# Population and Genetic Study of *Vibrio cholerae* from the Amazon Environment Confirms that the *WASA-1* Prophage Is the Main Marker of the Epidemic Strain that Circulated in the Region

**Lena Líllian Canto de Sá Morais[1]\***, **Daniel Rios Garza[1]**, **Edvaldo Carlos Brito Loureiro[2]**, **Elivam Rodrigues Vale[1]**, **Denise Suéllem Amorim de Sousa Santos[1]**, **Vanessa Cavaleiro Corrêa[1]**, **Nayara Rufino Sousa[1]**, **Tereza Cristina Monteiro Gurjão[1]**, **Elisabeth Conceição de Oliveira Santos[1]**, **Verônica Viana Vieira[3]**, **Erica Lourenço da Fonseca[3]**, **Ana Carolina Paulo Vicente[3]**

1 Environmental Section of the Evandro Chagas Institute, Ananindeua, Pará, Brazil, 2 Bacteriology Section of the Evandro Chagas Institute, Ananindeua, Pará, Brazil, 3 Laboratory of Molecular Genetics of Microorganisms, Oswaldo Cruz Foundation, Rio de Janeiro, Rio de Janeiro, Brazil

## Abstract

*Vibrio cholerae* is a natural inhabitant of many aquatic environments in the world. Biotypes harboring similar virulence-related gene clusters are the causative agents of epidemic cholera, but the majority of strains are harmless to humans. Since 1971, environmental surveillance for potentially pathogenic *V. cholerae* has resulted in the isolation of many strains from the Brazilian Amazon aquatic ecosystem. Most of these strains are from the non-O1/non-O139 serogroups (NAGs), but toxigenic O1 strains were isolated during the Latin America cholera epidemic in the region (1991-1996). A collection of environmental *V. cholerae* strains from the Brazilian Amazon belonging to pre-epidemic (1977-1990), epidemic (1991-1996), and post-epidemic (1996-2007) periods in the region, was analyzed. The presence of genes related to virulence within the species and the genetic relationship among the strains were studied. These variables and the information available concerning the strains were used to build a Bayesian multivariate dependency model to distinguish the importance of each variable in determining the others. Some genes related to the epidemic strains were found in environmental NAGs during and after the epidemic. Significant diversity among the virulence-related gene content was observed among O1 strains isolated from the environment during the epidemic period, but not from clinical isolates, which were analyzed as controls. Despite this diversity, these strains exhibited similar PFGE profiles. PFGE profiles were significant while separating potentially epidemic clones from indigenous strains. No significant correlation with isolation source, place or period was observed. The presence of the *WASA-1* prophage significantly correlated with serogroups, PFGE profiles, and the presence of virulence-related genes. This study provides a broad characterization of the environmental *V. cholerae* population from the Amazon, and also highlights the importance of identifying precisely defined genetic markers such as the *WASA-1* prophage for the surveillance of cholera.

**Competing interests:** The authors have declared that no competing interests exist.

\* E-mail: lenasa@iec.pa.gov.br

## Introduction

*Vibrio cholerae* is a genetically diverse inhabitant of the aquatic environment. Its lifestyle is intimately associated to the micro-fauna of brackish and estuarine waters, where it can be found as a free-living microbe or in diverse associations with zooplankton, phytoplankton, and other species [1-4]. Such diverse habitats in preferably warm and slightly alkaline environments constitute the species' natural reservoir [5]. Several studies have shown that great genetic diversity is observed at the intraspecific level [6,7]. This diversity reflects the dynamic existence of many *V. cholerae* populations in diverse aquatic environments throughout the world.

Pathogenic *V. cholerae* are also diverse groups of strains with relatively distinct genomic backgrounds. Nevertheless, epidemic and pandemic cholera belong to a few clusters of relatively clonal populations of independent origin [8,9]. Evolutionary studies that consider the complete genome sequences or housekeeping genes of *V. cholerae* strains generally agree that the seventh and ongoing cholera pandemic—which constitutes one of these population-clusters—originated from a lineage that can be distinctly identified [8,10,11]. This lineage is characterized by: (i) the O1 surface antigen; (ii) several gene systems which are related to the ability to colonize the human intestine, thus coordinately expressing virulence factors, and provoking cholera; and (iii) a number of genetic and physiological properties which define a biotype named El Tor [12]. In all of the outbreaks attributed to the seventh cholera pandemic, toxigenic O1 El Tor strains have been isolated from stool and, frequently, from the environment. This lineage is distinguishable through several typing methods from indigenous strains that are majorly non-toxigenic and of NAG serogroups [13-16].

The seventh cholera pandemic emerged in 1961 and since then its causative agent—the toxigenic O1 El Tor *V. cholerae*—has become endemic in areas of Africa, Asia and Latin America [17-19]. Epidemic outbreaks in endemic regions frequently exhibit periodicity that coincides with environmental variables such as yearly monsoons, tidal estuaries, and riverine systems [20-22]. Outbreaks of cholera in areas where it is not endemic are usually attributed to importation, while the extent of the epidemics in both endemic and epidemic settings are related to human transportation routes and risk factors encountered in the places where cholera reaches [23,24]. Poverty, malnutrition, crowding, migration, and poor sanitation, among other social and economic drivers, have been correlated to the extent and fatality-rate of cholera outbreaks. Other factors such as climate change and natural disasters also seem to play an important role in favoring the emergence and spread of cholera [25-27].

The origin of the last cholera epidemic in South America that started in Peru (1991) has not been completely elucidated. In this epidemic, a distinguishable toxigenic O1 El Tor clone was predominantly isolated from confirmed cholera cases. This clone exhibited important variations compared to other toxigenic O1 El Tor strains that were being isolated in other regions of the world [28]. Among these differences, the presence of a 50 kb prophage integrated into the alanine aminopeptidase gene was recently identified [8,29,30]. This prophage has been named *WASA-1*.

Before the introduction of cholera in 1991, environmental surveillance for the presence of potentially toxigenic *V. cholerae* strains was initiated in several Latin American countries. These initiatives followed the eminent risk of importation of the disease from countries that experienced cholera outbreaks [31]. Many environmental isolates were recovered in this period, including non-toxigenic O1s [32,33]. In the Brazilian Amazon, surveillance of wastewater in the city of Belem also isolated many *V. cholerae* NAG strains. During and after the epidemic, more extensive environmental surveillance has shown the presence of NAG strains in many rivers and streams, and toxigenic O1 isolates were also found during the epidemic period. Though the toxigenic O1 El Tor *V. cholerae* has failed to establish long-term endemicity in the Amazon, several potential risk factors, compared to the dynamics of cholera in other regions, highlight the need for continuous and effective surveillance of the bacterium's natural reservoir. Some of these factors are: (i) the presence of endemic populations of *V. cholerae* in many of the region's rivers and streams, which can undergo lateral gene transfer and constitute a natural reservoir of virulence genes [34]; (ii) an extensive human population that occupies water margins in both urban and rural areas, which lack appropriate sanitary infrastructure; (iii) periodic flooding of urban neighborhoods during the yearly rainy season; (iv) many sites of balneability and recreation that are potential sources of water contamination and of consumption of contaminated water; and (v) chronic inadequate supply of clean water even in urbanized areas, which represent high risk of fecal contamination.

The epidemiological importance of the Amazon region as a potential border for cholera in Brazil has been clearly indicated in two independent historical periods. During the third pandemic, cholera was introduced by a "caravella" coming from Europe into the country through the Amazon region. This epidemic extensively struck Brazil, infecting approximately 200 thousand people in 1855 [35]. Later, in the seventh pandemic, cholera again entered Brazil through the west-Amazon by the Peruvian border [36]. In this epidemic, the disease exhibited temporary endemicity with seasonal peaks for approximately six years [37]. The spread of cholera to other regions of the country in both of these occasions indicates the importance of the region and its environment concerning the risk of cholera in Brazil. This also indicates the importance of characterizing and monitoring the dynamic *V. cholerae* population in the Amazon environment.

