

The UniProt – Gene Ontology Annotation (UniProt-GOA) Project

The Gene Ontology (GO) (www.geneontology.org)

A set of structured, controlled vocabulary terms provided by the Gene Ontology Consortium, that describe functional information for a particular gene product.

Currently ~40,000 GO terms exist (October 2013) which describe:

- Cellular components (subcellular locations) where the gene products are located
- Molecular functions that gene products normally carry out
- Biological processes that gene products are involved in

Figure 1 shows the number of terms for each category

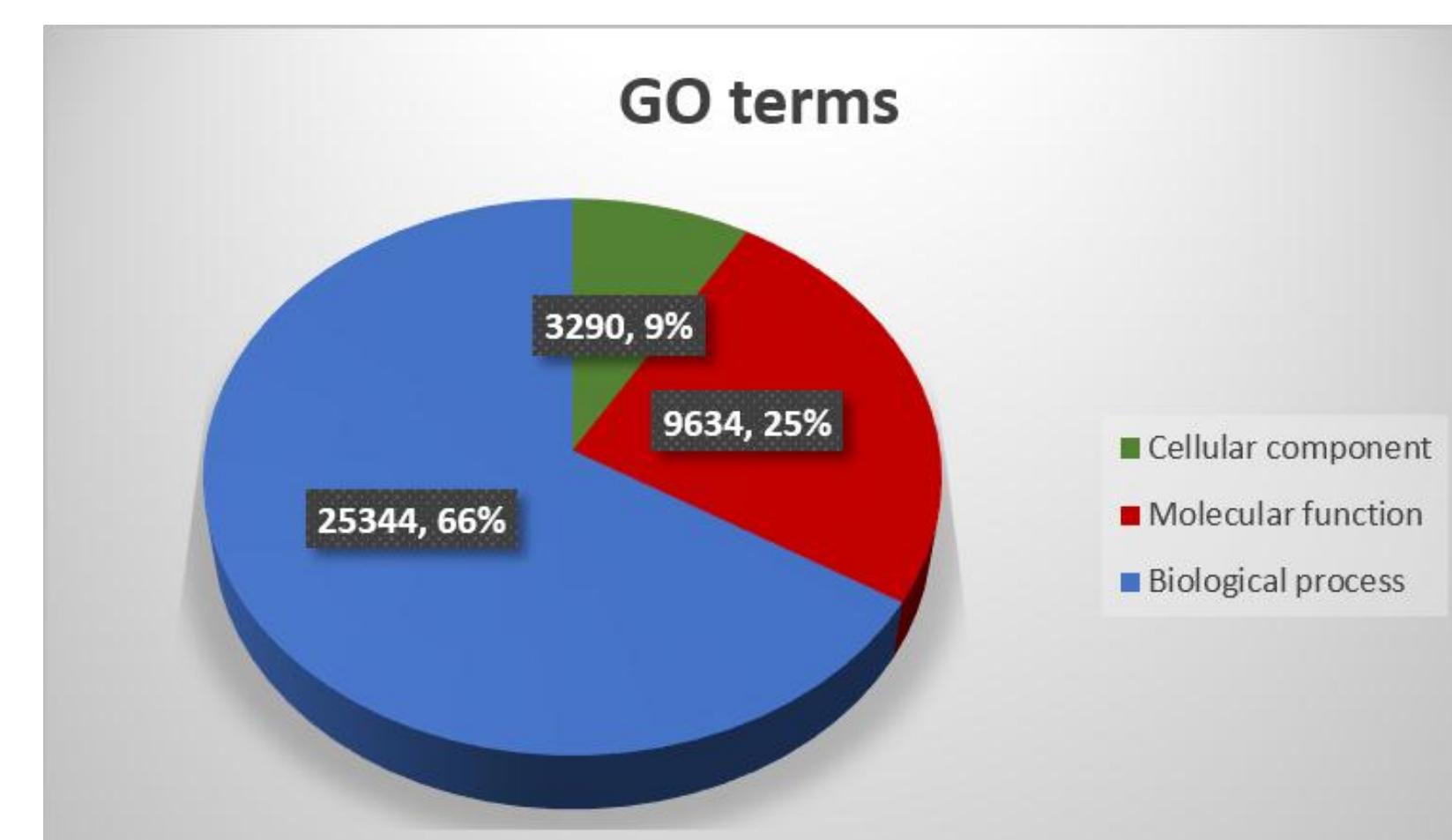


Fig. 1 Number of GO terms

The UniProt – Gene Ontology Annotation (UniProt-GOA) Project (www.ebi.ac.uk/GOA)

The UniProt-GOA project provides manually curated and electronically predicted GO annotations.

