



# Analysis of Microbial Genomes Using the COG Method

Nelson Richard\*

Malopolska Centre of Biotechnology, Poland

\*Corresponding Author's E-mail: [richardnelson@rediff.com](mailto:richardnelson@rediff.com)

**Received:** 01-Apr-2023, Manuscript No. IRJBB-23-96812; **Editor assigned:** 03-Apr-2023, PreQC No. IRJBB-23-96812 (PQ); **Reviewed:** 17-Apr-2023, QC No. IRJBB-23-96812; **Revised:** 22-Apr-2023, Manuscript No. IRJBB-23-96812 (R); **Published:** 28-Apr-2023, DOI: 10.14303/2250-9941.2022.47

## Abstract

The Clusters of Orthologous Genes (COG) database has been a well-liked resource for comparative genomics and annotation of microbial genomes for the past 20 years. Apart from simple functional annotation of sequenced genomes, the COG have been used for tasks like (i) unifying genome annotation in groups of related organisms; (ii) identifying missing and/or undetected genes in complete microbial genomes; (iii) analysis of genomic neighbourhoods, which in many cases allows prediction of novel functional systems; (iv) analysis of genomic neighbourhoods; Here, we go over the fundamentals of the COG technique and go over its main benefits and shortcomings when it comes to analysing microbial genomes.

**Keywords:** Comparative genomics, Enzyme evolution, Genome annotation, Orthologs, Paralogs

## INTRODUCTION

The study of genomes' structure, function, evolution, mapping, and editing is the focus of the interdisciplinary field of biology known as genomics. A genome is an organism's entire set of DNA, which includes all of its genes and the three-dimensional, hierarchical structural organisation they are organised into. Genomic science tries to characterise and quantify all of an organism's genes, their interactions, and influences on the organism as a whole, as opposed to genetics, which studies specific genes and their functions in inheritance. Enzymes and messenger molecules can work with genes to direct the creation of proteins (Olusegun KA et al., 2019). Proteins, in turn, build up bodily tissues and organs, regulate chemical processes, and transmit messages between cells. In order to assemble and analyse the structure and function of complete genomes, genomics also entails the sequencing and analysis of genomes using high throughput DNA sequencing and bioinformatics. Systems biology and discovery-based research have undergone a revolution as a result of advances in genomics, making it easier to comprehend even the most intricate biological systems, like the brain (Hend MT et al., 2014).

Since the human genome's sequence is complete, the key

challenge is figuring out how to decipher the data encoded in the DNA sequence. Despite the fact that many genome-wide investigations have already been carried out, it is still difficult to ascertain how genes, gene products, and their interactions work. Functional analysis is crucial for human health since changes to the human genome are likely to result in pathological diseases (Morteza RT et al., 2013). Functional genomic analysis has been carried out using a range of methods and tools for many years. High-throughput techniques, which range from conventional real-time polymerase chain reactions to more complicated systems, such next-generation sequencing or mass spectrometry, have, however, just recently undergone a quick revolutionising advancement (Mohamed SA 2017). Furthermore, for accurate bioinformatics analysis and solid scientific outcomes, laboratory investigation alone is not sufficient. These techniques allow for precise and thorough functional analysis including several academic disciplines, including genomes, epigenomics, proteomics, and interatomic (Nwangwa JN et al., 2016). This is necessary to close knowledge gaps regarding dynamic biological processes at the cellular and organismal levels. To achieve a successful study, it is important to consider both the advantages and limits of each approach before selecting the best one for a certain research. This is why the current review paper's goal

is to outline the most popular and often utilised techniques for a thorough functional analysis (Obembe AO et al., 2015). UCLA pre-doctoral students whose purpose is to do genomics research are supported by the Genomic Analysis Training Programme, which is financed by an NIH grant. In order for students to flourish in this interdisciplinary discipline, the programme is designed to ensure that they have a strong biological, computational, and statistical basis. The Genomic Analysis Training Programme offers stipends and support for tuition to its participants each year. Additionally, the award covers the cost of travel to the annual NHGRI research and training conference (Saif Q et al., 2015).

