The Role of *Vibrio cholerae* Genotyping in Africa

Rituparna De,1 Jayeeta Banerjee Ghosh,1 Sourav Sen Gupta,1 Yoshifumi Takeda,2 and G. Balakrish Nair3

1National Institute of Cholera and Enteric Diseases, Beliaghata, 2Collaborative Research Center of Okayama University for Infectious Diseases in India, Kolkata, and 3Translational Health Science and Technology Institute, Gurgaon, Haryana, India

**Toxigenic *Vibrio cholerae***, the causative agent of the disease cholera, is prevalent in the African continent from the 1970s when the seventh pandemic spread from Asia to Africa. In the past decade, cholera has caused devastating outbreaks in much of Africa, illustrated by the recent cholera epidemics in Zimbabwe and regions of central Africa. Given the extent of cholera in Africa, a robust and efficient surveillance system should be in place to prevent and control the disease in this continent. Such a surveillance system would be greatly bolstered by use of molecular typing techniques to identify genetic subtypes. In this review, we highlight the role that modern molecular typing techniques can play in tracking and aborting the spread of cholera.

**Keywords.** Africa; cholera; genotyping; surveillance; *Vibrio cholerae*.

*Vibrio cholerae* is a water- and food-borne pathogen that causes cholera, the severe watery diarrheal disease. According to the World Health Organization (WHO), there are an estimated 3–5 million cholera cases and 100 000–120 000 deaths due to cholera every year [1]. In May 2011, the World Health Assembly (WHA) recognized the reemergence of cholera as a significant global public health problem and adapted resolution WHA 64.15, calling for the implementation of an integrated comprehensive global approach to cholera control [2]. The grim picture of cholera in recent years is reflected by a 130% increase in the number of cases of cholera from 2000 to 2010 and an overall increase of 43% cholera cases from 2009 to 2010 [2]. In 2010, 48 countries reported the incidence of the disease.

**CHOLERA IN AFRICA**

In the 1960s, at the beginning of the seventh and current cholera pandemic, cholera had an exclusively Asian focus [3]. The current pandemic started in South Asia in 1961, reached Africa in 1970, where it has remained entrenched. In 2005, 31 (78%) of the 40 countries that reported indigenous cases of cholera to WHO were in sub-Saharan Africa [4]. The reported incidence of indigenous cholera in sub-Saharan Africa in 2005 (166 cases/million population) was 95 times higher than the reported incidence in Asia (1.74 cases/million population) and 16 600 times higher than in Latin America (0.01 cases/million population) [4]. In the same year, the cholera case fatality rate in sub-Saharan Africa (1.8%) was 3 times higher than that in Asia (0.6%); no cholera deaths were reported in Latin America [3]. In 2008, Africa witnessed one of the worst outbreaks, which started in mid-August in Zimbabwe and spread to cause cholera in the borders of Botswana, Mozambique, South Africa, and Zambia [5]. In 2010, Africa was again hit by a devastating wave of cholera, which affected regions of Central Africa—namely, Cameroon, Chad, Niger, Nigeria, and around the Lake Chad Basin [2]. In Africa, so far, cholera is caused by *V. cholerae* serogroup O1 El Tor. The other serogroup that causes cholera, O139, has not entered the African continent as yet.

Given the extent of cholera in Africa, a robust and efficient surveillance system should be in place to prevent and control the disease in this continent. The layer of laboratories that would allow an ideal surveillance system for cholera is shown in Figure 1. Such a system should focus on detecting the earliest cases of cholera in peripheral and remote areas and then rapidly institute control measures to contain the spread of the outbreak.

There are already a few disease surveillance programs for cholera and other preventable diseases that are operational in the African region. The Integrated Disease Surveillance...
and Response (http://www.cdc.gov/globalhealth/dphswd/idsr/) and the African Cholera Surveillance Network (http://www.africhol.org/) have been set up to strengthen the national public health surveillance and response systems in Africa. These multi-partner consortiums are essential for sustainable disease surveillance and can serve as early warning systems to significantly alleviate the number of cholera and other preventable disease cases in Africa. In this review, we present information on the role of different genotyping techniques as powerful tools in understanding the epidemiology of cholera in Africa.

