Conversely the concentration of *V. cholerae* specific DNA was nearly 2.5 times higher from the culture isolate spotted filter paper compared to the APW-enriched specimen spotted on filter paper, as per the median and the mean reported in Table 1. In both cases, the quantity of DNA was sufficient to perform WGS (Table 1). Only two APW-enriched specimen spotted samples, 600064 and 600068, displayed higher *V. cholerae* DNA concentration than their respective culture isolate counterpart.

The mapping coverage for all samples ranged from 8.8x to 500x with an average of 128.4x (Fig 1A). The percentage of *V. cholerae* N16961 reference genome covered by reads ranged from 19.71% to 98.33% with an average of 68.44% (Fig 1B).

In both APW-enriched specimen spotted filter papers and culture isolate spotted filter papers, Spearman correlation test showed a positive correlation between the quantity of *V. cholerae* DNA and the percentage of *V. cholerae* genome covered by short reads in DNA extracted from both APW-enriched and culture isolate spotted filter papers (Spearman correlation score = 0.42, p < 0.1) (S2 Fig). A minimum concentration of 0.02 ng/μL of *V. cholerae* specific DNA in the sample generated more than 75% coverage when mapped against the reference genome. 43% of all spotted filter papers that showed successful mapping contained a concentration of *V. cholerae* DNA greater than 0.02 ng/μL. The percentage of the reference genome mapped appeared to plateau between a concentration of 0.2 and 0.3 ng/μL.

When comparing the mapping data according to the sample preparation protocol used, APW-enriched specimen spotted filter papers generated slightly higher mapping quality scores, specifically 148.7 ±144.03 mean depth and 70.52% ±26.81 reference genome covered compared to 105.86 ±81.4 mean depth and 66.14% ±25.52 reference genome covered for isolate spotted filter papers (Fig 1). The two protocols, APW-enriched specimen and culture isolate spotted filter paper were not statistically significant (Wilcoxon rank test, p-value = 0.5518 and p-value = 0.1965 for mean depth and reference genome covered respectively).

De novo assemblies using SPAdes produced assemblies with < 1000 contigs for 13 out of 16 isolate DNA samples and 6 out of the 16 APW-enriched specimen spotted filter papers. The

Table 1. *V. cholerae* enriched and isolate filter paper specimens and DNA quantity.

Sample Name	Sample Type	Quantity Mean of *V.cholerae* (pg/ul)	Quantity Mean of Total DNA (pg/ul)
500289	Enriched	1.49	340
500289	Isolate	11.12	130
500291	Enriched	9.48	760
500291	Isolate	13.13	190
600052	Enriched	11.22	1400
600052	Isolate	41.87	480
600055	Enriched	0.10	1130
600055	Isolate	12.79	380
600057	Enriched	6.13	760
600057	Isolate	33.29	1510
600058	Enriched	20.31	1520
600058	Isolate	46.42	580
600059	Enriched	18.94	4490
600059	Isolate	30.50	130
600060	Enriched	13.09	700
600060	Isolate	34.73	300
600061	Enriched	17.80	3040
600061	Isolate	46.14	250
600064	Enriched	28.94	740
600064	Isolate	26.29	350
600065	Enriched	1.10	1120
600065	Isolate	25.69	160
600066	Enriched	2.15	1060
600066	Isolate	4.54	350
600067	Enriched	27.95	1230
600067	Isolate	38.04	220
600068	Enriched	19.85	1290
600068	Isolate	10.80	940
600069	Enriched	3.48	890
600069	Isolate	55.58	110
600071	Enriched	6.07	1770
600071	Isolate	19.49	820
Median	Enriched	10.35	1125
Mean	Enriched	11.76	1390
Median	Isolates	28.4	325
Mean	Isolates	28.15	431.25

best assembly with less than 100 contigs including some contigs larger than 50000bp and covering more than 97% of the genome was obtained from DNA extracted from the APW-enriched specimen spotted filter paper for specimen 600057 (Table S3).

APW-enriched specimen spotted filter paper exhibited larger contigs than culture isolate spotted filter paper assemblies (Fig 2A). The range in results was higher for the APW-enriched specimen spotted filter papers compared to the culture isolate spotted filter paper, as illustrated by the wide range of reference genome fraction covered by the assemblies of APW-enriched specimen spotted filter papers, varying from 0.007% to 97% (Fig 2B).

