Cross-species Mapping between Anatomical Ontologies: Terminological and Structural Support

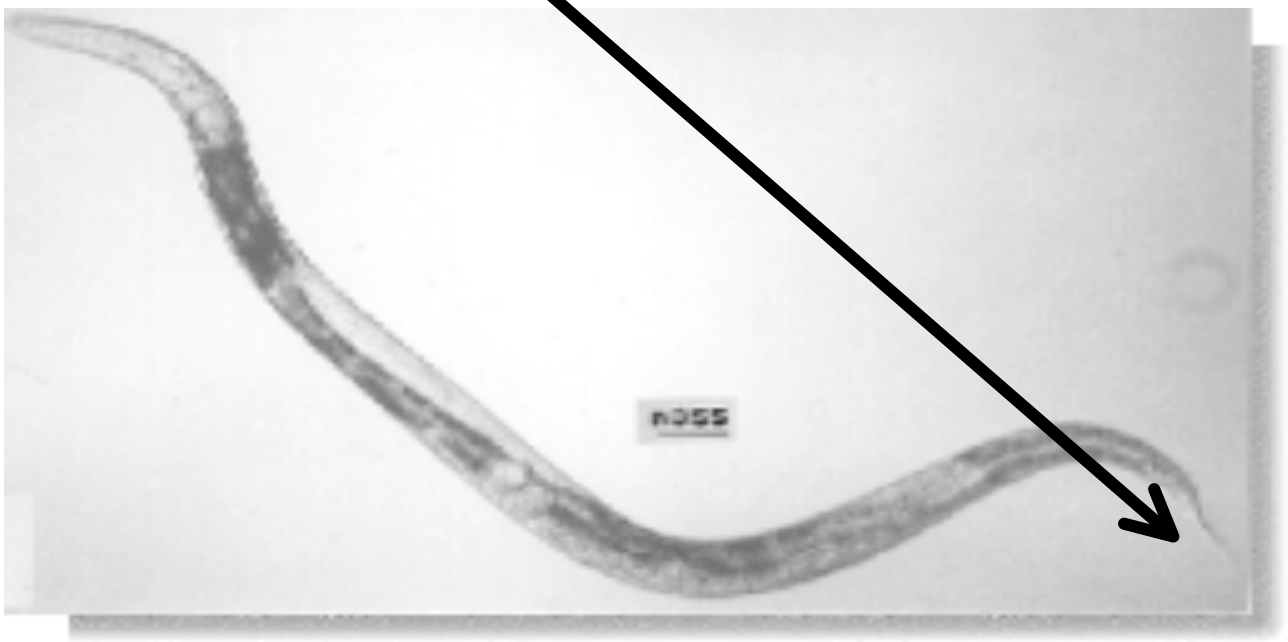
by
Sarah Luger, Stuart Aitken & Bonnie Webber
sluger@inf.ed.ac.uk, stuart@aiai.ed.ac.uk,
bonnie@inf.ed.ac.uk
XSPAN/University of Edinburgh

A Mapping Problem

- Mouse Tail

A Mapping Problem

- C. Elegans Tail



A Mapping Problem

- In mouse:

embryo **.** organ system **.** sensory organ **.** eye **.** optic stalk **.** *optic nerve*

- In drosophila:

larva **.** larval organ system **.** larval nervous system **.** larval central nervous system **.** larval brain **.** medulla anlage **.** *optic nerve*

A Mapping Problem

- Given
  - Mouse: 3559 anatomical parts
  - Drosophila: 506 anatomical parts
  - C. Elegans: 242 anatomical parts

- Can their terminologies and anatomical ontogolies suggest what parts may be similar (homologous)?

A Mapping Problem

- Mouse **tail** to C. elegans **tail**
  - Same name, different function
- Mouse **optic nerve** to drosophila **optic nerve**
  - Same name, same function
  - The ontologies show different paths.
- The goal is to suggest the anatomical parts that maybe similar. Does language suggest similarity? What clues can we use?

A Related Problem

- In two different models of human anatomy, do parts with similar names always denote similar tissues?
- In GALEN:

  Lobe of left lung

  Maps in FMA to:

  Upper Lobe of left lung
  Lower lobe of left lung
- Extrapolate from intra-human to inter-species comparisons.