**TAXONOMY AND SUBSPECIES CLASSIFICATION**

*V. cholerae* is a gram-negative, oxidase-positive, sucrose-fermenting γ-proteobacterium belonging to the family *Vibrionaceae*. So far, 210 serogroups have been discovered, of which only the O1 and O139 serogroups are the causative agents of cholera and have endemic and epidemic potential. O1 *V. cholerae* has 2 distinct biotypes (the Classical and EI Tor), and each biotype has 3 serotypes (Ogawa, Inaba, and Hikojima) (see Figure 2). The serogroups other than O1 and O139 are not associated with cholera and may cause sporadic cases and are collectively referred to as non-O1/non-O139 *V. cholerae*. Three non-O1 serogroups that have been associated with large outbreaks are O37 [6], O75 [7], and O141 [8]. Recent studies have shown that the O139 serogroup is a derivative of the prototype El Tor biotype strains [9]. Additional variants of the prototype El Tor, such as the Matlab strains [10], the Mozambique strains [11], and the altered El Tor strains [12], have been reported. The altered variant of El Tor strains in which the B-subunit of cholera toxin is of the classical type currently dominates globally.

![Figure 1. The capacity and expected outcome of different strata of laboratories required for a robust surveillance system and for prompt outbreak detection.](image1)

![Figure 2. Current classification of *Vibrio cholerae* and the derivatives of the El Tor biotype.](image2)
GENOTYPING METHODS

Not only is the identification of the etiologic agent important but the strains also require to be further characterized as diversity exists between strains of the same species. The efficiency of typing systems depends on typability, reproducibility, discriminatory ability, and ease to perform, as well as cost, efficiency, and time. The results produced by the typing system should be clear and unambiguous.

The strains isolated may be from a common source or from different sources, and they may be from the same or different geographical locations, so it becomes important for the epidemiologist to arrive at a conclusion based on the genetic relatedness of the various isolates. Strains that are genetically related are usually derived from a single cell; thus, its progeny expresses the same biochemical properties, virulence factors, and genomic traits. These strains are said to belong to the same clone [13]. Genotypic methods help in discriminating between clones, identifying phylogeny, tracking spread and transmission, and tracing the origin of the source of infection. Genotypic methods and their roles that have aided the molecular typing of sporadic, endemic, and epidemic V. cholerae are discussed in this review.

Plasmid Fingerprinting
The method of plasmid fingerprinting is based on cell lysis followed by gel electrophoresis. In case of very large plasmids, restriction digestion using a suitable endonuclease is employed to generate fragments that can be separated by gel electrophoresis. The pattern of the fragments is very distinct for strains and serves as a marker for differentiating between strains. However, because many fragments are produced, interpretation becomes difficult.

The limitation is that many strains lack plasmids; therefore, this technique cannot be applied [14]. In V. cholerae, plasmid fingerprinting can distinguish between the 2 biotypes of O1 serogroup as well as O1 from non-O1 non-O139. The 2 biotypes of V. cholerae O1 could be differentiated by using the 628 bp IS1004 element [15]. The pattern produced by non-O1 strains is distinguishable from that of epidemic strains.

Restriction Enzyme Analysis, Restriction Fragment Length Polymorphism, and Ribotyping
Restriction enzyme analysis (REA) exploits the presence of restriction endonuclease cutting sites called restriction sites in the bacterial genome. When a restriction endonuclease is used to digest total genomic DNA (chromosomal and plasmid), it fragments the DNA at specific restriction sites, yielding a large number of restriction fragments in the size range of 0.5 kilobase pairs (kb) to 50 kb in length that can be separated by agarose gel electrophoresis. Any single-nucleotide polymorphisms or insertions and deletions tend to abolish existing or create new restriction sites and as a result, the fragment size and number would also change accordingly. This pattern can convey to the epidemiologist about the genetic change that has taken place. However, larger fragments tend to overlap and hence the result gets indistinct and ambiguous.

Restriction fragment length polymorphism (RFLP), a modification of REA, employs hybridization methods and has been widely used to study the molecular epidemiology of V. cholerae. Fragments generated by restriction digestion are separated by conventional gel electrophoresis and then transferred onto a nitrocellulose or nylon membrane where they are hybridized using chemically or radioactively labeled probes. RFLP has been used to compare indigenous and imported cases of cholera and to analyze strains from outbreak and sporadic cases [16].

Ribotyping is a method closely related to RFLP in which chromosomal DNA preparations are hybridized with a ribosomal RNA probe [17]. Based on in silico genomics, it has been demonstrated that ribotype polymorphisms result from sequence variability in the neutral housekeeping genes flanking rRNA operons, with rRNA gene sequences serving solely as conserved, flank-linked tags [18]. However, this method has limited discriminatory ability.

Pulsed-Field Gel Electrophoresis
Pulsed-field gel electrophoresis (PFGE) emerged as a method to examine large eukaryotic chromosomes. Bacterial genomes that are 2000–5000 kb in size are digested using a rare cutter that generates 10–30 restriction fragments of 10–800 kb. NorI is currently the enzyme of choice particularly for V. cholerae. Strains can be distinguished on the basis of the banding pattern it generates. Analysis of PFGE is based on standardized criteria, which are followed worldwide so that there is no ambiguity about results produced when performed in different laboratories [19].