MetaSPAde generated assemblies for 15 of the 16 APW-enriched specimen spotted filter papers and 16 out of 16 culture isolate spotted filter papers, while SPAde only produced 24
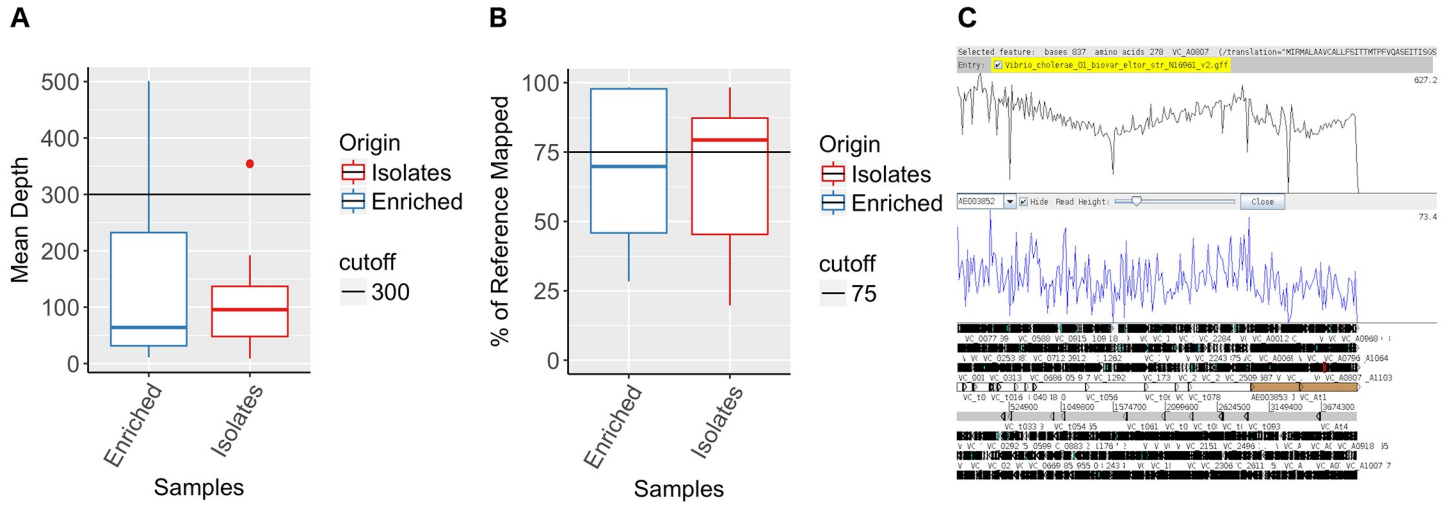
Fig 1. SMALT Mapping of short Illumina reads obtained from sequencing of DNA recovered from Whatman 903 filter cards. Average mean depth obtained from mapping short read Illumina sequences of DNA recovered from Whatman 903 cards (A). Percentage of *V. cholerae* reference genome N16961 covered by short Illumina reads mapped by SMALT (B). Artemis visualization of the short Illumina reads of sample 600066 mapped to the *Vibrio cholerae* reference genome N16961 (C).

assemblies out of 32 samples. One APW-enriched specimen sample, 500291, did not produce an assembly from either MetaSPAde analysis or SPAde analysis, likely due to limitations in the quantity of cholera-specific DNA available in the specimen. MetaSPAde analysis of APW-enriched specimen spotted filter papers allowed for the identification of the *V. cholerae* genome through sequence assemblies in a higher proportion of samples compared to culture isolate spotted filter papers.

Preliminary species diversity analyses of both APW-enriched specimen spotted filter papers and culture isolate spotted filter papers showed that *V. cholerae* reads represented one of the most abundant species in the majority of the samples (S3 Fig). Metagenomic analysis of APW-enriched specimen spotted filter paper showed the presence of a diverse microbial population (S3 Fig). As an example, bacteria belonging to the Enterobacteriaceae family such as various *Shigella* species or *Escherichia coli* strains, in addition to *V. cholerae* infection were identified in APW-enriched spotted filter papers. The use of MetaSPAde and MetaQUAST software allowed us to determine with higher accuracy the specific contribution of *V. cholerae* genome to the assemblies as well as the extent of bacterial diversity found in these samples (Fig 3). As expected, genomic diversity appears higher in enriched specimen spotted filter papers compared to culture isolate spotted filter papers. Similarly, *V. cholerae* genome is present in higher proportion in enriched specimen spotted filter papers compared to culture isolate spotted filter papers, which can be easily explained due to the lower quantity of DNA present in the isolate spotted filter papers (Fig 3).