The purpose of this study was to analyze a collection of environmental *V. cholerae* strains from the Amazon, spanning a large historical period which can roughly be divided into pre-epidemic (1977-1990), epidemic (1991-1996), and post-epidemic (1996-2007) periods. The presence of genes related to virulence phenotypes within the species was surveyed and the genetic relationship among the strains was analyzed through genomic restriction patterns. Using these variables and the information available concerning the strains, a Bayesian multivariate dependency model was employed to distinguish the importance of each variable on determining the others. Besides the broad characterization of the environmental *V. cholerae* population of the Amazon, this study reinforces the importance of identifying precisely defined genetic markers such as the *WASA-1* prophage for the surveillance of cholera.

## Materials and Methods

### *V. cholerae* strains and growth conditions

A total of 125 *V. cholerae* strains that were isolated from the environment at different sites in the Brazilian Amazon in the period of 1977 to 2007 were analyzed (Figure **1** and Table **1**). Four strains from stool samples of patients that were isolated in 1991-1994 during the Amazon cholera epidemic were also
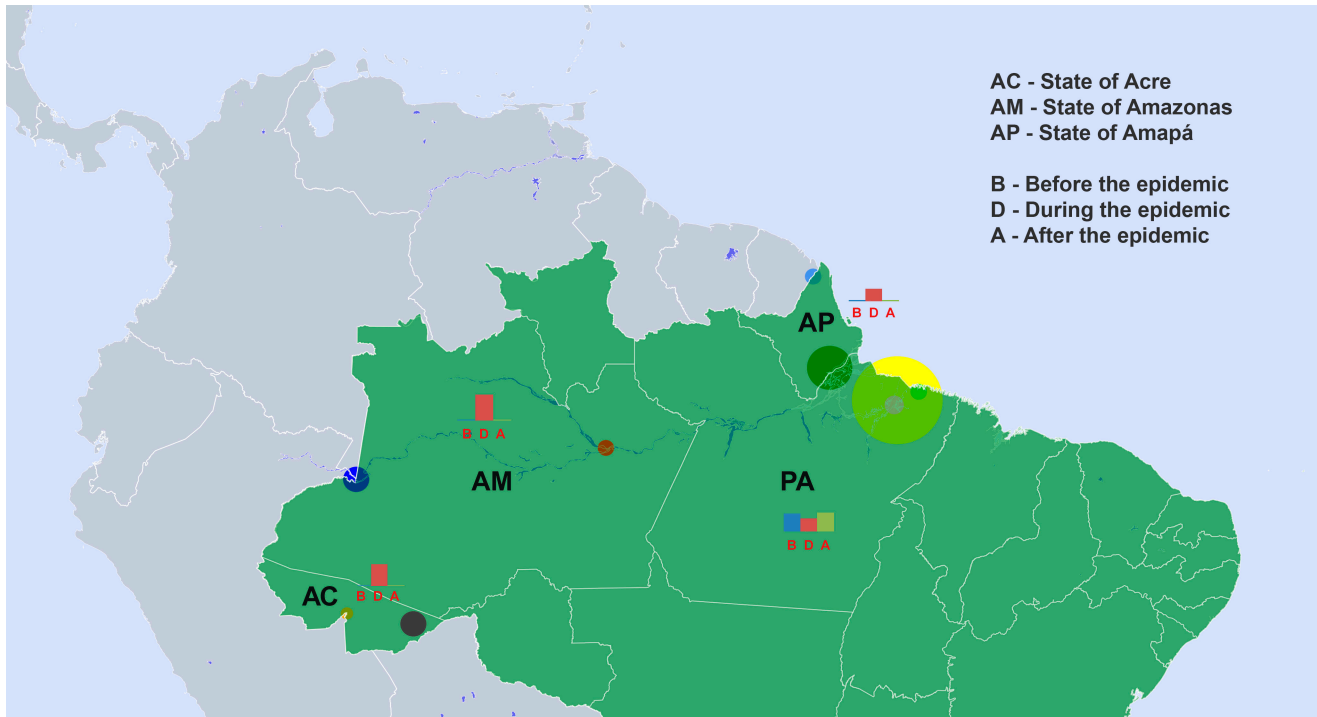
**Figure 1. Geographical distribution of *V. cholerae* isolates.** The geographical location of rivers, streams, and wastewater plants from where the strains that were used in this study were isolated are indicated in the map. The sizes of markers indicate the number of strains in each location, markers are centered in the cities where the strains were isolated (see Table S1). Belem (yellow), Barcarena (light green), Maruda (pink), Macapá (dark green), Oiapoque (light blue), Manaus (red), Tabatinga (light blue), Rio Branco (purple), and Santa Rosa (orange). Quantities of strains isolated in each period are indicated in the bar graphs.
doi: 10.1371/journal.pone.0081372.g001

**Table 1.** Environmental strains of *V. cholerae* isolated from the Amazon environment.

| | Pre-epidemic | Epidemic | Post-epidemic | Total |
|---|---|---|---|---|
| Number of strains | 33 | 62 | 35 | 130 |
| Serogroup (NAGs) | 33 | 37 | 34 | 104 |
| Serogroup (O1s) | 0 | 25 | 1 | 26 |
| Source (Copepod) | 0 | 0 | 2 | 2 |
| Source (Fish) | 0 | 3 | 0 | 3 |
| Source (Stool) | 0 | 5 | 0 | 5 |
| Source (River) | 0 | 29 | 20 | 49 |
| Source (Stream) | 0 | 9 | 10 | 19 |
| Source (Wastewater) | 33 | 16 | 3 | 54 |
| State (Pará) | 33 | 24 | 35 | 94 |
| State (Amapá) | 0 | 23 | 0 | 23 |
| State (Acre) | 0 | 7 | 0 | 7 |
| State (Amazonas) | 0 | 8 | 0 | 8 |

doi: 10.1371/journal.pone.0081372.t001

included as references; one clinical strain from a cholera-like outbreak that occurred in the Amazon in 1992 and was caused by a non-toxigenic O1 strain was also included; a classical strain isolated in Pakistan, 1986, and two Asian El Tor strains from India and Bangladesh, isolated in 1971 and 1973, respectively, were also used as references (Table **S1**).

All strains were from the bacterial collection of the Evandro Chagas Institute, Brazil, and were re-isolated and characterized following standard procedures for *V. cholerae*. Briefly, strains were cultured in alkaline peptone water (APW) and plated on thiosulphate citrate bile salt sucrose (TCBS). Colonies were further examined by biochemical assays, which

**Table 2.** Genes investigated by PCR.

| Gene | Abbrev. | Locus Tag | Ref. Genome (Ac. Number) |
|---|---|---|---|
| Cholera enterotoxin subunit A | *ctxA* | VC1456 | N16961 (NC_002505.1) |
| Cholera enterotoxin subunit B | *ctxB* | VC1457 | N16961 (NC_002505.1) |
| Zonula occludens toxin | *zot* | VC1458 | N16961 (NC_002505.1) |
| Accessory cholera enterotoxin | *ace* | VC1459 | N16961 (NC_002505.1) |
| Hypothetical protein | *orfU* | VC1460 | N16961 (NC_002505.1) |
| RTX toxin A | *rtxA* | VC1451 | N16961 (NC_002505.1) |
| RTX toxin activating protein | *rtxC* | VC1450 | N16961 (NC_002505.1) |
| Cholera toxin transcriptional activator | *toxR* | VC0984 | N16961 (NC_002505.1) |
| toxin co-regulated pilin | *tcp* | VC0828 | N16961 (NC_002505.1) |
| Phage replication protein Cri | *ptcl* | VC1469 | N16961 (NC_002505.1) |
| Heat-stable enterotoxin | *stn/sto* | M85198.1 | M85198.1 |
| Haemolysin | *hlyA* | VCA0219 | N16961 (NC_002506.1) |
| Lactonizing lipase | *hlyC* | VCA0021 | N16961 (NC_002506.1) |
| Hemolysin secretion protein | *hlyB* | VCA0220 | N16961 (NC_002506.1) |
| Phage deoxyribonucleoside kinase | *WASA-1* | O3Y_07175 | IEC224 (NC_016944.1) |
| Phage glycosyl hydrolase | *WASA-1* | O3Y_07005 | IEC224 (NC_016944.1) |
| Phage RNA polymerase | *WASA-1* | O3Y_07235 | IEC224 (NC_016944.1) |

doi: 10.1371/journal.pone.0081372.t002

included: oxidase; motility; fermentation on glucose, sucrose, and lactose; growth in 0% NaCl; lysine decarboxylase, arginine dehydrolase, ornithine decarboxylase, methyl-red, Voges-Proskauer, and indole tests. Serogouping was performed with a polyvalent O1 and O139 antisera, using the slide agglutination test (Mast Diagnostics Ltd). Species identification was further confirmed by PCR using primers for the species-specific gene *ompW* according to Nandi et al.(2000) [38].

