Vibrio cholerae El Tor strains linked to global cholera are homogeneous by pulsed-field gel electrophoresis

Fatema-Tuz Johura[1¶], Sahitya Ranjan Biswas[1¶], Shah M Rashed[1], Mohammad Tarequl Islam[1], Saiful Islam[1], Marzia Sultana[1], Anwar Huq[2], Nicholas R Thomson[3], Rita R. Colwell[2,4,5], Haruo Watanabe[6], and Munirul Alam[1*]

[1]icddr, b, Formerly International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh,

[2]Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland, USA

[3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

[4]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[5]Center for Bioinformatics and Computational Biology, University of Maryland Institute of Advanced Computer Studies, University of Maryland, College Park, Maryland 20742, USA

[6]National Institutes of Infectious Diseases (NIID), Tokyo, Japan

¶ These authors contributed equally to this work

[*]**Corresponding author**

Email: munirul@icddrb.org (MA)

Abstract

Vibrio cholerae O1 El Tor, causative agent of the ongoing seventh cholera pandemic, is native to the aquatic environment of the Ganges Delta, Bay of Bengal (GDBB). Recent studies traced pandemic strains to the GDBB and proposed global spread of cholera had occurred via intercontinental transmission. In the research presented here, *Not*I-digested genomic DNA extracted from *V. cholerae* O1 clinical and environmental strains isolated in Bangladesh during 2004 – 2014 was analyzed by pulsed-field gel electrophoresis (PFGE). Results of cluster analysis showed 94.67% of the *V. cholerae* isolates belonged to clade A and included the majority of clinical isolates of spatio-temporal origin and representing different cholera endemic foci. The rest of the strains were estuarine, all environmental isolates from Mathbaria, Bangladesh, and occurred as singletons, clustered in clades B and C, or in the small clades D and E. Cluster analysis of the Bangladeshi strains and including 157 El Tor strains from thirteen countries in Asia, Africa, and the Americas revealed 85% of the total set of isolates belonged to clade A, indicating all were related, yet did not form an homogeneous cluster. Overall, 15% of the global strains comprised multiple small clades or segregated as singletons. Three sub-clades could be discerned within the major clade A, reflecting distinct lineages of *V. cholerae* El Tor associated with cholera in Asia, Africa, and the Americas. The presence in Asia and the Americas of non-pandemic *V. cholerae* El Tor populations differing by PFGE and from strains associated with cholera globally suggests different ecotypes are resident in distant geographies.

Author Summary(words 154)

Cholera is a major health threat, especially in the Ganges Delta, Bay of Bengal (GDBB). *Vibrio cholerae,* causative agent of cholera, is native to the GDBB aquatic environment. Recent genomic studies suggest GDBB is the cholera hotspot where the disease spreads globally via human activity. Pulsed-field gel electrophoresis (PFGE) of *Not*I-digested genomic DNA from *V. cholerae* El Tor endemic cholera strains was done, including Bangladesh aquatic environment and clinical strains from distant geographical regions representing three cholera-prone continents. Results showed the majority of pandemic strains belonged to a major cluster, suggesting clonal relatedness. Ecotypes were detected, indicating geographically specific lineages. It is concluded that epidemic strains in Bangladesh and thirteen countries of Asia, Africa, and the Americas are geographically adapted, with independent evolution of the bacterium in respective geographical regions.

Introduction

Cholera, with seven pandemics reported to date, represents a significant chapter in human history and infectious disease. The acute form of diarrhea caused by *Vibrio cholerae* is related to production of a toxin that triggers the characteristic water loss and severe dehydration of cholera [1]. Cholera remains a threat in many countries, notably where access to safe drinking water is limited. Bangladesh is a developing country where cholera is endemic, with two annual peaks in some regions of the country [2]. The estimated global burden of cholera is 2.86 million cases and 95,000 deaths. In Bangladesh alone, approximately 100,000 million new cases are diagnosed each year, resulting in 4,500 deaths [3]. *V. cholerae* strains are widely distributed globally in many

coastal, estuarine, and brackish water ecosystems as free-living bacterial cells or associated with zooplankton, namely copepods [4-6]. Brackish waters in coastal areas support bacterial populations, with environmental stimuli favorable for bacterial growth prompting cholera outbreaks [5].

V. cholerae O1 is classified into two biotypes, classical and El Tor, based on genetic differences. The seventh and ongoing pandemic is attributed to the El Tor biotype of *V. cholerae* O1 [8]. Beginning in 1961 in Indonesia, the seventh pandemic of cholera included Africa in 1970, Latin America in 1991, and more recently Haiti and Yemen [9-12]. After the initial cases occurred, most of these regions continued to suffer episodes of cholera. There is debate whether the recurrent outbreaks of cholera in Africa and the Americas are caused by distinct intercontinental introduction or resurgence of indigenous clones. A few recent investigations investigated the genetic homogeneity of the 7th pandemic strains and connected both origin and recurrent transmission to a single source, the Bay of Bengal [13]. However, the co-existence of several local lineages, along with the pandemic clones and their regional evolution, has painted a very complex picture of bacterial population dynamics, especially in and out of endemic settings [14]. Thus, monitoring pandemic and local strains became a priority for some investigators, as a means of developing an effective public health model and strategy for controlling the current pandemic and preventing future pandemics of cholera.

Whole genome typing methods, e.g. multi-locus sequence typing (MLST), multi-locus variant analysis (MLVA), ribotyping, random amplification of polymorphic DNA (RAPD), and pulsed field gel electrophoresis (PFGE), have been employed to differentiate isolates and monitor transmission routes [15-18]. PFGE is a DNA fingerprinting method that can discriminate bacterial strains. Before introduction of whole genome sequencing (WGS), epidemiological studies of

cholera relied on PFGE [18]. An earlier study highlighted the value of PFGE in revealing clonality among isolates from two well-defined cholera outbreaks in Malaysia [19]. Intrinsic limitations include restriction digestion being skewed by mobile elements, hence restricted value for phylogeny. Although PFGE does not provide as high resolution as WGS, its stability and reproducibility allow rudimentary, yet comprehensive, analysis of ancestry [18].

