Genetic diversity and structure of the *Vibrio cholerae* population in Colombia

Riaño Pachón, Diego[1] (I); Valenzuela de Silva, Emilia (I); Mantilla Anaya, José (2); Agudelo, Carlos (3)

(1) Molecular Epidemiology Group, Biotechnology Institute, Universidad Nacional de Colombia (2) Pharmacy department, School of sciences, Universidad Nacional de Colombia (3) School of Medicine, Universidad Nacional de Colombia

Summary

Thirty-four *Vibrio cholerae* isolates from Colombian cholera epidemic between 1991 - 1996 were studied by pulsed field gel electrophoresis (PFGE). Three *V. cholerae* isolates not belonging to the epidemic in Colombia and a *V. fluvialis* isolate were also analysed. The genetic diversity observed for the 38 isolates typified was **0.95**, a very similar value to that obtained by multilocus enzyme electrophoresis (MLEE). No correlation was observed between the variables characterising these isolates (serotype, origin (clinical or environmental), state, electrophoretic type (MLEE) and restriction macro-fragment profile or pulsetype (PFGE)). No evidence of linkage disequilibrium was obtained on evaluating this population with the PFGE and MLEE genetic markers.

[1] Corresponding author. Mailing address: Instituto de Biotecnología, Edificio Manuel Ancízar, Universidad Nacional de Colombia, Bogotá D. C., Colombia. Phone 571-3165000 ext 16957. Fax 571-3165415. Electronic mail address: diriano@ibun.unal.edu.co

It is suggested that the *Vibrio cholerae* population in Colombia presents a sexual genetic structure. The lack of evidence for linkage disequilibrium and the absence of correlation between the epidemiological, biochemical and molecular variables available sustain such hypothesis. It is also sustained by the high value for genetic diversity obtained which could be reflecting the coexistence of several lines of descent having a high genetic flow amongst them.

Key words: *Vibrio cholerae* - pulsed-field gel electrophoresis, population genetics

Introduction

Cholera outbreaks cause the death of around 120,000 persons annually around the world and the number of sick people is even greater, most of them are children (Faruque *et al*. 1998).

An epidemic wave of cholera was presented in Colombia between 1991 and 1996. Some of the isolates involved in this epidemic were collected by the National Health Institute and they have been studied using different approaches (Bustamante 1999; Tamayo *et al.* 1997; Tamayo 1998).

The approaches used to study isolates involved in the Colombian cholera epidemic have usually set aside the conceptual framework of population genetics. In spite of the fact that when microbial genotypes are characterised for studying their epidemic dispersion, only population genetics can rigorously evaluate spatial and

temporal stability of these genotypes (Tibayrenc 1995a, Tibayrenc 1996). Epidemiological tipification is based on the supposition that multilocus genotypes used as epidemiological markers are stable. This is only true if the genetic flow is rare or absent in the species or population under study (Tibayrenc 1998). Knowledge concerning the genetic structure for those micro-organism populations having medical importance thus acquires great relevance since this knowledge can be used to know if the genotypes found in the tipification are space-temporal stable so that other characteristics of epidemiological interest can be mapped on them, or if, on the contrary, the genotypes are unstable and continuously they are diluted in the population's genetic pool.

The object of this work was to approach knowledge concerning the genetic structure of the *Vibrio cholerae* population causing the cholera epidemic in Colombia between 1991 and 1996.

Materials and Methods

Population to be studied

Thirty-four *Vibrio cholerae* O1 Colombian isolates taken from patients suffering from cholera during the Colombian epidemic from 1991 to 1996 were studied. The isolates were supplied by the National Health Institute of Colombia.

Three *Vibrio cholerae* O1 isolates having origin different to that of the epidemic cholera wave in Colombia were also studied. A *Vibrio fluvialis* isolate was also included, to use it as external group in genotyping analysis (Table I).

The isolates studied were characterised according to the following characteristics: serotype, origin (clinical or environmental), department, year (Agudelo *et al.* 1997) and electrophoretic type (ET) (Bustamante 1999), information obtained from previous studies by our group.

PFGE

The procedure described by Chang & Chui (1998) and by Rincón (1999) was followed, with some modifications.