### PCR Assays

Total genomic DNA was purified as previously described [39]. PCR assays were performed in all strains using specific primers for all genes listed in Table **2**. Platinum TAQ kit (Invitrogen) was used for DNA amplification following manufacturer's instructions. Specific annealing temperatures were used for each primer and results were assessed through electrophoresis (for primer sequences and annealing temperatures, see Table **S2**).

### Pulsed-Field Gel Electrophoresis (PFGE)

PFGE was performed according to PulseNet standardized protocol [40] with NotI restriction enzyme (Invitrogen). Plugs were submitted to electrophoresis using the CHEF-DR III equipment (BioRad). In each assay the Lambda PFGE marker (Biolabs) was used. Images of the pulsetypes were uploaded and analyzed using the BioNumerics software, Version 5.1 (Applied Maths). Profiles were compared using the Dice coefficient and clustered by the unweighted pair group method with arithmetic mean (UPGMA). Cluster confidence was assessed by bootstrap with $10^6$ replicates. Clusters were visualized and represented using the Fig Tree software, version 1.4 (tree.bio.ed.ac.uk/software/figtree/). A cladogram representation of the PFGE profiles, as well as images of representative gels can be found in Figure **S1**.

### Multivariate dependency model

Possible interrelations between the multivariate sets of discrete data which were available for each strain (presence of virulence-associated genes, isolation period, source, serogroup, city, and PFGE profiles) were investigated using a statistical learning approach, which is based on the construction of an optimal Bayesian Network (BN) [41]. For this purpose we used the open source internet service, B-Course (http://b-course.hiit.fi/bene) [42]. In this approach, the model was defined from the conditional probabilities (represented as arcs) between variables (represented as nodes), in which a best network was determined by iterating over different topologies and using an expectation maximization algorithm. A precise description of the algorithm implemented in B-Course is available in reference [43]. The strength of each arc is represented as the posterior probability of the model when the edge is removed.

## Results

### Gene content PCR profiles

The presence of 17 genes was investigated in all isolates (Table **2**). These included several genes related to virulence phenotypes in *V. cholerae*, such as: genes encoded by the genome of the CTXφ bacteriophage (*ctxAB*, *ace*, *zot*, and *orfU*) [44]; *tcpA*, which is the structural component of the toxin corregulated pilus [45]; *toxR*, a major regulator of many genes related to cellular functions, among which are the *V. cholerae* virulence program [46,47]; and the *WASA-1* prophage, a distinguished marker of the Latin American epidemic clone [30]. Other genes that are related to virulence phenotypes and genomic integration of virulence-related elements were also investigated (Table **2**).

Thirty-five different profiles (genotypes) were observed among the 104 NAG strains. The most frequent profile, present in 30.76% of the strains, was *rtxA, rtxC, toxR, hlyA, hlyC,* and *hlyB,* which were also, individually, the most frequent genes among NAGs of all three periods that were studied (Figure **2**). A temporal pattern was observed in the detection of the the CTXφ genes (Figure **2**). These genes were only found in strains isolated during and after the epidemic period. Nevertheless, none of these genes were co-amplified in any of the NAG strains. No NAG strains harbored *ctxAB, tcpA,* or the *WASA-1* prophage genes. The Cri replication protein, which is encoded by the toxin linked cryptic plasmid element (pTLC) exhibited a different pattern: a few strains (2.88%) from all the periods were found to harbor this element which is also reported to be a satellite phage (TLCφ) [48]. Similar to the Cri replication protein, toxR was amplified in strains from all three periods, but seemed to be more frequent among strains isolated during and after the epidemic period. Three strains (9.09%) from before the epidemic period and five strains (14.70%) from the post epidemic period harbored the *stn/sto* toxin gene, while none of the O1 strains were found to have this gene.

Among the O1 strains, all of the clinical El Tor from outbreaks in the Amazon harbored the same set of genes, namely *ctxAB, zot, ace, orfU, rtxA, rtxC, toxR, tcpA, pTLC, hlyA, hlyC, hlyB,* and the *WASA-1* genes (Figure **3**). The only difference from the clinical El Tor strains from Asia— 121 and N16961—was the presence of the *WASA-1* phage. Diverse virulence related gene contents were observed among the O1 strains that were isolated from the environment but seemed to be related to the epidemic clone. This suggests repeated events of gene loss in the environment. Lack of evidence for the presence of one, two, three, and even four of the genes tested was found in some of these strains, though all of them were positive for the three *WASA-1* prophage genes that were tested. We identified the O1 strains from the PFGE clusters 1, 2, 3, and 9 as being related to the epidemic clone (Figure **4**, discussed below). Among these strains, three of them tested negative for the *ctxAB* genes, while one of these was also negative for *tcpA*. *tcpA* amplification was also absent in two strains that contained *ctxAB* and other genes of the CTXφ.

The non-toxigenic O1 strain from the Amazonia lineage isolated from cholera-like outbreaks in the Northern Amazon [49] exhibited an identical gene content profile as the also non-toxigenic O1 strain isolated from the Amazon environment in 2007—LMA 3984-4 [50]. Though these strains exhibit identical PCR profiles, their genomic sequences indicate that they´re not clonal [6].

### PFGE profiles

Hierarchical clustering analysis of the PFGE profiles defined 23 distinct groups. All of the clinical and environmental O1 El Tor strains from the Amazon were clustered in the groups 1, 2, and 3 (Figure **4**). The N16961 strain was clustered with the strains from the Amazon of group 1, and the strain 121 from India clustered in group 3. The only *WASA-1* positive isolate that did not group within these three clusters was the strain V. 735, which clustered within group 9. The LMA 3984-4 strain

clustered with the strain 200 of the classical biotype which was isolated in Pakistan, reproducing previously published results [50].

A detailed analysis of the PFGE bands that were detected among O1 strains showed that groups 1, 2, and 3 were nearly identical, with only two and three different bands in between them when considering all of the individual cluster's bands (Figure **5, B**). Greater diversity was observed among polymorphic and conserved bands within each of these clusters (Figure **5, C-D**). Five bands were shared among all O1 strains, but only a small number of bands were exclusive of particular clusters, none of which were found in the groups 1 and 2 (Figure **5, E-F**).

Besides the clustering of clinical and environmental O1 El Tor strains, the PFGE profiles were not found to be informative of any other variable investigated in this study. As shown in Figure 4, the virulence genes that were predominant in the toxigenic O1 strains were detected in different clusters among NAG strains (Table **S1**). Additionally, no significant correlation was found between PFGE group, serogroup, period, city of isolation, and source (Figure **6**). In fact, only small groups of strains isolated in similar periods and cities clustered together, such as groups 4, 13, 16, 21, and 22, few of which were isolated from the same environmental sources (Figure **6** and Table **S1**).