In the study reported here, the objective was to understand both regional diversity and global distribution of *V. cholerae* El Tor representing seventh pandemic lineage. Therefore, strains in Bangladesh isolated over a decade were compared with strains from thirteen countries across Asia, Africa, and the Americas. Both environmental and clinical isolates were included since the local environment can influence strains, persisting and becoming epidemic, as well as providing access to autochthonous strains of *V. cholerae* El Tor.

Results

The *Not*I restriction enzyme digested genomic DNA of the test strains into 20 to 23 fragments and the molecular sizes of the DNA fragments ranged from 20 to 350 kb. Digested genomic DNAs of different spatiotemporal origin and resulting different biotype categorizations were subjected to PFGE. The resulting band patterns were analyzed by Dice similarity coefficient and UPGMA clustering methods to determine genetic and ancestral relatedness.

Local diversity and distribution of the Bangladesh isolates

In a dendogram obtained by UPGMA analysis of DNA band patterns, the Bangladesh isolates comprised five different clades, A, B, C, D, and E (Figs 1 and 2). Of the 169 isolates, 160

clustered in clade A, suggesting a single lineage. A single clinical isolate, collected in 1991 from Matlab, clustered with clade A, a clade of predominantly clinical isolates from endemic sites, including estuary villages in Bangladesh. While a few environmental isolates were found to cluster in clade A, other environmental isolates collected in 2012 from Mathbaria exhibited different PFGE patterns and did not join clade A (S1 Fig). Isolates with different pulsotypes included three singletons and a small clade [B(1); C(1); D(6); E(1)] (Fig 1). Singletons in clades B, C, and E were isolates from the aquatic environment. It should be noted that clade D included isolates of both clinical and environmental origin. All clade D isolates possessed *rstR* classical biotype, a characteristic limited to this clade (S1 Table). Most of the Chhatak isolates (27 of 31) had the same pulsotype closely related to *V. cholerae* isolates from Dhaka and Mathbaria (S1 Fig).

Global distribution of clones

When the PFGE banding patterns of *V. cholerae* O1 isolates from Bangladesh were compared with those of 157 isolates collected from thirteen countries across Asia, Africa, and Latin America, the isolates could be differentiated into 16 clades; A through P (Fig 3). Clade A isolates from Bangladesh clustered with 133 of the 293 strains, comprising a majority of isolates from 14 countries and three continents (Figs 3 and 4). Hence, a majority of the *V. cholerae* El Tor isolates from different geographical regions revealed a similar PFGE pattern and fell into a major clade, with country-specific sub-clustering, i.e., subclades within the major clade A. The sub clades of A included *V. cholerae* El Tor strains from Vietnam (n=15), Zambia (n=9), Haiti (n=3), India (n=2), Pakistan (n=3), Sri Lanka (n=1), and most of the strains from Zimbabwe (8 of 12) and Nepal (27 of 39), Fig 4. Three subclades within clade A reflected a broader spatial distinction and 202 of 273 (74%) Asian and African isolates comprising subclade Ia and 60 of 273 (22%) in

subclade Ib. Latin American isolates (16 of 21) in clade A comprised subclade II. Interestingly, the Latin American isolates were predominantly *V. cholerae* prototype ET (S1 Table). A few isolates from Bangladesh, Thailand and one from Zimbabwe fell into sub-clade II (Fig 4). Subclade II *V. cholerae* El Tor strains from Mexico, isolated during 1992 to 1999, were located in a branch, separating them from isolates collected between 2004 and 2008. Three *V. cholerae* El Tor strains isolated during the 2010 Haitian cholera outbreak comprised subclade Ia, with Asian and African strains. Two *V. cholerae* O1 EL Tor strains isolated in Bangladesh during 2011 joined with *V. cholerae* ET strains from Peru, Brazil, and Mexico, based on PFGE, Figs 3 and 4.

Aside from clade A, the clade E isolates from Bangladesh shared PFGE pattern with an ET isolate from Peru and two from Zimbabwe (Fig 5). As with the Bangladesh isolates, locally restricted diversity was observed for isolates from Mexico, reflected by 11 distinct groups in addition to clade A. Groups F-P comprised isolates from Mexico, notably those collected during 2000 to 2004 (Fig 4).

Discussion

In many epidemiological studies, pulsed-field gel electrophoresis (PFGE) is used to discern source attributes of strains from different outbreaks. With the advent of whole genome sequencing and comparative genomics, the once gold standard PFGE is less appreciated as a DNA fingerprinting tool of epidemiological implication. In the study reported here, PFGE analysis of *V. cholerae* O1 El Tor isolates associated with endemic cholera in Bangladesh and thirteen countries of Asia, Africa and the America showed the strains to be related genetically, but not homogenous globally. A majority of the strains comprised a major clade, with divergence

158 noted for non-pandemic and environmental isolates from the Bay of Bengal, Bangladesh and
159 strains from the Gulf of Mexico. According to a recent WGS-based study of a restricted subset of
160 *V. cholerae* clones, those strains were responsible for epidemic cholera worldwide [20-21]. In
161 this study, PFGE was used to analyze clinical strains of *V. cholerae* from different geographical
162 locations and with different genetic and phenotypic characteristics.

163 *V. cholerae* El Tor biotype has dominated clinically over the classical biotype since 1961, the
164 latter having last been isolated in Bangladesh in 1992 [22-23]. Observed co-existence of the two
165 biotypes for such a long time likely resulted in the hybrid characteristics of El Tor with classical
166 biotype attributes, as observed in Bangladesh [24-25]. In this study, the majority of *V. cholerae*
167 El Tor isolates from clinical and environment sources comprised a major clade, suggesting
168 similarity of strains from both sources and associated with epidemics in Bangladesh. Some of the
169 environmental strains did not fall into clade A, suggesting those to be genetically divergent
170 pulsotypes present in a diverse population existing in the environment. Environmental *V.*
171 *cholerae* ET in our study, with a few exceptions, were similar to clinical isolates in clade A.
172 Previous epidemiological surveillance conducted in the Bay of Bengal estuary has shown
173 pathogenic strains can be detected in aquatic habitats, either in the culturable or non-culturable
174 state, depending on the season [26].

