

# Bioinformatics and microbial biodiversity: analysis of vibrios by the GenEnv system

A. PAPARINI, D. SANTONI, V. ROMANO SPICA

Department of Health Sciences, Public Health Unit; University Institute of Movement Sciences (IUSM), Rome, Italy

## Key words

Vibrio • Bioinformatics • Public health

## Summary

Sequence-based approaches to prokaryotic systematics and typing represent a modern and promising strategy in epidemiology and environmental microbiology. GenEnv, a database-driven system for bacterial typing, was developed in order to provide user friendly tools for supporting biomolecular analysis of bacteria. The family Vibrionaceae represents a heterogeneous taxon of aquatic microorganisms, harbouring a plethora of

genomes currently analyzed by different molecular techniques. Under the query "Vibrio", GenEnv retrieved 256 organisms, included in a total number of 19 families. Overall, 548 sequences, comprising 16S rRNA ( $n = 402$ ), *rpoB* ( $n = 1$ ), *gyrB* ( $n = 145$ ) were available. In addition, GenEnv system allowed primer design, homology analysis and restriction maps, for immediate applications to the study of Vibrionaceae.

## Introduction

*Vibrio fischeri*, *Vibrio splendidus* and *Photobacterium phosphoreum* are among the earliest known non pathogenic species of vibrios. They were described by the dutch microbiologist Martinus Willem Beijerinck, around 1900. The genus *Vibrio* includes several taxa, not all detrimental to humans. Overall, more than 70 species are included into the class of Gammaproteobacteria and four different families, namely Enterovibrionaceae, Photobacteriaceae, Salinivibrionaceae and Vibrionaceae. Vibrios are gram negative, usually motile rods, prevalent in the aquatic environment. Noteworthy, some species display ability to establish interesting forms of commensalism and symbiosis with marine animals such as the Hawaiian squid *Euprymna scolopes* and the aquatic arthropod *Gerris spinolae*; others are found associated with plants, corals, shrimps, seagrass, sponges, zooplankton, fish and molluscs [1-3].

Cholera is an old known, infectious disease caused by the aquatic bacterium *Vibrio cholerae*. The disease is characterised by acute diarrhoeic episodes with massive loss of fluids and electrolytes, vomiting and leg cramps. It is generally recognised that infection spreads by ingestion of waters and foods contaminated with faeces containing the bacterium or by accidental swallowing of microorganisms, present onto various environmental matrices. When cholera hit Florence, in 1854, during the Asiatic Cholera Pandemic of 1846-1863, the Italian physician Filippo Pacini, undertook meticulous histological observations that eventually

led him to describe a comma-shaped bacillus, which he called "Vibrio". His discovery was granted prominence only several years later when, in 1965, the international committee on nomenclature adopted the name *Vibrio cholerae* Pacini 1854 for the etiologic agent of cholera. During the last 185 years, seven, nearly global pandemics have been described. The *Vibrio* responsible for the seventh pandemic, now in progress, is known as *Vibrio cholerae* O1, biotype El Tor. The beginning of the current pandemic is reckoned associated with an outbreak in Indonesia, in 1961; the pandemic has then moved to Africa and South America [4]. In 1992, a new serogroup, O139, appeared in southern Asia where it has then become endemic. According to some Authors, this novel serogroup threatens to start the next pandemic [5]. Other than *cholerae*, reckoned human pathogenic vibrios are: *V. parahaemolyticus*, *V. vulnificus* and, secondarily, *Grimontia hollisae*, *Photobacterium damsela*, *V. alginolyticus*, *V. cincinnatiensis*, *V. fluvialis*, *V. furnisii*, *V. harveyi*, *V. metschnikovii*, *V. mimicus*. Most of these species were found to contaminate commonly eaten seafood, like shrimp, squid, crab, cockles and mussels [6].