XSPAN Project

Heriot-Watt University
School of Mathematical and Computer Sciences

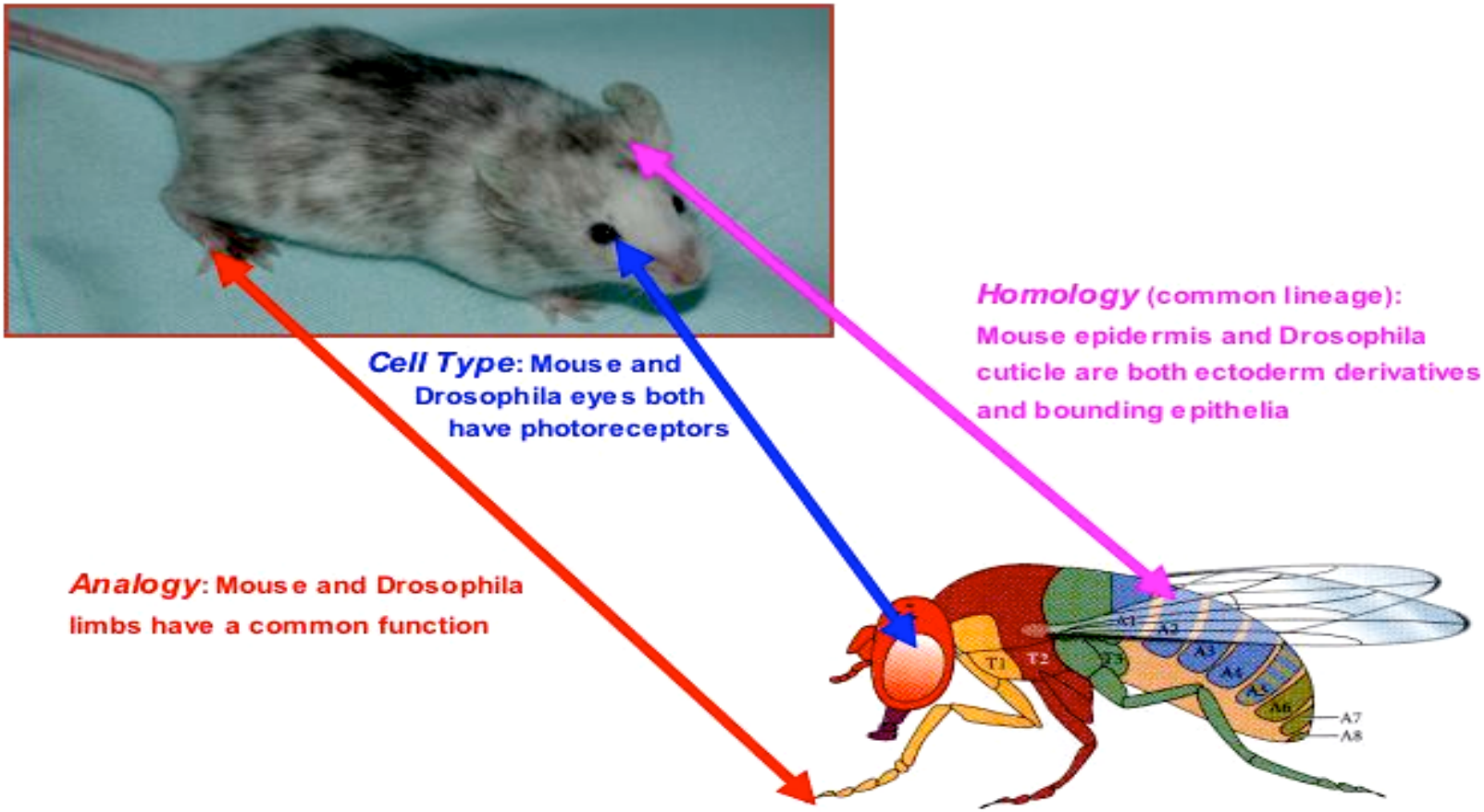
University of Edinburgh
School of Biomedical Sciences
School of Informatics

XSPAN

- A framework for recording expert knowledge about anatomy.