175 A major genome-based study postulated the pandemic *V. cholerae* strain originated in Bay of
176 Bengal villages of Asia and transmitted world-wide in three different waves [13]. It was
177 concluded that *V. cholerae* O1 ET has the ability to travel inter-continentally and adapt to its
178 place of introduction by sharing niches with existing microflora in coastal and estuarine regions,
179 including the Gulf coast of Mexico [27]. Notwithstanding the fact that outbreaks occurring after
180 introduction can be attributed to *V. cholerae* and the pathogen can be introduced repeatedly or

adapt locally, either is possible. Whole genome sequencing based studies linked epidemics in Africa and the Americas to multiple introduction events, rather than preexisting pathotypes [8-9]. The PFGE banding patterns observed for the majority of *V. cholerae* O1 included in this study support an intercontinental transmission hypothesis [28], but only in a global context. The observation of country-based subclades indicates an independent evolution of the pandemic pathogen. Genetic changes were reported among initially homogeneous *V. cholerae* O1 ET initiating the Haitian cholera epidemic in 2010 [28]. In this context, *V. cholerae* ET strains associated with the Haitian cholera in 2010 were observed to be closely related to isolates from other Southeast Asian countries [28].

While cholera had not been reported in the Americas for more than a century before 1991, the observed presence of *V. cholerae* classical biotype and diverse *V. cholerae* ET lineages in Mexico was uncharacteristic for a region outside of Asia or Africa, suggesting a capricious nature of the bacterium [29]. Most *V. cholerae* ET isolates in Mexico that diverged separately from the pandemic clones lacked CTX phage [30] and were not related to the non-toxigenic isolates from Thailand [31]. Previously, we had shown that CTX phage negative isolates dominated clinical cases in Mexico during 2001-2004 [30] and studies posited the isolates to be ancestors of the *V. cholerae* responsible for the sixth and seventh pandemics [29] .

While the observed relatedness of PFGE patterns of *V. cholerae* El Tor associated with cholera epidemics in Asia, Africa, and the Americas supports the potential for global transmission of the pandemic pathogen [13], the divergence of strains and their region-specific signatures also support independent evolution of *V. cholerae* locally. Clearly the overall picture is complex and warrants regular monitoring to assist designing effective intervention models to counter future

pandemics. In any case PFGE data, as presented in this study show this technology can be effective for analysis and source-tracking of cholera outbreaks.

Materials and Methods

Geographical profile of isolates

A total of 169 strains were collected between 1991 and 2014 from four endemic sites in Bangladesh: Mathbaria (n=99); Dhaka (n=38); Chhatak (n=31); and Matlab (n=1). Of these, 119 and 50 were of clinical and environmental origin, respectively. To investigate phylogenetic relationships, an additional 157 strains (150 clinical and 7 environmental) were collected from 13 countries across Asia, Africa, and Latin America [Nepal 39, Thailand 32, Vietnam 15, Pakistan 3, India 2, Sri Lanka 1, Zambia 9, Zimbabwe 12, Mozambique 2, Mexico 34, Brazil 2, Peru 3, and Haiti 3] (Table 1). All strains were confirmed as *V. cholerae* O1 El Tor biotype by culture, and serotype and biotype specific genotype. Detailed information for the isolates is provided in Tables 1 and 2, and S1 Table.

Table 1. Geographic source of *V. cholerae* El tor isolates included in this study.

		Source		Serotype	
Country	No. of strains	Environmental	Clinical	Inaba	Ogawa
Bangladesh	169	50	119	33	136
Sri Lanka	1	-	1	-	1
Pakistan	3	-	3	1	2
India	2	-	2	2	-
Nepal	39	6	33	-	39
Vietnam	15	-	15	-	15
Thailand	32	1	31	17	15

Zambia	9	-	9	-	9
Zimbabwe	12	-	12	4	8
Mozambique	2	-	2	-	2
Mexico	34	-	34	17	17
Brazil	2	-	2	2	-
Haiti	3	-	3	-	3
Peru	3	-	3	-	3
	n= 326				

Table 2. *V. cholerae* genotypes based on *ctxB, rstR, and tcpA* genes.

Countries	*ctxB*			*rstR*				*tcpA*			
	Classical	El Tor	Negative	Classical	El Tor	Both	Negative	Classical	El Tor	Both	Negative
Bangladesh	167	2	-	7	162	-	-	-	169	-	-
Nepal	39	-	-	-	39	-	-	-	39	-	-
Thailand	27	-	5	9	18	-	5	-	32	-	-
India	2	-	-	-	2	-	-	-	2	-	-
Pakistan	3	-	-	-	3	-	-	-	3	-	-
Sri Lanka	1	-	-	-	1	-	-	-	1	-	-
Vietnam	15	-	-	-	15	-	-	-	15	-	-
Zambia	9	-	-	-	9	-	-	-	9	-	-
Mozambique	2	-	-	2	-	-	-	-	2	-	-
Zimbabwe	12	-	-	-	12	-	-	-	12	-	-
Mexico	5	16	13	-	16	5	13	1	26	4	3
Brazil	-	2	-	-	2	-	-	-	2	-	-
Peru	-	3	-	-	3	-	-	-	3	-	-
Haiti	3	-	-	-	3	-	-	-	3	-	-

Pulsed-field gel electrophoresis (PFGE)

Whole agarose-embedded genomic DNA was prepared from each isolate. PFGE was carried out using a contour-clamped homogeneous electrical field (CHEF-DRII) apparatus (Bio-Rad), following procedures described previously [32]. Conditions for separation were as follows: 2 to 10s for 13 h, followed by 20 to 25 s for 6 h. An electrical field of 6 V/cm was applied at an included field angle of 120°. Genomic DNA of the test strains was digested by *Not*I restriction enzyme (Gibco-BRL, Gaithersburg, MD) and *Salmonella enterica* serovar Braenderup was digested using *Xba*I, with fragments employed as molecular size markers. Restriction fragments were separated in 1% pulsed-field-certified agarose in 0.5X TBE (Tris-borate-EDTA) buffer. Post-electrophoresis gel-treatment included gel staining and de-staining. DNA was visualized using a UV transilluminator and images were digitized using a one-dimensional gel documentation system (Bio-Rad).

