

# FIGfams

RossOverbeek

Fellowship for Interpretation of Genomes

## Introduction

FIGfams - yet another set of protein families

Definition of a FIGfam protein family:

a set of proteins that are “globally similar” and in which all members share a **common function**

## Isofunctional Homologs

Gerlt and Babbitt (2001):

- “Those of us who study the relationships between sequence, structure, and function should discontinue the use of ‘ortholog’ and ‘paralog’, unless we want to focus on the speciation and gene duplication events that produced functional diversity in homologs...”
- They proposed “Isofunctional homologs” for genes that are end-to-end similar and have the same function.

[www.nmpdr.org](http://www.nmpdr.org)

## Protein Similarity

Members of a family are “end-to-end” similar

- “End-to-end”: the sequences are easily recognizably similar over **at least 70%** of the length of the protein sequences
- Similarity is measured **using BLAST**

[www.nmpdr.org](http://www.nmpdr.org)

## Common function

- FIGfams are an attempt to form sets of proteins performing the same **cellular function**
- The most reliable FIGfams are based on our **manually curated Subsystems**
- Our two other ways used to place proteins into the same FIGfam:
  - Aligning two very **similar genomes** with confidence establish a **correspondence between genes** in a region
  - If **proximity** on the chromosome has been **preserved over many genomes**, we believe the proteins in that relationship play the same functional role

[www.nmpdr.org](http://www.nmpdr.org)

## FIGfams in the RAST Server

- FIGfams are used in the RAST (Rapid Annotation using Subsystem Technology) for functional annotation
- it is important that we **avoid false-positives**
- ☞ **correctly** identify the functions of the proteins encoded by as many genes as possible
- set the thresholds to **minimize the false-positives**
- ☞ some genes that could be annotated using FIGfams are not

[www.nmpdr.org](http://www.nmpdr.org)