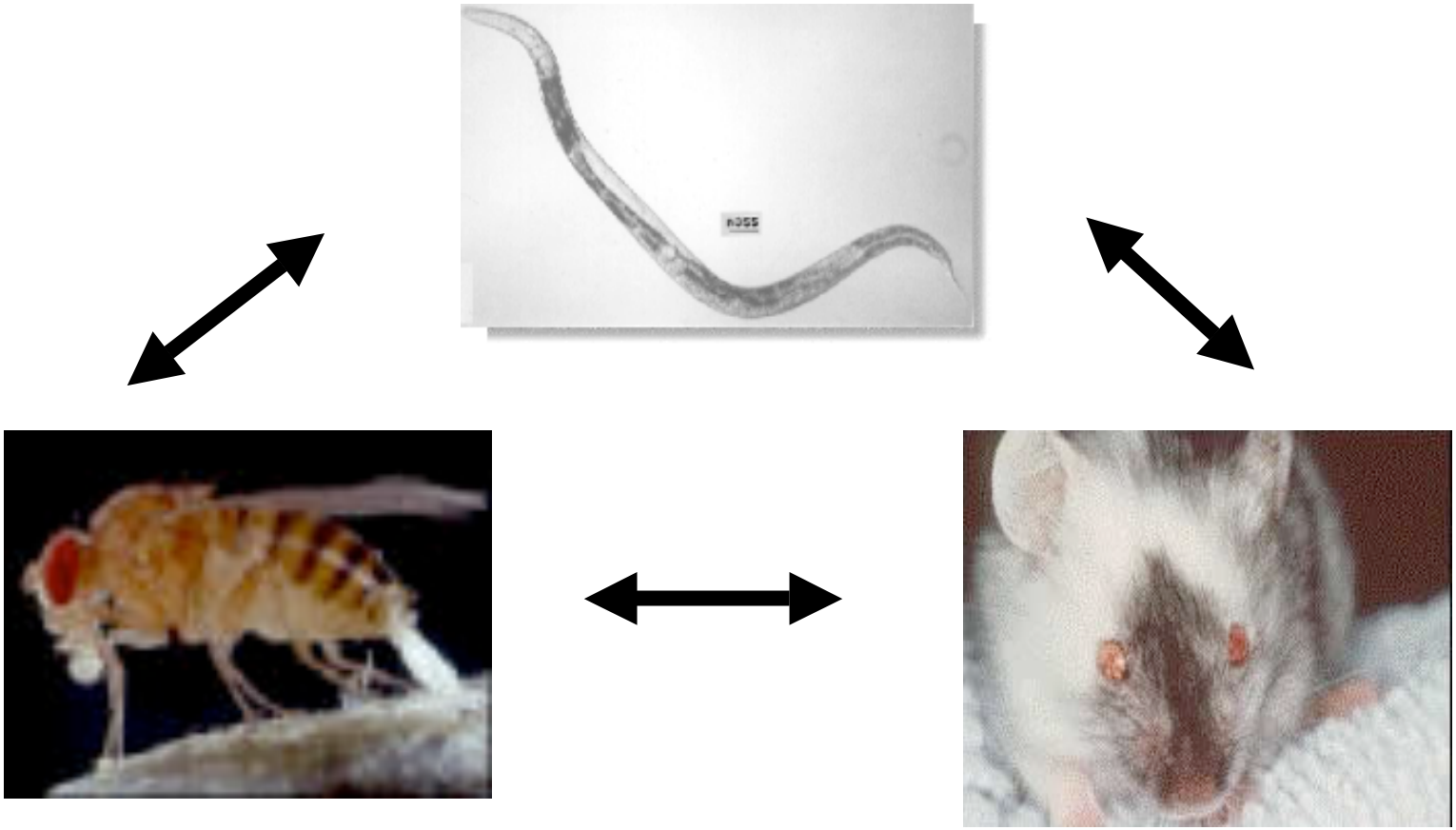
- A Web server with information about evolutionary, functional, developmental and cellular anatomy:
  - Homology relationships
  - Functional similarities
  - Lineage relationships
  - Cell types