16S rRNA, flanking sequences of the cytotoxin-hemolysin virulence gene *vvhA*, pR72H DNA fragment and 16S-23S rRNA intergenic spacer region have been proven effective for the classification of vibrios [7-9]. Biomolecular, genomic analyses, of either clinical or environmental isolates, are performed, at present, by amplified fragment length polymorphism (AFLP), fluorescence *in situ* hybridisation (FISH), amplified ribo-

somial DNA restriction analysis (ARDRA), random amplified polymorphic DNA (RAPD), repetitive extragenic palindromes (REP), restriction fragment length polymorphism (RFLP), multilocus enzyme electrophoresis (MLEE) and multilocus sequence typing (MLST) and analysis [3, 10-17]. Of specific importance are techniques like FISH and amplification-based methods, allowing typing of nonculturable microbial strains. The viable but nonculturable (VBNC) state of many bacterial species of *Vibrio*, may cause restrictions in biotyping, especially whenever cultivation-dependent techniques are required [18, 19]. Ethidium bromide monoazide was recently utilized to selectively allow amplification of a target sequence, in viable cells of *Vibrio vulnificus* [20]. Detection and typing of vibrios by DNA microarrays may also represent an advanced effective strategy [21]. Development of new detection methods is relevant for species like *Vibrio vulnificus*, whose prevalence in waters and shellfish shows no correlation with conventional faecal indicator organisms [22, 23]. MLEE and MLST base on analogous principles, with diverse discriminatory capabilities. MLEE involves the investigation of polymorphic housekeeping enzymes, whereas MLST relies on several fragments from housekeeping genes [24, 25]. Real time PCR combines sensibility, accuracy and rapidity, and represents an emergent technique [20, 26]. For phylogenetic studies of vibrios, ribosomal sequences, like 5s rRNA, 16s rRNA, 23s rRNA, *recA*, *rpoA*, and *pyrH* have been used. The 16s rRNA is considered one of the most effective tools for genotyping and classifying vibrios, within the family *Vibrionaceae* [3, 17, 27, 28]. GenEnv is a database of selected sequences, accessible via the World Wide Web [29, 30]. It is provided by the University Institute of Movement Sciences (IUSM), Rome, Italy, via dedicated links available on the website ([www.bioigene.it](http://www.bioigene.it)). The database is constantly edited and updated, and represents part of a system, conceived to provide a user-friendly bioinformatic tool, for supporting detection and typing of bacteria. Its main application fields are public health and environmental microbiology. However, GenEnv is a flexible system with interesting applications also for basic and applied research. The system includes the database of selected genes queryable through a graphical interface. The interface also supports the use of user friendly tools for primer design, homology analysis and restriction maps handling and retrieval. Contemplated sequences are 16S rRNA, *rpoB* and *gyrB*, imported from Genbank and other databases, after a meticulous sorting, aimed at minimization of redundancy.

## Methods

The online instructions were followed to obtain the information on vibrios available on the GenEnv database (Fig. 1). The browser window was accessed from GenEnv home page, following the apposite link. Through the specific form, the query “*Vibrio*” was sub-