### Multivariate dependency model

The possible correlation between the presence of genes, PFGE profiles and other data available from the strains (Table **S1**) was estimated using a statistical learning tool to build a best-topology Bayesian Network (Figure **7**). The constructed model shows significant relations which can be logically explained, adding to the knowledge about the population of environmental *V. cholerae* from the Amazon in the periods that were studied (strains used as references were not included in the constructed model). Strong dependency was observed between period, city, and source which were influenced by the composition of the samples we had available for this study. The PFGE profile was not important in determining any of the variables, while the *WASA-1* prophage correlates strongly with: (i) the definition of serogroups, (ii) PFGE profiles, and (iii) *tcpA* and *ace* (removing arcs reduce the model probability $>10^9$ times). The *WASA-1* prophage was also the most significant determinant of *pTLC*, and *zot*. Furthermore, strong arcs were also observed between *hlyB* and *rtxA*, and between *rtxA* and *rtxC*.

### Discussion

The composition and population structure of environmental *V. cholerae* strains from the Amazon region have not yet been the subject of extensive studies. Though previous works have included clinical and environmental O1 and NAG strains from this region isolated in different periods [51-53], the present study is, to our knowledge, the first time that strains spanning a large time and space interval—temporarily including pre-epidemic, epidemic, post-epidemic isolates—have been evaluated. Previous studies of environmental strains from other
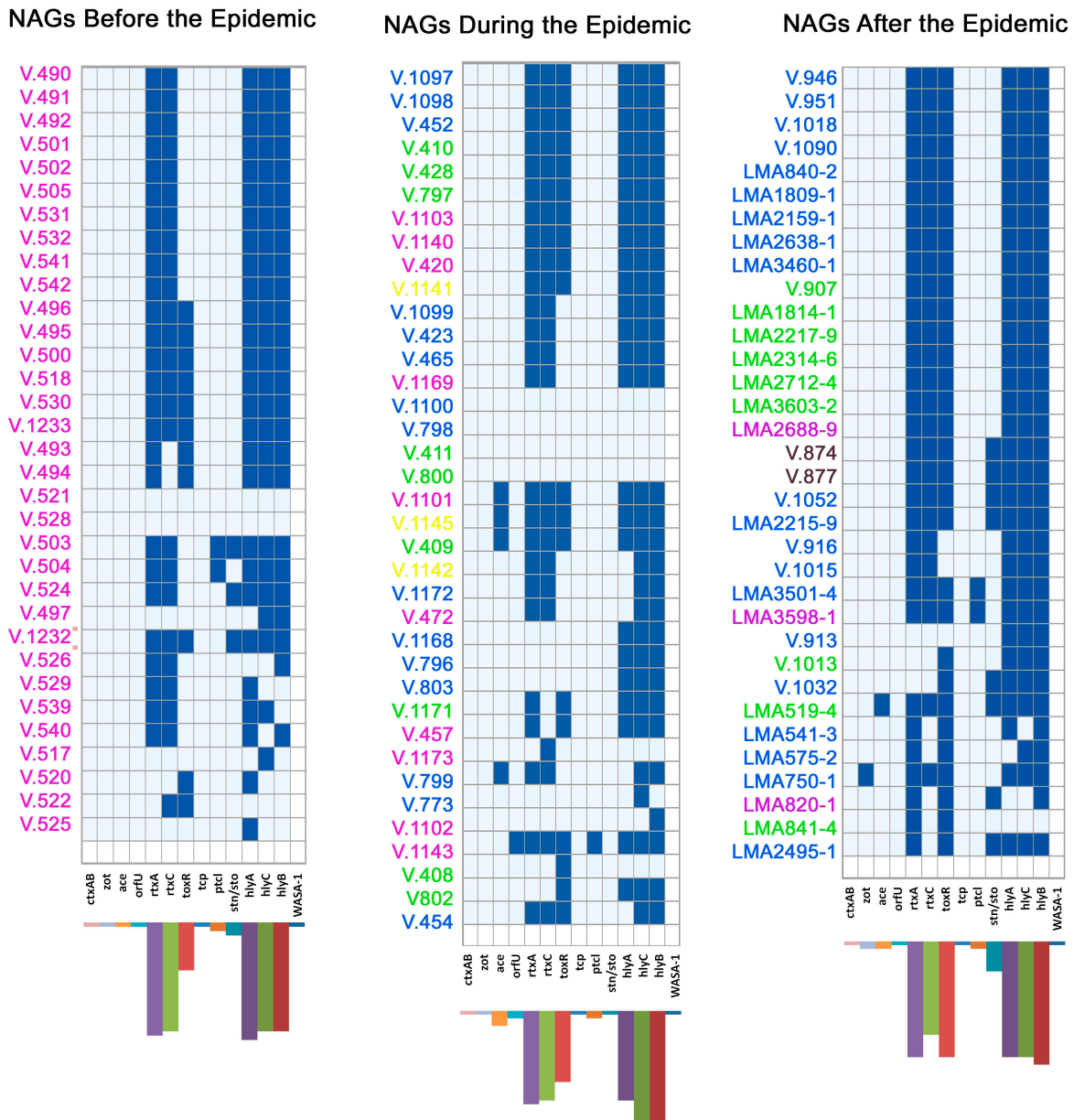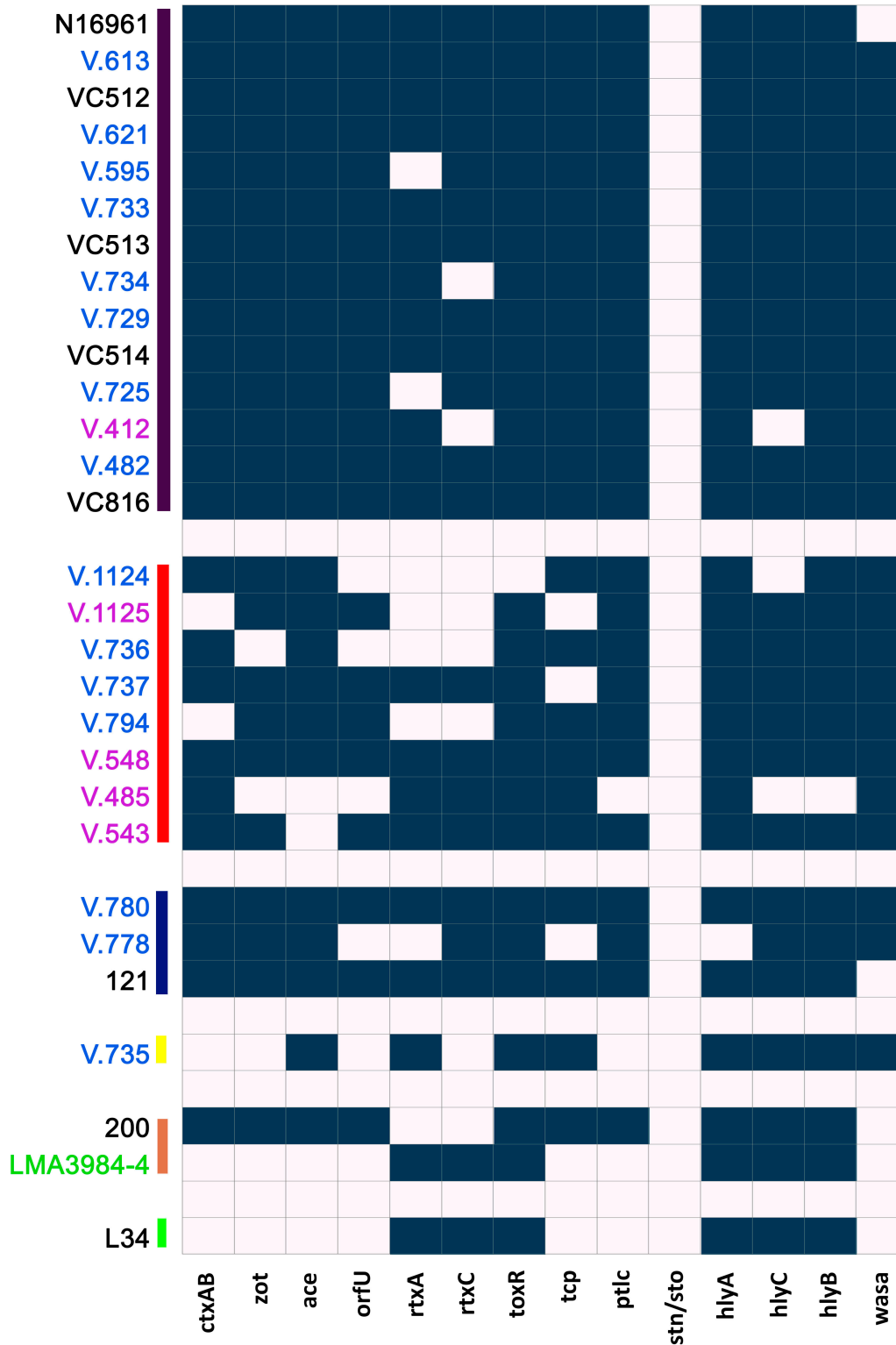
**Figure 2. Distribution of genotypes among NAG strains.** The presence or absence of virulence-related genes are represented, respectively, by blue and white squares. The histogram below each figure correspond to the frequency of each gene. The colors highlighting the strains' keys correspond to the isolation sources. Strains highlighted pink were isolated from wastewater, blue from superficial water, green from superficial stream water, yellow from fish, and brown from copepods.