XSPAN: Background

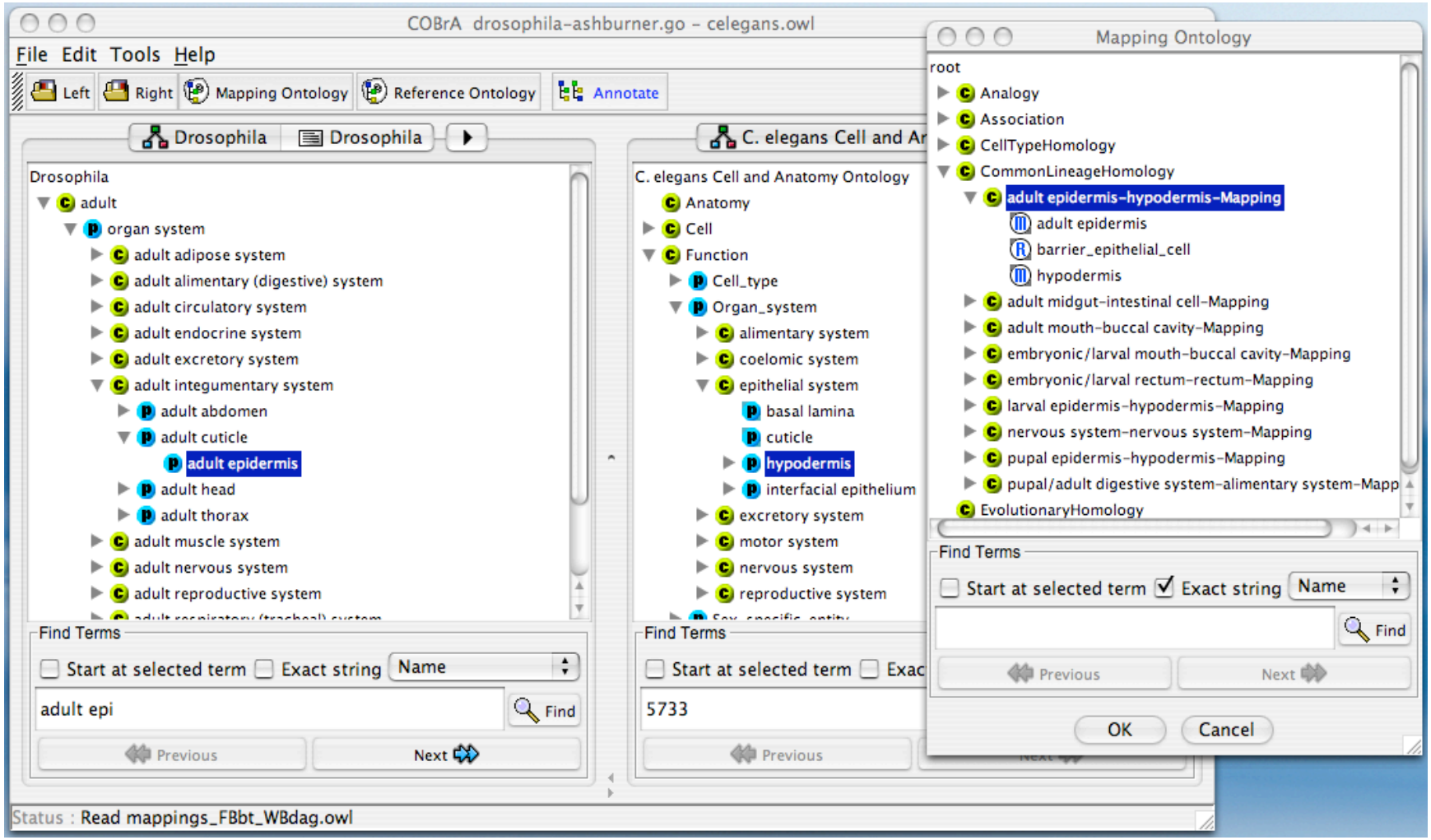


Examples of anatomical relationships between mouse and Drosophila

Current Species Comparisons



Example Ontologies: COBrA



Comparison Examples

Earlier I gave the impression that comparison between terms was based on a “short form.” That is not exactly true. The short form needs to be understood in context.

mouse **.** embryo **.** organ system **.** sensory organ **.** ear **.** external ear **.** pinna **. mesenchyme**

mouse **.** embryo **.** organ system **.** visceral organs **.** alimentary system **.** gut **.** foregut **.** pharynx **. associated mesenchyme**

mouse **.** embryo **.** organ system **.** nervous system **.** central nervous system **.** brain **.** forebrain **.** telencephalon **.** corpus striatum **.** caudate nucleus **. head**

Comparison Motivation

- Two motivations for using more than the leaf label:
  - Context is important as terms are not uniquely denotated across the ontologies.
  - The important terms are spread across the labels of the path, they are not restricted to the leaf terms.
    - This reflects the choices biologists made in grouping and structure.