Manually curated annotations are captured on the basis of published literature by various curation groups worldwide, whereas predictions are created by groups such as HAMAP, InterPro, Ensembl Compara, etc. using sequence and structure similarity as well as phylogenetic relationships.

As shown in Figure 2, in October 2013 there were ~196,000,000 GO annotations to ~30,000,000 proteins, covering ~431,000 taxonomic groups.

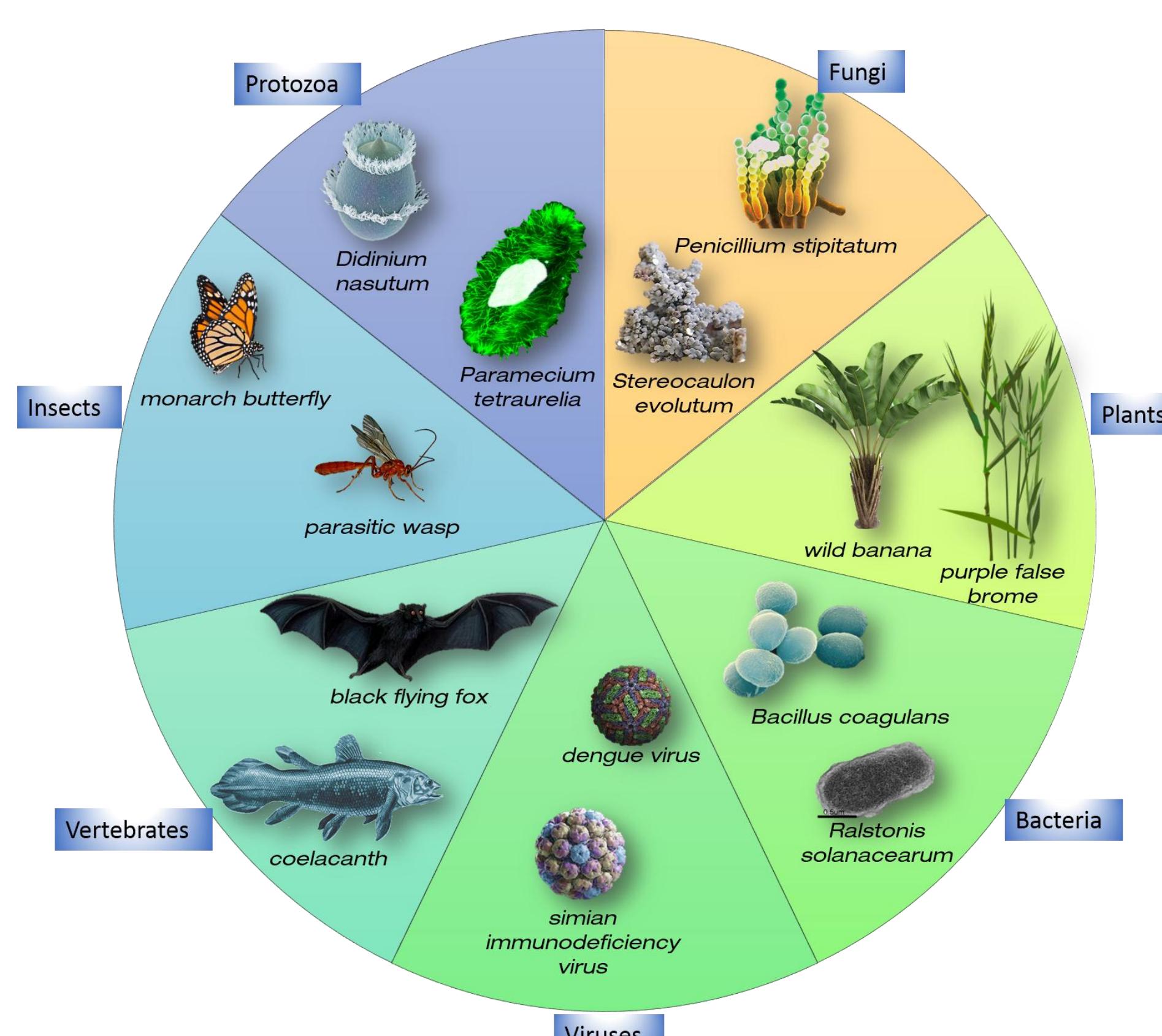


Fig. 3 Examples of the species with only automatic GO annotations

Manual annotation benefits

Manual annotation provides the most specific and detailed annotations for gene products.