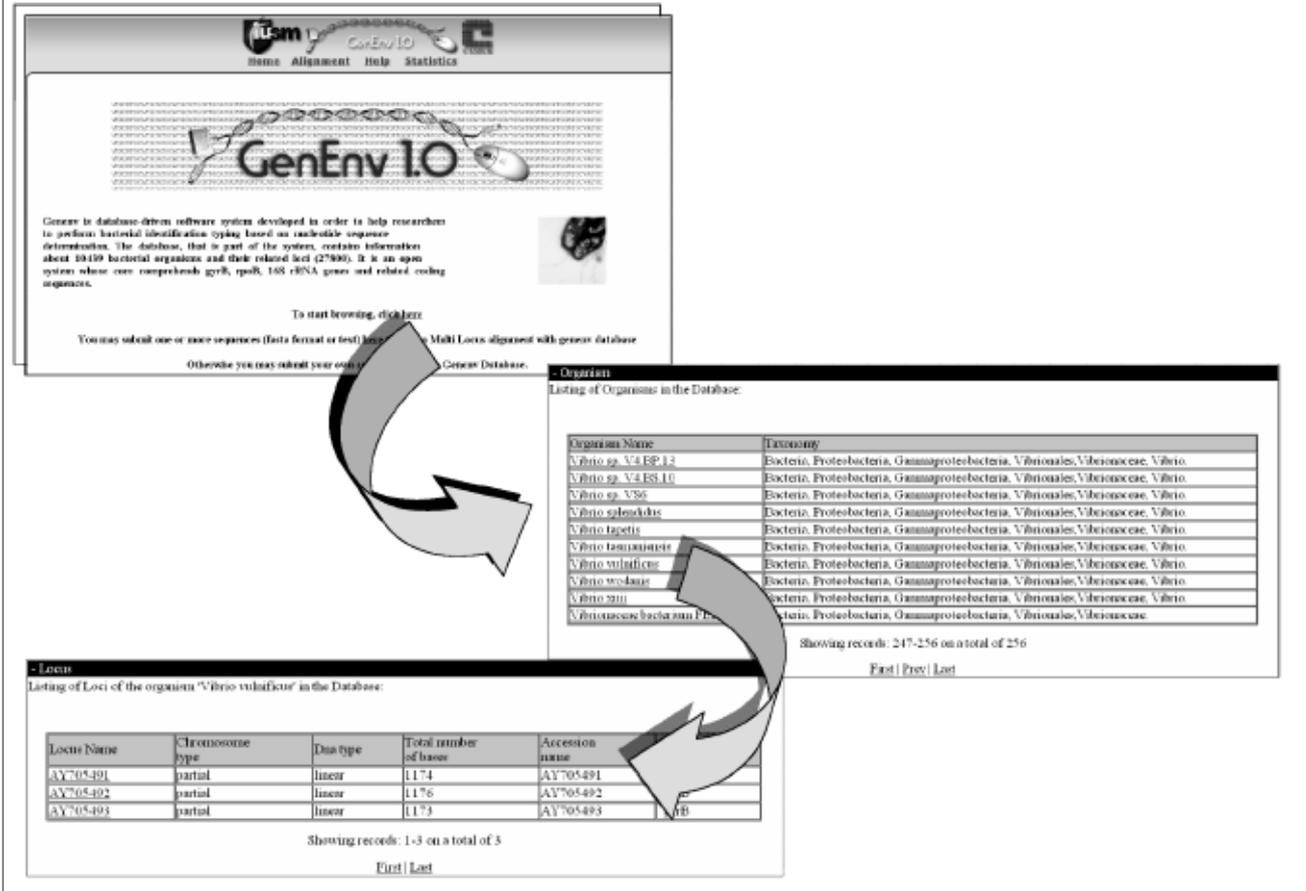
mitted to GenEnv, by clicking on the “search” button. At this point GenEnv provided the complete list of organisms, present in the database. The total number of records was obtained from the bottom of the result window. A searchable table listed the name and taxonomy associated to each organism. Additional information, such as locus name, chromosome type, total number of bases, accession name and locus associated genes, was obtained by clicking on each individual organism. Bacterial loci were randomly verified by following the related links. Primer pairs for PCR amplification, as well as restriction enzymes for sequence analysis, were obtained from the “histogram” window. Once retrieved, the output for the query “*Vibrio*” was compared to all the database content through the “statistics” link, which provided information on the total number ( $n = 27,800$ ) of available CDS, loci and organisms. Subsequently, the output was parsed by Perl scripts, to find the families and CDSs acquired through dynamic queries.

