Comparison Between Complete Genome Sequences of *Vibrionaceae* Species

Patrícia Pilisson Côgo, João Meidanis {patricia.cogo, meidanis}@ic.unicamp.br

University of Campinas, Institute of Computing

Introduction

We describe a methodology employed to compare six genomes of five *Vibrionaceae* species that have been completely sequenced.

The analysis of these data is based on “Genome Rearrangements”, i.e., mutations involving the movement of large segments, instead of classical methods of genome analysis by string comparison looking for point mutations.

The distance matrices obtained will be employed to build phylogenetic trees of these species. We expect that with these data we will gain a deeper understanding of the evolution of vibrios and improve the taxonomic classification based on traditional methods.

Biologic Characteristics

Vibrionaceae is a heterogenous family that comprises, according to recent research, species from the genuses *Vibrio*, *Photobacterium*, *Enterobacterium* and others.

Vibrios, i.e., *Vibrionaceae* strains, are gram negative, γ-Proteobacteria. They are abundantly found in aquatic environments or in associatiation with eucaryots. Although some associations could be symbiotic, severe pathogens of humans and other species belong to this group [1].

They have two chromosomes, one larger and one smaller. The majority of genes that endode cell functions and pathogenic factors are located in the large one. The small chromosome usually contains genes for environmental adaptation.

Six strains have had their genomes completely sequenced. They are:

•***Photobacterium profundum***: A high-pressure adapted bacterium, sequenced by University of Padova, Italy, in 2004.

•***Vibrio choleare***: The etiologic agent of cholera, a severe disease in developing countries. There are pathogenic and nonpathogenic strains of this species and it was the first vibrio completely sequenced, by TIGR, USA, in 2000.

•***Vibrio fisheri***: A marine, nonpathogenic bacterium, symbiont in the light-emitting organs of certain squids and fish, sequenced by Kewalo Marine Laboratory, Hawaii, USA, in 2005.

•***Vibrio parahaemolyticus***: Human pathogen that causes gastroenteritis by invading or interacting with intestinal epithelial cells. It was sequenced by Osaka University, Japan, in 2003.

•***Vibrio vulnificus* CMCP6 and YJ016**: Important etiologic agents of infections and septicemia in humans, acquired by wound or contamined seafood. They were sequenced by Chonnan University, South Korea, and NHRI, Taiwan, respectivelly, in 2003.

Some details of these genomes are showing in table 1.

	Chromosome 1		**Chromosome 2**	
	Size (MBPs)	Proteins	Size (MBPs)	Proteins
P. profundum	4.0853	3416	2.23794	2008
V. choleare	2.19115	2742	1.07231	1093
V. fisheri	2.90618	2575	1.33202	1172
V. parahaemolyticus	3.288556	3080	1.87721	1752
V. vulnificus CMCP6	3.2194	2962	1.84485	1562
V. Vulnificus YJ016	3.35451	3259	1.85707	1696

Table 1: Details of *Vibriocanea* complete genomes

Methodology

Our approach is strongly based on the series of steps tradicionally followed when mesuring evolution in Bioinformatics, namely:

•Define the model;
•Identify homolog structures;
•Employ the model and these structures to find the most parcimonious sequence of events that can explain the differences between the genomes;
•Based on the most parcimonious scenario, estimate real mutation rates.

Initially, our model was based on the ocurrence of *Block Interchange* events, i.e, swapping of two not adjacent segments of any length [2]. Although the power of this rearrangement operation is to explain large genomic differences in few events, it has serious drawbacks. It does not take into account the ocurrence of translocations (the exchange of segments between different chromosomes – vibrios have two) and reversals (a fundamental operation that occurs frequently during the evolution of species), as can be observed in figure 1.

The identification of homolog structures is done in two different ways:

•Breaking the genomes in large blocks through mechanisms to find similarity, as in the software tool Mauve [3];
•Using the original gene sequence with specific criteria to define homology. The criterium adopted is based on high connectivity of protein similarity as described by McLysaght el. al. [4]. However, its adequacy to this study is still under analysis.

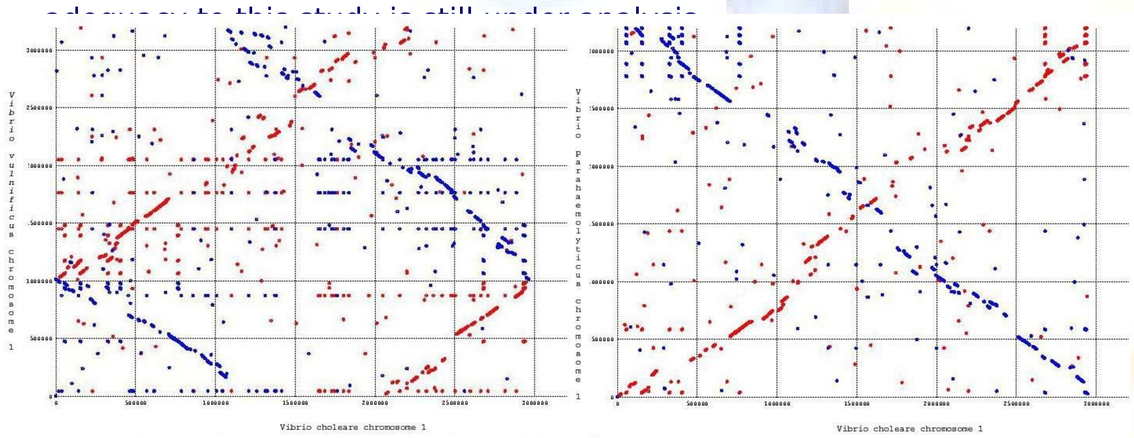

Fig. 1: Dot plots that provide a visual comparison between genomes. From left to right: *V. vulnificus* CMCP6 (vertical) x *V. choleare* and *V. parahaemolyticus* x *V. choleare.* Revelsals are shown in blue.

Estimation of real rearrangement rates has received little attention in the literature and is until this moment an open question.

There are many free tools that ccan be used to build phylogenetic trees from distance matrices, as we are doing in this work. We investigate some of them, spefically the softwares Fitch, Kitsch and Neighbor, avaliable in the PHYLIP package (*http://evolution.genetics.washington.edu/phylip*).

Acknowledgments

Fabiano Thompson and Marcelo da Silva Reis, consultants in biologic and practical aspects.

References

[1] F. L. Thompson, T. Iida, and J. Swings. Biodiversity of vibrios. *Microbiology and Molecular Biology Reviews*, pages 403–431, 2004.
[2] D. A. Christie. Sorting permutations by blockinterchange. *Information Processing Letters*, 60:165–169, 1996.
[3] A. C. Darling, B.Mau, F. R. Blattner, and N. T. Perna. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14(7):1394–140, 2004.
[4] A. Mclysaght, P. F. Baldi, and B. S. Galt. Extensive gene gain associated with adaptive evolution of poxviruses. *Proceedings of the National Academy of Sciences of the USA*, 100(26):15655–15660, 2003.