Whole genome and phylogenetic analysis of *V. cholerae* genomes

Of the 32 APW-enriched specimen and culture isolate spotted filter papers samples sequenced, 20 failed to provide coverage above 50% or generate assemblies of greater than 1000 contigs. Therefore, these samples did not meet the criteria for further analysis and were excluded. Assemblies from both the APW-enriched specimen and culture isolate spotted filter papers samples of the same patient were simultaneously aligned with the *V. cholerae* reference genome N16961. Alignment comparison demonstrated that assemblies of all APW-enriched specimen spotted filter papers contained contigs outside of the *V. cholerae* genome and include sequences that show a high degree of similarity with other bacterial genomes (Fig 3).
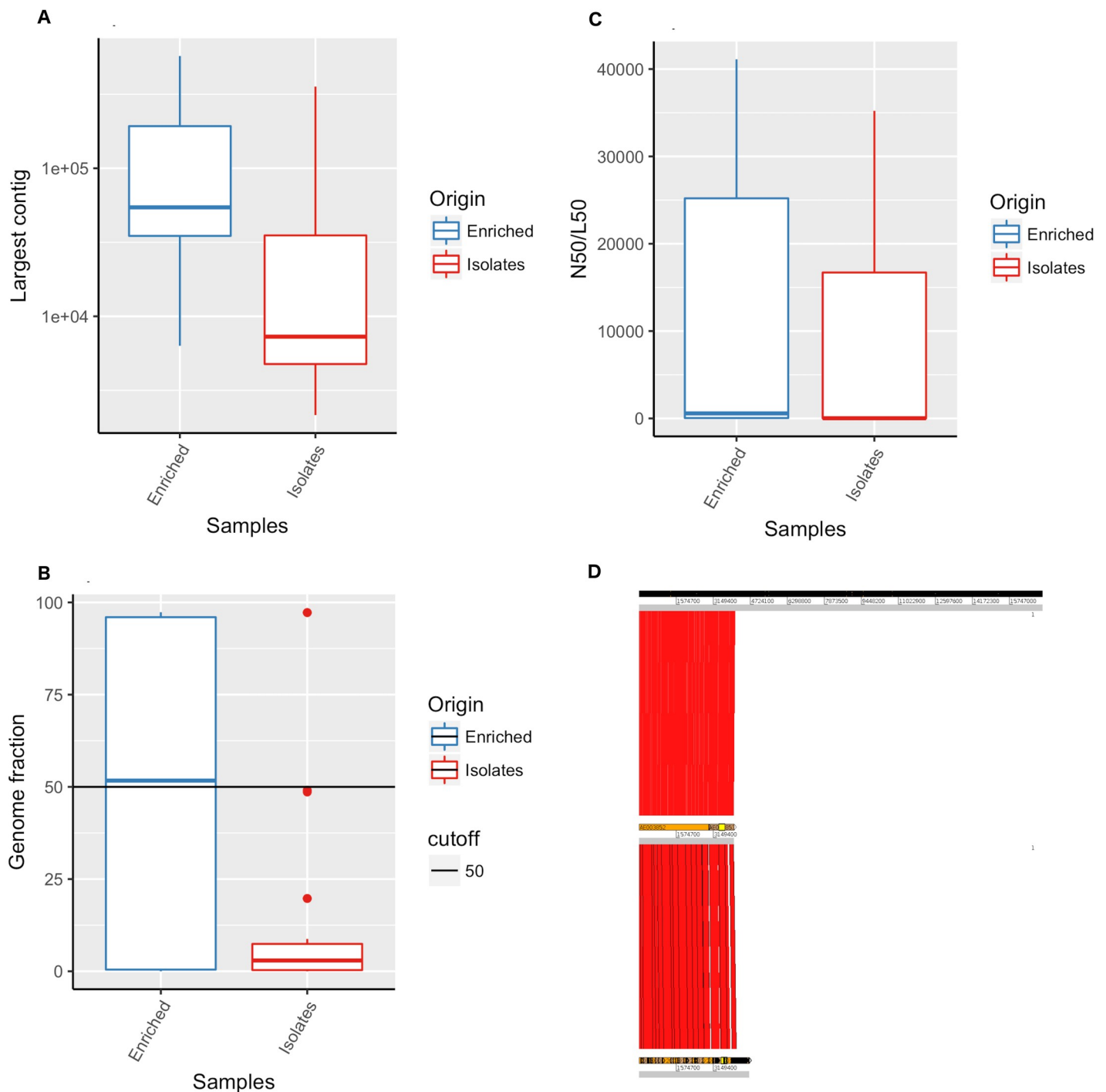
Fig 2. Assembly of genomes obtained from sequencing of DNA recovered from Whatman 903 filter paper cards spotted with APW-enriched specimens and culture derived isolates from patients infected with *Vibrio cholerae* O1. *De novo* spade assembly quality assessment obtained from short read Illumina sequences of DNA recovered from Whatman 903 filter paper cards. Largest contig (A), N50 (B) and percentage of genome fraction (C) of reference genome N16961 covered by metaSPAde assemblies from APW-enriched specimen spotted filter papers versus culture isolate spotted filter papers. ACT comparison of SPAde assemblies obtained from short read Illumina sequences of DNA recovered from Whatman 903 filter paper cards of APW-enriched specimen versus culture isolate spotted filter papers (D).
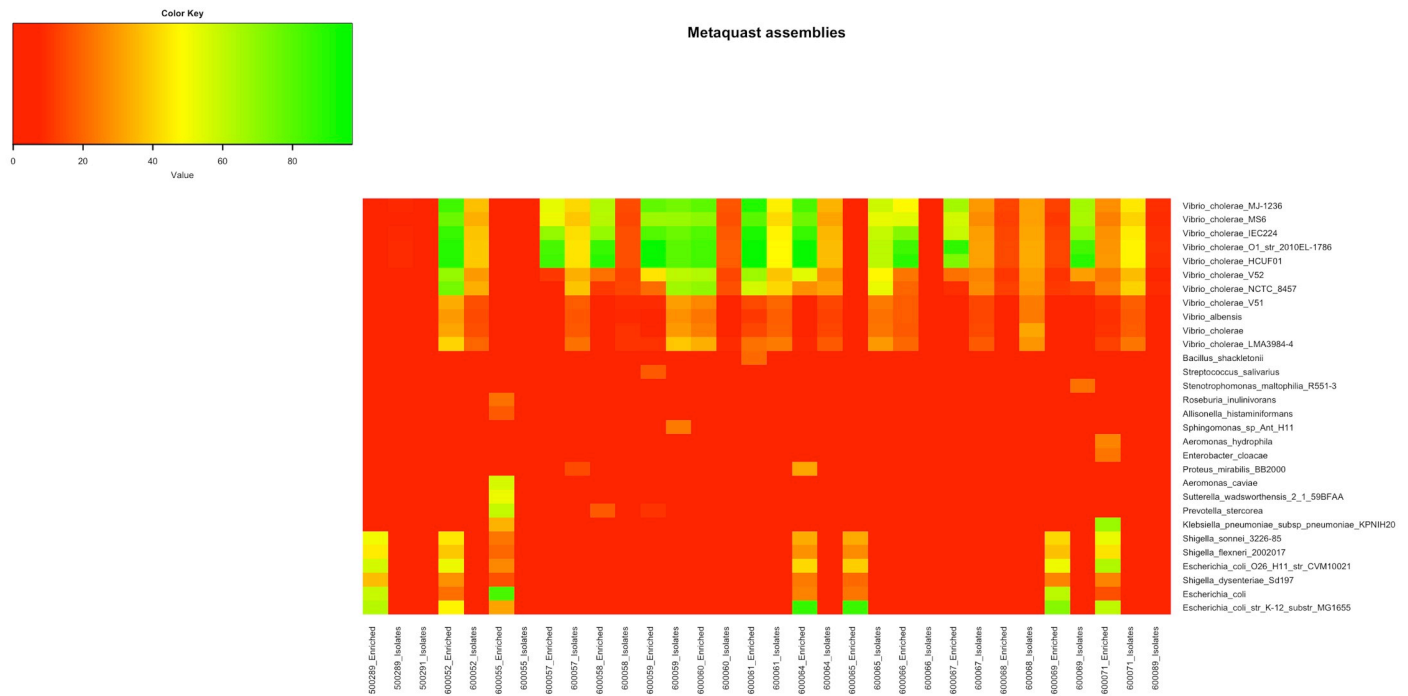
Fig 3. MetaSPAde assemblies of genomes obtained from sequencing of DNA recovered from Whatman 903 filter paper cards. Genomes fraction covered by metaSPAde genome assemblies obtained from short read Illumina sequences of DNA extracted from APW-enriched specimen spotted filter paper and culture isolate spotted filter papers.

