Mini Review

It Medical Team
https://www.itmedicalteam.pl/

Archives of Clinical Microbiology
1989-8436

2022
Vol. 13 No. 9: 201

Microbial Genomics: Genome Sequencing

Dr. Komalikha Rizavi*

Pakistan women Medical College, RR University of Microbiology, Pakistan.

Corresponding author:
Dr. Komalikha Rizavi

komarizavi@gmail.com

Pakistan women Medical College, RR University of Microbiology, Pakistan.

Citation: Rizavi K (2022) Microbial Genomics: Genome Sequencing. Arch Clinic Microbio, Vol. 13 No. 9: 201.

Abstract

Bacterial genomes are generally smaller and less variant in size among species when compared with genomes of eukaryotes. Bacterial genomes can range in size anywhere from about 130 kbp to over 14 Mbp. A study that included, but was not limited to, 478 bacterial genomes, concluded that as genome size increases, the number of genes increases at a disproportionately slower rate in eukaryotes than in non-eukaryotes [1]. Thus, the proportion of non-coding DNA goes up with genome size more quickly in non-bacteria than in bacteria. This is consistent with the fact that most eukaryotic nuclear DNA is non-gene coding, while the majority of prokaryotic, viral, and organellar genes are coding. Right now, we have genome sequences from 50 different bacterial phyla and 11 different archaeal phyla. Second-generation sequencing has yielded many draft genomes (close to 90% of bacterial genomes in GenBank are currently not complete); third-generation sequencing might eventually yield a complete genome in a few hours. The genome sequences reveal much diversity in bacteria. Analysis of over 2000 Escherichia coli genomes reveals an E. coli core genome of about 3100 gene families and a total of about 89,000 different gene families. Genome sequences show that parasitic bacteria have 500–1200 genes, free-living bacteria have 1500–7500 genes, and archaea have 1500–2700 genes [2]. A striking discovery by Cole et al. described massive amounts of gene decay when comparing Leprosy bacillus to ancestral bacteria. Studies have since shown that several bacteria have smaller genome sizes than their ancestors did. Over the years, researchers have proposed several theories to explain the general trend of bacterial genome decay and the relatively small size of bacterial genomes. Compelling evidence indicates that the apparent degradation of bacterial genomes is owed to a deletional bias.

Received: 02-Sep-2022, Manuscript No. IPACM-22-13057; **Editor assigned:** 05-Sep-2022, Pre-QC No IPACM-22-13057 (PQ); **Reviewed:** 23-Sep-2022, QC No. IPACM-22-13057; **Revised:** 26-Sep-2022, Manuscript No. IPACM-22-13057 (R); **Published:** 30-Sep-2022; DOI: 10.36648/1989-8436-10.9-201

Methods and Techniques

As of 2014, there are over 30,000 sequenced bacterial genomes publicly available and thousands of metagenome projects. Projects such as the Genomic Encyclopedia of Bacteria and Archaea (GEBA) intend to add more genomes. The single gene comparison is now being supplanted by more general methods. These methods have resulted in novel perspectives on genetic relationships that previously have only been estimated. A significant achievement in the second decade of bacterial genome sequencing was the production of metagenomic data, which covers all DNA present in a sample. Previously, there were only two metagenomic projects published [3].

Genomes of Bacteria

Log-log plot of the total number of annotated proteins in genomes submitted to GenBank as a function of genome size. Based on data from NCBI genome reports. Bacteria possess compact genome architecture distinct from eukaryotes in two important ways: bacteria show a strong correlation between genome size and number of functional genes in a genome, and those genes are structured into operons. The main reason for the relative density of bacterial genomes compared to eukaryotic genomes (especially multicellular eukaryotes) is the presence of noncoding DNA in the form of intergenic regions and introns. Some notable exceptions include recently formed pathogenic bacteria [4]. This was initially described in a study by Cole et al. in which Mycobacterium leprae was discovered to have a significantly higher percentage of pseudogenes to functional genes (~40%) than its free-living ancestors. Furthermore, amongst species of bacteria, there is relatively little variation in genome size when compared with the genome sizes of other major groups of life.

Genome size is of little relevance when considering the number of functional genes in eukaryotic species. In bacteria, however, the strong correlation between the number of genes and the genome size makes the size of bacterial genomes an interesting topic for research and discussion. The general trends of bacterial evolution indicate that bacteria started as free-living organisms. Evolutionary paths led some bacteria to become pathogens and symbionts [5]. The lifestyles of bacteria play an integral role in their respective genome sizes. Free-living bacteria have the largest genomes out of the three types of bacteria; however, they have fewer pseudogenes than bacteria that have recently acquired pathogenicity. Facultative and recently evolved pathogenic bacteria exhibit a smaller genome size than free-living bacteria, yet they have more pseudogenes than any other form of bacteria. Obligate bacterial symbionts or pathogens have the smallest genomes and the fewest pseudogenes of the three groups. The relationship between life-styles of bacteria and genome size raises questions as to the mechanisms of bacterial genome evolution. Researchers have developed several theories to explain the patterns of genome size evolution amongst bacteria.

