Garrine, Marcelino; Mandomando, Inacio; Vubil, Delfino; Nhampossa, Tacita; Acacio, Sozinho; Li, Shan; Paulson, Joseph N; Almeida, Mathieu; Domman, Daryl; Thomson, Nicholas R; +2 more... Alonso, Pedro; Stine, Oscar Colin; (2017) Minimal genetic change in Vibrio cholerae in Mozambique over time: Multilocus variable number tandem repeat analysis and whole genome sequencing. PLOS NEGLECTED TROPICAL DISEASES, 11 (6). ISSN 1935-2735 DOI: https://doi.org/10.1371/journal.pntd.0005671

Downloaded from: http://researchonline.lshtm.ac.uk/4650923/

DOI: https://doi.org/10.1371/journal.pntd.0005671

Usage Guidelines:

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
Minimal genetic change in *Vibrio cholerae* in Mozambique over time: Multilocus variable number tandem repeat analysis and whole genome sequencing

Marcelino Garrine¹, Inácio Mandomando¹,², Delfino Vubil¹, Tacilta Nhampossa¹,², Sozinho Acacio¹,², Shan Li³, Joseph N. Paulson⁴,⁵, Mathieu Almeida⁴, Daryl Domman⁶, Nicholas R. Thomson⁶, Pedro Alonso¹,⁷, Oscar Colin Stine³*

¹ Centro de Investigación en Saúde de Manhica (CISM), Maputo, Mozambique, ² Instituto Nacional de Saúde (INS), Ministério da Saúde, Maputo, Mozambique, ³ Department of Epidemiology and Public Health, University of Maryland Baltimore, Baltimore, Maryland, United States of America, ⁴ Institute for Applied Computer Sciences, University of Maryland, College Park, Maryland, United States of America, ⁵ Department of Biostatistics, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, ⁶ Infection Genomics, Wellcome Trust Sanger Institute, Hinxton, England, United Kingdom, ⁷ ISGlobal Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clínic / Universitat de Barcelona, Barcelona, Spain

*ostin001@umaryland.edu

**Abstract**

Although cholera is a major public health concern in Mozambique, its transmission patterns remain unknown. We surveyed the genetic relatedness of 75 *Vibrio cholerae* isolates from patients at Manhica District Hospital between 2002–2012 and 3 isolates from river using multilocus variable-number tandem-repeat analysis (MLVA) and whole genome sequencing (WGS). MLVA revealed 22 genotypes in two clonal complexes and four unrelated genotypes. WGS revealed i) the presence of recombination, ii) 67 isolates descended monophyletically from a single source connected to Wave 3 of the Seventh Pandemic, and iii) four clinical isolates lacking the cholera toxin gene. This Wave 3 strain persisted for at least eight years in either an environmental reservoir or circulating within the human population. Our data raises important questions related to where these isolates persist and how identical isolates can be collected years apart despite our understanding of high change rate of MLVA loci and the *V. cholerae* molecular clock.

**Author summary**

Cholera is a deadly disease caused by the bacterium *Vibrio cholerae*. The ancestral home of cholera is around the Bay of Bengal, but recently cholera has moved to Africa. In Africa, cholera occurs in sporadic outbreaks. In order prevent cases of cholera, we want to understand the transmission of cholera in Africa, does it stay in one place or does it move around. To gain insight into these questions, we have examined the DNA of the bacteria. The DNA provides a identity for each isolate and we can infer how the isolates are related...
to each other based on the number and type of DNA changes. In our study, we examined the DNA of cholera isolates from southern Mozambique. We were surprised how similar all the Mozambique isolates were even though that were collected up to eight years apart. Based on previous work, we would have expected much more change in the DNA. Our data raises several important questions that relate to where these cholera isolates persist, possibly in local refuges, and how seemingly identical isolates can be collected years apart despite our understanding of the high rate of change of the molecular clock.