The substantial level of coverage facilitated the alignment of short reads to *V. cholerae* O1 reference genome, confirming the molecular epidemiological characterization of these strains as *V. cholerae* O1. Further, several biologically relevant genes and genomic features of the *V. cholerae* genome could be identified in APW-enriched specimen spotted filter papers as well as in culture isolate spotted filter papers. Such examples of genes are *ctxA* and *ctxB* cholera toxin-encoded genes embedded into the integrated CTXΦ prophage; and *tcpA*, a gene of the *Vibrio* pathogenicity island. The strains were not part of O139 serogroup, demonstrated by the absence of genes including *rstA*, or *wfbA* of the *rfb* region (Fig 1C and S4 Fig)[35–38][35–38][34–37].

Variant calling was performed following SMALT mapping in order to identify single nucleotide variant sites (SNV) and sequence diversity within the *V. cholerae* genomes of each sample. Variant calling was not performed in samples with a DNA concentration below 0.02 ng/µL (S5A Fig).

Compiling all analyses, the highest quality samples were selected based on assembly criteria such as *V. cholerae* genome fraction covered ($>$ 50%), number of contigs ($>$ 100), largest contig ($>$ 5000), total assembly length ($>$ 2.2Mb) and NG50 $>$ 500 (S3 Table). Genome distance subsequently estimated using MASH on the eight best spade assemblies of this study were analysed in the context of sequences published by Weill et al. representing *V. cholerae* samples spanning over the past century [2]. This data was clustered using a Neighbour-Joining tree demonstrating the characteristic waves reflective of the global phylogeny of the 7th pandemic as shown in (Fig 4) and https://microreact.org/project/S1OfV91PG. The phylogenetic analysis confirmed the affiliation of the Cameroonian *V. cholerae* strains extracted from APW-enriched and culture isolate spotted filter papers to the 7th pandemic El Tor. Importantly, the
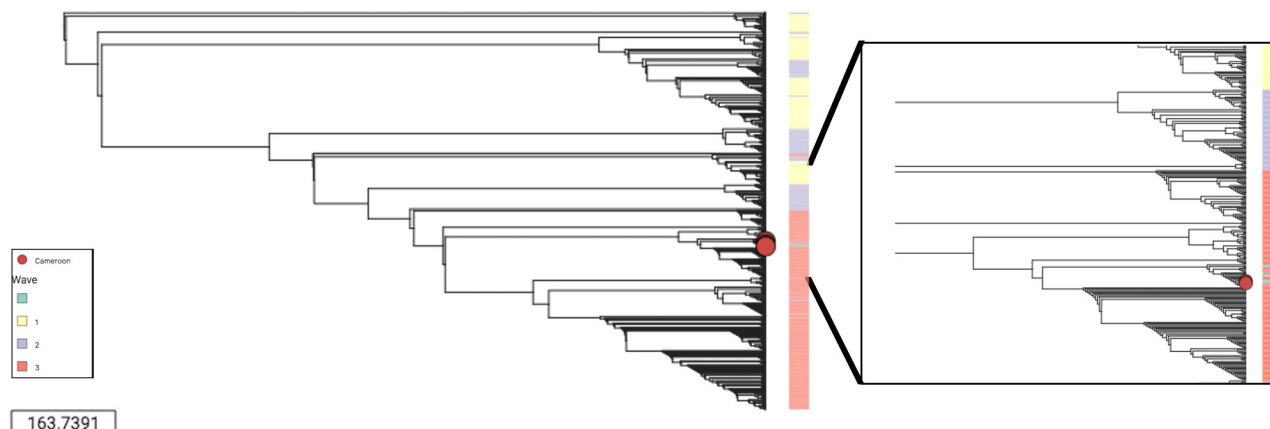
Fig 4. Neighbor-Joining tree representing MASH generated distance matrix based on high quality Spade assemblies.