doi: 10.1371/journal.pone.0081372.g002

regions of Brazil have shown several patterns regarding the presence and distribution of virulence-related genes and population diversity accessed through several strain-typing methods [33,51-53].

Though different studies have used different gene sets to characterize the virulence-related genotype of *V. cholerae* strains in Brazil, some of the common genes and genomic elements can be compared. In a study by Theophilo et al., 2006 [51], the presence of *ctxA*, *zot*, *ace*, and *tcpA* was

**Figure 3. O1 genotypes.** The presence and absence of virulence-related genes are represented, respectively, by blue and white squares. The strains are grouped in colored bars according to their PFGE cluster (Fig. 4): from top to bottom are groups 1 (purple), 2 (red), 3 (blue), 9 (yellow), 6 (orange), and 8 (green). The colors highlighting the strain keys correspond to the isolation sources. Strains highlighted pink were isolated from wastewater, blue from superficial water, green from superficial stream water, and black from clinical sources.
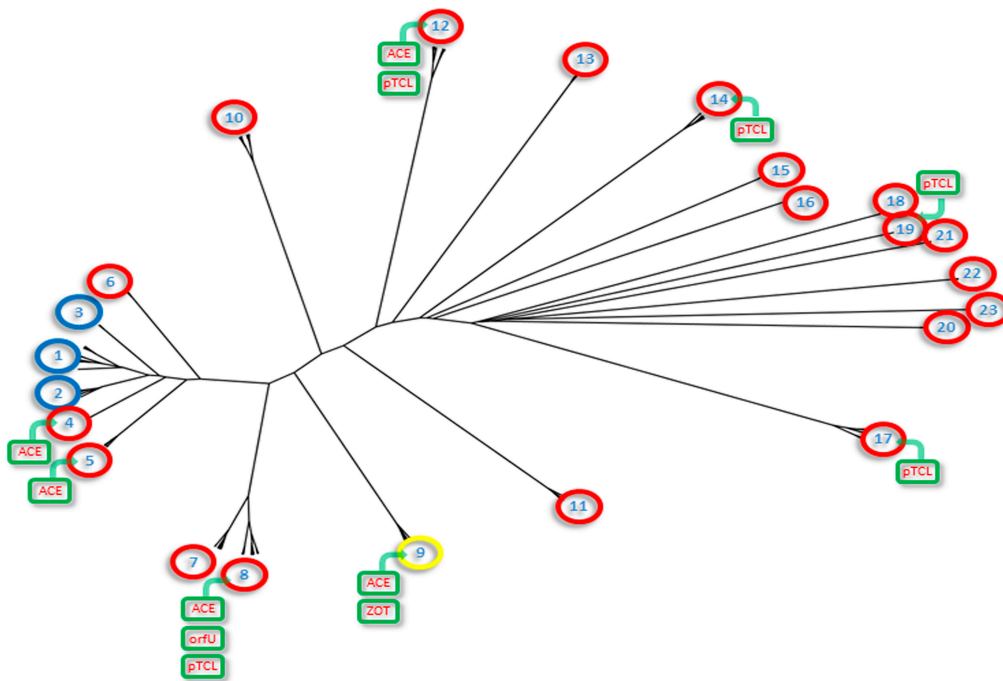
**Figure 4. Hierarchical clustering of PFGE profiles.** Relative distances are based on the pairwise comparison of band patterns through the dice coefficient and clustering through UPGMA. Blue groups are formed by the epidemic clones, while red groups contain environmental NAGs. The yellow group has one strain (V. 735) which is identified as a possible epidemic clone (see text). The integration of genes which are related to epidemic clones is indicated by squares in each cluster.

doi: 10.1371/journal.pone.0081372.g004

investigated in clinical and environmental strains from different states of Brazil. These strains were isolated between 1991 and 2000. *ctxA* was found in clinical and environmental NAG strains from Northeastern states of Brazil, but not in any strains from the Amazon. *zot*, on the other hand, was the most common gene among these strains, and was amplified individually in some of the strains, and was also co-amplified with *ctxA*, *tcpA*, and/or *ace* in some of the other strains. This reveals a great diversity in the integration and profile of the CTXφ genes among these strains. This diversity has also been observed in other countries, where *ctxAB* was found in strains that were negative for tcpA [34]. *tcpA* encodes the receptor for the integration of the *CTXφ* [45].

NAG *V. cholerae* strains positive for *ctxA* but negative for *zot*, both of which are positive or negative for *tcpA* have also been reported from environmental sources from the Brazilian state of São Paulo [33]. Among these strains—similar to the environmental strains from the Amazon here reported—*hlyA* and *toxR* were the most frequently amplified genes. The same study evaluated O1 strains from clinical and environmental sources isolated during the epidemic period in São Paulo. In

contrast to the environmental O1s from the Amazon, all of these strains were positive for *ctxA*, *zot*, and *tcp*. Significant diversity was found among the CTXφ genes and *tcpA* in a more recent study of environmental O1 isolates from the aquatic basin of state of Pernambuco [54]. O1 strains with partial sets of CTXφ genes with and without tcpA were identified, similar to what we found among the environmental O1 isolates from the Amazon.

