MANTECH Publications
Journal of Research in Microbiology and Immunology
Volume 7, Issue 2, May-August, 2025

Gene Ontology (Go)-Review

Dr. S. Sreeremya
External Faculty of Pharmacology
Department of Pharmacology
Crescent College OF Nursing, Palakkad, Kerala, India
E-mail Id: *sreeremyasasi@gmail.com*

ABSTRACT

The Gene Ontology (GO) project precisely garners a comprehensive source for functional genomics. The project is a collaborative effort that creates the evidence-supported annotations to describe the biological roles of the individual genomic products (e.g. the genes, proteins (a macromolecule), ncRNAs, and complexes) by classifying them using our ontologies.

KEYWORDS: *Gene Ontology (GO), genomic products, genes, proteins (a macromolecule), ncRNAs, annotations.*

INTRODUCTION

That is, graph structures comprised of the classes for molecular functions, the biological processes these contribute to, the cellular locations where these occur (the cellular components), and the relationships linking these, in a species-independent manner. A „GO annotation" describes the association between a class from the ontology and a gene product, as well as the references to the evidence supporting the association (Roncaglia et al.,2013). Nearly two decades of the efforts make the GO an integrated resource of the functional information or data for genes from over 460 000 species (including strains) covering the plants, animals, and the microbial world. The work of the Gene Ontology Consortium (GOC) addresses the need for the consistent descriptions and elaborate understanding of gene products across biological databases, providing not only comprehensive coverage of the biological concepts but also communitywide agreement on how those should be availed to describe gene functions across all organisms. There are three separate aspects to this effort: (i) the development and the maintenance of the ontology, (ii) the annotation of gene products

(Van Dam et al., 2013), and (iii) the development and the continuous improvement of tools and training that facilitate the creation, maintenance, and availing of the ontologies. Here researchers describe the latest improvements to the tools and resources of the GOC. Ontologies, annotations, and the bio tools are freely available via the Internet at http://www.geneontology.org (Mungall et al., 2014).

NEW FEATURES AND STRATEGIES FOR IMPROVEMENTS

Shared vocabularies are an important or efficient step toward unifying biological databases, yet as knowledge changes, the vocabularies and their use necessarily change, resulting in the individual curators evaluating data differently. To address the concerns of inconsistent data representation, the GOC continuously garners enhancements to its tools, resources, and policies, improving the annotation consistency and ensuring that the annotations reflect the current state of biological knowledge (Mungall et al., 2012).

The GO knowledgebase is large and widely dynamic for applications that use the components of the GO knowledgebase, it is quiet crucial that the ontology and associated annotations represent the current states of knowledge and are not just an archive of all public data. Therefore, all the aspects of the GO knowledgebase are dynamic (ontology, annotations, GO-CAMs, links to the external ontologies, etc.), and citable, versioned updates are released on a monthly basis. Researchers describe each component of the knowledgebase, focusing on recent changes made to improve the resource during the past couple of years. Statistics and descriptions given here are based on the GO release 2022–11–03 (http://release. geneontology.org/2022-11-03, doi:10.5281/zenodo.7407024)(Hastings et al.,2013).

ONTOLOGY-THE BRANCH OF SCIENCE

The ontology component of the GO knowledgebase consists of the terms availed to describe functional characteristics of gene products, which are linked together by the relations into a labeled directed acyclic graph (like a hierarchy but with the multiple parentages allowed). It also encompasses term definitions, synonyms, and relations to terms from external ontologies. The GO is available in different editions, encompassing (1) the “basic” edition, which includes only core relationship types; (2) the core ontology, encompassing additional relationship types; and (3) the “go-plus” edition which also encompasses relationships to terms in other ontologies. These editions are explained on the GO downloads page

http://geneontology.org/ docs/download-ontology/. The ontology contains precisely 43,303 terms, specifically linked together by 88,099 relationships in the basic edition. When relationships to external terms are encompassed, there are121, 698 relationships; release statistics can be viewed at the site named - http:// geneontology.org/stats(D önitz et al.,2012). The GO is subject to constant review and the revision to most accurately model the current biological knowledge. Revision of the ontology encompasses the addition or obsoletion of terms and reorganization of the relationship structure. New GO terms are precisely added to represent concepts previously missing from the GO in response to the published findings, or when a branch of GO is revised (Natale et al., 2011). Terms may be obsolete when unused or inconsistently availed in annotation, when they are redundant with the other terms, or during revision of specific branches of the ontology (Fig-1)