Phylogeny

As single-gene comparisons have largely given way to genome comparisons, phylogeny of bacterial genomes have improved in accuracy. The Average Nucleotide Identity (ANI) method quantifies genetic distance between entire genomes by taking advantage of regions of about 10,000 bp. With enough data from genomes of one genus, algorithms are executed to categorize species. This has been done for the Pseudomonas avellanae species in 2013 and for all sequenced bacteria and archaea since 2020 [6].

To extract information about bacterial genomes, core- and pan-genome sizes have been assessed for several strains of bacteria. In 2012, the number of core gene families was about 3000. However, by 2015, with an over tenfold increase in available genomes, the pan-genome has increased as well. There is roughly a positive correlation between the number of genomes added and the growth of the pan-genome. On the other hand, the core genome has remained static since 2012 [7]. Currently, the E. coli pan-genome is composed of about 90,000 gene families. About one-third of these exist only in a single genome. Many of these, however, are merely gene fragments and the result of calling errors. Still, there are probably over 60,000 unique gene families in *E. coli* [8].

Microbial gene

The identification of genes in prokaryotic genomes has advanced to the stage at which nearly all protein-coding regions can be identified with confidence. Computational gene finders using Markov modelling techniques now routinely find more than 99% of protein-coding regions5 and RNA genes6. Once the protein-coding genes have been located, the most challenging problem is to determine their function [9]. Typically, about 40–60% of the genes in a newly sequenced bacterial genome display a detectable sequence similarity to protein sequences whose function is at least tentatively known. This sequence similarity is the primary basis for assigning function to new proteins, but the transfer of functional assignments is fraught with difficulties [10].

Conclusion

So far, studies in genomics have only scratched the surface of microbial diversity and have revealed how little is known about microbial species. In the next few years, more than 100 projects for sequencing microbial genomes should be completed, providing the scientific community with information on more than 300,000 predicted genes. A significant number of these genes will be novel and of unknown function. These novel genes represent exciting new opportunities for future research and potential sources of biological resources to be explored and exploited. The benefits of comparative genomics in understanding biochemical diversity, So far, studies in genomics have only scratched the surface of microbial diversity and have revealed how little is known about microbial species. In the next few years, more than 100 projects for sequencing microbial genomes should be completed, providing the scientific community with information on more than 300,000 predicted genes. A significant number of these genes will be novel and of unknown function. These novel genes represent exciting new opportunities for future research and potential sources of biological resources to be explored and exploited. The benefits of comparative genomics in understanding biochemical diversity, virulence and pathogenesis, and the evolution of species have been unequivocally demonstrated and the usefulness of comparative techniques will improve as more genomes become available. One of the major challenges is to develop techniques for assessing the function of novel genes on a large scale and integrating information on how genes and proteins interact at the cellular level to create and maintain a living organism. It is not unreasonable to expect that, by expanding our understanding of microbial biology and biodiversity, great strides can be made in the diagnosis and treatment of infectious diseases and in the identification of useful functions in the microbial world that could be applied to agricultural and industrial processes.

References

1 Behjati S, Tarpey PS (2013) what is next generation sequencing. Arch Dis Child Educ Pract Ed 98: 236-238.

2 Chmielecki J, Meyerson M (2014) DNA sequencing of cancer: what have we learned. Annu Rev Med 65: 63-79.

3 Abate AR, Hung T, Sperling RA, Mary P, Rotem A, et al. (2013) DNA sequence analysis with droplet-based microfluidics. Lab on a Chip 13: 4864-4869.

4 Pekin D, Skhiri Y, Baret JC, Le Corre D, Mazutis L, et al. (2011) Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab on a Chip 11: 2156-2166.

5 Olsvik O, Wahlberg J, Petterson B, Uhlén M, Popovic T, et al. (1993) Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1 strains. J Clin Microbiol 31: 22-25.

6 Pettersson E, Lundeberg J, Ahmadian A (2009) Generations of sequencing technologies. Genomics 93: 105-111.

7 Callaway, Ewen (2021) Million-year-old mammoth genomes shatter record for oldest ancient DNA Permafrost-preserved teeth, up to 1.6 million years old, identify a new kind of mammoth in Siberia. Nature 590: 537-538.

8 Castro Christina, Marine Rachel, Ramos Edward, Ng Terry Fei Fan (2019) the effect of variant interference on de novo assembly for viral deep sequencing. BMC Genomics 21: 421.

9 Mahé P, El Azami M, Barlas P, Tournoud M (2019) A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis. PeerJ 7: 6857.

10 Moréra S, Larivière L, Kurzeck J, Aschke-Sonnenborn U, Freemont PS, et al. (2001) High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding. J Mol Biol 311: 569-577.