Forty-ml LB broth were inoculated with 0.5ml from a 16-hour culture. This was incubated at 37^{o}C until reaching an optical density of around 0.6 at 610 nm.

The cells were harvested by centrifugation at 4000g and suspended in washing buffer (1M NaCl, 10mM Tris-HCl, pH8.0, 10mM EDTA, pH8.0) obtaining a 1.9x10^{9}UFC/ml concentration. 500µl of this cell suspension were mixed 1:1 with 1.5%[wt/vol] low melting point agarose. This was allowed to gel at 4^{o}C for 5min. Plugs obtained were then immediately lysed in lysis buffer (1M NaCl, 6mM Tris-HCl, pH8.0, 100mM EDTA, pH8.0, 0.5%[wt/vol] N-lauril sodium sarcocinate, 0.2%[wt/vol] deoxycholate, 0.5%[vol/vol] triton X-100, 1mg/ml lysozyme) at 37^{o}C for 24 hours. They were then deproteinised with ESP buffer (l%[wt/vol] Lauril Sodium Sarcocinate, 500mM EDTA, pH8.0, 1mg/ml Proteinase K) for 48 hours at 50^{o}C. The cell debris were removed with successive TE (10mM Tris-HCl, pH8.0, 1mM EDTA, pH8.0) washes. Plugs were equilibrated in restriction-enzyme buffer for 1 hour prior digestion.

Digestion was carried out with the *Not* I restriction endonuclease, using 10U of the enzyme for each digestion at 37^{o}C for 16-18 hours.

The pulsed field electrophoresis was done in the Gene Navigator™, Pharmacia LKB system with 1%[wt/vol] agarose gels in 0.5X TBE buffer using the following pulse program: 2s for 4h, 4s for 4h, 8s for 4h, 16s for 4h, 32s for 4h, 70s for 4h. A constant 180V voltage was applied and temperature kept at 12^{o}C.

Once electrophoresis ended, the gels were dyed in a 1µg/ml Ethidium Bromide solution and the gel was digitalised using the documenting equipment GelDoc 2000 (Bio-Rad 1999).

Statistical analysis

The Quantity One-Discovery Series (Bio-Rad 1999) program was used for gel normalisation, estimating each fragment's molecular weight. The inter and intra gel error was evaluated using the Molecular Fingerprint Analyzer program (Stanford Center for Tuberculosis Research 1998) to assure correct comparison between the different electrophoretic patterns, obtained in different gels (this is useful when aligning the bands). The similarity between isolate pairs was evaluated according to the following expression:

$$d_{xy} = 1 - \frac{2n_{xy}}{\left(n_x + n_y\right)}$$

where $\boldsymbol{n_{xy}}$ is the number of fragments shared between electrophoretic lines ***x*** and ***y***, and $\boldsymbol{n_x}$ and $\boldsymbol{n_y}$ are the number of fragments exhibited by each one of the lines

(Lynch 1990). The Molecular Fingerprint Analyzer program (Stanford Center for Tuberculosis Research, 1998) was used to calculate the distances.

A dendrogram was generated from the calculated distances, following the UPGMA procedure implemented in the MEGA program (Kumar *et al.* 1993).

The population's genetic diversity was estimated according to:

$$H = \left(1 - \sum x_i^2\right)\left(\frac{n}{n-1}\right)$$

where x_i is the ith type (pulsetype) frequency, and n is the number of types.

Multiple correspondence analysis was used to examine correspondence between variables and was also employed to determine the relationships between the different categorical variables characterising *Vibrio cholerae* isolates involved in the cholera epidemic in Colombia, i.e. state, origin, year, ET (MLEE), serotype and pulsetype, as determined in this work.

Linkage disequilibrium in the Colombian *Vibrio cholerae* population was evaluated by the ***g*** test proposed by Tibayrenc (Tibayrenc 1995, Tibayrenc 1996, Tibayrenc 1998) consisting of measuring the correlation between two groups of independent markers. In this case, ET (MLEE, Bustamante 1999) and pulsetype (PFGE, obtained in this work) were employed as independent genetic markers. The Mantel program version 2.0 (Liedloff 1999) and Mantel Test version 1.2.1 (Briers 1999) were used to conduct this test.