Image analysis

The fingerprint pattern in each gel was analyzed using a computer software package, Bionumeric (Applied Maths, Belgium). After background subtraction and gel normalization, the fingerprint patterns were typed according to banding similarity and dissimilarity, using the Dice similarity coefficient and unweighted-pair group method employing average linkage (UPGMA) clustering, as recommended by the manufacturer. The results were graphically represented as dendrograms.

Institutional approval

All the experimental protocols were reviewed and approved by the Research Review Committee (RRC), and Ethics Review Committee(ERC) of the International Centre for Diarrhoeal Disease

Research, Bangladesh (research grant: 1R01A139129-01 and protocol: PR-14017). All methods were conducted in accordance with the guidelines of the RRC and ERC.

Conflict of interest

The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgement

This research was partially supported by Japan Food Hygiene Association through National Institutes of Infectious Diseases (NIID), Tokyo, Japan, and the NIH (USA) research grant 1R01A139129-01 under collaborative agreements between the Johns Hopkins Bloomberg School of Public Health, icddr,b, and the University of Maryland. Authors gratefully acknowledge Richard Bradley Sack, Johns Hopkins Bloomberg School of Public Health, USA, and K-M Kam and Cindy KY Luey, Public Health Laboratory Center, Hong Kong, and the icddr,b hospital and laboratory staff for their support. icddr,b gratefully acknowledges the following donors who provide unrestricted support: Governments of Bangladesh, Canada, Sweden and the UK.

Author Contributions

Conceptualization and study design: FTJ, MA

Data curation: FTJ, SRB

Formal analysis: FTJ, SRB

Investigation: FTJ, SMR, MTI, SI, MS

Methodology: FTJ, SMR, MTI, SI, MA

Supervision: MA

Writing – original draft: FTJ, SRB

Writing—review and editing: TA, AH, NRT, RRC, MA

All authors contributed to the article and approved the submitted version

References

1. Harris J. F, Ryan, ET, and Calderwood, SB. Cholera. Lancet. 2012;379:2466-76.
2. Alam M, Islam A, Bhuiyan N, Rahim N, Hossain A, Khan GY, et al. Clonal transmission, dual peak, and off-season cholera in Bangladesh. Infection ecology & epidemiology. 2011 Jan 1;1(1):7273.
3. Ali M, Nelson AR, Lopez AL, Sack DA. Updated global burden of cholera in endemic countries. PLoS neglected tropical diseases. 2015 Jun 4;9(6):e0003832.
4. Colwell RR, Huq A. Environmental Reservoir of Vibrio cholerae The Causative Agent of Cholera a. Annals of the New York Academy of Sciences. 1994 Dec;740(1):44-54.
5. Huq A, Sack RB, Nizam A, Longini IM, Nair GB, Ali A, et al. Critical factors influencing the occurrence of Vibrio cholerae in the environment of Bangladesh. Applied and environmental microbiology. 2005 Aug;71(8):4645-54.
6. Lutz C, Erken M, Noorian P, Sun S, McDougald D. Environmental reservoirs and mechanisms of persistence of Vibrio cholerae. Frontiers in microbiology. 2013 Dec 16;4:375.

7. Hu D, Liu B, Feng L, Ding P, Guo X, Wang M, et al. Origins of the current seventh cholera pandemic. Proceedings of the National Academy of Sciences. 2016 Nov 29;113(48):E7730-9.
8. Weill FX, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, et al. Genomic history of the seventh pandemic of cholera in Africa. Science. 2017 Nov 10;358(6364):785-9.
9. Domman D, Quilici ML, Dorman MJ, Njamkepo E, Mutreja A, Mather AE, et al. Integrated view of Vibrio cholerae in the Americas. Science. 2017 Nov 10;358(6364):789-93.
10. Eppinger M, Pearson T, Koenig SS, Pearson O, Hicks N, Agrawal S, et al. Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. MBio. 2014 Nov 4;5(6):e01721-14.
11. Qadri F, Islam T, Clemens JD. Cholera in Yemen—an old foe rearing its ugly head. New England Journal of Medicine. 2017 Nov 23;377(21):2005-7.
12. Choi SY, Rashed SM, Hasan NA, Alam M, Islam T, Sadique A, et al. Phylogenetic diversity of Vibrio cholerae associated with endemic cholera in Mexico from 1991 to 2008. MBio. 2016 Mar 15;7(2):e02160-15.
13. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature. 2011 Sep;477(7365):462-5.
14. Codeço CT. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. BMC Infectious diseases. 2001 Dec;1(1):1-4.
15. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic

309 microorganisms. Proceedings of the National Academy of Sciences. 1998 Mar
310 17;95(6):3140-5.
311 16. Okada K, Roobthaisong A, Nakagawa I, Hamada S, Chantaroj S. Genotypic and
312 PFGE/MLVA analyses of Vibrio cholerae O1: geographical spread and temporal changes
313 during the 2007–2010 cholera outbreaks in Thailand. PLoS One. 2012 Jan 24;7(1):e30863.
314 17. Leal NC, Sobreira M, Leal-Balbino TC, de Almeida AM, da Silva MJ, Mello DM, et al.
315 Evaluation of a RAPD-based typing scheme in a molecular epidemiology study of Vibrio
316 cholerae O1, Brazil. Journal of applied microbiology. 2004 Mar;96(3):447-54.
317 18. Rahaman M, Islam T, Colwell RR, Alam M. Molecular tools in understanding the evolution
318 of Vibrio cholerae. Frontiers in microbiology. 2015 Oct 6;6:1040.
319 19. Mahalingam S, Cheong YM, Kan S, Yassin RM, Vadivelu J, Pang T. Molecular
320 epidemiologic analysis of Vibrio cholerae O1 isolates by pulsed-field gel electrophoresis.
321 Journal of clinical microbiology. 1994 Dec;32(12):2975-9.
322 20. Lan R, Reeves PR. Pandemic spread of cholera: genetic diversity and relationships within the
323 seventh pandemic clone of Vibrio cholerae determined by amplified fragment length
324 polymorphism. Journal of Clinical Microbiology. 2002 Jan;40(1):172-81.
325 21. Moore S, Thomson N, Mutreja A, Piarroux R. Widespread epidemic cholera caused by a
326 restricted subset of Vibrio cholerae clones. Clinical Microbiology and Infection. 2014 May
327 1;20(5):373-9.
328 22. Kaper JB, Morris JG Jr, Levine MM. Cholera. Clin Microbiol Rev. 1995 Jan;8(1):48-86
329 23. Faruque SM, Abdul Alim AR, Rahman MM, Siddique AK, Sack RB, Albert MJ. Clonal
330 relationships among classical Vibrio cholerae O1 strains isolated between 1961 and 1992 in
331 Bangladesh. Journal of clinical microbiology. 1993 Sep;31(9):2513-6.

24. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, Sack DA. New variants of Vibrio cholerae O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. Journal of clinical microbiology. 2002 Sep;40(9):3296-9.

25. Nair GB, Qadri F, Holmgren J, Svennerholm AM, Safa A, Bhuiyan NA, Ahmad QS, Faruque SM, Faruque AS, Takeda Y, Sack DA. Cholera due to altered El Tor strains of Vibrio cholerae O1 in Bangladesh. Journal of clinical microbiology. 2006 Nov;44(11):4211-3.

26. Alam M, Sultana M, Nair GB, Siddique AK, Hasan NA, Sack RB, et al. Viable but nonculturable Vibrio cholerae O1 in biofilms in the aquatic environment and their role in cholera transmission. Proceedings of the National Academy of Sciences. 2007 Nov 6;104(45):17801-6.

27. Kaper JB, Bradford HB, Roberts NC, Falkow ST. Molecular epidemiology of Vibrio cholerae in the US Gulf Coast. Journal of Clinical Microbiology. 1982 Jul;16(1):129-34.

28. Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. New England Journal of Medicine. 2011 Jan 6;364(1):33-42.

29. Boucher Y. Sustained local diversity of Vibrio cholerae O1 biotypes in a previously cholera-free country. MBio. 2016 May 3;7(3):e00570-16.

30. Alam M, Rashed SM, Mannan SB, Islam T, Lizarraga-Partida ML, Delgado G, et al. Occurrence in Mexico, 1998–2008, of Vibrio cholerae CTX+ El Tor carrying an additional truncated CTX prophage. Proceedings of the National Academy of Sciences. 2014 Jul 8;111(27):9917-22.

354 31. Tapchaisri P, Na-Ubol M, Tiyasuttipan W, Chaiyaroj SC, Yamasaki S, Wongsaroj T, et al.
355 Molecular typing of Vibrio cholerae O1 isolates from Thailand by pulsed-field gel
356 electrophoresis. Journal of health, population, and nutrition. 2008 Mar;26(1):79.
357 32. Cameron DN, Khambaty FM, Wachsmuth IK, Tauxe RV, Barrett TJ. Molecular
358 characterization of Vibrio cholerae O1 strains by pulsed-field gel electrophoresis. Journal of
359 clinical microbiology. 1994 Jul;32(7):1685-90.

360

361 Supporting information

362

363 **S1 Fig. Cluster analysis of isolates from Bangladesh.** The area and year of isolation are shown
364 in color codes.

365 **S1 Table. *Vibrio cholerae* El Tor strains included in this study with source and year of**
366 **isolation.**

367

368 **Figure captions**

369 **Fig 1. Clonal diversity and geographical distribution of isolates from Bangladesh.**
370 Isolates belonging to clade A were found in all four areas. Clade A contained both clinical and
371 environmental isolates. In addition to clade A, strains of other clades were found, but only in
372 Mathbaria.
373
374 **Fig 2. PFGE analysis of the isolates collected from Bangladesh.**
375 A total of 169 strains were analyzed which resulted in 5 clades. Clade A contained 94.67% of the
376 isolates. The rest 9 isolates which were only found in Mathbaria formed the other 4 groups (B, C,
377 D and, E).
378
379 **Fig 3. Comparison of band patterns between isolates from Bangladesh and other countries.**
380 The isolates from Bangladesh were compared with 157 additional isolates collected from 13
381 other countries. The resulting phylogenetic tree represents 16 groups. Clade A contained 89.88%
382 of the total isolates including 160 isolates of Bangladesh.
383
384 **Fig 4. Global phylogeny of isolates based on PFGE pattern.**
385 Countries are represented by color. 160 isolates of Bangladesh formed clade A, with 133 strains
386 isolated from other countries. Three subclades were observed in clade A: Subclade Ia and Ib

387 contained strains mostly of African and Asian origin. Subclade II comprised predominantly Latin
388 American strains. Country subclusters were also observed.
389
390 **Fig 5. Global distribution of clones.**
391
392
393
394
395
396