samples fit within the third wave of the global phylogenetic tree of *V. cholerae*, in close proximity to other Cameroonian samples of recent years such as those dated from 2005, and from 2010 and 2011.

Based on these observations, we have concluded that the following quality criteria need to be fulfilled for correct phylogenetic interpretation, namely a proportion of reference genome covering greater than 50%, a mean depth greater than 20x, *V. cholerae* DNA concentration greater than 0.02 ng/μL, and examination of assemblies (NG50).

Discussion

In this study, we successfully sequenced DNA from two types of samples spotted onto filter papers for preservation. Not only were these preservation methods proven safe but effective for WGS quality standards after collection, storage and transport.

DNA was recovered from all Cameroonian spotted filter papers and WGS proved to be successful for all samples despite low quantity of DNA recovered. WGS results allowed for mapping short reads for the majority of APW-enriched specimen spotted filter papers and all but one of the culture isolate spotted filter papers. Despite high heterogeneity, the quality of the mapping for APW-enriched specimen spotted filter papers, when compared to culture isolate spotted filter papers, proved to be of variable but satisfactory quality. Quality was illustrated by several criteria including mean depth, proportion of reference genome covered, DNA concentration, and NG50. Mapping confirmed the identification of specific virulence genes and the absence of genes implicated in important biological pathways of *V. cholerae*, providing critical molecular epidemiological information to characterize cholera outbreaks in remote and/or unstable areas [39–41][38–40][39–41][35–37]. Furthermore, successful assemblies obtained from WGS of these samples were instrumental in identifying gene context and gene organisation within reconstituted genomes. Analyses suggested that the use of APW enrichment of stool rather than the more laborious selection of isolate spotted filter paper might be more efficient for reconstitution of *V. cholerae* assemblies. This data provides a source of information to develop informed experimental hypotheses that may reveal new biological mechanisms of *V. cholerae* bacteria.

The use of metagenomics software tools showed bacterial diversity in *V. cholerae* infected samples and highlighted the prospect for using spotted filter paper for routine metagenomics analysis. This finding highlights the presence of co-occurrences of potential gut bacterial

colonisation or co-infection with other diarrheal pathogens. Novel biological interaction mechanisms may also be explored at the bacterial population level, such as the complexity of the gut microbiota in cholera infected patients [42]. Extracting DNA from APW-enriched specimen spotted filter papers revealed the potential for studying multiple bacterial populations through WGS. The ability to study the diversity of bacterial populations from spotted filter papers will facilitate the study and understanding of the microbiome in low-resource settings, not only as it pertains to cholera.

Mapping based on quality criteria such as the proportion of reference genome covered, mean depth, original concentration of DNA, and high-quality assembly criteria such as NG50, number and size of contigs allowed us to restrict our analysis to high quality assemblies only. This high-quality assembly data could be used to understand the genetic distance between *V. cholerae* strains and place the analyzed sample within a general phylogenetic context in the global history of cholera transmission. Pairwise mutation distance-based clustering results confirmed the low level of diversity expected in *V. cholerae* clinical samples of an epidemic outbreak. Phylogeny is a critical tool that has been proven to contribute to characterizing outbreaks and to provide evidence for global and local transmission. This tool will be of specific value in remote and resource-constrained settings such as regions where cholera is endemic or regions with elevated risk of cholera epidemics. It will facilitate testing and verification of experimental hypotheses related to the biology of *V. cholerae* in controlled laboratory settings where the opportunity may not be otherwise possible. This is a result of difficulties in specimen processing and preservation in remote and austere settings where cholera is often endemic. However, with the increasing affordability of sequencing and the recent development of affordable and compact sequencing technologies, such as ISeq and Oxford Nanopore MinION, access to these technologies in countries at high-risk for cholera is increasing. Together the use of dried specimens in combination with more affordable resources in country may facilitate informed decision-making for a timely response to cholera outbreaks in remote and low-resource areas.

There are several areas of improvement to be considered in future studies. First, since sequencing was not the original intention of the specimen preservation, we did not preserve non-enriched/direct stool samples on the filter paper. Currently, we are working to collect crude, APW-enriched specimen and culture isolate spotted filter papers in tandem to facilitate the comparison of DNA quality as well as sequence results across all potential specimen preservation types to optimize the method most applicable in a low resource setting. Second, the DNA extraction method is an important limitation in the comparison of APW-enriched versus isolate filter paper sequences in this paper. The use of ethanol precipitation likely greatly reduced the final DNA concentration available for sequencing from the isolate spotted filter paper, therefore a direct extraction comparison is warranted. Efforts are currently underway to actively improve the quality and quantity of DNA extracted from filter paper cards through protocol refinement for all specimen types. Subsequent to this study, we have also employed duplicated spotting of specimens at all of our study sites as it has shown to be advantageous providing additional specimen available at minimal cost for these advanced molecular studies. Finally, the wide array of filter paper technologies available will present options for further consideration to determine the specific type of filter paper best for use in subsequent work to optimize DNA preservation on filter paper.