Results

Restriction macrofragment analysis

A typical gel obtained by pulsed field electrophoresis for use in the characterisation of 38 *Vibrio* genus isolates is shown in figure 1. Only fragments greater than 20 Kpb were recorded for the analysis. Table II presents a summary of the results obtained from *Vibrio cholerae* isolates genome digestion and *Vibrio fluvialis* isolates with the *Not* l restriction enzyme. These first results agree with the number and size of those fragments generated with *Not* l reported for *V. cholerae* in other studies (Arakawa *et al.* 2000, Khetawat *et al.* 1998).

The genome average size estimated for all *Vibrio cholerae* isolates envisaged in this study was 2,298,545.5 pb ± 185,827,3 pb.

Recently reported *Vibrio cholerae* genome size is 4,033,460 pb (Heidelberg *et al*. 2000). There is about 42% discrepancy between the estimate for *V. cholerae* genome size found in this work and the size reported by Heidelberg *et al.* 2000. This discrepancy can be attributed to the fact that some bands can be constituted by several similar size fragments even though originating from different regions of the genome. As can be seen in Figure 1, the intensity of the bands is variable and those presenting greater intensity and amplitude could comprise this type of band. (Khetawat *et al.* 1998, Trucksis *et al*. 1998).

Genetic distance and cluster analysis

The average for different bands between isolate pairs was 7, with a maximum of 17 and a minimum of 0 (identical isolates); most isolates differed in 5 bands. The maximum genetic distance was 0.36 and the minimum 0; genetic distance average was 0.18. From the dendrogram present in figure 2 shows differences between species at 0.18 distance.

Two clearly distinguishable groups were found at a 0.13 distance. All those isolates from the epidemic recovered from patients are in group I; group II includes those isolates from Brazil and the Colombian environmental isolates.

Subgroups **A, B** and **C,** in group I, and subgroups **D** and **E,** in group II, are formed at a 0.11 distance. Group **A** contains around 74% of the isolates studied, followed by group **C** having 10% and then groups **B** and **E,** each with a 5%.

Genetic diversity was estimated on groups formed at a 0.05 distance (Figure 2). This distance was chosen to be the maximum level at which isolate identity was maintained according to the error in estimating restriction macro-fragment size. According to this, the population's genetic diversity (H) was **0.95**. A **0.92** value was obtained on calculating electrophoretic type diversity (the analogue for pulse-type genetic diversity) for the same group of isolates but with the MLEE data obtained by Bustamante (1999).

Both values are very similar, which could indicate special processes to those to which these isolates are submitted. This high diversity could be the product of a sexual genetic structure so that it is practically possible to recover all the multilocus

genotypes. Linkage disequilibrium estimation and multiple correspondence analysis were carried out was to verify this.

According to this analysis, it was found that there is a strong correlation between the characteristics: San Andrés (a tropical island), state; 21, ET (according to Bustamante 1999); 11-E-a, Pulsetype; Environmental Origin. This relationship is interesting, as it indicates that this environmental isolate's genotype is, apparently, different to pathogen genotype, since it can be distinguished by MLEE-ET, as well as by pulsetype. However, it is still necessary to include more environmental isolates, gathered at different times and from different regions to assure that this correlation is due to different genotypes and not to a sampling bias. Except for this correlation, there is no correlation between those variables with which *V. cholerae* isolates of involved in the Colombian cholera epidemic are characterised.

Evaluation of linkage disequilibrium

Linkage disequilibrium evaluation of the population obtained in this study (employing MLEE-ET and pulsetype as independent genetic markers) confirmed that there is no correlation between the markers. There is no evidence of linkage disequilibrium. The Mantel test significance value (with 10^4 permutations) was 0.2956. We also calculated the standard coefficient of correlation which, in spite of the fact that it is not perfectly rigorous, provides commendable results (Tibayrenc 1995b); its value was 0.0463, with reliability greater than 99.99%.

All this indicates that there is no evidence of linkage disequilibrium in the population being studied; rather more, it is evidence of a high degree of recombination. The high diversity calculated would be an effect of this fact.