Lexical Analysis

- Normalize terms to limit the effect of different descriptive styles including dealing with American and English variants.
- Compare content words by removing stop words.
- Ensure comparable forms of words by stemming and lemmatizing.
- Results are then treated as an unordered set.

Lexical Analysis Examples

- Use example pairing for comparison:

1) arch of aorta
2) aortic sinus
3) visceral muscle of larval heart

1') arch aort
2') aort sinu
3') viscer muscl larval heart

Lexical Analysis Examples

mouse **.** embryo **.** organ system **.** cardiovascular system **.** heart **. aortic sinus**

drosophila **.** embryo **.** embryonic organ system **.** embryonic circulatory system **.** embryonic **.** larval dorsal vessel **.** embryonic **.** larval heart **. visceral muscle of larval heart**

1. *Node* comparison or leaf node in a tree.

   aortic sinus to **visceral muscle of larval heart**

2. *Path-based* comparison or sequence of node labels from root to leaf.

Methodology

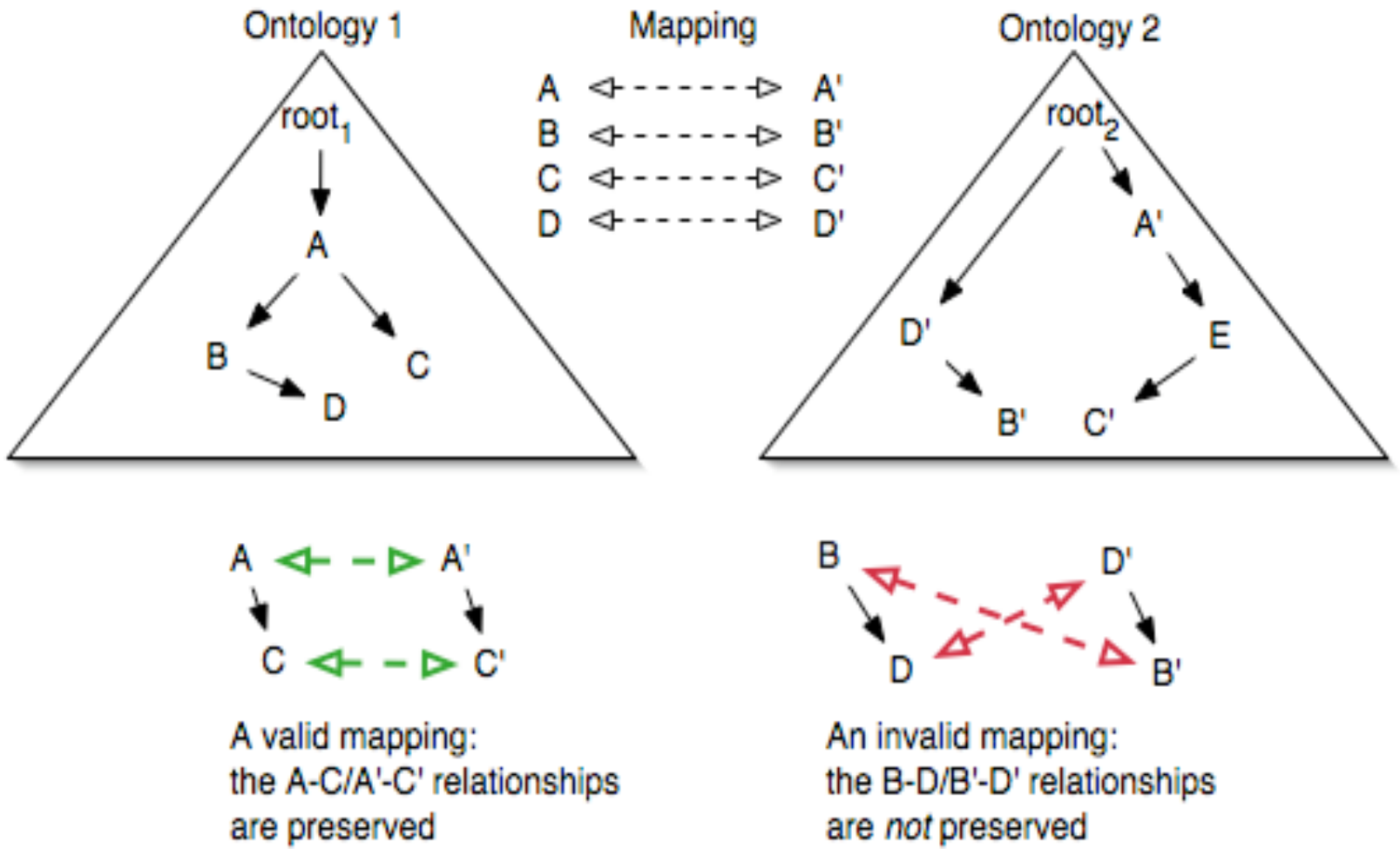
- Tissue pairs assessed structurally.
- Use a similarity threshold to limit the number of results.
- Resultant pairs have one to many mappings:

