INVITED: **Comparative Microbial Genomics**

Giri Narasimhan
Bioinformatics Research Group (BioRG)
School of Computing and Information Sciences
Florida International University
Miami, FL 33199, USA.
Email: giri@cs.fiu.edu

***Abstract*—As microbial sequencing data starts to pour in at an increasing rate, comparative genomics holds the keys to decipher and mine this wealth of information. We discuss the diverse ways in which the availability of comparative genomics data allows us to answer more questions with greater ease and greater confidence.**

***Keywords*-microbes; comparative genomics;**

I. INTRODUCTION

Microbes display a staggering level of diversity in their phenotypes, biological processes, cellular organizations, and functional capabilities, often in order to adapt to their harsh environments, compete with other microbes, or to reproduce. While there do exist microbes with enormous genomes, most microbes have smaller and simpler genomes and are a great place to start to tackle the mysteries of life on earth and to attempt to comprehend their complexities. Microbes are relatively inexpensive to sequence and easy to experiment with in the laboratory. It is therefore small wonder that model systems such as the bacterium, *Escherichia coli*, is one of the best studied organisms.

One of the consequences of the dramatic decrease in costs and time for sequencing is an unprecedented accumulation of genomic sequence information on multiple closely related pathogenic microbes. For example, complete genome sequences for at least 18 *Listeria*, 25 *Burkholderia*, and 21 *Vibrio* species/strains are available. The resulting challenge is to develop meaningful methodologies and perform systematic and comprehensive comparative studies of these genomes. Comparative genomics has emerged as a powerful approach to study the molecular basis for the behavior and evolution of related organisms. In short, **comparative genomics helps to understand the genomic context of the phenotypic differences** between the organisms in question. For pathogenic bacteria that have virulent and non-virulent variants, comparative genomics assists in understanding the genes, proteins, and processes that dictate their pathogenicity. These studies can also provide clues to comprehend antibiotic resistance mechanisms, evolution (in the environment and in the host), hotspots in genomes, horizontal gene transfers, and much more.

When comparing a set of genomes, it is possible to use the data and existing tools to "pluck the low hanging fruit" (e.g., size comparison, ortholog tables, presence and absence of individual genes and proteins, and more). However, the challenge is in asking the right questions, and teasing out difficult correlations, to build the right models of relationships, and to visualize or present the information appropriately. Another important challenge is to design tools that can scale as problem sizes grow in the future. Finally, we recognize the need for more smart databases that can not only store all the information, but provide an environment for hypothesis generation.

II. APPLICATIONS

Several important questions in the area of comparative microbial genomics remain unsolved. First, existing tools do a poor job in helping us detect and understand *genome rearrangements* such as inversions, translocations, and large repeat regions. Second, the evolution of co-habiting bacterial species and strains remains largely a mystery. We only have a limited understanding of the nature, scale, and significance of *horizontal gene transfers*. Third, to understand the *evolution* of cellular processes, it is important to study the evolution of transcription factors and their binding sites, regulatory networks and mechanisms, and protein-protein interaction networks. Finally, from a human health perspective, it is critical to understand the differential *pathogenicity* of individual microbes and communities of microbes in terms of their virulence factors, antibiotic resistance, quorum sensing, ability to form biofilms and coexistence in diverse microbial communities.

A variety of comparative genomics projects in the Bioinformatics Research Group (BioRG) are designed to address these questions on microbial organisms.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of NIH Grants P01 DA15027-01 and NIH/NIGMS S06 GM008205, FREA Award from NICHD/EARDA grant G11HD038341, and a grant from the Florida Department of Health.