Discussion

Seventeen restriction macro-fragment profiles or pulsetypes were determined among 38 *Vibrio* spp. isolates. Two of these pulsetypes grouped more than 44% of the isolates. In spite of this, great genetic variability was detected (H=**0.95**); this value is similar to some diversity measurements which have been estimated on this population. The high value could mean the coexistence of several different strains or clones, it can suggest that the Colombian epidemic was originated by more than one clone. Or, supposing that a single clone originated it, this clone has diversified much over time and there has been great genetic interchange between all variants.

This suggestions are sustained by the lack of evidence for correlation between variables and genetic markers, as can be seen from multiple correspondence analysis and Mantel test between the two genetic markers (electrophoretic-type and pulsetype). This indicates that all, or nearly all, possible genotypes can be recovered and that there is no association between genotypic nor phenotypic characteristics. This can be interpreted as being evidence of a great genetic exchange level between *Vibrio cholerae* strains circulating in Colombia.

This result contrasts totally with that obtained by Tamayo (Tamayo *et al.* 1997, Tamayo 1998); this researcher arrived at the conclusion that the isolates from the Colombian epidemic originated from a clone, being supported by results obtained from ribotyping. However, a different result was obtained in her study from looking at the population through randomly amplified DNA polymorphism, with which greater variability than that suggested by ribotipification was observed.

At the moment (without having the whole of this study's data available) one could venture to use it to support our hypothesis, since these markers (PFGE, MLEE, ribotipification and RAPD) taken together offer the same vision. They are not correlated, as different conclusions are reached on using each one of them (v.g. different dendrograms). If genetic structure of this population was clonal, with all the markers used, one would have to arrive at this conclusion. As this is not happening it can be suggested that the *Vibrio cholerae* population in Colombia suffers from a strong genetic interchange, or exhibits a sexual genetic structure.

This result does not disagree with the proposition that the Latin American epidemic originated from a single clone, but teaches us that the factors to which microorganisms are submitted to in Colombia are different and, thanks to evolutionary molecular genetics, we can gain knowledge of their particularities in this habitat.

