# PAINT

- Curation tool to allow propagation of annotations
- PAINT products
  - GAF file with *inferred* annotations
    - Annotations to extant proteins (RefGenome & non-RG)
    - Annotations to ancestral nodes
  - Evidence file
    - Reasoning behind annotations
    - Questions
  - These and some other accessory files are deposited in CVS.
  - Results are also visible at PanTree.org (not .com!)

Let's take a look

# Two types of PAINT curation

- "Complete" curation
- "Fast pass"

# Curation Guidelines

- Start with MF, then CC, then BP
- Select the most specific GO term possible.
- Apply the term as widely as possible.
- When necessary, suggest corrections or changes to annotations or ontologies
  - Suggest more specific annotations only if it will change what you can propagate.
- Evidence files have a suggested format (next slide)

# Evidence file format

- Boilerplate: provided in PAINT

- Description of phylogeny

- Annotation notes and reasoning, by ontology

- Questions for MOD curators, by ontology (from wiki)

- Questions for ontology curators, by ontology (from wiki)

# Curation Guidelines, continued: Don't get bogged down

- Don't need to address *every* EXP annotation.
  - "protein binding"
  - Biochemical fractions (membrane fraction, synaptosome, etc.)
  - "response to" IEP's
  - HTP papers
  - Downstream effects (IMP's)
- Prioritize the most important EXP annotations (next slide)
- If an annotation doesn't make any sense, add a question to the wiki page
- (For synchronized curation projects) Impose deadlines: Keep pace with the literature curation. Don't PAINT for longer than the curation period.

# Establish priorities: General approach

- Look for tautologies ("duh" annotations)
  - Wnt → Wnt signaling pathway
  - Topoisomerase → topoisomerase complex
- Most of the problems will be in BP
  - Work outwards, bottom-to-top
    - Molecular processes
    - Cellular processes
    - Organismal processes
      - Organ development
      - Embryonic development
      - Pattern formation

# Establish Priorities
# Example: the Wnt family

1) Signaling pathways (especially those mentioning "Wnt" by name, such as GO:0060070 "canonical Wnt receptor signaling pathway")
2) Signaling pathway/developmental pathway cross-products
3) Regulation of transcription as a result of signaling
4) Developmental terms
   a) organs and tissues
   b) high-level pattern formation

Let's try a few

# Suggestions

- Read the existing Evidence files for examples.

- AmiGO Annotation Tracker can be used to get all EXP annotations for a family.

- EBI QuickGO is a convenient way to visualize relationships between terms.

  - I'm taking suggestions for easier/more powerful interfaces.

- Ask Mike or Rama for written instructions.

# Fast Pass curation

- Try to capture only the most important aspects of the biology.

- Don't worry about the more distant relationships.

- Do not read past the abstract of papers.

- Don't worry about suggesting more specific annotations (unless they pop out at you).

- DO retain the structure of the Evidence file, including the description of the phylogeny. (It'll pay off.)

- If you still have questions, put them in the wiki and move on.

# Philosophizing

- These are inferences.  Hence, the standards are different from literature curation.  Do not hesitate to use your background knowledge.

- When you propagate an annotation, you are saying that that function existed in that ancestral protein.

- All else being equal, the simplest explanation is probably best.

# Future notes

- Taxon constraints are coming, but try not to say that bacteria have nuclei.

- Has_part will eventually address many Function-Process relationships automatically, but PAINT does not (yet).