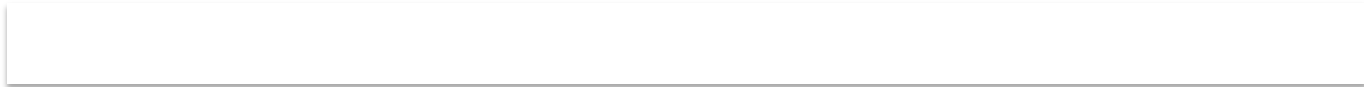


# Topic: COGs and Comparative Genomics

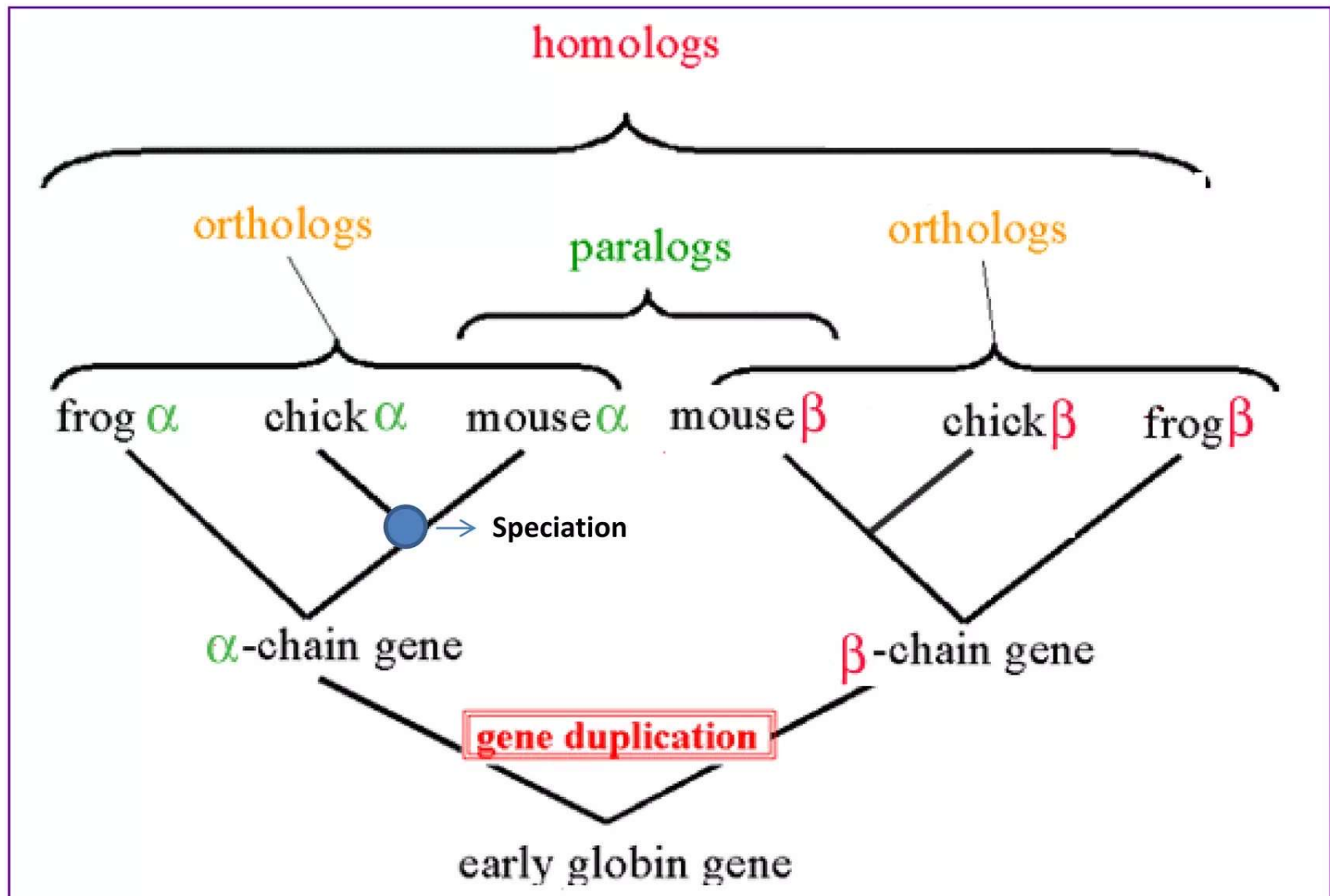


## Some important terminologies:

**Orthologs** are genes in different species that evolved from a common ancestral gene by speciation. Normally, **orthologs** retain the same function in the course of evolution. Identification of **orthologs** is critical for reliable prediction of gene function in newly sequenced genomes.

**Paralogs** are genes related by duplication within a genome. Orthologs retain the same function in the course of evolution, whereas **paralogs** evolve new functions, even if these are related to the original one.

**Speciation** is the origin of a new species capable of making a living in a new way from the species from which it arose. As part of this process it has also acquired some barrier to genetic exchange with the parent species.