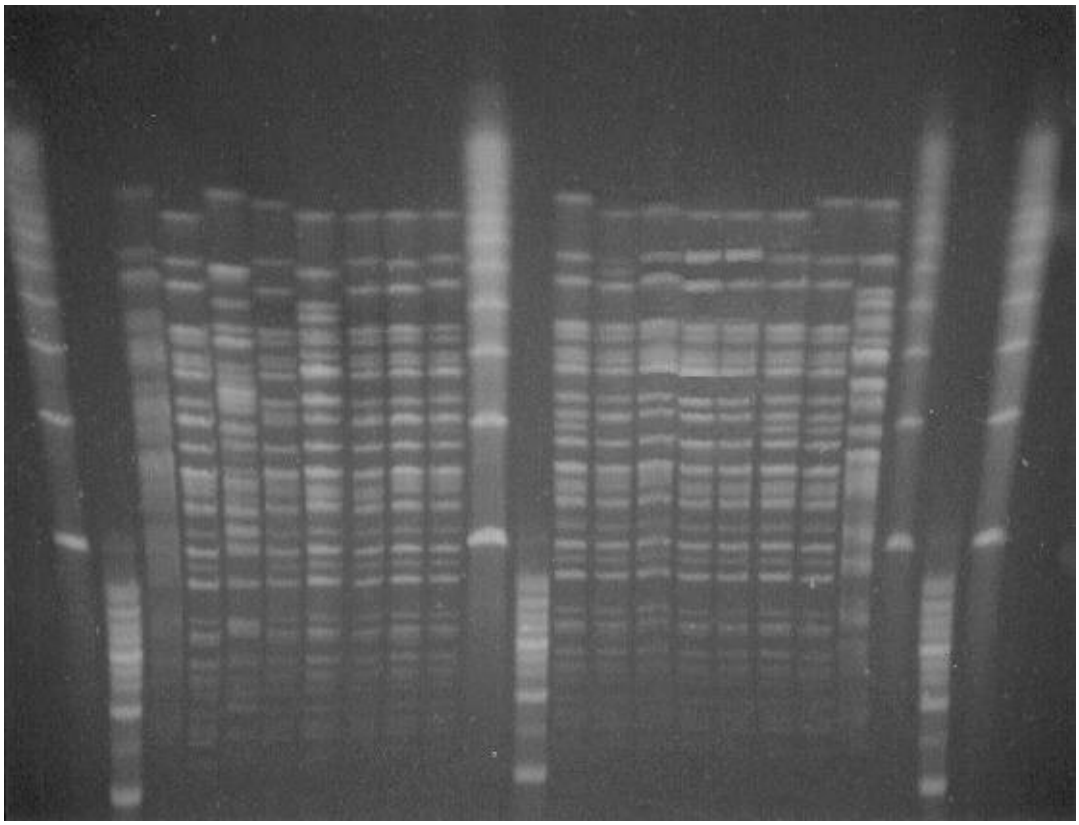

Figure 1. Photograph of a tipical gel from Pulsed-Fiel Gel Electrophoresis. Restriction fragments generated by *Not* I. 24 h. Electrophoresis at 12 °C. Pulses: 2 sec. for 4h; 4 sec. for 4h; 8 sec. for 4h; 16 sec. for 4h; 32 sec. for 4h: 70 sec. for 4h.