Cameron et al [20] first showed the usefulness of PFGE for V. cholerae and its higher discriminatory prowess compared to other existing methods like multilocus enzyme electrophoresis (MLEE) and ribotyping. PFGE was able to discriminate strains more efficiently as compared with MLEE and ribotyping.

Randomly Amplified Polymorphic DNA
Polymerase chain reaction (PCR)–based techniques like AP-PCR (arbitrarily primed PCR), also called randomly amplified polymorphic DNA (RAPD) assay have been useful to identify and characterize strains within a short time. In this method, a single short, 10-base-pair primer set, which is not specific for any particular gene and randomly hybridizes at various locations on the chromosome, is used to amplify the genomic DNA. The resulting PCR products represent a variety of different-sized DNA fragments, which are separated by agarose gel electrophoresis. However, this technique lacks uniform and standardized interpretation guidelines. It is also susceptible to technical variation.
Multilocus Sequence Typing
In the methods described so far, genetic variation could be observed in the form of variation of banding patterns, but these methods are insufficient to show where in the nucleotide sequence the variation has arisen. In sequencing methods, the nucleotide sequence of the gene or plasmid or whole genome is available; hence, a more detailed analysis can be done. Multilocus sequence typing (MLST), developed in 1998 [21], is a method that is based on the nucleotide sequence variation directly. Here, multiple loci of housekeeping genes of ideally 450–500 bp are simultaneously examined by DNA sequencing.

Olsvik et al [22] sequenced the ctxB gene and showed that there are distinct nucleotide sequences for the 2 biotypes. If this underlying information was not revealed, then the El Tor variants, which have been isolated since 1994 and phenotypically resemble the El Tor prototype strains, would not have been classified as a variant [23]. Due to the application of DNA sequencing, the whole genome sequence [24] and, thereafter, differences in the nucleotide sequence of a host of genes (such as tcpA, hlyA, rstR, and many more) could be determined, as there are separate biotype specific sequences of these genes. This determination has been immensely useful in understanding the genetic composition of newly evolved strains like the variant strains. Recently, a third mutation in the ctxB gene has been reported. According to this finding, histidine at the 20th amino acid position (which was present in all available sequences of the V. cholerae ctxB gene) has been replaced with asparagine [25]. This mutation assumed significance when it was determined that the cholera epidemic in Haiti since October 2010 was caused by strains of V. cholerae O1 carrying the third mutation. How such changes make an impact on the epidemiology of V. cholerae O1 still remains unclear.

MLST has been shown to be a better typing technique than PFGE as far as discriminatory ability is concerned [26] and is probably the best suited for population genetic studies, although it is expensive. Another drawback is its occasional limited discriminatory power, as it is only based on sequences of housekeeping genes, which are highly conserved. In order to overcome this drawback, multivirulence-locus sequence typing, based on highly polymorphic virulence gene loci, was developed.

Molecular typing techniques have revealed extremely valuable information on the epidemiology of V. cholerae over the decades. However, PFGE in particular has been adopted as a well-suited technique by a number of investigators as part of a global laboratory network (PulseNet International; http://www.cdc.gov/pulsenet/) for typing food-borne pathogens, including V. cholerae, allowing real-time comparison of pathogen profiles through a dynamic database. Standardized protocols used across all laboratories allow comparison of PFGE profiles globally, aiding in determining outbreak source, disease clusters, and pathogen transmission.

ROLE OF GENOTYPING METHODS IN EPIDEMIOLOGY OF CHOLERA
From a translational research point of view, genotyping methods would yield important knowledge on rapid genetic alterations and therefore enable the detection of emergence of new subtypes. Genotyping methods would also yield important information on antimicrobial resistance. This information would contribute to improving methods of diagnosis and treatment of cholera.

Listed below are some of the important roles of genotyping methods in epidemiology of cholera:

Differentiating Strains of Different Origin and Identifying Clusters
Molecular typing techniques have been successfully used to determine whether strains of the same or different clones circulate and cause outbreaks in a given setting. These techniques have aided investigation by enabling the comparison of strains from different geographical areas and examining if there is any genetic relatedness among them. Molecular typing techniques like PFGE and MLST have been used to group strains into different clusters and examine the distribution of these clusters and their role in outbreaks [26]. During outbreaks, isolated strains have been characterized using techniques such as RAPD to reveal variations to identify clusters [27]. Dalsgaard et al [28] conducted a study to compare strains of V. cholerae O1 isolated in the years 1987, 1994, and 1995 that caused outbreaks of cholera in Guinea-Bissau in Africa. They used ribotyping and arrived at the conclusion that there were 2 different clusters responsible for the outbreaks. One cluster was exclusively associated with strains from the 1987 outbreak, while the other cluster caused outbreaks in 1994 and 1995, and the 2 ribotype patterns from the 2 clusters were not identical [28].