In our study, all of the epidemic O1 strains from clinical sources were positive for all four CTXφ genes and *tcpA*. In the strains from environmental sources, a significant diversity was observed regarding the presence or absence of the above genes, which suggests mechanisms of gene loss and possibly attenuated virulence among these strains. The clinical manifestations of cholera are dependent in many aspects on the products of the CTXφ genes [44], specially the cholera toxin, which is a product of the *ctxAB* cluster. The colonization of the small intestine depends on the products of the tcpA gene [45,47]. Other aspects of the *V. cholerae* virulence program are also related to *ace* and *zot* [55,56]. For this reason, the isolation of strains potentially lacking a partial set of these
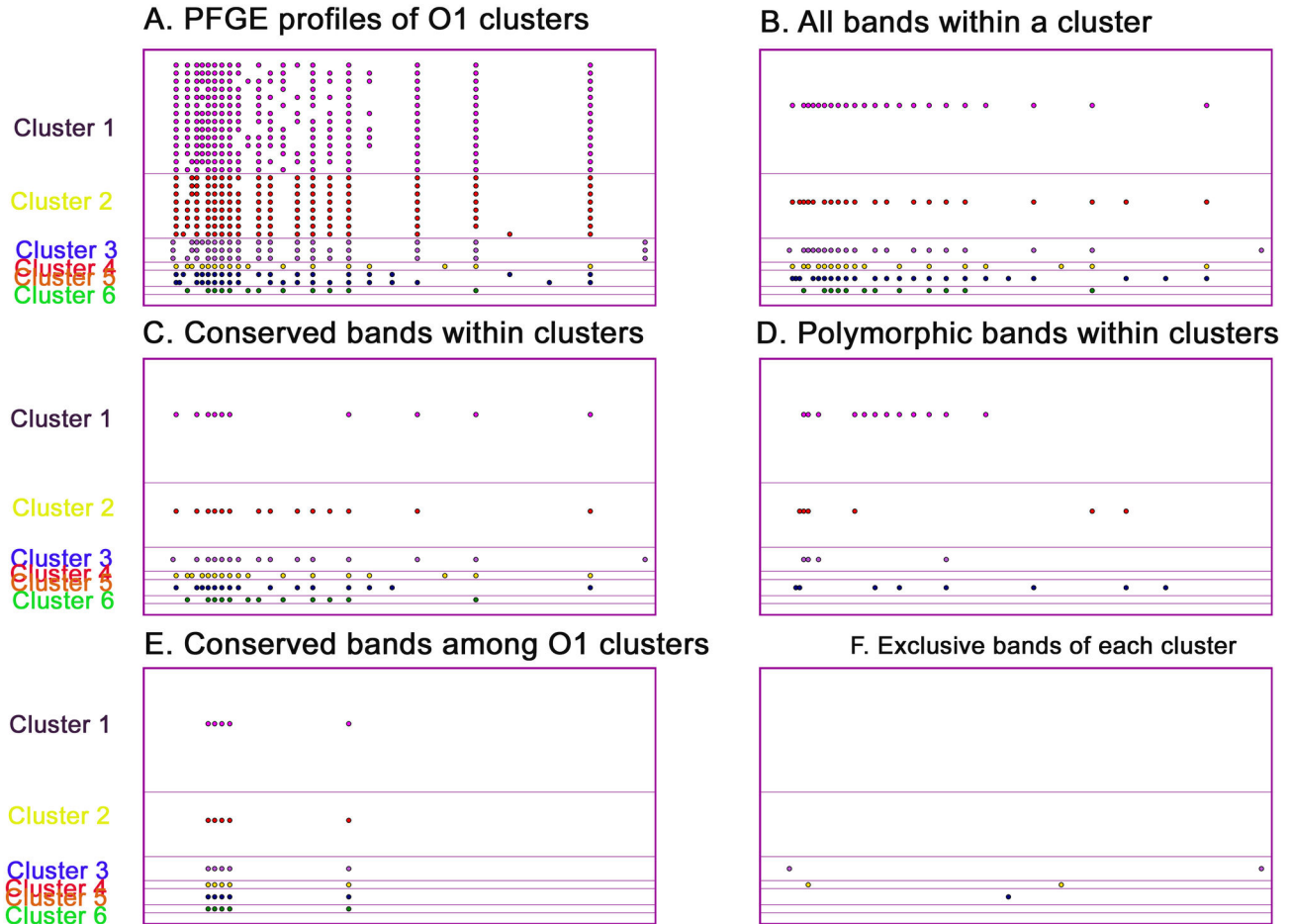
**Figure 5. O1 PFGE clusters.** Plot of restriction-site band for the six clusters of O1 strains, indicating all bands within a cluster (B), conserved bands within all strains of a cluster(C), bands that vary among strains of the same cluster (D), bands that are conserved among all O1 strains(E), and bands that are unique to any single cluster (F).
doi: 10.1371/journal.pone.0081372.g005
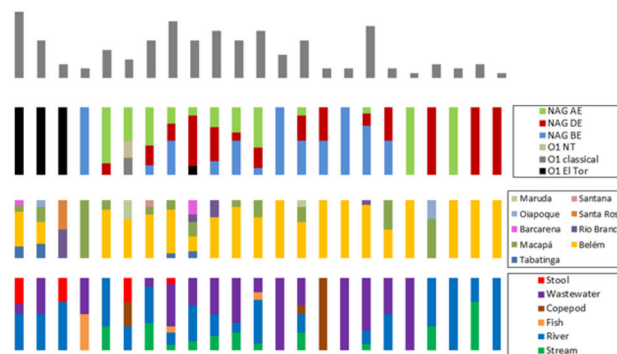


**Figure 6. Correlation between PFGE profiles and data from strains.** Each column represents a PFGE-profile group. The superior bars indicate the number of strains, while the colors (according to legend) represent the frequency of each of the variables.
doi: 10.1371/journal.pone.0081372.g006

genes suggests that their virulence could be attenuated. The fact that these strains share similar PFGE profiles with clinical

epidemic isolates suggests that they are not different populations that co-circulated with epidemic isolates, but were
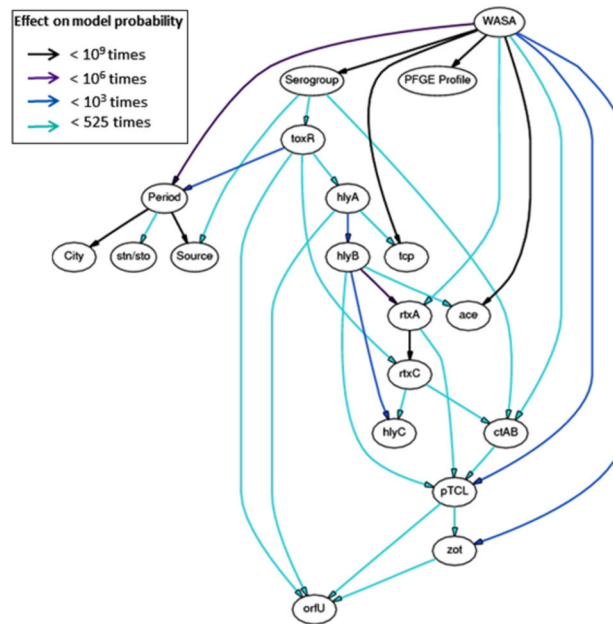
**Figure 7. Dependency model of multivariate data from strains.** Bayesian network representing conditional probabilities of variables that were available for the strains. Arcs are colored according to the impact in the posterior probability of the model when the arc is removed. The network represents the end result of the evaluation of $4.5 * 10^7$ different topologies, in which the last $1.4 * 10^7$ evaluations did not yield a better model. The network was constructed using the online B-Course software [42].

indeed originally epidemic strains that lost genes while struggling in the environment. Nevertheless, PCR evidence is not sufficient to conclusively determine the absence of these genes. Complete genome sequences, which our group is currently analyzing, will provide further evidence about what genomic events occurred in the non-amplified regions of these strains.

Two important issues can be distinguished from comparative analysis of the strains in our study. The first is the virulence profile and virulence potential that the strains exhibit regarding their gene composition. The second is the relation between different populations, which can suggest patterns regarding their origin and dispersion. Genome restriction-pattern analysis through PFGE is limited to elucidating recent clonal divergence and generally fails in reconstructing the genetic relation between populations that diverged over a longer period of time [57]. Since identity and diversity are measured according to similar genome fragment sizes that are within two restriction sites, both cryptic similarity and an overestimation of diversity are possible. Knowledge of the strains and the use of other methods are invaluable to identify and correct possible inconsistencies.

In our study, all of the toxigenic O1 El Tor strains were clustered within three groups which were the most similar to each other (Figure **4**). One of these groups was shared with other O1 El Tor strains that were *ctxAB* negative but harbored other CTXφ genes, suggesting gene-loss events rather than a misclustering of a separated group. All of these strains were positive for the *WASA-1* prophage, which reinforces their

identity as epidemic strains [30]. One of the O1 strains (V.735), which was isolated during the epidemic period, was *ctxAB* negative and clustered with other environmental NAGs rather than with the three aforementioned mentioned O1 El Tor clusters. This strain was positive for the *WASA-1* prophage, and harbored the *tcpA* gene. A closer look at its band profile shows that only two of its bands were not shared by any other strains of the El Tor clusters, strongly suggesting that this was likewise originally an epidemic strain.

Attempts to draw conclusions regarding the abundance or true diversity of the *V. cholerae* population in the Brazilian Amazon during the studied period is limited because only a few strains were available. Modeling the interrelations of the multivariate data through a Bayesian network proved to be an efficient manner of structuring the data in a way that facilitates expert analysis and conclusions. In a sense, the data was reduced to a smaller set of interrelations exhibiting quantified uncertainty between them. In this analysis, the *WASA-1* prophage, when compared to other genes that were investigated, exhibited greater importance in discriminating several of the variables that were evaluated.