## Results and Discussion

Under the query “*Vibrio*” GenEnv provided 256 organisms and a total number of 19 families (Tab. I). The total number of available sequences, retrievable from GenEnv, under the query “*Vibrio*”, was 548 and included: 16S rRNA ( $n = 402$ ), *rpoB* ( $n = 1$ ), *gyrB* ( $n = 145$ ). The GenEnv output contained 5 uncultured species of vibrios, with at least one 16S rRNA CDS sequence available, in particular: uncultured *Cellvibrio* sp. belonging to the family of the *Pseudomonadaceae*, uncultured *Desulfovibrio* sp. belonging to the family of *Desulfovibrionaceae*, uncultured *Desulfovibrionaceae* bacterium belonging to the family of *Desulfovibrionaceae*, uncultured *Photobacterium* sp. belonging to the family of *Vibrionaceae*, uncultured *Vibrio* sp. belonging to the family of *Vibrionaceae*. Retrieved data, under the query “*vibrio*”, were compared to the previous contents of the database. In one year (from November 04 to November 05), the total number of retrievable organisms, grew from 192 to 256 (+ 33%), including the new uncultured organism, *Photobacterium* sp. and the family *Actinomycetaceae*, showing the increasing interest in this microbial *taxon*.

Microrganisms belonging to uncultured species arouse particular interest, in ecological studies. It is widely recognized that as-yet-uncultured microrganisms represent the vast majority of organisms in most environments. This evidence mainly derives from the analyses of 16S rRNA gene sequences, amplified directly from environmental matrices, by culture-independent methods. These strategies are acquiring increasing relevance, so that a new term was coined “metagenomics”, indicating environmental and community genomics analysis by direct extraction and cloning of DNA from heterogeneous collections of environmental samples [31, 32]. The family *Vibrionaceae* includes a heterogeneous group of microrganisms, with different ecological properties, that can be successfully

**Fig. 1.** GenEnv Home Page (A). By clicking on the specific link, the user can access the database browser and insert, in the specified field, the key to query the database. In this example the key "Vibrio" provided the organisms shown in the result window (B). By clicking on the organism name, GenEnv provides the list of available associated loci. In this example *Vibrio vulnificus* was selected (C), allowing access to available sequences and further analyses.



investigated by the use of bioinformatics tools such as the GenEnv system.

Bioinformatics offers advanced tools to archive, manage and retrieve an overwhelming wealth of steadily growing information. Its wide applicability covers several scientific fields, including also environmental microbiology. In particular, bioinformatics relies on the availability of exploitable genetic data. As a result, filtered databases, with reduced or minor redundancy, can effectively improve the efficiency in detecting and typing environmental species. In this context, the GenEnv system was conceived and developed to represent an easy and effective online instrument, to study bacteria, starting from a panel of three selected genes: 16S rRNA, *rpoB* or *gyrB*.

A major advantage of the GenEnv consists in its applicability to any bacterial isolate, either clinical or collected from the environment, and either not previously described or uncultured. GenEnv approach is not strictly constrained to a fixed set of mutations belonging to a specific microorganism, as are multilocus-based strategies [25]. MLST database is very specific, but available only for already characterised species such as *Vibrio cholerae* or *Vibrio vulnificus*. It provides effective typing especially when submitting seven or

**Tab. I.** GenEnv output, under the query "Vibrio".

Family	Nr. of organisms
Acidaminococcaceae	1
Actinomycetaceae	4
Bdellovibrionaceae	5
Chromatiaceae	2
Clostridiaceae	1
Deferribacteraceae	2
Desulfohalobiaceae	2
Desulfomicrobiaceae	10
Desulfonatrumaceae	2
Desulfovibrionaceae	38
Lachnospiraceae	4
Neisseriaceae	1
Nitrospiraceae	1
Undefined	2
Pseudomonadaceae	9
Rhodocyclaceae	2
Rhodospirillaceae	1
Succinivibrionaceae	4
Vibrionaceae	165
Total	256

more *loci*, amplified and fully sequenced. The MLST approach cannot be adopted when dealing with species whose MLST database has not been yet realised, such as *Desulfovibrio desulfuricans*, *Enterovibrio norvegicus*, *Vibrio aestuarianus*, *Vibrio splendidus* or unculturable vibrio spp. GenEnv is less focused on each single species, but can potentially provide immediate information on any bacterial isolate. In addition, several user-friendly tools allow identification of PCR primers, restriction maps, graphical diagrams of con-

served regions. During the last decade, several microorganisms have been studied and sequenced and a total of 281 complete genomes are presently retrievable from the Entrez database, including several vibrios such as *Vibrio parahaemolyticus*, *Vibrio fischeri*, *Vibrio vulnificus* and *Vibrio cholerae*. Availability of full genomic sequences and advances in bioinformatics will provide enticing and promising perspectives for basic research, epidemiology and biological risk management.