Introduction

Cholera remains a public health concern in developing countries with an estimated burden of 1.2–4.3 million cases and 28,000–142,000 deaths per year, worldwide [1]. South Asia and sub-Saharan Africa account for the majority of cases and deaths. Between 01 January and 03 June 2013, a total of 25,762 cholera cases and 490 deaths were reported from 18 African countries. Mozambique accounted for 7% (1,861/25,762) of cases and 4% (19/490) of deaths, being the third most affected country after the Democratic Republic of the Congo and Angola [2]. The most recent cholera outbreak in Mozambique started in December 2014 in Nampula [3], where there were 8,835 cases with case fatality rate of 0.7% (65 deaths) in 5 northern and central provinces in five months [4,5]. The most recent cholera outbreak in the south was reported in 2011 [6]. The peak of a cholera epidemic is often preceded by increasing prevalence of the pathogenic strains in the environment [7] where *Vibrio cholerae* are harbored in aquatic reservoirs during extended periods between outbreaks [8].

*Vibrio cholerae* O1 is associated with most epidemic and pandemic outbreaks [9]. Whole genome sequence (WGS) analysis of *V. cholerae* isolates from around the world [9–12] demonstrated that the current (Seventh) Pandemic is monophyletic and originated from a single source with a clonal expansion of the lineage, with a strong temporal signature [13]. The seventh pandemic has been divided into 3 waves beginning in 1952, 1981 and 1988, respectively [13]. Each wave appears to be a selective sweep, as was the second wave of *V. cholerae* that swept across Haiti after the initial introduction [14]. In Africa, isolates from all three waves have been identified [13] with wave 3 isolates forming two distinct clades in Kenya between 2005 to 2010 [15].

The epidemiology and transmission patterns of outbreaks of *V. cholerae* have been explored using both multilocus variable number tandem repeat analysis (MLVA) and WGS [13,16]. In rural Bangladesh, MLVA revealed that multiple genetic lineages of *V. cholerae* occur naturally in the environment with geographic and seasonal genetic variation and identical genotypes can be found in the environment and humans [12]. In Kenya, MLVA demonstrated that several distinct genetic lineages emerged simultaneously during outbreaks in a single cholera season, linked to local environmental reservoirs [17]. WGS revealed all lineages were part of the Seventh Pandemic expansion [15]. In central and western Africa, MLVA revealed distinct clusters of isolates from different countries: Democratic Republic of the Congo, Zambia, Togo, and Guinea [18]. A second study in Guinea demonstrated spread and differentiation of *V. cholerae* during an outbreak [19].

In Mozambique, previous analysis of *V. cholerae* O1 showed these isolates had the typical traits of the El Tor biotype overall except that they carried a tandem array of classical CTX prophage [20] located on the small chromosome [21,22]. However, there are no data available on the genetic relatedness of *V. cholerae* circulating in Manhiça in southern Mozambique. Here, we characterized clinical and environmental *V. cholerae* isolates from Mozambique using
MLVA and WGS to determine the genetic relatedness of strains isolated from patients with diarrhea in Manhiça District Hospital.

**Methodology**

**Site description**

Manhiça District Hospital (MDH) is a 110 bed referral health facility for Manhiça District, a rural area of Maputo Province in southern Mozambique. The characteristics of the area have been described in detail elsewhere [23,24]. Briefly, the climate is subtropical with two distinct seasons: a warm, rainy season between November and April, and a cool and dry season during the rest of the year. Manhiça has 160,000 inhabitants, who are mostly subsistence farmers or workers in two large sugar- and fruit-processing factories. The Manhiça Health Research Centre (Centro de Investigação em Saúde da Manhiça[CISM]) is adjacent to the MDH and has been conducting continuous demographic surveillance for vital events and migrations since 1996 [23], currently covering 165,000 individuals.

**Ethics statement**

The strain collection of *V. cholerae* described in the study was isolated from cholera surveillance and other studies conducted in the Manhiça community by CISM approved by the National BioEthic Committee (CNBS). Any and all patient data were anonymized/de-identified. The IRB at University of Maryland School of Medicine approved the use of anonymized strains.

**Bacterial isolates**

A total of 75 *V. cholerae* isolates were collected from MDH, between 2002 and 2012. The strains were isolated from stool of patients admitted to the MDH with suspicion of cholera, presenting with watery diarrhea. In addition, three isolates collected from the Incomati River were included. *V. cholerae* isolates were identified by standard biochemical tests; and confirmed by API-20E biochemical test strips (bioMérieux SA, Marcy-l’Etoile, France). Serotypes were determined using commercially available poly- and mono-clonal slide agglutination antisera (Mast Group Ltd., Merseyside, UK) according to the manufacturer’s instructions. All the isolates were stored at -80˚C in tryptone soya broth (TSB) with 15% glycerol, and retrieved at the time for molecular characterization.