In conclusion, we present a proof of concept for WGS of DNA extracted from APW-enriched specimen and culture isolate spotted filter papers specifically targeted, but not limited, to *V. cholerae* strains preserved on dried filter papers. We have determined the minimum methodological requirements allowing for successful WGS that would allow for the retrieval of biologically relevant genomic information. Until sequencing is widely accessible and affordable, the optimization of this method provides high-level molecular information at low cost

and limited difficulty to countries at-risk. In conjunction with new sequencing technologies that may soon be available in low-resource settings, we may soon understand transmission patterns in-real time rather than post-outbreak characterization. The optimization of filter paper preservation for WGS will pave the way towards a better understanding of *V. cholerae* transmission and outbreak dynamics globally.

Supporting information

S1 Fig. Protocol summary of *Vibrio cholerae* DNA harvest from specimens preserved on filter paper cards.
(JPEG)

S2 Fig. The quantity of DNA recovered from Whatman 903 filter cards spotted with APW-enriched specimens and culture isolates from patients infected with *Vibrio cholerae* O1 is critical for the quality of mapping. Positive Spearman correlation between the quantity of *Vibrio cholerae* DNA in the material recovered from the Whatman 903 filter cards and the mapping quality as measured by percentage of *Vibrio cholerae* reference genome N16961 covered by short Illumina reads mapped by SMALT and mean depth of short Illumina reads.
(TIF)

S3 Fig. Metagenomic analysis of DNA sequences recovered from Whatman 903 filter cards spotted with APW-enriched specimens and culture derived isolates from patients infected with *Vibrio cholerae* O1. Proportion of reads specific to one species over all reads obtained from Kraken analysis of short read Illumina sequences of DNA recovered from Whatman 903 filter cards of APW-enriched specimen spotted filter papers versus culture isolates spotted filter papers (Selective threshold > 0.1%)
(TIF)

S4 Fig. Artemis visualization of the short Illumina reads mapped to the *Vibrio cholerae* reference genome N16961 at the *TcpA* gene locus and at *RstA* gene locus.
(TIF)

S5 Fig. SNP calling from mapping Illumina short reads to reference genome N16961 from DNA recovered from Whatman 903 filter cards spotted with APW-enriched specimens and culture derived Isolates from patients infected with *Vibrio cholerae* O1. SNP calling based on SMALT mapping of short read Illumina sequences of DNA recovered from Whatman 903 filter cards of APW-enriched specimen versus culture isolate spotted filter papers samples (A). Comparison of SNP between APW-enriched specimen and culture isolate among samples with higher than 75% of *Vibrio cholerae* reference genome N16961 mapped(B) and among samples with higher than 50% of *Vibrio cholerae* reference genome N16961 mapped, higher than 0.02ng/µL *Vibrio cholerae* DNA and higher than 20x mean depth (C).
(TIF)

S1 Table. SMALT mapping statistics of short Illumina reads obtained from sequencing of DNA recovered from APW-enriched specimen and culture isolate spotted Whatman 903 filter papers.
(DOCX)

S2 Table. SPAde assembly Quast results of short Illumina reads obtained from sequencing of DNA recovered from APW-enriched specimen and culture isolate spotted Whatman 903 filter papers.
(DOCX)

S3 Table. Accession numbers of samples used in this study. All data is available at https://www.ebi.ac.uk/ena.
(DOCX)

Author Contributions

Conceptualization: Angèle H. M. Bénard, Maria Fookes, Stanislas Rebaudet, Jacques Boncy, Renaud Piarroux, David A. Sack, Nicholas Thomson, Amanda K. Debes.

Data curation: Angèle H. M. Bénard, Maria Fookes.

Formal analysis: Angèle H. M. Bénard, Maria Fookes.

Funding acquisition: David A. Sack, Nicholas Thomson.

Investigation: Angèle H. M. Bénard, Maria Fookes, Jerome Ateudjieu, Watipaso Kasambara, Stanislas Rebaudet, Jacques Boncy, Paul Adrien, Renaud Piarroux, David A. Sack, Amanda K. Debes.

Methodology: Angèle H. M. Bénard, Etienne Guenou, Maria Fookes, Jerome Ateudjieu, Watipaso Kasambara, Matthew Siever, Amanda K. Debes.

Project administration: Angèle H. M. Bénard, Etienne Guenou, Maria Fookes, Jerome Ateudjieu, Watipaso Kasambara, Stanislas Rebaudet, Paul Adrien, Renaud Piarroux, David A. Sack, Amanda K. Debes.

Resources: Angèle H. M. Bénard, Etienne Guenou, Maria Fookes, Jerome Ateudjieu, Watipaso Kasambara, Matthew Siever, Stanislas Rebaudet, Jacques Boncy, David A. Sack, Nicholas Thomson, Amanda K. Debes.