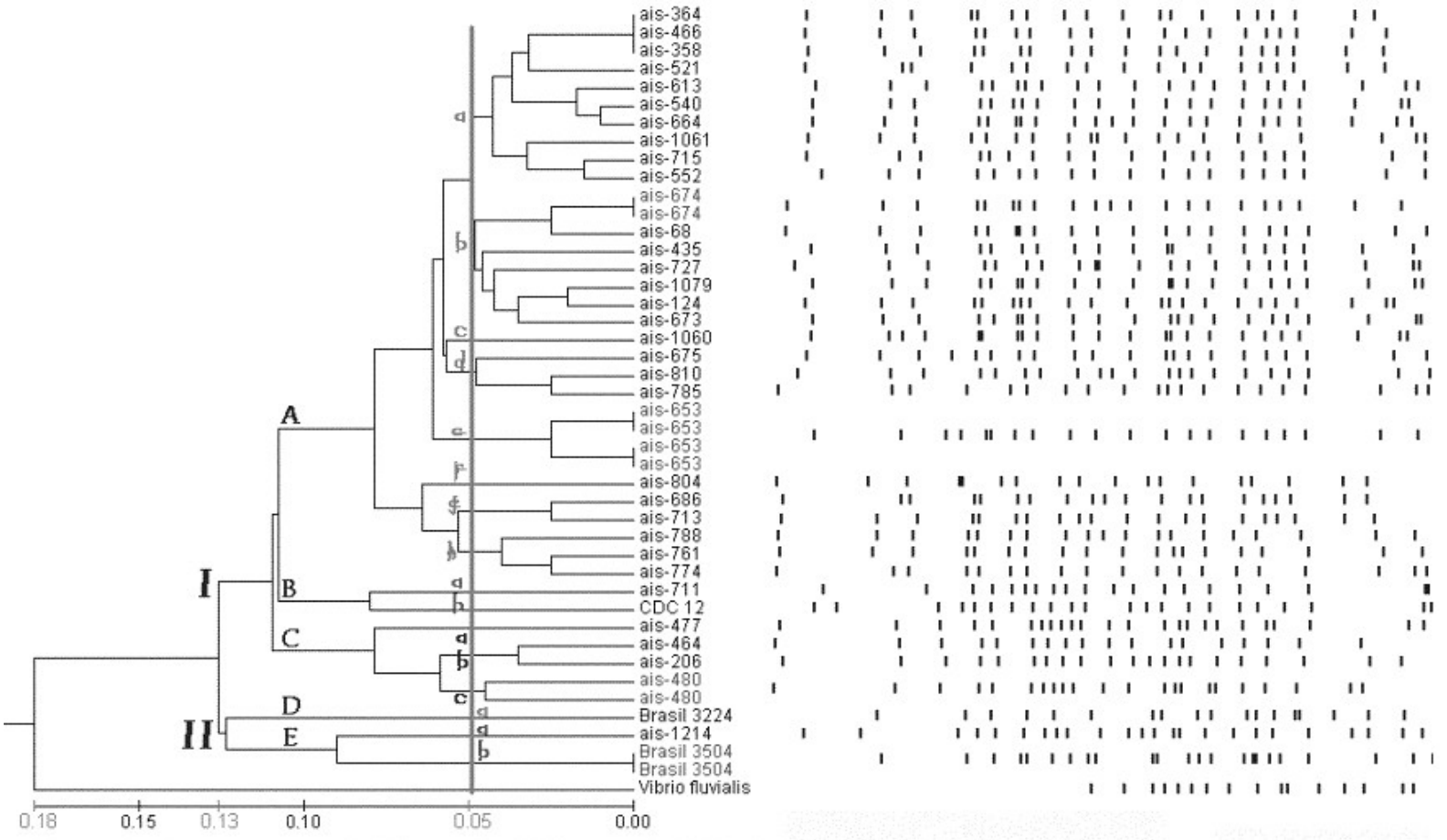

Figure 2. Dendogram obtained from distance matrix for the proportion of different bands between isolates, generated by UPGMA algorithm. Restriction macrofragments profile obtained is shown for each isolate. Isolates in red indicate repeated isolates.

Table I. Data regarding *V. Cholerae* isolates genotipified in this study.

Isolate	TE[a]	Serotype	Origin(Clinical/Environmental)[b]	Departament	Year
68	26	Inaba	Clinical	Amazonas	1991
124	11	Inaba	Clinical	Amazonas	1991
206	37	Inaba	Clinical	Nariño	1991
358	37	Inaba	Clinical	Bolívar	1991
364	36	Inaba	Clinical	Amazonas	1991
435	37	Inaba	Clinical	Tolima	1991
464	27	Inaba	Clinical	Guajira	1991
466	27	Inaba	Clinical	Cauca	1991
477	37	Inaba	Clinical	Magdalena	1991
480	37	Inaba	Clinical	Valle	1991
521	13	Inaba	Clinical	Antioquia	1991
540	17	Ogawa	Clinical	Bolívar	1992
552	10	Inaba	Clinical	Valle	1992
613	14	Inaba	Clinical	Cauca	1992
653	24	Inaba	Clinical	Tolima	1992
664	29	Inaba	Clinical	Antioquia	1992
673	30	Inaba	Clinical	Guajira	1992
674	30	Inaba	Clinical	Magdalena	1992
675	30	Inaba	Clinical	Antioquia	1992
686	32	Inaba	Clinical	Amazonas	1992
711	37	Ogawa	Clinical	Guajira	1993
713	20	Ogawa	Clinical	Amazonas	1993
715	6	Ogawa	Clinical	Amazonas	1993
727	5	Ogawa	Clinical	Amazonas	1993
761	36	Ogawa	Clinical	Nariño	1994
774	19	Ogawa	Clinical	Valle	1994
788	19	Ogawa	Clinical	Amazonas	1995
785	38	Inaba	Clinical	Nariño	1995
804	33	Ogawa	Clinical	Cauca	1995
810	38	Inaba	Clinical	Chocó	1995
1060	33	Ogawa	Clinical	Tolima	1996
1061	33	Ogawa	Clinical	Tolima	1996
1079	19	Ogawa	Clinical	Nariño	1996
1214	21	Ogawa	Environmental	San Andrés	
V. fluviallis	*Vibrio fluviallis*				
Brasil	*Vibrio cholerae* O1 Brasil 3224				
Brasil	*Vibrio cholerae* O1 Brasil 3504				
CDC 12	*Vibrio cholerae* O1 Biotype El Tor, serotype Ogawa				

a. TE: Electrophoretic type (MLEE) data obtained by Bustamante, 1999.
b. This characteristics refers to whether the isolate was obtained from a patient (clinical) or from the environment (environmental)

Table II. Summary of results from digestion with *Not* I and pulsed-field electrophoresis

Isolates	Number of fragments	Range of fragment numbers	Range of fragment sizes (bp[a])
V. cholerae epidemic	21	19-25	20,143.4 – 415,172.2
V. cholerae non-epidemic	22	18-24	21,110.9 – 343,152.6
V. fluvialis	17	NA[b]	20.887.6 – 101,868.1

a. bp: Base pairs; b. Does not apply

References

Agudelo C, Mantilla Anaya JR, Valenzuela de Silva EM, Castañeda E, Tamayo M, Peñarete L, Ramírez E, Cabello F 1997. Epidemiología Molecular del Cólera en Colombia. Informe Técnico Científico. Universidad Nacional de Colombia - Instituto Nacional de Salud. Bogotá, Colombia. 39 pp.