## References

- [1] Ruby EG, McFall-Ngai MJ. *A squid that glows in the night: development of an animal-bacterial mutualism*. J Bacteriol 1992;174:4865-70.
- [2] Shukla BN, Singh DV, Sanyal SC. *Attachment of non-culturable toxigenic Vibrio cholerae O1 and non-O1 and Aeromonas spp. to the aquatic arthropod Gerris spinolae and plants in the River Ganga, Varanasi*. FEMS Immunol Med Microbiol 1995;12:113-20.
- [3] Thompson FL, Iida T, Swings J. *Biodiversity of vibrios*. Microbiol Mol Biol Rev 2004;68:403-31.
- [4] Sack DA, Sack RB, Nair GB, Siddique AK. *Cholera*. Lancet 2004;363:223-33.
- [5] Bhattacharya SK, Bhattacharya MK, Nair GB, Dutta D, Deb A, Ramamurthy T, et al. *Clinical profile of acute diarrhoea cases infected with the new epidemic strain of Vibrio cholerae O139: designation of the disease as cholera*. J Infect 1993;27:11-5.
- [6] Elhadi N, Radu S, Chen CH, Nishibuchi M. *Prevalence of potentially pathogenic Vibrio species in the seafood marketed in Malaysia*. J Food Prot 2004;67:1469-75.
- [7] Hervio-Heath D, Colwell RR, Derrien A, Robert-Pillot A, Fournier JM, Pommepuy M. *Occurrence of pathogenic vibrios in coastal areas of France*. J Appl Microbiol 2002;92:1123-35.
- [8] Lee CY, Pan SF, Chen CH. *Sequence of a cloned pR72H fragment and its use for detection of Vibrio parahaemolyticus in shellfish with the PCR*. Appl Environ Microbiol 1995;61:1311-7.
- [9] Campbell MS, Wright AC. *Real-time PCR analysis of Vibrio vulnificus from oysters*. Appl Environ Microbiol 2003;69:7137-44.
- [10] Austin B, Dawyndt P, Gyllenberg M, Koski T, Lund T, Swings J, et al. *Sliding window discretization: a new method for multiple band matching of bacterial genotyping fingerprints*. Bull Math Biol 2004;66:1575-96.
- [11] Lan R, Reeves PR. *Pandemic spread of cholera: genetic diversity and relationships within the seventh pandemic clone of Vibrio cholerae determined by amplified fragment length polymorphism*. J Clin Microbiol 2002;40:172-81.
- [12] Maluping RP, Ravelo C, Lavilla-Pitogo CR, Krovacek K, Romalde JL. *Molecular typing of Vibrio parahaemolyticus strains isolated from the Philippines by PCR-based methods*. J Appl Microbiol 2005;99:383-91.
- [13] Kenzaka T, Yamaguchi N, Tani K, Nasu M. *rRNA-targeted fluorescent in situ hybridization analysis of bacterial community structure in river water*. Microbiology 1998;144:2085-93.
- [14] Hernandez G, Olmos J. *Molecular identification of pathogenic and nonpathogenic strains of Vibrio harveyi using PCR and RAPD*. Appl Microbiol Biotechnol 2004;63:722-7.
- [15] Farfan M, Minana D, Fuste MC, Loren JG. *Genetic relationships between clinical and environmental Vibrio cholerae isolates based on multilocus enzyme electrophoresis*. Microbiology 2000;146:2613-26.
- [16] Kotetishvili M, Stine OC, Chen Y, Kreger A, Sulakvelidze A, Sozhamannan S, et al. *Multilocus sequence typing has better discriminatory ability for typing Vibrio cholerae than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness*. J Clin Microbiol 2003;41:2191-6.