**Preparation of bacterial DNA**

A pure culture of *V. cholerae* was plated in Thiosulfate Citrate Bile Sucrose (TCBS) agar and incubated overnight at 37˚C. DNA was extracted using the Qiagen QIAamp DNA Mini Kit (Hilden, Germany). The DNA template was sent to the University of Maryland Baltimore, Baltimore, Maryland, USA for molecular typing by MLVA and WGS.

**MLVA**

DNA from each isolate was amplified by PCR using the conditions and primers previously described for 5 loci containing variable length tandem repeats [11]. The amplified products were separated and detected using a model 3730xl Automatic Sequencer (ABI) and their sizes were determined using internal lane standards (Liz600; ABI, Foster City, CA) with the Gene Mapper v4.0 program (ABI). Genotypes were determined according to the published formulas to calculate the number of repeats from the length of each allele and identify the alleles at the 5 loci [11]. The 5 loci, in order, are VC0147, VC0436–7 (intergenic), VC1650, VCA0171, and...
VCA0283; thus, the genotype 9,4,6,19,11 indicates that the isolate has alleles of 9, 4, 6, 19, and 11 repeats at the 5 loci, respectively. Relatedness of the strains was assessed by eBURSTv3 (http://eburst.mlst.net), in which genetically related genotypes were defined as those possessing at least 4 identical alleles of the 5 loci. An alternative analysis was performed using Network 2.x (http://www.fluxus-engineering.com/sharefaq.htm).

Whole genome sequencing

The DNA concentration was quantified by NanoDrop 2000 Spectrophotometer (Thermofisher Scientific, Waltham, MA, USA) and only specimens with sufficient concentration (n = 71) were submitted to WGS. DNA was prepared for Illumina sequencing using the KAPA High Throughput Library Preparation Kit (KapaBiosystems, Wilmington, MA). DNA was fragmented with the Covaris E210. Libraries were prepared using a modified version of manufacturer’s with-bead protocol (KapaBiosystems, Wilmington, MA). The libraries were enriched and barcoded by ten cycles of PCR amplification step with primers containing an index sequence seven nucleotides in length. The libraries were sequenced on a 100 bp paired-end run on an Illumina HiSeq2500 (Illumina, San Diego, CA).

The quality of the 101-base paired-end reads was confirmed using Fastqc (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Kmergenie (http://kmergenie.bx.psu.edu/) was used for choice of the best peak and the assembly was performed using the SPAdes software [25]. CSI Phylogeny 1.0a (http://cge.cbs.dtu.dk/services/CSIPhylgeny/) was used to generate a tree of genetic relatedness based on high quality nucleotide variants and then compared to V. cholerae O1 El Tor reference strain N16961 (NCBI accession numbers AE003852 and AE003853). Splitstree [26] was used to determine networks. Previous work in our lab demonstrated that this pipeline produces identical results to SMALT [27].

A high resolution SNP based phylogeny for the 67 7th pandemic strains was placed in context of a globally representative collection of 274 isolates (S1 Table) by mapping the reads to the V. cholerae O1 El Tor reference N16961 using SMALT (http://www.sanger.ac.uk/resources/software/smal) as previously described [28]. Gubbins [29] was used to simultaneously remove regions of high SNP density and putative recombination sites in the alignment and infer the phylogenetic tree. The pre-seventh pandemic strain M66 (NCBI accession numbers CP001233 and CP001234) was used to root the phylogenetic tree.

To accurately place the four non-01 V. cholerae into a phylogenetic context we calculated a core genome alignment of 1093 genes (1,055,747 bp) using Roary [30] from a set of diverse V. cholerae genomes along with genomes of the closely related species Vibrio metoecus and Vibrio parilis. The resulting alignment was used to reconstruct the phylogenetic relationship using RAxML v. 7.8.6 [31] under the GTR model with 100 bootstrap replicates. The resulting phylogenetic trees were visualized using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Among the 78 isolates (Table 1), 26 distinct genotypes were identified by MLVA using all five VNTR loci. The numbers of distinct alleles at loci VC0147, VC0437, VC1650, VCA0171, and VCA0283 were 10, 2, 4, 7 and 7, respectively.