Software: Angèle H. M. Bénard, Nicholas Thomson.

Supervision: Angèle H. M. Bénard, Etienne Guenou, Jerome Ateudjieu, Matthew Siever, Stanislas Rebaudet, Renaud Piarroux, David A. Sack, Nicholas Thomson, Amanda K. Debes.

Validation: Angèle H. M. Bénard, Matthew Siever, Amanda K. Debes.

Visualization: Angèle H. M. Bénard, Amanda K. Debes.

Writing – original draft: Angèle H. M. Bénard, Matthew Siever, Amanda K. Debes.

Writing – review & editing: Angèle H. M. Bénard, Etienne Guenou, Maria Fookes, Jerome Ateudjieu, Matthew Siever, Stanislas Rebaudet, Paul Adrien, Renaud Piarroux, David A. Sack, Nicholas Thomson, Amanda K. Debes.

References

1. A New Way to Calculate Costs of Infectious Diseases [Internet]. [cited 2018 Mar 23]. Available from: http://www.centerforhealthsecurity.org/about-the-center/pressroom/press_releases/2013-05-07_ID_Cost_Calculator.html

2. Weill F-X, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, et al. Genomic history of the seventh pandemic of cholera in Africa. Science. 2017 Nov 10; 358(6364):785–9. https://doi.org/10.1126/science.aad5901 PMID: 29123067

3. Domman D, Quilici M-L, Dorman MJ, Njamkepo E, Mutreja A, Mather AE, et al. Integrated view of Vibrio cholerae in the Americas. Science [Internet]. 2017 Nov 10 [cited 2018 Dec 7]; 358(6364):789–93. Available from: http://science.sciencemag.org/content/358/6364/789 https://doi.org/10.1126/science.aao2136 PMID: 29123068

4. Albiger B, Revez J, Leitmeyer KC, Struelens MJ. Networking of Public Health Microbiology Laboratories Bolsters Europe's Defenses against Infectious Diseases. Front Public Health [Internet]. 2018 [cited 2019 Mar 1];6. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2018.00046/full

5. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2010 Jan 22; 327(5964):469–74. https://doi.org/10.1126/science.1182395 PMID: 20093474

6. Hasnain SE. Molecular epidemiology of infectious diseases: a case for increased surveillance. Bull World Health Organ. 2003; 81(7):474. PMID: 12973637

7. Dengo-Baloi LC, Semá-Baltazar CA, Manhique LV, Chitio JE, Inguane DL, Langa JP. Antibiotics resistance in El Tor Vibrio cholerae 01 isolated during cholera outbreaks in Mozambique from 2012 to 2015. PloS One. 2017; 12(8):e0181496. https://doi.org/10.1371/journal.pone.0181496 PMID: 28792540

8. Desai SN, Pezzoli L, Alberti KP, Martin S, Costa A, Perea W, et al. Achievements and challenges for the use of killed oral cholera vaccines in the global stockpile era. Hum Vaccines Immunother. 2017 04; 13(3):579–87.

9. Azman AS, Luquero FJ, Salje H, Mbaïbardoum NN, Adalbert N, Ali M, et al. Micro-Hotspots of Risk in Urban Cholera Epidemics. J Infect Dis. 2018 Aug 24; 218(7):1164–8. https://doi.org/10.1093/infdis/jiy283 PMID: 29757428

10. Lessler J, Moore SM, Luquero FJ, McKay HS, Grais R, Henkens M, et al. Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales. Lancet Lond Engl. 2018 12; 391(10133):1908–15.

11. Debes AK, Ateudjieu J, Guenou E, Ebile W, Sonkoua IT, Njimbia AC, et al. Clinical and Environmental Surveillance for Vibrio cholerae in Resource Constrained Areas: Application during a 1-Year Surveillance in the Far North Region of Cameroon. Am J Trop Med Hyg. 2016 Mar 2; 94(3):537–43. https://doi.org/10.4269/ajtmh.15-0496 PMID: 26755564

12. Debes AK, Ateudjieu J, Guenou E, Lopez AL, Bugayong MP, Retiban PJ, et al. Evaluation in Cameroon of a Novel, Simplified Methodology to Assist Molecular Microbiological Analysis of V. cholerae in Resource-Limited Settings. PLoS Negl Trop Dis. 2016 Jan 6; 10(1):e0004307. https://doi.org/10.1371/journal.pntd.0004307 PMID: 26735969

13. Rebaudet S, Moore S, Normand A-C, Koivogui L, Garnotel E, Jambai A, et al. Direct Dried Stool Sampling on Filter Paper for Molecular Analyses of Cholera. Am J Trop Med Hyg. 2016 Jul 6; 95(1):251–2. https://doi.org/10.4269/ajtmh.16-0246a PMID: 27385669

14. Nandi B, Nandy RK, Mukhopadhyay S, Nair GB, Shimada T, Ghose AC. Rapid method for species-specific identification of Vibrio cholerae using primers targeted to the gene of outer membrane protein OmpW. J Clin Microbiol. 2000 Nov; 38(11):4145–51. PMID: 11060082

