MEDWIN PUBLISHERS
Committed to Create Value for Researchers

International Journal of Zoology and Animal Biology
ISSN: 2639-216X

Comparative Genomics and Our Understanding of the Evolutionary Relationship

Viorica C*

University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca, Romania

***Corresponding author:** Cosier Viorica, University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca, Romania, Email: viorica.cosier@gmail.com

Short Communication

Volume 4 Issue 6

Received Date: December 21, 2021

Published Date: December 31, 2021

DOI: 10.23880/izab-16000344

Abstract

The development of molecular techniques and major advances in automating DNA sequencing and developing computer programs to analyse large amounts of sequence data, made the sequencing of large genome a real possibility in the mid-1980. The first results of these data were used for analysing genes and gene expression that has revolutionized experimental biology. Comparative genomics approaches have become powerful tools as multiple genome sequences are deciphered fully. Comparison of the entire genomes (or part of genomes) of different species, strains, or individuals, provides a high detailed view of our understanding of functions and evolutionary relationship of each genome.

Keywords: Comparative Genomics; Evolutionary Relationship; Sequence Data

Introduction

The first complete nonviral genome sequenced was the mithocondrion's circular genome in 1981, with 16,159 bp. In fact, the first complete eukaryotic genome sequenced was the genome of *S. cerevisae,* reported in 1996 [1], followed by the genome of *Caenorhabditis elegans* in 1998 [2]. With the human complete genome sequence, finalised in 2003, and nearly 3 billion base pair deciphered, the era of genomics begun [3]. Some milestones in genomic sequencing were achieved first in 2000 with the fruit fly genome, with 118,4 Mb in the euchromatic region (14,015 genes) and another 60 Mb of highly repetitive sequences, finalised in march 2000 [4], followed by the flowering plant *Arabidopsis thaliana*, also an important model organism, finalised in 2010, with a genome of 120 Mb and 25,900 genes; the mouse genome with 2,7 Mb and 22,000 protein coding genes [5]; and the dog genome (*Canis lupus familiaris*) (~ 2,500 Mb) with over 22,000 genes and 3,200 gene coding for RNAs [6]. All these final genome sequences were used as refences for the resequencing and annotation of new genomes, after comparative analyses. From 2010 until today 442,402 sequencing projects have been completed. (https://gold.jgi.doe.gov/statistics). Sequence analysis gives the most unambiguous evidence for the relationship among species; which organisms or viruses are present in a sample, the presence of a homologous gene in another organism, provide a valuable way to determine the function of human genes, to study the cancer-related gene variations within populations or to improve the capacity of natural resources management to protect species. In this review a brief description of few comparative genomics studies and uses are described.

Understanding of Functions and Evolutionary Relationship of Each Genome

The availability of multiple genomes offers the prospect of improved gene prediction by comparing genomic regions from multiple species and finding conserved (relatively unchanged) regions [7]. Common features of two organisms will often be encoded within the DNA that is conserved between the species [8]. In closely related species such as

humans and mice, exons of protein-coding genes tend to change substantially slower that the surrounding noncoding DNA. The DNA sequences controlling the expression of genes that are regulated similarly in two related species should also be conserved. Conversely, sequences that encode (or control the expression of) proteins and RNAs responsible for differences between species will themselves be divergent. The chimpanzee genome was compared with the genome of the mouse and rat to find regions where at least 96 of 100 bp were perfect matches and over 30,000 regions were found. These regions were then compared with human genome to find most dissimilar sequences. One of the genes identified in analysis was HAR-1 for human accelerated region 1. For chimpanzee HAR-1 gene is nearly identical to the chicken gene with only two bases differences, that has changed in 310 million years of evolution. But comparing the chimpanzee HAR-1 gene with human gene eight bases were different, suggesting that this region has clearly changed in last 6 million years, and positive selection in coding sequences is characteristic for evolution of human genes implicated in other functions, such as immune system and olfaction [9]. Several other key human genes have been identified in comparative genomics studies, like FOXP2 and ASPM, which plays important roles in speech and regulates brain size [10,11].

Virus Detection

Traditional viral detection techniques such as *in vitro* viral cultures, immunologic assays, and PCR can identify only one or just a few specific viral targets in a single test. Identification of many viruses that cause viral infections are challenging, but virus identification can be made simpler and more effective by using the comparative genomics and DNA microarray techniques. In the analysis, conserved viral sequences at the genus level are used; the virus probes are used not only to identify known viruses but also for discerning the genera of emerging or uncharacterized ones [12].

Metagenomics Analysis

It involves the analysis of the genomes in entire communities of microbes isolated from the environment, a DNA sample that derived from bacteria, viruses, protest and fungi. At the core of metagenomics analysis is the whole genome shotgun sequencing, and sequences are reassembled using complex computer algorithms. Each of the reassembled sequences can be compared with a DNA sequence in the database. The high throughput sequencing in characterizing the microbial communities can be used also in gut microbiome deciphering to understand how it changes with the health of the host.

References

1. Goffeau A, Barrell BG (1996) Life with 6000 genes. Science 274(546): 563-547.
2. Stein LD, Bao Z (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. Science 282(5396): 2012-2018.
3. Collins F, Green E, Guttmacher A (2003) A vision for the future of genomics research. Nature 422(6934): 835-847.
4. Adams MD, Celniker SE (2000) The genome sequence of Drosophila melanogaster. Science 287(5431): 2185-2195.
5. Waterston RH, Lindblad Toh K (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.
6. Lindblad Toh K, Wade CM (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438(7069): 803-819.
7. Miller W, Makova KD, Nekrutenko A, Hardison RC (2004) Comparative genomics. Annu Rev Genomics Hum Genet 5: 15-56.
8. Hardison RC (2003) Comparative Genomics. PLoS Biol 1(2): e58.
9. Haygood R, Babbitt CC, Fedrigo O, Wray GA (2010) Contrasts between adaptive coding and noncoding changes during human evolution. Proc Natl Acad Sci 107(17): 7853-7857.
10. Pollard KS, Salama SR, King B, Kern AD, Dreszer T, et al. (2006) Forces shaping the fastest evolving regions in the human genome. PLoS Genet 2(10): e168.
11. Levchenko A, Kanapin A, Samsonova A, Gainetdinov RR (2018) Human Accelerated Regions and Other Human-Specific Sequence Variations in the Context of Evolution and Their Relevance for Brain Development. Genome Biol Evol 10(1): 166-188.
12. Chou CC, Lee TT, Chen CH (2006) Design of microarray probes for virus identification and detection of emerging viruses at the genus level. BMC Bioinformatics 7: 232.
13. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, et al. (2018) Current understanding of the human microbiome. Nat Med 24(4): 392-400.