Two clonal complexes of genetically related genotypes (each complex comprising genotypes that differed by an allelic change at a single locus) and four singleton genotypes unrelated to any other (differed by ≥2 loci) were identified (Fig 1). The first clonal complex (CC1) consisted of twenty of the twenty-six genotypes and comprised 91% (71/78) of the isolates (including all 3 isolates from river water). In CC1 the most common genotype, identified as the “founder” genotype (defined by eBURST as the genotype that differed from the largest number
Table 1. Vibrio cholerae isolates and their genetic characterization.

<table>
<thead>
<tr>
<th>specimen ID</th>
<th>sample</th>
<th>collection date</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total reads</th>
<th># Contigs</th>
<th>Total length</th>
<th>N50</th>
<th># predicted genes (&gt; = 300 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0005</td>
<td>stool</td>
<td>12-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1582246</td>
<td>124</td>
<td>4042725</td>
<td>235749</td>
<td>3232</td>
</tr>
<tr>
<td>0010</td>
<td>river</td>
<td>14-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1517086</td>
<td>153</td>
<td>4162168</td>
<td>235741</td>
<td>3345</td>
</tr>
<tr>
<td>0014</td>
<td>river</td>
<td>14-May-02</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>17042970</td>
<td>145</td>
<td>4041694</td>
<td>235689</td>
<td>3231</td>
</tr>
<tr>
<td>0015</td>
<td>stool</td>
<td>20-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1517086</td>
<td>116</td>
<td>4040728</td>
<td>168110</td>
<td>3229</td>
</tr>
<tr>
<td>0016</td>
<td>stool</td>
<td>20-May-02</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>12570276</td>
<td>126</td>
<td>4041525</td>
<td>235753</td>
<td>3234</td>
</tr>
<tr>
<td>0017</td>
<td>stool</td>
<td>20-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1695398</td>
<td>129</td>
<td>4039562</td>
<td>246625</td>
<td>3229</td>
</tr>
<tr>
<td>0018</td>
<td>stool</td>
<td>20-May-02</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>19</td>
<td>21</td>
<td>12570276</td>
<td>0</td>
<td>4041525</td>
<td>235753</td>
<td>3234</td>
</tr>
<tr>
<td>0019</td>
<td>stool</td>
<td>20-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1517086</td>
<td>116</td>
<td>4040728</td>
<td>168110</td>
<td>3229</td>
</tr>
<tr>
<td>0020</td>
<td>stool</td>
<td>20-May-02</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>12570276</td>
<td>0</td>
<td>4041525</td>
<td>235753</td>
<td>3234</td>
</tr>
<tr>
<td>0021</td>
<td>stool</td>
<td>20-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1695398</td>
<td>129</td>
<td>4039562</td>
<td>246625</td>
<td>3229</td>
</tr>
<tr>
<td>0022</td>
<td>stool</td>
<td>20-May-02</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>19</td>
<td>21</td>
<td>12570276</td>
<td>0</td>
<td>4041525</td>
<td>235753</td>
<td>3234</td>
</tr>
<tr>
<td>0023</td>
<td>stool</td>
<td>20-May-02</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1517086</td>
<td>116</td>
<td>4040728</td>
<td>168110</td>
<td>3229</td>
</tr>
</tbody>
</table>

(Continued)
of other genotypes at a single locus), comprised 41% (29/71) of the isolates. In this complex, the founder genotype radiated into 9 other genotypes, and 5 of those differentiated further. Two of the three river water isolates shared a common genotype with the clinical isolates, one of which was of the founder genotype. Interestingly we found isolates with identical MLVA genotypes up to 8 years apart. In general, isolates from the same year were genetically related, all four genotypes found in 2002 were single locus variants, as were the two genotypes in 2008, two in 2009, and three genotypes from 2010. In contrast to similarity within a year, in months (May 2002, March 2003, April 2003, May 2003, June 2010) where multiple cases presented with *V. cholerae*, we found that isolates belonged to more than one genotype.