Arakawa E, Murase T, Matsushita S, Shimada T, Yamai S, Ito T, Watanabe H 2000. Pulsed-Field Gel Electrophoresis-Based Molecular Comparison of *Vibrio cholerae* O1 Isolates from Domestic and Imported Cases of Cholera in Japan. J Clin Microbiol. 38(1):424-426

Bio-rad 1999. Quantity One ® User Guide for Version 4.1Windows and Macintosh. USA. x + 364 pp.

Briers RA 1999. Mantel Test version 1.2.1 Macro for Excel 97

Bustamante SL 1999. Caracterización por Electroforesis de Enzimas Multilocus de *Vibrio cholerae*. Trabajo de Grado. Universidad Nacional de Colombia. Carrera de Biología. Santafé de Bogotá D. C. Colombia. iv + 82 pp.

Chang N, Chui L 1998. A Standarized Protocol for the Rapid Preparation of Bacterial DNA for Pulsed-Field Gel Electrophoresis. Diagn Microbiol Infect Dis 31:275-279

Faruque SM; Albert MJ; Mekalamos JL. 1998. Epidemiology, Genetics, and Ecology of Toxigenic *Vibrio cholerae*. Microbiol Mol Biol Rev 62:1301-1314

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD,

Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter C, Fraser CM 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406: 477-484

Khetawat G, Bhadra RK, Kar S, Das J 1998. *Vibrio cholerae* O139 Bengal: Combined Physical and Genetic Map and Comparative Analysis With the Genome of V. Cholerae O1. J Bacteriol 180:4516-4522

Kumar S, Tamura K, Nei M 1993. Molecular Evolutionary Genetics Analysis. Pennsylvania State University. http://www.megasoftware.net

Liedloff A. 1999. Mantel Version 2.0 Mantel Non-Parametric Test Calculator. School of Natural Resource Sciences. Queensland University of Technology, Brisbane, Australia

Lynch M. 1990. The Similarity Index and DNA Fingerprinting. Mol Biol Evol 7:478-484

Rincón G 1999. Análisis Genético de Cepas de *Staphylococcus epidermidis* Mediante Electroforesis de Campo Pulsado. Tesis de Grado. Universidad Nacional de Colombia. Posgrado Interfacultades de Microbiología. Santafé de Bogotá D. C. Colombia. 90 pp.

Stanford Center for Tuberculosis Research 1998. Molecular Fingerprint Analizer (SCTR-MFA) versión 0.4.1 (beta) available at http://molepi.stanford.edu/free_software.html

Tamayo MM 1998. Determinación del Origen Clonal de Aislamientos Colombianos de *Vibrio cholerae* O1. Tesis de Grado. Pontificia Universidad Javeriana. Posgrado de Microbiología. Santa Fe de Bogotá D. C. Colombia

Tamayo MM, Koblavi S, Grimont F, Castañeda E, Grimont PAD 1997. Molecular Epidemiology of *Vibrio cholerae* O1 Isolates From Colombia. J Med Microbiol 46:611-616

Tibayrenc M 1995a. Populations Genetics of Parasitic protozoa and Other Microorganisms. Adv Parasitol 36:47-115

Tibayrenc M 1995b. Population Genetics and Strain Typing of microorganims: How to Detect Departures from Panmixia Without Individualizing Alleles and Loci?. C R Acad Sci III 318:135-139

Tibayrenc M 1996. Towards a Unified Evolutionary Genetics of Microorganisms. Ann Rev Microbiol 50:401-429

Tibayrenc M 1998. Genetic epidemiology of Parasitic protozoa and Other Infectious Agents: The Need for an Integrated Approach. Int J Parasitol 28:85-104

Trucksis M, Michalski J, Kang Deng Y, Kaper JB 1998. The Vibrio cholerae Genome Contains Two Unique Circular Chromosomes. Proc Natl Acad Sci USA 95:14464-14469