## DISCUSSION

Reliable genome annotation—that is, the exact identification of the genes, including the accurate determination of gene borders and functional annotation of the gene product(s)—is essential to the success of the overall genomic enterprise. Proteins from entire microbial genomes can be categorised phylogenetically using the Clusters of Orthologous Groups of Proteins (COGs) database. Despite the COG system's expansion through time, it has always been the intention for each COG to stand for a family of orthologous protein-coding genes (Therese MG et al., 2019). The straightforward definition of orthology as a one-to-one relationship, however, does not accurately capture the evolutionary relationships between these genes when the compared genomes are separated by great evolutionary distances and have significantly different numbers of genes because of such evolutionary processes as lineage-specific gene duplication and loss as well as horizontal gene transfer. Due to the complexity of the interactions between genes that have developed over time, the COGs have developed into families of co-orthologous genes that represent both one-to-many and many-to-many links (Yunusa H et al., 2018). Thus, the term "orthologous groups" (of proteins) was developed to encompass these more intricate evolutionary interactions between genes and to make it easier to ascribe (generic) roles to genes and their offspring. The COGs have changed their name to Clusters of Orthologous Genes as the genomic community has come to accept the idea of co-orthologous links between genes. The association and correlation analysis paradigm is the current standard for genomic research of complicated disorders. Genome-wide association studies (GWAS) have made great progress in understanding the genetic architecture of complicated diseases, yet the genetic variants they have discovered can only partially account for the heritability of complex disorders. The majority of genetic variations are still unknown. The ability of association analysis to identify the underlying causes of complicated disorders is restricted. The paradigm of genetic analysis has to change from association analysis to causal inference (Celestina A et al., 2021).

## CONCLUSION

As a platform for comparative genomic research, the

COG technique for identifying orthologous genes was created not long after the first few microbial genomes had been sequenced. One would have predicted that in 20 years, evolutionary techniques will entirely replace this straightforward strategy based on sequence similarity hierarchy. This is not the case, however, in large part because of the limited extent of lineage-specific paralogy, differential gene loss, and domain shuffling, as well as the extended orthology conjecture, which states that bidirectional best hits between genomes correspond to orthologs, and the latter have equivalent functions. The COG method for identifying orthologous genes was developed shortly after the first few microbial genomes had been sequenced, and it serves as a platform for comparative genomic research. One would have assumed that this simple strategy based on sequence similarity hierarchy will be completely replaced by evolutionary techniques in 20 years. However, this is not the case, largely due to the limited extent of lineage-specific paralogy, differential gene loss, and domain shuffling, as well as the extended orthology conjecture, which states that bidirectional best hits between genomes correspond to orthologs, and that these latter have equivalent functions.

## REFERENCES

1. Olusegun KA, Emmanuel BO, Gbadebo EA, Jelili AB, Adedija DW (2019). Mitigation of Aluminium Phosphide-induced Hematotoxicity and Ovarian Oxidative Damage in Wistar Rats by Hesperidin. *American Journal of Biochemistry*. 9: 7-16.
2. Hend MT, Omnia EK, Hekmat MT, Amira A (2014). Potential anti-inflammatory effect of lemon and hotpepper extracts on adjuvant-induced arthritis in mice. *The Journal of Basic Applied Zoology*.
3. Morteza RT, Alireza B, Afshin S, Maryam NN, Hannan E, et al (2013). Alteration in Liver Enzymes in Aluminum Phosphide Poisoning a Retrospective Study. *Iranian Journal of Toxicology*. 7:1-4.
4. Mohamed SA (2017). Nephroprotective Effect of Melatonin against Aluminum Phosphide Induced Renal Tissue Damage in Rats. *Journal of Bioscience and Applied Research*. 3: 252-272.
5. Nwangwa JN, Udefa AL, Anyaibe SC, Iragunima AS, Eze NC (2016). Comparative Effect of Chilli Pepper (*Capsicum frutescens*) Extract and Capsaicin on Some Haematological Parameters and Serum Electrolytes in Albino Wistar Rats. *International Blood Research Reviews*. 6: 1-8.
6. Obembe AO, Ofutet EO, Okpo-ene AI, Okon VE, Eyong EE (2015). The Effect of Aqueous Extract of *Xylopiya Aethiopica* (Nigro Pepper) On Some Haematological Parameters in Albino Rats. *World Journal of Pharmaceutical Research*. 4: 2576-2583.
7. Saif Q, Ruhi K, Aparna S (2015). Aluminium phosphide induced acute kidney injury. *The Egyptian Journal of Internal Medicine*. 27: 115.
8. Therese MG, Emmanuel OA, Phumzile M, Busie M (2019). Cowpea (*Vigna unguiculata* (L.) Walp) for food security: an evaluation of end-user traits of improved varieties in Swaziland. *Sci Rep*. 9: 1-6.

9. Yunusa H, Hassan Z, Deepika V (2018). Preserving or Poisoning: A Case of Dried-Beans from Nigeria. *International Journal of Management Technology and Engineering*. 7: 2249-7455.
10. Celestina A, Makanju DA, Eunice A, TomiwJO (2021). Histological and proximate analysis of ash produced from sesame palm fruit and Parkia on albino rats. *International Journal of Biological and Pharmaceutical Sciences Archive*. 02: 026-030.