The other *V. cholerae* isolates consisted of a second clonal complex CC2 and four singletons (Fig 1). CC2 had two genotypes and comprised only 4% (3/78) of the analyzed isolates. The four singleton genotypes contained one isolate each corresponding to 5% (4/78) of the isolates. Three of them were isolated in 2008 and the fourth in 2012.

---

**Table 1. (Continued)**

<table>
<thead>
<tr>
<th>specimen ID</th>
<th>sample</th>
<th>collection date</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total reads</th>
<th># Contigs</th>
<th>Total length</th>
<th>N50</th>
<th># predicted genes (≥ 300 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300506</td>
<td>stool</td>
<td>27-Apr-08</td>
<td>12</td>
<td>12</td>
<td>3</td>
<td>17</td>
<td>21</td>
<td>9274148</td>
<td>173</td>
<td>400231</td>
<td>339296</td>
<td>3217</td>
</tr>
<tr>
<td>302015</td>
<td>stool</td>
<td>16-Feb-09</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>23</td>
<td>10980664</td>
<td>118</td>
<td>4214484</td>
<td>246476</td>
<td>3393</td>
</tr>
<tr>
<td>302029</td>
<td>stool</td>
<td>5-Mar-09</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>11199308</td>
<td>174</td>
<td>4229711</td>
<td>274301</td>
<td>3394</td>
</tr>
<tr>
<td>326</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>20</td>
<td>15654426</td>
<td>124</td>
<td>4165971</td>
<td>246637</td>
<td>3348</td>
</tr>
<tr>
<td>347</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>15286606</td>
<td>146</td>
<td>4051922</td>
<td>246633</td>
<td>3327</td>
</tr>
<tr>
<td>374</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>8013166</td>
<td>141</td>
<td>4043250</td>
<td>299283</td>
<td>3225</td>
</tr>
<tr>
<td>375</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>19</td>
<td>21</td>
<td>17973118</td>
<td>173</td>
<td>401898</td>
<td>215270</td>
<td>3229</td>
</tr>
<tr>
<td>382</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>1726956</td>
<td>130</td>
<td>406077</td>
<td>173229</td>
<td>3234</td>
</tr>
<tr>
<td>398</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>16</td>
<td>21</td>
<td>17282880</td>
<td>443</td>
<td>4102857</td>
<td>114507</td>
<td>3259</td>
</tr>
<tr>
<td>399</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>15570318</td>
<td>99</td>
<td>4038494</td>
<td>274422</td>
<td>3227</td>
</tr>
<tr>
<td>420</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>18048184</td>
<td>161</td>
<td>4040168</td>
<td>225617</td>
<td>3234</td>
</tr>
<tr>
<td>505</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>16442290</td>
<td>119</td>
<td>4047614</td>
<td>246645</td>
<td>3234</td>
</tr>
<tr>
<td>510</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>16619422</td>
<td>117</td>
<td>4041499</td>
<td>216033</td>
<td>3229</td>
</tr>
<tr>
<td>511</td>
<td>stool</td>
<td>NA</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>19004672</td>
<td>136</td>
<td>4043523</td>
<td>235737</td>
<td>3233</td>
</tr>
<tr>
<td>655630.3</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>14338294</td>
<td>178</td>
<td>4070175</td>
<td>246637</td>
<td>3247</td>
</tr>
<tr>
<td>655664.3</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>22</td>
<td>13650350</td>
<td>142</td>
<td>4175603</td>
<td>166703</td>
<td>3357</td>
</tr>
<tr>
<td>655665.0</td>
<td>stool</td>
<td>NA</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td>17379870</td>
<td>113</td>
<td>4044785</td>
<td>246633</td>
<td>3235</td>
</tr>
<tr>
<td>710180.8</td>
<td>stool</td>
<td>NA</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>16</td>
<td>22</td>
<td>16589666</td>
<td>111</td>
<td>4201582</td>
<td>235757</td>
<td>3385</td>
</tr>
<tr>
<td>740115.4</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>22</td>
<td>16126366</td>
<td>117</td>
<td>4043435</td>
<td>298463</td>
<td>3233</td>
</tr>
<tr>
<td>769845.7</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>23</td>
<td>15919268</td>
<td>124</td>
<td>4204461</td>
<td>299303</td>
<td>3385</td>
</tr>
<tr>
<td>770067.4</td>
<td>stool</td>
<td>NA</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>24</td>
<td>9609488</td>
<td>133</td>
<td>4201831</td>
<td>235749</td>
<td>3385</td>
</tr>
<tr>
<td>770180.8</td>
<td>stool</td>
<td>NA</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>16</td>
<td>22</td>
<td>15532364</td>
<td>116</td>
<td>4198654</td>
<td>212820</td>
<td>3386</td>
</tr>
<tr>
<td>780298.0</td>
<td>stool</td>
<td>NA</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>23</td>
<td>14033966</td>
<td>136</td>
<td>4203224</td>
<td>167915</td>
<td>3386</td>
</tr>
<tr>
<td>S/Nida</td>
<td>stool</td>
<td>NA</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>24</td>
<td>15541740</td>
<td>140</td>
<td>4205161</td>
<td>208032</td>
<td>3388</td>
</tr>
<tr>
<td>0092</td>
<td>stool</td>
<td>15-Mar-03</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>20</td>
<td>15375291</td>
<td>143</td>
<td>4094492</td>
<td>232090</td>
<td>3279</td>
</tr>
<tr>
<td>0006</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0020</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0093</td>
<td>stool</td>
<td>NA</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>stool</td>
<td>NA</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>271</td>
<td>stool</td>
<td>NA</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>333</td>
<td>stool</td>
<td>NA</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15375291</td>
<td>143</td>
<td>4094492</td>
<td>232090</td>
<td>3279</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0005671.t001
We successfully sequenced the whole genomes of 71 \textit{Vibrio cholerae} O1 isolates (Table 1). The average number of high quality reads was 15,375,291 which upon assembly produced an average of 143 contigs (range: 99 to 443) or scaffolds with an average depth of 374. The assembled genome was 4.09 Mb (3.95 to 4.23 Mb) in length and had an average N50 of 232,090 bp (144,507 to 339,296 bp).