EMAPA: 16039	FBbt: 00000052
EMAPA: 16039	FBbt: 0000111
EMAPA: 16039	FBbt: 00006005
EMAPA: 16069	FBbt: 00001056
EMAPA: 16103	FBbt: 0000125

Structural Analysis

- Evaluate structural similarity by taking the ontologies as graphs with directed but unlabeled edges.
  - First examine the intra-species relationships
  - Check to see if the relative positions are consistent between species.
  - There may not be evidence.

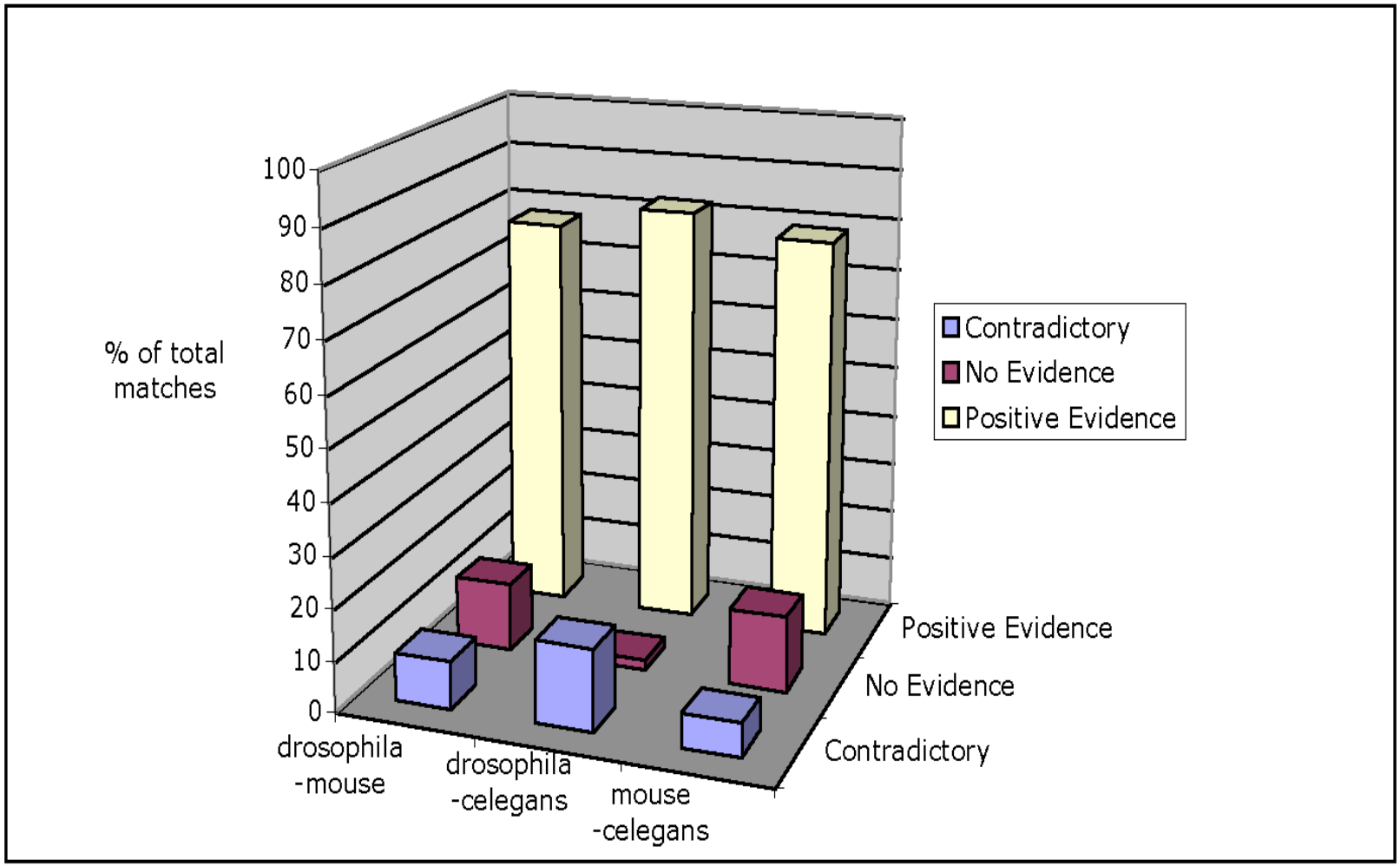
Structural Analysis



Results

- Node-based comparisons
  - Approximately 80% of lexical mappings have support from the ontology.
  - Less than 16% of proposed mappings have either no evidence for or against, or are contradictory across the three comparisons.
- Path-based comparisons
  - With lexical mappings at 75% similarity, the number of contradictory matches was reduced to zero.

Results

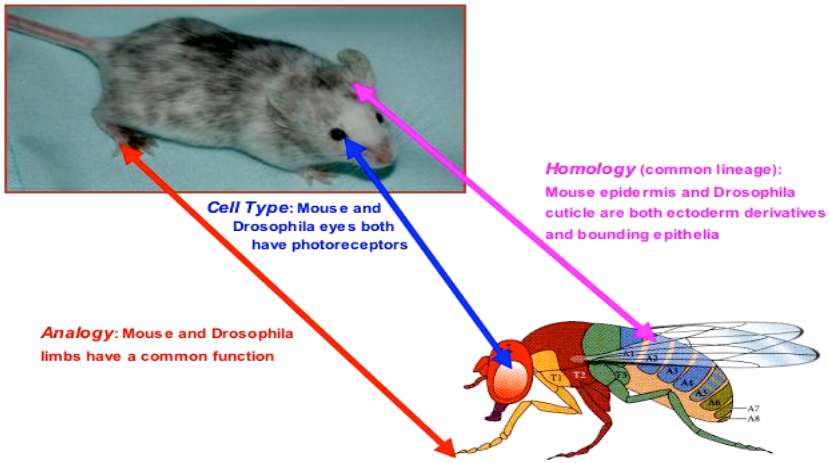

% of total matches
100
90
80
70
60
50
40
30
20
10
0
drosophila -mouse
drosophila -celegans
mouse -celegans
Contradictory
No Evidence
Positive Evidence
Positive Evidence
No Evidence
Contradictory

Pairwise Results

	C. elegans	**2732**
Mouse	79% positive	2121
	15% no evidence	358
	6% contradictory	254
	C. elegans	**1625**
Drosophia	82% positive	1337
	2% no evidence	32
	16% contradictory	256

Pairwise Results

	Drosophila	**2732**
Mouse	78% positive	2121
	13% no evidence	358
	9% contradictory	254