# COGs

- Cluster of orthologous genes.
- Clusters of Orthologous Groups, are groups of three or more **ortholog** genes, meaning they are direct evolutionary counter parts and are considered to be part of an 'ancient conserved domain'. A COG is defined as three or more proteins from the genomes of distant species that are more similar to each other than to any other protein within the individual genome.
- COGs can be used to predict the function of homologous proteins in poorly studied species and can also be used to track the evolutionary divergence from a common ancestor, hence providing a powerful tool for functional annotation of uncharacterized proteins.
- Important in comparative genomics studies

# Application of COG

- The most straightforward application of the COGs is for the prediction of functions of individual proteins or protein sets, including those from newly completed genomes.

## The COG database

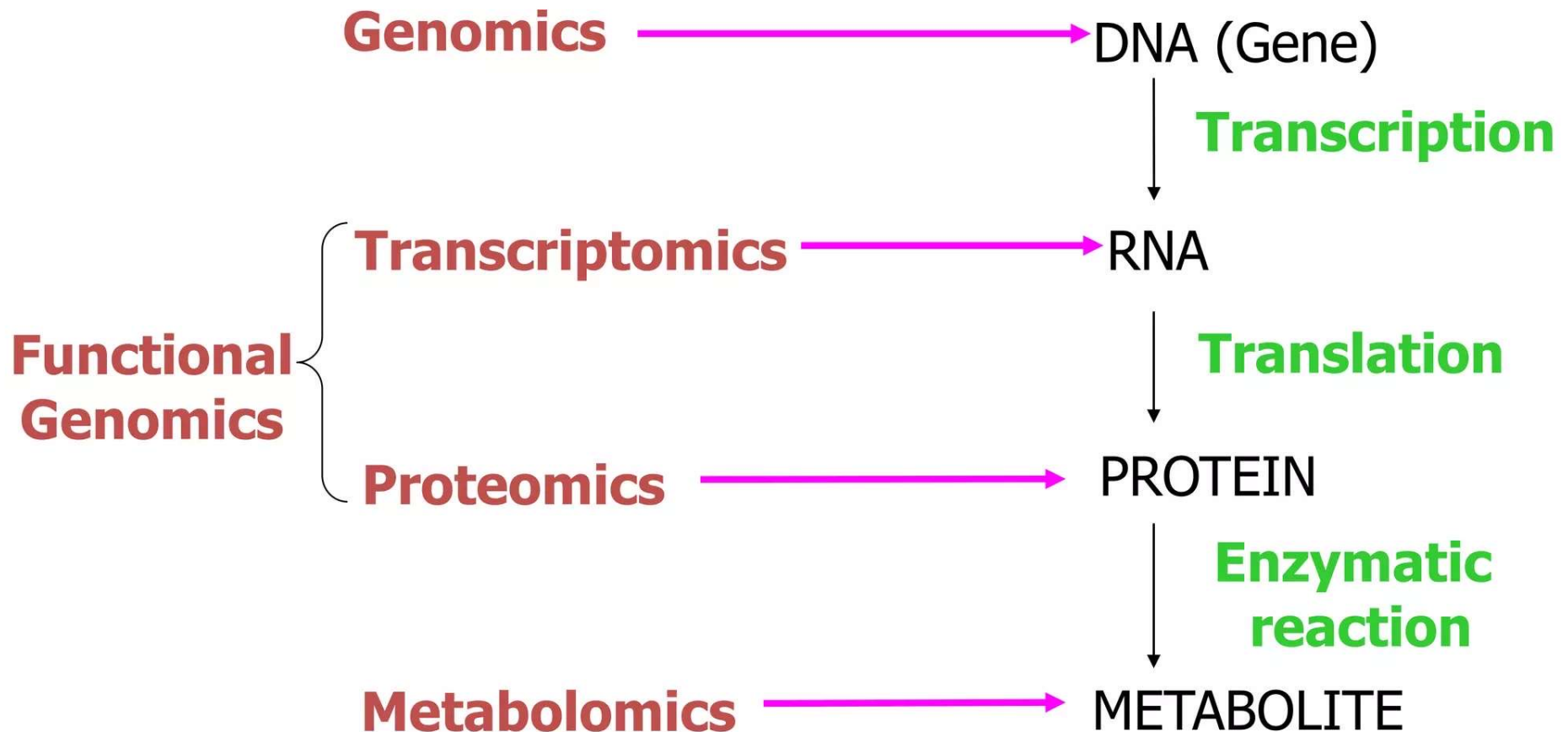
NCBI provides a [COG database](#) that consists of 4,873 COGs that code for over 136,000 proteins from the genomes of 50 bacteria, 13 archaea and 3 unicellular eukaryotes. This database uses completely sequenced genomes to classify proteins using the orthology concept.

# **What are some questions that comparative genomics can address?**

- How has the organism evolved?
- What differentiates species?
- Which non-coding regions are important?
- Which genes are required for organisms to survive in a certain environment?

# What is Comparative Genomics?

It is the comparison of one genome to another.