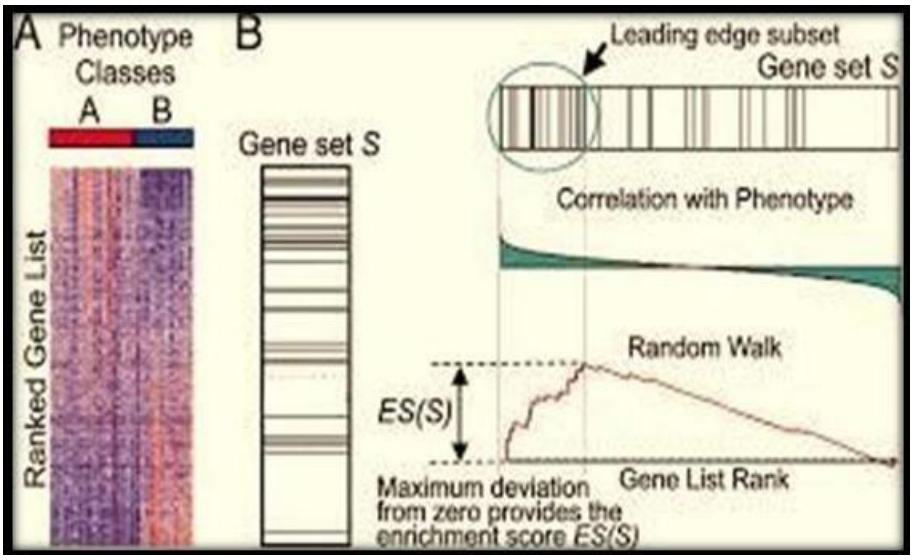

Figure: 1

ANNOTATIONS

A GO annotation is a statement asserting that a particular gene or the gene product has a particular functional characteristic (GO term). New annotations are continually added to the knowledgebase or the database or information base. In the past two years, experimentally supported gene function annotations have been added from over 10,000 scientific papers including both research and review papers. As of November 2022, the GO knowledgebase comprises experimental knowledge from almost 173,000 papers (Gaudet et al., 2011). GO annotations derived from the experimental data are added primarily by the annotation groups in the GO Consortium, which typically curate the biological knowledge by organism. GO

annotations are also regularly reviewed and may be edited or precisely removed from the knowledgebase for various reasons, particularly when the ontology terms are revised (see "Ontology" section above) or when annotations are invalidated by later experimental data.

The Annotations to terms that will be obsolete are manually reviewed and the annotations are made to a different term whenever possible (Huntleyet al., 2014). For example, when researchers edited the ontology for histone modifications, more than 2,000 annotations to the obsolete terms were manually reviewed, and the histone modifying enzymes were reannotated to the appropriate MF term, while annotations from the indirect effects were either removed or reannotated to different, appropriate GO terms. More minor annotation typically reviews occur regularly. The Phylogenetic Annotation with the GO project involves an integrated biocurator review and analysis of annotations that has provided additional quality control (Dimmer et al., 2012). The GO user community also plays a pivotal role in identifying incorrect annotations. Because each annotation can be traced to the published paper comprising the underlying evidence or describing a method used to infer the annotation, users can quickly verify the accuracy of the given annotation. Potential errors can be reported by clicking on the "Help" link at the top of the GO homepage (http:// geneontology.org). In addition, authors of a paper availed to create GO annotations can easily retrieve and review all the annotations from a given paper and suggest changes.

PHYLOGENETIC ANNOTATION

Phylogenetic Annotation availing GO (PAN-GO) project creates a set of biocurator-reviewed, selected GO annotations. The PAN-GO process is delineated in detail by few researchers. Briefly, using the PAINT software tool, a biocurator reviews all experimentally supported GO annotations collected for all members of a protein family, in the contexts of a phylogenetic tree from the PANTHER bio resource. They then select the most informative and nonredundant GO terms that represent the gene‟s functional characteristics and specifications. Biocurators then model the evolution of these characteristics in the tree by specifying branches along which the GO terms were gained or lost, taking into the account events such as duplications, the mutations, horizontal gene transfers, and taxonomic specificity. This allows for different members of the same family to be precisely annotated with different GO terms when justified by the experimental data or information. All PAN-GO annotations can be traced to the experimental evidence in one or more related genes. To date,

a total of 8,196 protein families (out of 11,720 families with experimental data) have been curated

FUNCTIONAL CONSERVATION REQUIRES A COMMON LANGUAGE OR LINGUISTICS FOR ANNOTATION

The impact of the grand biological unification more evident than in the eukaryotes, where the genomic sequences of the three model systems are already available (budding yeast, Saccharomycescerevisiae, completed in 1996 by few researchers; the nematode worm Caenorhabditiselegans, completed in 1998 by few researchers; and the fruit fly Drosophila melanogaster , completed earlier this year) and the two more (the flowering plant Arabidopsis thaliana and the fission yeast Schizosaccharomycespombe) are imminent. The complete genomic sequence of the human genomes is expected in a year or two, and the sequence of the mouse (Musmusculus) will likely follow for analysis shortly thereafter. The first comparison between two complete the eukaryotic genomes (budding yeast and worm5) revealed that a surprisingly a much large fraction of the genes in these two organisms displayed evidence of the orthology. About 12% of the worm genes(~18,000) encode proteins whose biological roles and efficiency could be inferred from their similarity to their putative orthologuesinyeast, comprising about 27% of the yeast genes (~5,690). Most of these proteins have been found to have a main role in the „core bio-logical processes" common to all eukaryotic cells, such as DNA replication, the transcription and metabolism. A three-way comparison among budding yeast, worm and the fruitfly shows that this relationship can be extended; the same subset of the yeast genes generally have recognizable homologues in the fly genome.