- [17] Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S, Hoste B, et al. *Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis*. Appl Environ Microbiol 2005;71:5107-15.
- [18] Wong HC, Shen CT, Chang CN, Lee YS, Oliver JD. *Biochemical and virulence characterization of viable but nonculturable cells of Vibrio parahaemolyticus*. J Food Prot 2004;67:2430-5.
- [19] Warner JM, Oliver JD. *Randomly amplified polymorphic DNA analysis of starved and viable but nonculturable Vibrio vulnificus cells*. Appl Environ Microbiol 1998;64:3025-8.
- [20] Wang S, Levin RE. *Discrimination of viable Vibrio vulnificus cells from dead cells in real-time PCR*. J Microbiol Methods 2005.
- [21] Vora GJ, Meador CE, Bird MM, Bopp CA, Andreadis JD, Stenger DA. *Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic Vibrio spp*. Proc Natl Acad Sci USA 2005;102:19109-14.
- [22] Dziejman M, Serruto D, Tam VC, Sturtevant D, Diraphat P, Faruque SM, et al. *Genomic characterization of non-O1, non-O139 Vibrio cholerae reveals genes for a type III secretion system*. Proc Natl Acad Sci USA 2005;102:3465-70.
- [23] Harwood VJ, Gandhi JP, Wright AC. *Methods for isolation and confirmation of Vibrio vulnificus from oysters and environmental sources: a review*. J Microbiol Methods 2004;59:301-16.
- [24] Vieira VV, Teixeira LF, Vicente AC, Momen H, Salles CA. *Differentiation of environmental and clinical isolates of Vibrio mimicus from Vibrio cholerae by multilocus enzyme electrophoresis*. Appl Environ Microbiol 2001;67:2360-4.
- [25] Chan MS, Maiden MC, Spratt BG. *Database-driven multi locus sequence typing (MLST) of bacterial pathogens*. Bioinformatics 2001;17:1077-83.
- [26] Gubala AJ. *Multiplex real-time PCR detection of Vibrio cholerae*. J Microbiol Methods 2005.
- [27] Thompson CC, Thompson FL, Vandemeulebroecke K, Hoste B, Dawyndt P, Swings J. *Use of recA as an alternative phylogenetic marker in the family Vibrionaceae*. Int J Syst Evol Microbiol 2004;54:919-24.
- [28] MacDonell MT, Swartz DG, Ortiz-Conde BA, Last GA, Colwell RR. *Ribosomal RNA phylogenies for the vibrio-enteric group of eubacteria*. Microbiol Sci 1986;3:172-5,178.
- [29] Romano-Spica V, Santoni D, Papanini A, Orsini M, Castrignanò T. *Il database GenEnv: un nuovo strumento di bioinformatica per la sanità pubblica*. J Prev Med Hyg 2004;45:160.

- [30] Romano-Spica V, Santoni D, Papparini A, Orsini M, Giammanco G, Castrignanò T. *Bioinformatica e Database GenEnv nella gestione del rischio biologico*. *Igiene e Sanità Pubblica* 2004;6:403-22.
- [31] Handelsman J. *Metagenomics: application of genomics to uncultured microorganisms*. *Microbiol Mol Biol Rev* 2004;68:669-85.
- [32] Riesenfeld CS, Schloss PD, Handelsman J. *Metagenomics: genomic analysis of microbial communities*. *Annu Rev Genet* 2004;38:525-52.

■ Received on November 29, 2005. Accepted on March 6, 2006.

■ Correspondence: Prof. V. Romano-Spica, IUSM, Public Health Unit, p.zza Lauro De Bosis 6, 00194 Rome, Italy - Tel./Fax +39 06 36733247 - E-mail: vrs@iusm.it.