Of the 71 isolates, the genomes of sixty-seven differed by less than 100 SNVs. Consistent with this, the core genome of all the wave 3 isolates shared 2802 genes. To look for evidence of recombination in these sequences we examined 2543 single copy conserved genes. We removed all of the invariant nucleotides, any SNVs in locally collinear blocks smaller than 200 bp, and examined the remaining variant sites in Clustal W (Fig 2). These data showed evidence...
of homoplasy by visual examination: for example there were multiple examples of dinucleotide sequences at the same site in different isolate genomes being present in all four possible combinations (e.g. AA, AG, GA, GG). This was despite the high level of pairwise nucleotide conservation between isolate genomes, usually differing by only 18 to 62 nucleotides. Among the 148 SNVs, 360 pairs were determined to have both alleles in all possible combinations. The probability of this occurring by mutation alone is vanishingly small (the probability of the same mutation on two different genomes to the 360th power). The most parsimonious explanation for this observation is recombination within this population of *V. cholerae* O1s.

The genetic relatedness, including the recombinant loci, was estimated from WGS using a phylogenetic network [26]. As shown in Fig 3, while all of the genomes were distinct, there was some clustering by year of isolation, four of the five isolated in 2010 clustered together as did the four from 2008 and the two from 2009. The largest number of isolates in our collection were collected in 2003. It is evident from Fig 3 that they occupy positions throughout the network. This was also true for the three river isolates. In the network, every sequence had some nucleotide variants that distinguish it from every other sequence, thus in contrast to the MLVA network, there are no central genotypes that might be construed as the founder.

Fig 4 shows the phylogenetic tree with genomic sequences from a global collection of 274 previously published sequences (S1 Table) including the 67 genomic sequences generated in this study. These data show that the sequences of the Manhiça, Mozambique isolates cluster in a monophyletic group distinct by 39 nucleotides from the backbone of the third wave of the Seventh Pandemic phylogeny (S2 Table). All Seventh Pandemic isolates harbor virulence associated genes, such as the CTX prophage, the genomic islands VSP-I and VSP-II, the toxin-correlated pilus, the toxic linked cryptic element, and the integrative conjugative element SXT harboring multidrug resistance genes. The CTX phage harbours the ctxBcla allele within an otherwise El Tor biotype CTX phage sequence and is typical of isolates referred to as “atypical” El Tor biotypes [32].

