

Combinatoric Approaches to Bio-Ontologies

**Cliff Joslyn
Susan Mniszewski
Karin Verspoor**

**Computer and
Computational Sciences
Los Alamos National
Laboratory**

**Andy Fulmer
Gary Heaton**

Procter & Gamble Corp.



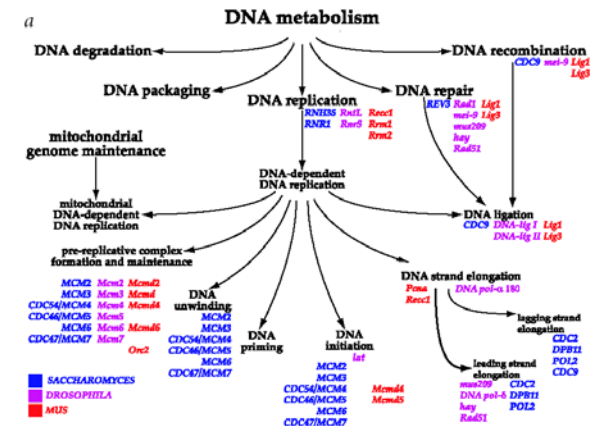
**Modeling, Algorithms,
and Informatics (CCS-3)**



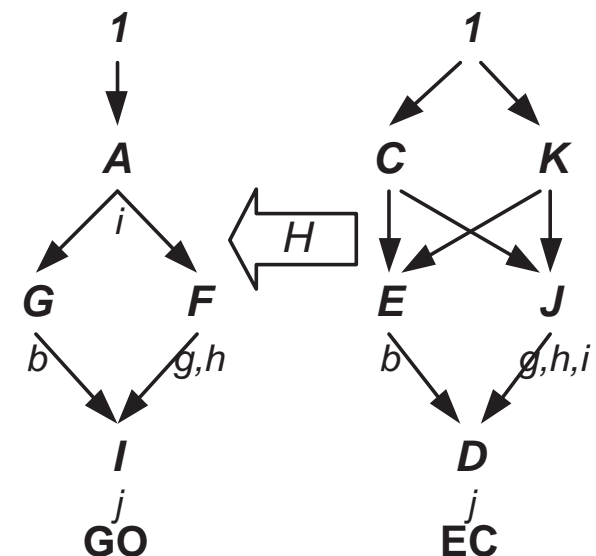
First ISCB Rocky Mountain Regional Bioinformatics Meeting
December, 2003

APPROACH TO BIO-ONTOLOGIES

- Bio-ontologies as taxonomically organized data objects: GO, EC, MeSH
- Combinatoric scientific computing:
 - Semantic hierarchies as large **partially ordered sets**
 - Math, algorithms, data structures, tools, visualization, statistics
- GO as source of semantic relations for **lexical semantic** approaches (Verspoor poster)
- **Categorization** of gene lists in GO
- **Hybrid ontology management and ontology matching**: order preserving mappings amongst ontologies
- Semi-automatic **ontology induction**



Ashburner Nat. Gen. 2000



CATEGORIZATION IN THE GO

- Represent GO as poset $\langle P, \leq \rangle$ with gene labels $X, F: X \mapsto 2^P$
- Given labels (genes) $c, e, i \in X$
- What node(s) in $P = \{A, B, C, \dots, K\}$ are best to “pay attention to”?
- Pseudo-distances of comparable nodes $p_i \leq p_j$
- Scores rank nodes wrt/gene locations, balancing:
 - **Coverage:** Covering as many genes as possible
 - **Specificity:** But at the “lowest level” possible
- “Cluster” based on non-comparable high score nodes