# Difference is in Scale and Direction

## Other “omics”

One or several genes compared against all other known genes.

Use genome to inform us about the entire organism.

## Comparative

Entire Genome compared to other entire genomes.

Use information from many genomes to learn more about the individual genes.



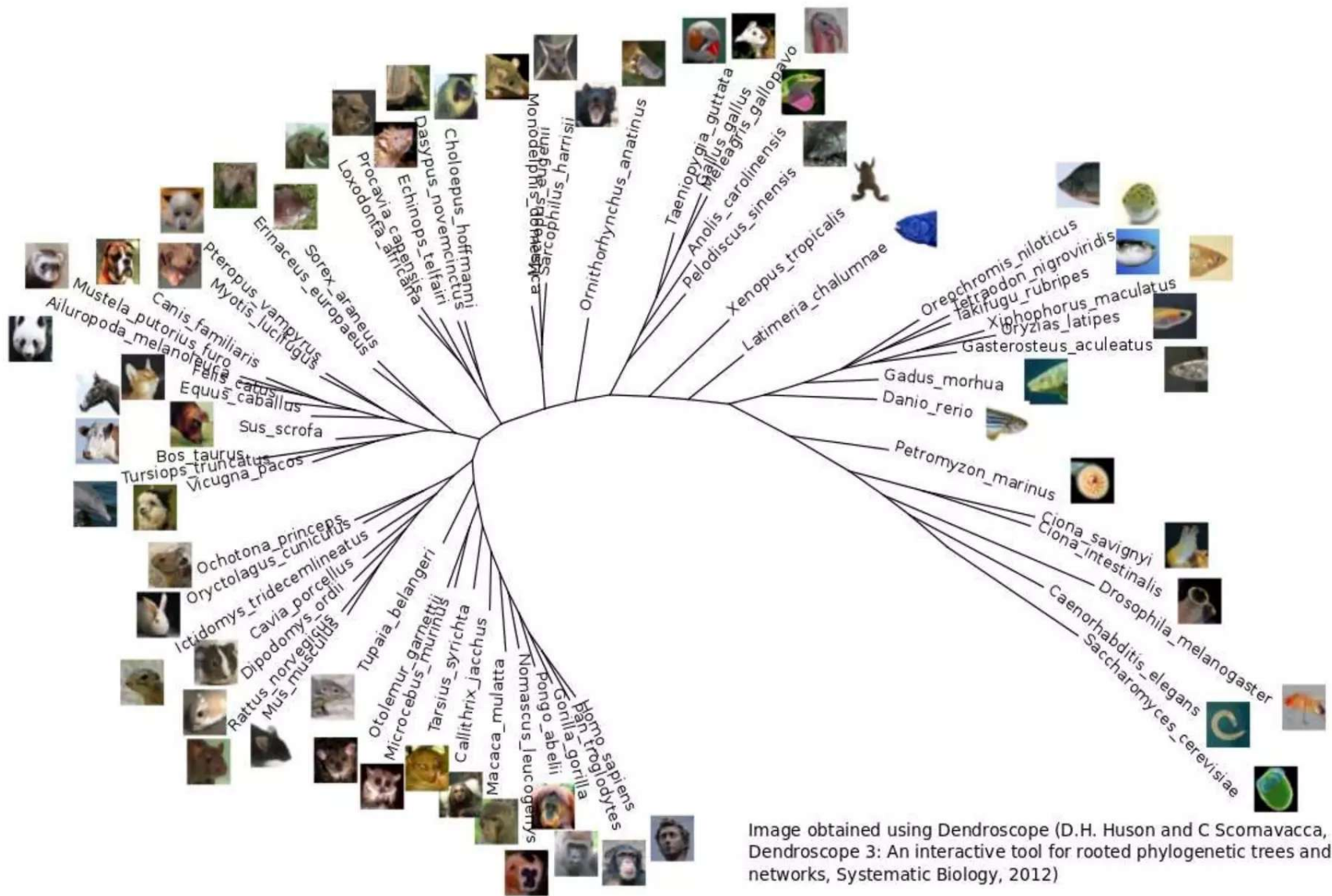
# Comparative genomics

- Discover what lies hidden in genomic sequence by comparing sequence information.
- Main areas
  - Whole genome alignment
  - Gene prediction
  - Regulatory element prediction
  - Phylogenomics
  - Pharmacogenetics

# Comparative Genomics

**Comparative genomics** is a field of biological research in which the **genomic** features of different organisms are compared. The **genomic** features may include the DNA sequence, genes, gene order, regulatory sequences, and other **genomic** structural landmarks

Comparative genomics is a powerful tool for identifying the features and dissecting the functions of genomes. The approach is based on selection for the gene or regulatory region constraining the evolution of the sequence. Comparison with other genomes has become an integral part of the analysis of the human genome sequence and is one of the most effective methods for identifying genes (*Batzoglou et al. , 2000; Roest Crollius et al. , 2000*)



**Figure: Species tree of different organisms**

THANK YOU