To consider an environmental strain as being of epidemic origin, at least three of the following conditions were considered: (i) it was isolated during an epidemic period; (ii) it is of the O1 serotype; (iii) it harbors CTXφ genes and *tcpA*; and (iv) it has a similar PFGE pattern as clinical isolates from the epidemic. None of these conditions, alone, would be enough to determine the others. For instance, in epidemic periods indigenous and epidemic strains were isolated from the

environment; O1 isolates from non-epidemic lineages were isolated from clinical and environmental sources; strains with CTXφ genes were found among NAGs, and strains from epidemic origin that were negative for *tcpA* were also identified and; one strain with three of the aforementioned conditions clustered with indigenous strains rather than isolates from the epidemic. All of the strains that harbor the *WASA-1* prophage genes meet three or all four of the conditions, while all strains that are negative for this element do not. Evidence so far shows that investigating the presence of the *WASA-1* prophage among environmental isolates from the Amazon is the best way of determining if the strain is from an epidemic origin or not. Similarly consistent genetic markers from epidemic strains of other places have been identified and suggested for use as molecular identifiers of the strains in clinical and environmental sources. The VCA phage from the toxigenic US Gulf Coast strain and the CTX variant from the Haitian strains are examples of these molecular markers [28,58].

## Supporting Information

**Table S1. Detailed profiles of strains used in this study.**

(PDF)

**Table S2. List of primers used in this study.**
(PDF)

**Figure S1. PFGE cladogram of strains used in this study.**
(PDF)

## References

1. Colwell RR, Kaper J, Joseph SW (1977) *Vibrio cholerae*, *Vibrio parahaemolyticus*, and other vibrios: occurrence and distribution in Chesapeake Bay. Science 198: 394-396. doi:10.1126/science.198.4315.394-a. PubMed: 910135.
2. Huq A, Small EB, West PA, Huq MI, Rahman R et al. (1983) Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. Appl Environ Microbiol 45: 275-283. PubMed: 6337551.
3. Kirschner AK, Schlesinger J, Farnleitner AH, Hornek R, Süss B et al. (2008) Rapid growth of planktonic *Vibrio cholerae* non-O1/non-O139 strains in a large alkaline lake in Austria. Appl Environ Microbiol 74: 2004-2015. doi:10.1128/AEM.01739-07. PubMed: 18245230.
4. Worden AZ, Seidel M, Smriga S, Wick A, Malfatti F et al. (2006) Trophic regulation of *Vibrio cholerae* in coastal marine waters. Environ Microbiol 8: 21-28. doi:10.1111/j.1462-2920.2005.00863.x. PubMed: 16343318.
5. Huq A, West PA, Small EB, Huq MI, Colwell RR (1984) Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. Appl Environ Microbiol 48: 420-424. PubMed: 6486784.
6. Hasan NA, Choi SY, Eppinger M, Clark PW, Chen A et al. (2012) Genomic diversity of 2010 Haitian cholera outbreak strains. Proc Natl Acad Sci U S A 109: E2010-E2017. doi:10.1073/pnas.1207359109. PubMed: 22711841.
7. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci U S A 106: 15442-15447. doi:10.1073/pnas.0907787106. PubMed: 19720995.
8. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH et al. (2011) Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477: 462-465. doi:10.1038/nature10392. PubMed: 21866102.
9. Karaolis DK, Lan R, Reeves PR (1995) The sixth and seventh cholera pandemics are due to independent clones separately derived from environmental, nontoxigenic, non-O1 *Vibrio cholerae*. J Bacteriol 177: 3191-3198. PubMed: 7768818.
10. Cho YJ, Yi H, Lee JH, Kim DW, Chun J (2010) Genomic evolution of *Vibrio cholerae*. Curr Opin Microbiol 13: 646-651. doi:10.1016/j.mib.2010.08.007. PubMed: 20851041.
11. Feng L, Reeves PR, Lan R, Ren Y, Gao C et al. (2008) A recalibrated molecular clock and independent origins for the cholera pandemic clones. PLOS ONE 3: e4053. doi:10.1371/journal.pone.0004053. PubMed: 19115014.
12. Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB (2012) Cholera. Lancet 379: 2466-2476. doi:10.1016/S0140-6736(12)60436-X. PubMed: 22748592.
13. Salles CA, Momen H (1991) Identification of *Vibrio cholerae* by enzyme electrophoresis. Trans R Soc Trop Med Hyg 85: 544-547. doi:10.1016/0035-9203(91)90251-S. PubMed: 1755069.
14. Popovic T, Bopp C, Olsvik O, Wachsmuth K (1993) Epidemiological application of a standardized ribotype scheme for *Vibrio cholerae* O1. J Clin Microbiol 31: 2474-2482. PubMed: 7691876.
15. Coelho A, Vicente AC, Baptista MA, Momen H, Santos FA et al. (1995) The distinction of pathogenic *Vibrio cholerae* groups using arbitrary primed PCR fingerprints. Res Microbiol 46: 671-683.
16. Faruque SM, Roy SK, Alim AR, Siddique AK, Albert MJ (1995) Molecular epidemiology of toxigenic *Vibrio cholerae* in Bangladesh studied by numerical analysis of rRNA gene restriction patterns. J Clin Microbiol 33: 2833-2838. PubMed: 8576328.
17. Colwell RR (1996) Global climate and infectious disease: the cholera paradigm. Science 274: 2025-2031. doi:10.1126/science.274.5295.2025. PubMed: 8953025.
18. Bompangue Nkoko D, Giraudoux P, Plisnier PD, Tinda AM, Piarroux M et al. (2011) Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978-2008. Emerg Infect Dis 17: 2026-2034. PubMed: 22099090.
19. Borroto RJ, Martinez-Piedra R (2000) Geographical patterns of cholera in Mexico, 1991-1996. Int J Epidemiol 29: 764-772. doi:10.1093/ije/29.4.764. PubMed: 10922357.
20. Alam M, Hasan NA, Sadique A, Bhuiyan NA, Ahmed KU (2006) Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environments of Bangladesh. Appl Environ Microbiol 72: 4096-40104. doi:10.1128/AEM.00066-06. PubMed: 16751520.
21. Ruiz-Moreno D, Pascual M, Bouma M, Dobson AP (2007) Cholera seasonality in Madras (1901-1940): dual role for rainfall in endemic and epidemic regions. Ecohealth 4: 52-62. doi:10.1007/s10393-006-0079-8.
22. Constantin de Magny G, Murtugudde R, Sapiano MR, Nizam A, Brown CW et al. (2008) Environmental signatures associated with cholera epidemics. Proc Natl Acad Sci U S A 105: 17676-17681. doi:10.1073/pnas.0809654105. PubMed: 19001267.
23. Colombara DV, Cowgill KD, Faruque AS (2013) Risk factors for severe cholera among children under five in rural and urban Bangladesh, 2000-2008: a hospital-based surveilance study. PLOS ONE 8: e54395. doi:10.1371/journal.pone.0054395. PubMed: 23349875.

24. Rosewell A, Addy B, Komnapi L, Makanda F, Ropa B et al. (2012) Cholera risk factors, Papua New Guinea, 2010. BMC Infect Dis 12: 287. doi:10.1186/1471-2334-12-287. PubMed: 23126504.

25. Sidley P (2008) Floods in southern Africa results in cholera outbreak and displacement. BMJ 336: 471. doi:10.1136/bmj.39503.700903.DB.

26. Constantin de Magny G, Guégan JF, Petit M, Cazelles B (2007) Regional-scale climate-variability synchrony of cholera epidemics in West Africa. BMC Infect Dis 7: 20. doi:10.1186/1471-2334-7-20. PubMed: 17371602.