One of our aims is to undertake focused annotation projects, to improve both the ontology and its association to gene products. Recent examples of this include annotation of proteins involved in kidney and heart development, apoptosis, necroptosis and proteins found in the peroxisome. Manual curation not only significantly improves the quality of annotations, but is also beneficial for the analysis of a complete dataset.

An example is shown in Figure 4, which is taken from one of our recent publications and shows the comparison between biological processes enriched either only in human (A) or yeast peroxisomal protein set (B). This work highlights that despite some differences between the two species, the majority of biological processes that occur in the peroxisome are similar between yeast and human.

Uses of GO Annotation



Combining the GO structure and specific associations between terms and gene products provides a powerful way to capture, query and analyse functional information independent of species.

Using a variety of third party tools you can:

- assess experimental investigations or computational predictions (e.g. for a subcellular fractionation assay)
- summarize the functional characteristics shared by a proteome, genome or group of related genes/proteins
- find the processes, functions or subcellular location(s) which a set of interacting proteins have in common
- formulate hypotheses for changes in gene or protein expression levels

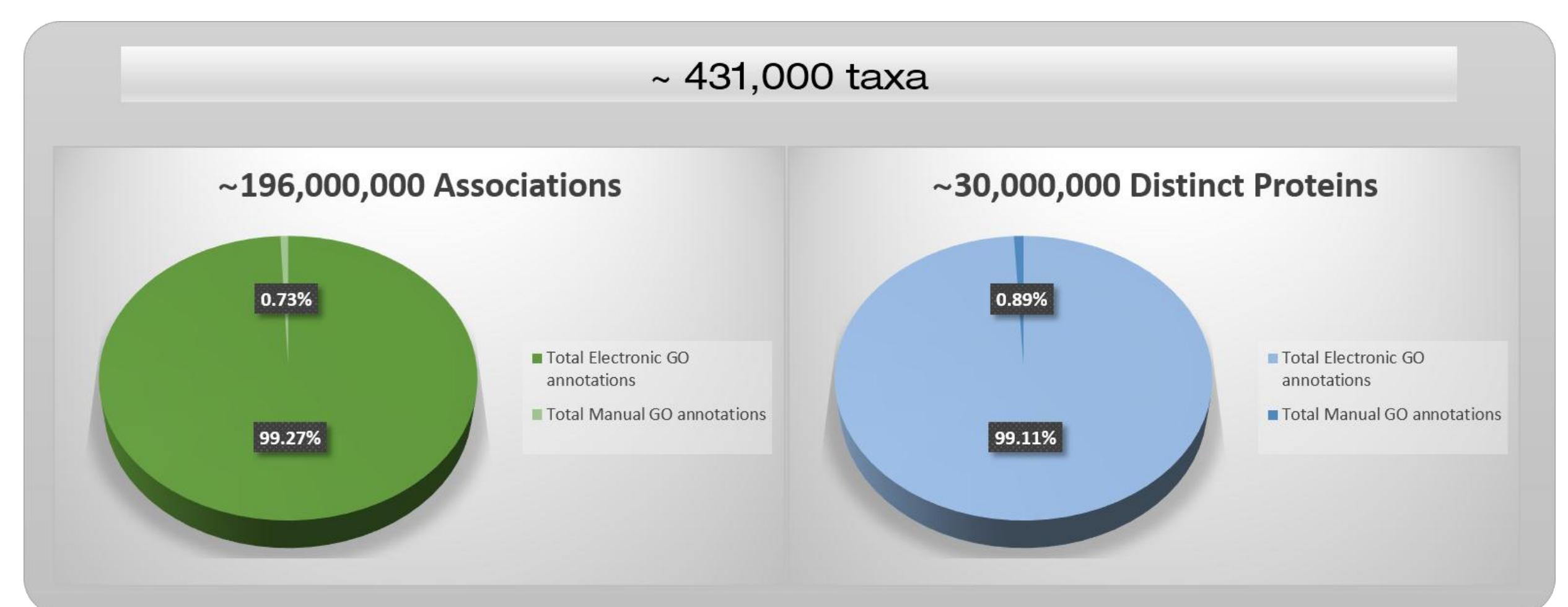


Fig. 2 Number of automatic and manual annotations

Automatic annotation benefits

One of the strengths of GO annotation is the ability to provide information for proteins which have not been experimentally characterized, when their orthologues are well studied. In light of quickly developing sequencing techniques that result in an exponential growth of available genomes, accurate automatic annotation becomes more and more important every day.