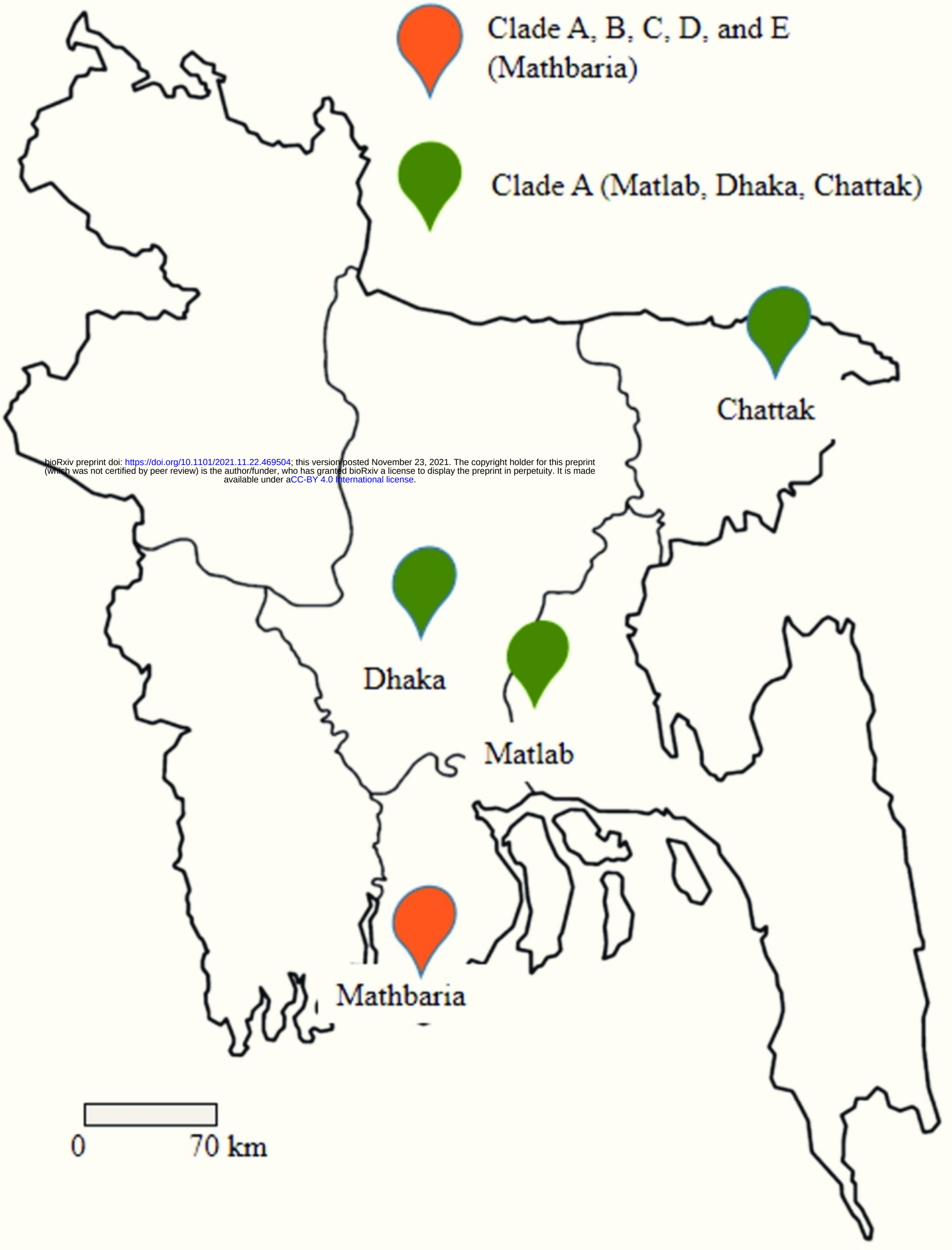


Fig1

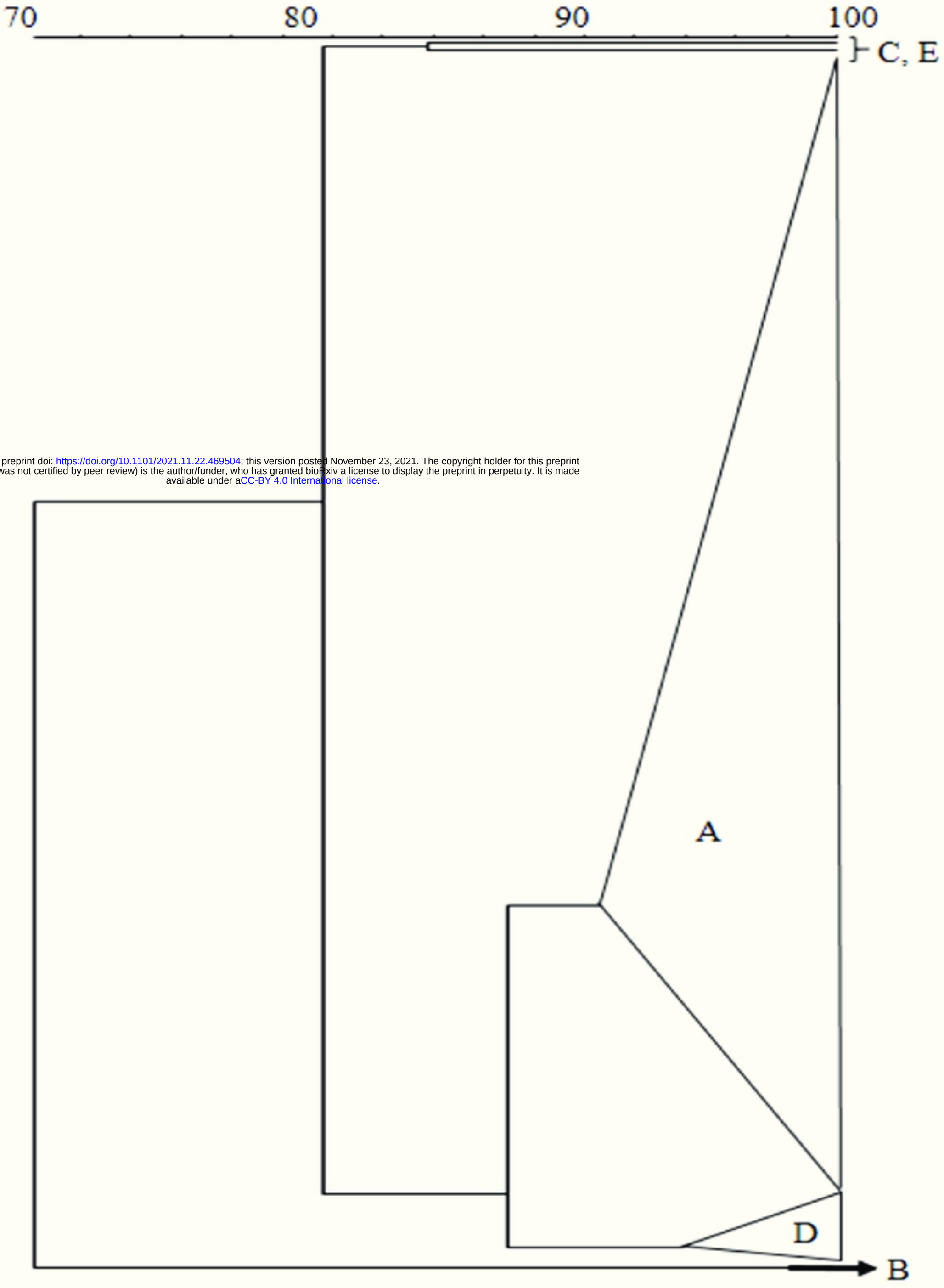


Fig2

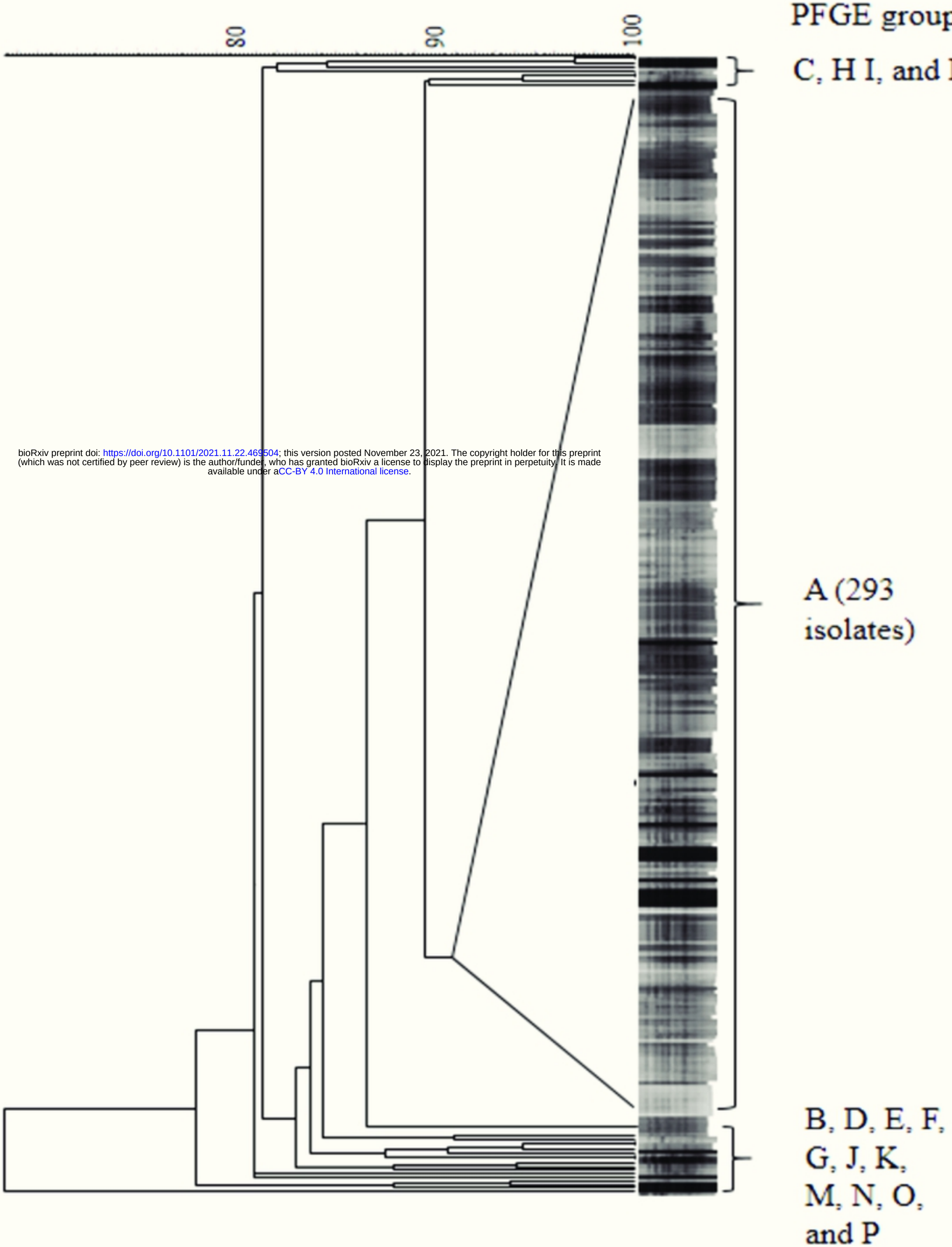