The Estimates of sequence and functional conservation between the genes of these model systems and those of the mammals are less reliable, as no mammalian genome sequence is yet known in its entirety. Nevertheless, it is clear that a high level of this quenceandbio functional conservation will extend to all eukaryotes, with the likelihood that genes and the proteins that carry out the core biological processes will again be probable orthologues (Tripathi et al., 2013).

Furthermore, since the late 1980s, many research and experimental confirmations of functional conservation between mammals and model organisms (commonly yeast) have

been specifcally published. His astonishingly high degree of sequence and functional conservation presents both the opportunities and challenges. The main opportunity lies in the possibility of the automated transfer of biological annotations from the experimentally tractable model organisms to the less tractable organisms based on the gene and protein sequence similarity. Such information can be used to improve human health or the agriculture. The challenge lies in meeting the requirements for a largely or entirely computational system for precisely comparing or transferring annotation among different species (Wick et al., 2014).

CONCLUSION

By suing several biological tools, gene banks and several connected database like sequence database, microarray database, the technique of gene ontology and annotation can be made systematic and efficient.

REFERNCES

1. Roncaglia,P., Martone,M., Hill,D., Berardini,T., Foulger,R. ,Imam,F., Drabkin,H., Mungall,C. and Lomax,J. (2013) The Gene Ontology (GO) Cellular Component Ontology: integration with SAO (Sub cellular Anatomy Ontology) and other recent developments. J. Biomed. Semant., 4.
2. Van Dam,T., Wheway,G., Slaats,G., Group,S.S., Huynen,M. andGiles,R. (2013) The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. Cilia,
3. Mungall, C.J., Dietze, H. and Osumi-Sutherland,D. (2014) Use ofOWL within the Gene Ontology. In: Proceedings of the 11thInternational Workshop on OWL: Experiences and Directions. Rivadel Garda, Italy, pp. 25–36.
4. Mungall, C., Torniai,C., Gkoutos,G., Lewis,S. and Haendel,M. (2012) Uberon, an integrative multi-species anatomyontology. Genome Biol., 13, R5.
5. Hastings, J., de Matos,P., Dekker,A., Ennis,M., Harsha,B., Kale,N.,Muthukrishnan,V., Owen,G., Turner,S., Williams,M. et al. (2013)The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res.,41,D456–D463

6. D önitz ,J. and Wingender,E. (2012) The ontology-based answers(OBA) service: A connector for embedded usage of ontologies in applications. Front. Genet., 3, doi:10.3389/fgene.2012.00197.
7. Natale,D.A., Arighi,C.N., Barker,W.C., Blake,J.A., Bult,C.J.,Caudy,M., Drabkin,H.J., D"Eustachio,P., Evsikov,A.V., Huang,H.et al. (2011) The Protein Ontology: a structured representation of protein forms and complexes. Nucleic Acids Res., 39, D539–D545.
8. Gaudet,P., Livstone, M.S., Lewis,S.E. and Thomas,P.D. (2011) Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief. Bio inform., 12, 449–462.
9. Huntley,R., Harris,M., Alam-Faruque,Y., Blake,J., Carbon,S., Dietze,H., Dimmer,E., Foulger,R., Hill,D., Khodiyar,V. et al. (2014)A method for increasing expressivity of Gene Ontology annotations using a compositional approach.BMC Bioinformatics, 15, 155.
10. Dimmer,E.C., Huntley,R.P., Alam-Faruque,Y., Sawford,T.,O"Donovan,C., Martin,M.J., Bely,B., Browne,P., Mun Chan,W.,Eberhardt,R. et al. (2012) The UniProt-GO annotation database in2011. Nucleic Acids Res., 40, D565–D570.
11. Tripathi,S., Christie,K.R., Balakrishnan,R., Huntley,R., Hill,D.P., Thommesen,L., Blake,J.A., Kuiper,M. and Lægreid,A. (2013) Gene Ontology annotation of sequence-specific DNA binding transcription factors: setting the stage for a large-scale curation effort. Database,2013, bat062, doi:10.1093/database/bat062.
12. Wick,H., Drabkin,H., Ngu,H., Sackman,M., Fournier,C., Haggett,J.,Blake,J., Bianchi,D. and Slonim,D. (2014) DFLAT: functional annotation for human development. BMC Bioinformatics, 15, 45.