Examples of anatomical relationships between mouse and Drosophila

Future Work

	Average path length in nodes
Mouse	7.9
Drosophila	6.4
C. elegans	6.0

Weighting paths helps normalize specificity.

Future Work

mouse **.** organ systems **.** circulatory system **.** heart **.** valve muscle
1 2 4 8 16

The last term is weighted more than all prior terms combined. This filters out the “garbage” when comparing similar root-to-leaf paths with vastly different levels of specificity.

drosophila **.** organ system **.** upper torso **.** circulatory system **.**
1 2 4 8
valve network **.** heart **.** valves **.** valve muscles
16 32 64 128

Future Work

- Augment the three m.o. datasets with synonyms and abbreviations.
  - Some are provided in the anatomies, but not systematically or consistently
  - Introduce synonyms from an anatomical reference.
- Establish the effect of additional information on previous results.

References

- Aitken, J.S., B.L. Webber and J.B.L. Bard. Part-of Relations in Anatomical Ontologies: A Proposal for RDFS and OWL Formalisations. *Proc PSB 04*, 9:166-177(2004)
- Zhang, S., and O. Bodenreider. Investigating Implicit Knowledge in Ontologies with Application to the Anatomical Domain. *Proc PSB 04*, 9:250-261(2004)
- Zhang, S., P. Mork and O. Bodenreider. Lessons Learned from Aligning Two Representations of Anatomy. To be published in KR-MED 2004.