The four sequences that differed by >24,000 nucleotides among the 3,237,973 basepairs conserved in all our sequenced genomes did not cluster with sequences from the Seventh Pandemic (Fig 5) and are quite distantly related to each other, but all were more closely related to the other *V. cholerae* sequences compared to other species. These isolates were taken from patients admitted to hospital with suspected cholera indicating that these divergent lineages...
are capable of causing clinical symptoms, as has been reported previously [13]. Unlike the Seventh Pandemic strains, these four strains do not contain any of the aforementioned virulence associated genes, although they contain a few genes, such as the RTX toxin gene cluster and the hemolysin \( hlyA \), from some of the virulence associated islands.

The network based on WGS included isolates of both CCs (CC1 and CC2), however while the isolates of CC1 was distributed through the network, the isolates of CC2 clustered together (Fig 3). In the same way, the four singletons isolates by MLVA did not cluster in the network and were quite distantly related by WGS analysis.

Discussion

Cholera remains important public health problem in Mozambique. We characterized \( V. \) cholerae isolates using MLVA and WGS to determine the genetic relatedness and transmission dynamics of cholera outbreaks in the Manhiça District. Our analyses of WGS data revealed that 94% of the isolates were a monophyletic group in the third wave of the Seventh Pandemic. These isolates have formed a locally evolving population that has persisted for at least eight
Fig 4. Phylogeny of 7th pandemic isolates of *V. cholerae*. The branch lengths are proportional to the number of variable nucleotides. All of the isolates from Manhiça cluster together and are most closely related to isolates from India, Nepal, Kenya, Haiti and Bangladesh. Other isolates from Mozambique and Kenya are more distantly related.

https://doi.org/10.1371/journal.pntd.0005671.g004
years, either in a local environmental reservoir or circulating within the human population, and sporadically caused recurrent disease in southern Mozambique. It is yet to be shown if these isolates are representative of those circulating outside of the study site across Mozambique.

Previous studies have reported the role of environmental factors, such as seasonal fluctuations, that influence the dynamics of *V. cholerae* in environmental reservoirs [33,34]. Of note, all of our isolates were collected during the first half of the year, January through July, with a peak in May. However, our study does not allow establishing a clear relationship between environmental strains and those causing cholera outbreak. The presence of four isolates that are not part of the Seventh Pandemic (differing by ~ 0.75% of the genome sequence) demonstrates that a diverse set of *V. cholerae* not linked to the Seventh Pandemic are causing a background, low level of sporadic disease in southern Mozambique, despite the absence of many of the major virulence determinants. This has been seen elsewhere including in the Gulf coast of the USA particularly in the 1980’s and 1990’s [32].

We detected the presence of recombination in our WGS data. Most analyses of WGS data detect recombination as a large number of SNVs in short sequence of DNA. Our method of detection, the four-gamete test applied to haploid genomes, removes the constraint of the SNVs occurring in a small region by simply looking at dinucleotide pairs located at any
distance from each other. Since mutations occur at random, a dinucleotide sequence, say GG, can mutate at either of two positions for example: AG and GA. In order to further mutate to form AA, one of the positions must mutate a second time. However, in this instance it also possible to replace the original sequence in a single recombination event. We found 360 pairs of nucleotide positions across all our genomes at which all four combinations of alleles in the dinucleotides were found. The probability of this occurring by mutation is extremely small (the rate of mutation at the same nucleotide to the 360th power). Our sample is unusual because it was collected in small area and in a short time frame. The geographical and temporal proximity of isolates is a prerequisite for recombination. *V. cholerae* meets other necessary prerequisites like having an intact mechanism for uptake of DNA and integration into the chromosome [38]. The clearest examples of the effectiveness of recombination can be found in serotype switching, a process known to be accelerated by chitin [39]. The limited amount of variation in our sample is consistent with the recombinant events occurring within this population of *V. cholerae* O1s.