Transmission: Trace Back and Tracking Spread of Strains
Researchers at the University of Sydney [29] used amplified fragment length polymorphism (AFLP) to understand the relationship among different strains from African isolates of pandemic V. cholerae during the 1970s; although the majority of them were in the same subcluster, there were at least 2 separate introductions of cholera during that period. Although there was no epidemiological evidence, their study showed that epidemics in north and east Africa originated from different sources. They found that isolates from west Africa (Sierra Leone and Senegal) were identical, but were different from east Africa (Ethiopia) in 5 of the 50 informative AFLP bands and from the isolates from north Africa (Morocco). Although the isolates from north Africa were expected to be closer to those from east Africa, they were actually closer to the west African isolates from north Africa.
isolates, with only 2 differences in informative bands. Cholera remained at low levels in Africa after this epidemic, and only came back at high levels in 1991 with 2 main epidemic foci: in south and east Africa (Zambia, Mozambique, Malawi, and Angola) and in west Africa [30]. Comparing AFLP profiles of isolates from this epidemic to those of the 1970s and later, it could be confirmed that strains causing outbreaks in west Africa were clearly from Asia and have been present in the continent for at least 3 years prior to the epidemic in 1991[29]. A continuous flow of cholera into Africa from regions of endemcity such as Asia, and the power of AFLP to trace the source of pandemic clones, were exhibited. Another study using whole-genome sequencing of 154 globally and temporally represented V. cholerae isolates to identify high-resolution markers (single-nucleotide polymorphisms) found that the seventh pandemic had spread from the Bay of Bengal in at least 3 independent but overlapping waves with several transcotinental transmission events [31], with a common ancestor in the 1950s. This study was able to construct a global high-resolution phylogeny for pandemic V. cholerae that not only pointed at a single source of origin for the current seventh pandemic, but also provided a framework independent of analysis of subgenomic regions, such as mobile elements, for future epidemiological and phenotypic analysis of V. cholerae, including transmission-tracking and typing.

Phylogenetic Analysis
PFGE and MLST have been used to study the phylogenetic relationship among different strains by studying different loci or by means of dendogram analysis. These techniques have also been used to study the relationship between epidemic and nonepidemic strains [26] and also to study the relationship between clinical and environmental isolates and relatedness among different serogroups [32]. Environmental and epidemic clusters could be easily identified on the basis of plasmid fingerprinting. This study showed that strains of different serogroups may possess same fingerprint pattern and vice versa [15]. Byun et al [33] confirmed the phylogenies inferred from the MLEE studies by comparative nucleotide sequence analysis of housekeeping genes, which detects synonymous as well as nonsynonymous substitutions in chromosomally located genes for pair-wise comparisons and construction of phylogenetic trees.

Identification of Genetic Markers
New genetic regions that are inserted or deleted can be examined with the help of molecular typing techniques [34]. In V. cholerae O1 seventh-pandemic strains, 2 such new regions that were missing in the pre-seventh pandemic O1 El Tor strains and also absent in Classical biotype strains are Vibrio seventh pandemic–I and Vibrio seventh pandemic–II. Only on genetic analysis of the different regions of V. cholerae genome could this information be revealed [35]. New genetic elements that are inserted or deleted can be studied to trace the origin of strains [36] and can even serve as genetic markers for differentiating strains on the basis of presence or absence of these markers.

Study of Genetic Rearrangement
Genetic rearrangement in different parts of the genome can be examined to study the evolution of new clones from existing ones [37]. In Calcutta, India, the O139 Bengal strain that appeared in 1992 was replaced by O1 El Tor strains for 6 months. These El Tor strains that reemerged had different genetic constitution compared to the preexisting ones [38]. Again, when O139 reemerged after a short period of disappearance, their characterization revealed genetic rearrangement—as a result, the new ribotype pattern for these strains isolated in 1995 was different from the ribotype of the previous O139 Bengal strains. In addition, they were different from the previous O139 strains in terms of resistance to antibiotics like trimethoprim, sulfamethoxazole, and streptomycin, encoded by the SXT element, a conjugative, self-transmissible, integrating element. On molecular analysis it was found that all the strains with the new ribotype had a deletion of 3.6 kb in the SXT element; thus, unlike the previous O139 strains, they were sensitive to these antibiotics [39].