27. Speelmon EC, Checkley W, Gilman RH, Patz J, Calderon M et al. (2000) Cholera incidence and El Niño-related higher temperature. JAMA 283: 3072-3074. doi:10.1001/jama.283.23.3072-a.

28. Wachsmuth IK, Bopp CA, Fields PI, Carrillo C (1991) Difference between toxigenic *Vibrio cholerae* O1 from South America and US gulf coast. Lancet 337: 1097-1098. doi:10.1016/0140-6736(91)91744-F. PubMed: 1673518.

29. de Sá Morais LL, Garza DR, Loureiro EC, Nunes KN, Vellasco RS et al. (2012) Complete genome sequence of a sucrose-nonfermenting epidemic strain of *Vibrio cholerae* O1 from Brazil. J Bacteriol 194: 2772. doi:10.1128/JB.00300-12. PubMed: 22535947.

30. Garza DR, Thompson CC, Loureiro EC, Dutilh BE, Inada DT et al. (2012) Genome-wide study of the defective sucrose fermenter strain of *Vibrio cholerae* from the Latin American cholera epidemic. PLOS ONE 7: e37283. doi:10.1371/journal.pone.0037283. PubMed: 22662140.

31. Goodgame RW, Greenough WB (1975) Cholera in Africa: a message for the West. Ann Intern Med 82: 101-106. doi: 10.7326/0003-4819-82-1-101. PubMed: 1235746.

32. Levine MM, Black RE, Clements ML, Cisneros L, Saah A (1982) The pathogenicity of nonenterotoxigenic *Vibrio cholerae* serogroup O1 biotype El Tor isolated from sewage water in Brazil. J Infect Dis 145: 296-299. doi:10.1093/infdis/145.3.296. PubMed: 7061878.

33. Rivera IN, Chun J, Huq A, Sack RB, Colwell RR (2001) Genotypes associated with virulence in environmental isolates of *Vibrio cholerae*. Appl Environ Microbiol 67: 2421-2429. doi:10.1128/AEM. 67.6.2421-2429.2001. PubMed: 11375146.

34. Chakraborty S, Mukhopadhyay AK, Bhadra RK, Ghosh AN, Mitra R et al. (2000) Virulence genes in environmental strains of *Vibrio cholerae*. Appl Environ Microbiol 66: 4022-4028. doi:10.1128/AEM. 66.9.4022-4028.2000. PubMed: 10966424.

35. Cooper DB (1975) Brazil´s long fight against epidemic disease, 1849-1917, with emphasis on yellow fever. Bull N Y Acad Med 51: 672-696. PubMed: 1093591.

36. Vicente ACP, Coelho AM (2005) 1990s *Vibrio cholerae* epidemic, Brazil. Emerg Infect Dis 11: 171-172. PubMed: 15714663.

37. Waldman EA, Antunes JL, Nichiata LY, Takahashi RF, Cacavallo RC (2002) Cholera in Brazil during 1991-1998: socioeconomic characterization of affected areas. J Health Popul Nutr 20: 85-92. PubMed: 12022164.

38. Nandi B, Nandy RK, Mukhopadhyay S, Nair GB, Shimada T, Ghose AC (2000) Rapid method for species-specific identification of *Vibrio cholerae* using primers targeted to the gene of outer membrane protein OmpW. J Clin Microbiol 38: 4145-4151. PubMed: 11060082.

39. Sambrook J, Russell DW (2001) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press.

40. Cooper KL, Luey CK, Bird M, Terajima J, Nair GB et al. (2006) Development and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio cholerae*. Foodborne Pathog Dis 3: 51-58. doi:10.1089/fpd.2006.3.51. PubMed: 16602979.

41. Jensen FV (2001) Bayesian networks and decision graphs. Springer-Verlag.

42. Myllymaki P, Silander T, Tirri H, Uronen P (2002) B-Course: a web-based tool for Bayesian and casual data analysis. Int J Artif Intell T 3: 369-387.

43. Silander T, Myllymaki P (2012) A simple approach for finding the globally optimal Bayesian network structure. arXiv: 1206.6875

44. Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272: 1910-1914. doi:10.1126/science.272.5270.1910. PubMed: 8658163.

45. Taylor RK, Miller VL, Furlong DB, Mekalanos JJ (1987) Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc Natl Acad Sci U S A 84: 2833-2837. doi:10.1073/pnas.84.9.2833. PubMed: 2883655.

46. Peterson KM, Mekalanos JJ (1988) Characterization of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonization. Infect Immun 56: 2822-2829. PubMed: 2902009.

47. Reidl J, Klose KE (2002) *Vibrio cholerae* and cholera: out of the water and into the host. FEMS Microbiol Rev 26: 125-139. doi:10.1111/j. 1574-6976.2002.tb00605.x. PubMed: 12069878.

48. Hassan F, Kamruzzaman M, Mekalanos JJ, Faruque SM (2010) Satellite phage TLCφ enables toxigenic conversion by CTX phage through dif site alteration. Nature 467: 982-985. PubMed: 20944629.

49. Coelho A, Andrade JR, Vicente AC, Salles CA (1995) New variant of *Vibrio cholerae* O1 from clinical isolates in Amazonia. J Clin Microbiol 33: 114-118. PubMed: 7535309.

50. Sá LLC, Vale ERV, Garza DR, Vicente ACP (2012) *Vibrio cholerae* O1 from superficial water of the Tucunduba stream, Brazilian Amazon. Braz J Microbiol 43: 635-638. doi:10.1590/ S1517-83822012000200027. PubMed: 24031874.

51. Theophilo GN, Rodrigues DP, Leal NC, Hofer E (2006) Distribution of virulence markers in clinical and environmental *Vibrio cholerae* non-O1/ non-O139 strains isolated in Brazil from 1991 to 2000. Rev Inst Med Trop Sao Paulo 48: 65-70. doi:10.1590/S0036-46652006000200002. PubMed: 16699625.

52. Thompson FL, Thompson CC, Vicente AC, Theophilo GN, Hofer E et al. (2003) Genomic diversity of clinical and environmental *Vibrio cholerae* strains isolated in Brazil between 1991 and 2001 as revealed by fluorescent amplified fragment length polymorphism analysis. J Clin Microbiol 41: 1946-1950. doi:10.1128/JCM.41.5.1946-1950.2003. PubMed: 12734232.

53. Vital Brazil JM, Alves RM, Rivera IN, Rodrigues DP, Karaolis DK et al. (2002) Prevalence of virulence-associated genes in clinical and environmental *Vibrio cholerae* strains isolated in Brazil between 1991 and 1999. FEMS Microbiol Lett 215: 15-21. doi:10.1111/j. 1574-6968.2002.tb11364.x. PubMed: 12393195.

54. Leal NC, de Araújo Figueiroa AC, Cavalcanti VO, da Silva SC, Leal-Balbino TC et al. (2008) Characterization of *Vibrio cholerae* isolated from the aquatic basins of the state of Pernambuco, Brazil. Trans R Soc Trop Med Hyg 102: 272-276. doi:10.1016/j.trstmh.2007.12.008. PubMed: 18258274.

55. Fasano A, Baudry B, Pumplin DW, Wasserman SS, Tall BD et al. (1991) *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. Proc Natl Acad Sci U S A 88: 5242-5246. doi: 10.1073/pnas.88.12.5242. PubMed: 2052603.

56. Trcksis M, Conn TL, Wasserman SS, Sears CL (2000) *Vibrio cholerae* ACE stimulates Ca(2+)-dependent Cl(-)/HCO(3)(-) secretion in T84 cells in vitro. Am J Physiol Cell Physiol 279: C567-C577. PubMed: 10942706. Available online at: PubMed: 10942706

57. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE et al. (1995) Interpreting chromosomal DNA restriction pattern produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol 33: 2233-2239. PubMed: 7494007.

58. Talkington D, Bopp C, Tarr C, Parsons MB, Dahourou G et al. (2011) Characterization of toxigenic *Vibrio cholerae* from Haiti, 2010-2011. Emerg Infect Dis 17: 2122-2129. PubMed: 22099116.