Fig3

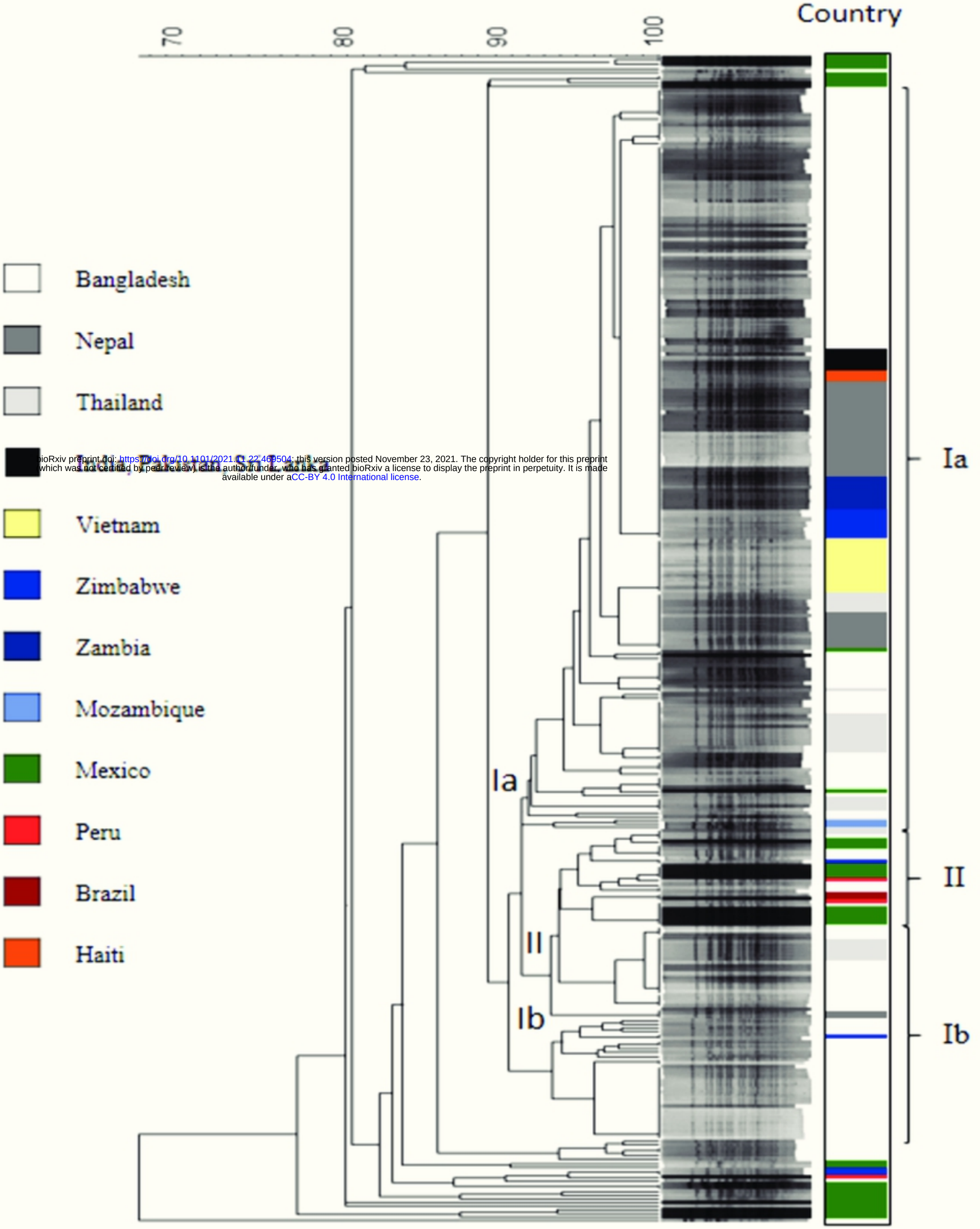
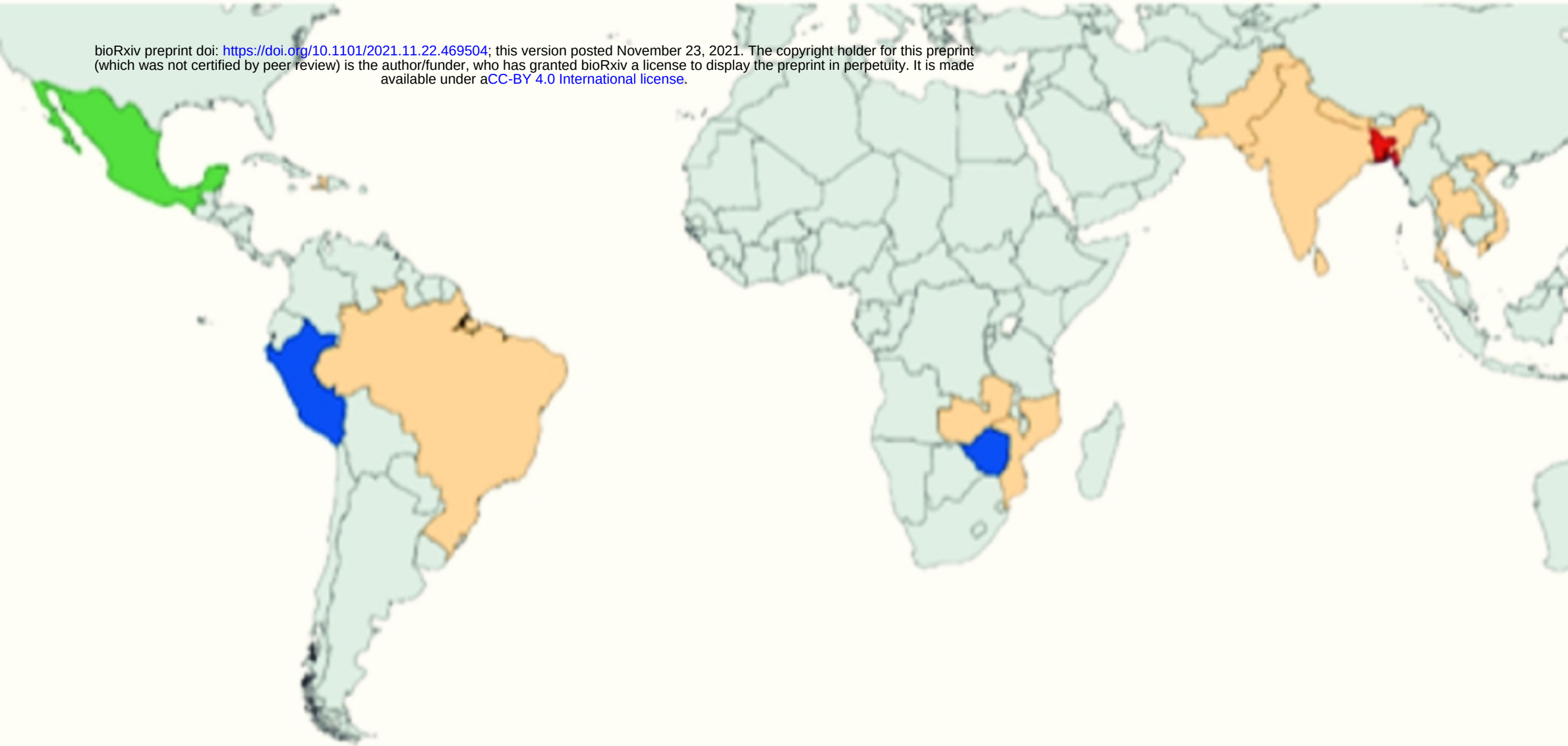


Fig4



Clade (country)

A-E (Bangladesh)

A, F-P (Mexico)

A (India, Pakistan, Sri Lanka, Nepal, Mozambique, Zambia)

A, E (Peru, Zimbabwe)

Fig5