15. Hoshino K, Yamasaki S, Mukhopadhyay AK, Chakraborty S, Basu A, Bhattacharya SK, et al. Development and evaluation of a multiplex PCR assay for rapid detection of toxigenic Vibrio cholerae O1 and O139. FEMS Immunol Med Microbiol. 1998 Mar; 20(3):201–7. https://doi.org/10.1111/j.1574-695X.1998.tb01128.x PMID: 9566491

16. Wood EJ. Molecular cloning. A laboratory manual by T Maniatis, E F Fritsch and J Sambrook. pp 545. Cold Spring Harbor Laboratory, New York. 1982. $48 ISBN 0-87969-136-0. Biochem Educ [Internet]. 1983 [cited 2019 Jan 18];11(2):82–82. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1016/0307-4412%2883%2990068-7

17. SMALT | Wellcome Sanger Institute [Internet]. [cited 2018 Mar 9]. Available from: http://www.sanger.ac.uk/science/tools/smalt-0

18. Picard Tools—By Broad Institute [Internet]. [cited 2018 Mar 9]. Available from: http://broadinstitute.github.io/picard/

19. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15; 30(14):2068–9. https://doi.org/10.1093/bioinformatics/btu153 PMID: 24642063

20. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 2012 Jan 1; 40(D1):D130–5.

21. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

22. Bcftools by samtools [Internet]. [cited 2018 Mar 9]. Available from: https://samtools.github.io/bcftools/

23. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinforma Oxf Engl. 2000 Oct; 16(10):944–5.

24. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: the Artemis comparison tool. Bioinformatics. 2005 Aug 15; 21(16):3422–3. https://doi.org/10.1093/bioinformatics/bti553 PMID: 15976072

25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol. 2012 May; 19(5):455–77.

26. ABACAS [Internet]. [cited 2018 Mar 9]. Available from: http://abacas.sourceforge.net/

27. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res [Internet]. 2017 Mar 15 [cited 2018 Mar 9]; Available from: http://genome.cshlp.org/content/early/2017/04/07/gr.213959.116

28. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies. Bioinforma Oxf Engl. 2016 01; 32(7):1088–90.

29. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015 Jul; 25(7):1043–55. https://doi.org/10.1101/gr.186072.114 PMID: 25977477

30. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014 Mar 3; 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46 PMID: 24580807

31. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008 Feb 8; 9:75. https://doi.org/10.1186/1471-2164-9-75 PMID: 18261238

32. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016 Jun 20; 17:132. https://doi.org/10.1186/s13059-016-0997-x PMID: 27323842

33. Katz L. mashtree: Create a tree using Mash distances [Internet]. 2018 [cited 2018 Mar 9]. Available from: https://github.com/lskatz/mashtree

34. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genomics. 2016 Nov; 2(11): e000093.

35. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, Sack DA. New Variants of Vibrio cholerae O1 Biotype El Tor with Attributes of the Classical Biotype from Hospitalized Patients with Acute Diarrhea in Bangladesh. J Clin Microbiol. 2002 Sep; 40(9):3296–9. https://doi.org/10.1128/JCM.40.9.3296-3299.2002 PMID: 12202569

36. Quinones M, Davis BM, Waldor MK. Activation of the Vibrio cholerae SOS Response Is Not Required for Intestinal Cholera Toxin Production or Colonization. Infect Immun. 2006 Feb; 74(2):927–30. https://doi.org/10.1128/IAI.74.2.927-930.2006 PMID: 16428736

37. Davis BM, Moyer KE, Boyd EF, Waldor MK. CTX Prophages in Classical Biotype Vibrio cholerae: Functional Phage Genes but Dysfunctional Phage Genomes. J Bacteriol. 2000 Dec; 182(24):6992–8. https://doi.org/10.1128/jb.182.24.6992-6998.2000 PMID: 11092860

38. Chatterjee SN, Chaudhuri K. Lipopolysaccharides of Vibrio cholerae: II. Genetics of biosynthesis. Biochim Biophys Acta BBA—Mol Basis Dis. 2004 Oct 14; 1690(2):93–109.

39. Ghosh R, Sharma NC, Halder K, Bhadra RK, Chowdhury G, Pazhani GP, et al. Phenotypic and Genetic Heterogeneity in Vibrio cholerae O139 Isolated from Cholera Cases in Delhi, India during 2001–2006. Front Microbiol. 2016; 7:1250. https://doi.org/10.3389/fmicb.2016.01250 PMID: 27555841

40. Roobthaisong A, Okada K, Htun N, Aung WW, Wongboot W, Kamjumphol W, et al. Molecular Epidemiology of Cholera Outbreaks during the Rainy Season in Mandalay, Myanmar. Am J Trop Med Hyg. 2017 Nov; 97(5):1323–8. https://doi.org/10.4269/ajtmh.17-0296 PMID: 28820711

41. Rahaman MH, Islam T, Colwell RR, Alam M. Molecular tools in understanding the evolution of Vibrio cholerae. Front Microbiol. 2015; 6:1040. https://doi.org/10.3389/fmicb.2015.01040 PMID: 26500613

42. Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, et al. Metagenomic profile of gut microbiota in children during cholera and recovery. Gut Pathog [Internet]. 2013 Feb 1 [cited 2018 Dec 18]; 5:1. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574833/ https://doi.org/10.1186/1757-4749-5-1 PMID: 23369162