Predicting Outbreaks
Bik et al [15] demonstrated 2 examples of non-O1 strains with a fingerprint resembling that of epidemic O1 strains. These were the O139 Bengal strain and an O37 strain. Bik et al showed that the O37 Sudan strain [6] is genetically closely related to classical O1 strains. Similar to O139 Bengal, O37 Sudan lacked most of the O1 antigen cluster but did contain flanking genes. Thus, O37 Sudan represents a second example of an epidemic V. cholerae strain carrying non-O1 antigens. Their study underlines the importance of genotypic methods for the differentiation of V. cholerae strains and for the recognition of strains with epidemic potential.

Developing Databases
Developing databases for networking and organizing genotyping data from different national and regional laboratories are essential for comprehensive and real-time understanding and tracking of cholera epidemics and outbreaks. These databases could prove to be extremely useful for early warning of disease outbreaks and understanding the transmission of various strains during epidemic situations. The success of such databases lies in utilizing standardized genotyping methods across all laboratories sharing and utilizing information derived from these databases. One such initiative is PulseNet International, which partners globally with different laboratories. The database contains standardized PFGE profiles along with molecular epidemiological information of a large number of food-borne pathogens, including a large number of V. cholerae isolates.
These PFGE typing profiles can be used for unambiguous assessment of transmission routes, and to assess sources of infection and changing genetic patterns for timely public health recognition and response to disease outbreaks.

**BENEFITS OF TYPING**

Taking into account the burden of cholera in Africa, it is of prime importance to understand the pattern of outbreaks and the type of strains that cause these outbreaks, and to trace their origin and spread. Without genotyping techniques, it would be difficult to get sufficient discriminating information and thereby would impede epidemiological investigation. This, in turn, would affect dissemination of information to the sentinel centers, and would prolong the epidemic and delay treatment measures. Conventional methods can undoubtedly detect the causative agent in case of sporadic cases or outbreaks. However, taking into consideration the rapid genesis of antibiotic resistance and changes at the genomic level due to the high level of recombination in aquatic reservoirs and in the host, the conventional and immunochromatographic methods [40] would fail to detect the most recent changes at the genetic level.

**LIMITATIONS OF MOLECULAR TYPING METHODS**

Typing methods were initially expensive, laborious, and complex [41], due to which they were not adopted by less equipped laboratories. However, with time, these techniques have become part and parcel of most microbiology laboratories.

On comparing the limitations of the different techniques, one can arrive at a conclusion about the feasibility and practical applicability of the technique. For example, MLST is believed to be better than PFGE in terms of discrimination and speed. MLST examines the entire genetic fragment that is being sequenced, while PFGE can study variation only at specific restriction sites and variations arising due to large insertions or deletions of DNA. These methods are costly, requiring expensive and sophisticated apparatus and technical expertise to perform and analyze the results. Such requirements may hinder the universal application of these techniques in the epidemiological scenario.

**CONCLUSION**

It was during the seventh pandemic that Africa was for the first time exposed to cholera. It spread easily among the population as they lacked immune protection against the disease, which was due to the prototype El Tor biotype strains. Since then, no part of Africa has been spared by the menace of cholera, and since its entry into the continent, it has become a chronic problem. This disease is responsible for the high rates of mortality and morbidity, and making Africa the worst affected continent, even in recent times. Epidemiological programs need to predict epidemics; in case one arises out of the failure of the socioeconomic system, the immediate goal must be to prevent deaths by proper management and use of proper antibiotics to curtail its spread.

The surveillance system suggested for cholera (Figure 1) will allow for rapid detection of cholera cases and outbreaks and better management of the disease. It will enhance the preparedness and response capabilities to epidemics as well as provide a rationale for decision making and implementation of timely public health interventions. A robust surveillance on cholera in a country or continent would perhaps be the most useful activity to stem or abort the spread of an outbreak of cholera. It is important to prevent a cholera outbreak from flaring up, and to contain it as close as possible to the epicenter of the outbreak. A national surveillance system is essential to promote the flow of information within the different levels of the health system, and also to strengthen the capacity and skill of existing laboratories. Another important area that the proposed surveillance setup will bring forth is the involvement of the community in first detailing the public health burden of cholera in Africa and then assessing the success of various interventions and measures.

**Notes**

**Financial support.** This work was supported by the Japan Initiative for Global Research Network on Infectious Diseases (J-GRID); Ministry of Education, Culture, Sports, Science and Technology of Japan; and Indian Council of Medical Research (ICMR), New Delhi, India.

**Potential conflicts of interest.** All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**