Over 400,000 species have only automatically predicted GO annotations. Several examples are shown in Figure 3.

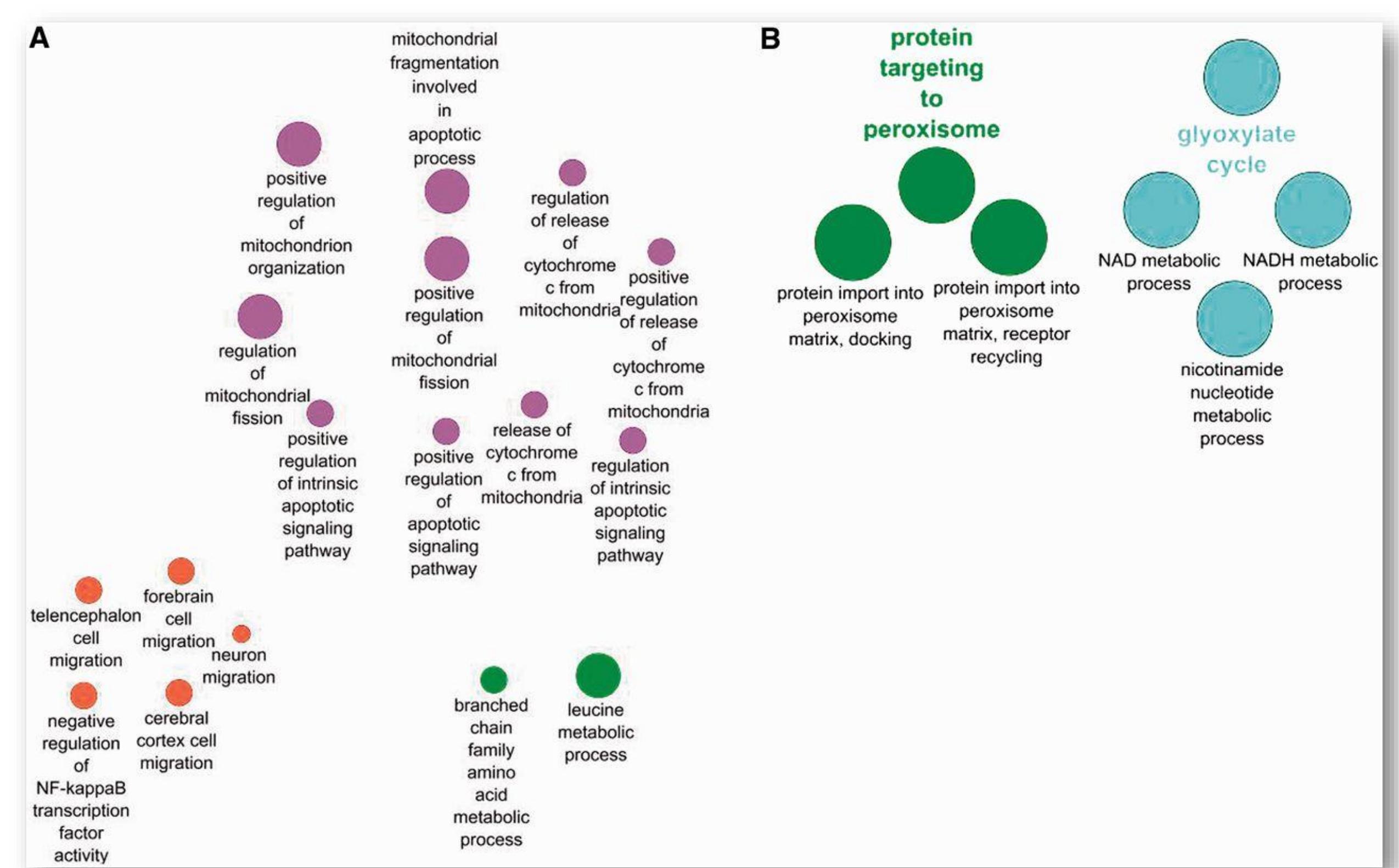


Fig. 4 Comparison between biological processes enriched only in human (A) and yeast (B) peroxisomal protein sets. Taken from Mutowo-Muellenet et al., 2013