Most isolates (91%) were distributed among the 20 genotypes of CC1 and 41% had the founder genotype, supporting the hypothesis of a common ancestor which subsequently differentiates into additional genotypes. In a study of 187 isolates conducted in Haiti, only 9 MLVA genotypes clustered in a single clonal complex and 53% had the founder genotype [35].

We found more alleles (ten) at the first locus (VC0147) on large chromosome than at the two loci on small chromosome (VCA0171 and VCA0283) with seven alleles for each one. Our findings are in contrast to previous descriptions that show the three loci on the large chromosome varying at a slower rate than those on the small chromosome [11]. Extensive genetic variation on both chromosomes has been reported [36], but not more variation at the large chromosome loci [35–37].

WGS and MLVA patterns were performed to discriminate *V. cholerae* isolates from various geographic locations and distinct populations [17,40]. But, none of the 26 MLVA genotypes that we found in this study has been reported in previous studies from Haiti, Thailand, Bangladesh, India, Vietnam and Mozambique [10,11,35,36,40,41]. However, the three loci on large chromosome are likely to be considered the best for estimating across large distances [35]. Thus, when we considered the MLVA profiles in terms of the three loci (VC0147, VC0437, and VC1650) on large chromosome, the isolates with profiles 8,4,6, (50%, 39/78) and 9,4,6, (10%, 8/78) were related to the Haiti and Bangladesh strains, respectively [11,35]. Our WGS analysis demonstrates that these Mozambican isolates formed a distinct lineage within a clade that includes El Tor variants from Bangladesh, China, Haiti, Nepal, India, and Kenya that belong to the current global radiation of the Seventh Pandemic. The Manhiça, Mozambique isolates differ from the wave 3 backbone by 39 nucleotides. Furthermore, previous reports demonstrated Wave 2 strains in Mozambique during the same time period [21,42]. Taken together with our analysis, it is clear that there were multiple, independent *V. cholerae* lineages from Wave 2 and Wave 3 which were circulating within Mozambique during this period of time. In Manhiça, although the relationships between isolates differed in detail between MLVA and WGS, both analyses demonstrated the isolates were very closely genetically related, even if they are collected several years apart.

**Conclusion**

Our findings further define the molecular epidemiology of *V. cholerae* in Mozambique. Our study demonstrates that Wave 3 isolates of the Seventh Pandemic have become established in Manhiça, Mozambique and have persisted in this region over the time of this study alongside *V. cholerae* Wave 2 strains in other regions of the country. The subsequent radiation of
genotypes has been enriched by the process of recombination as detected in the WGS data. Our data raises several important questions that relate to where these *V. cholerae* isolates persist and how seemingly identical isolates can be collected years apart despite our understanding of the high rate of change of the MLVA loci and the *V. cholerae* molecular clock. Although the environmental triggers for the emergence of cholera are unknown in Manhiça, it is important to be vigilant to prevent an emergence from becoming an outbreak.

**Supporting information**

S1 Table. Isolates for 7th pandemic tree.
(XLSX)

S2 Table. List of 39 mutations that distinguish isolates from Manhiça, Mozambique from the Wave 3 backbone.
(XLSX)

**Acknowledgments**

The authors thank all study participants to the study and all workers of the Centro de Investigação em Saúde da Manhiça for their help in obtaining the data. The authors finally thank the district health authorities for their collaboration in the research activities ongoing in the Manhiça District. Lili Fox Velez edited various versions of the manuscript.

**Author Contributions**

Conceptualization: MG IM OCS.

Data curation: MG IM DV TN SA SL.

Formal analysis: MG SL JNP MA DD OCS.

Funding acquisition: IM NRT PA OCS.

Investigation: MG IM DV TN SA SL.

Methodology: MG JNP MA DD OCS.

Project administration: IM NRT PA OCS.

Resources: MG IM DV TN SA SL JNP MA DD.

Software: MG SL JNP MA DD.

Supervision: IM NRT PA OCS.

Validation: MG IM DV TN SA SL JNP MA DD NRT OCS.

Visualization: MG JNP MA DD OCS.

Writing – original draft: MG OCS.

Writing – review & editing: MG IM DD NRT OCS.

**References**


Minimal genetic change in Vibrio cholerae in Mozambique over time: MLVA & WGS


