

## Databases and ontologies

## COBrA: a bio-ontology editor

Stuart Aitken<sup>1,\*</sup>, Roman Korf<sup>1</sup>, Bonnie Webber<sup>1</sup> and Jonathan Bard<sup>2</sup><sup>1</sup>School of Informatics and <sup>2</sup>Biomedical Science, The University of Edinburgh, Edinburgh EH8 9LE, UKReceived on June 1, 2004; revised on September 23, 2004; accepted on October 11, 2004  
Advance Access publication October 28, 2004

## ABSTRACT

COBrA is a Java-based ontology editor for bio-ontologies that distinguishes itself from other editors by supporting the linking of concepts between two ontologies, and providing sophisticated analysis and verification functions. In addition to the Gene Ontology and Open Biology Ontologies formats, COBrA can import and export ontologies in the Semantic Web formats RDF, RDFS and OWL.

**Availability:** <http://www.xspan.org>

**Contact:** [stuart@aiai.ed.ac.uk](mailto:stuart@aiai.ed.ac.uk)

Bio-ontologies play a crucial role in the indexing of experimental data—providing both unique IDs for aspects of anatomy, phenotype, process, cellular structure and molecular function (GO Consortium, 2000), and conceptual abstractions for aggregating results (Bard and Rhee, 2004). Biological ontologies have grown in size, and their encoding languages have become more sophisticated, with the result that tools for creating, editing, verifying and maintaining them (e.g. version control, meta-data attribution, provenance, etc.) have become essential. This note presents COBrA, an ontology editor and mapping tool.

In parallel with expanding the range of domains being captured and the number of terms in key resources such as the Gene Ontology (GO), researchers have been examining the formal and conceptual bases underlying ontology languages and modelling principles (Smith *et al.*, 2003). Initially constructed on an intuitive basis, many bio-ontologies are being scrutinized with regard to their underlying principles, and their support of inference—this being critical for automated verification. Ontologies of the same or similar conceptual domains are also being examined with respect to how they map to one another. The languages of the Semantic Web have a role to play as they provide standards, tools and techniques. For example, the Web Ontology Language (OWL, <http://www.w3.org/TR/owl-ref>) has a specified syntax and semantics designed for the sharing and reuse of ontologies over the Web. Utilising reasoners for OWL-Lite, OWL-DL and fragments of OWL-Full, OWL provides the mechanisms to address outstanding issues in bio-ontologies. For the ontology editor described here, OWL provides solutions to the problems of concept mapping and ontology verification.

COBrA is a product of the XSPAN project (<http://www.xspan.org>) which uses concept mapping to express judgements of homologies and analogies between tissues across different anatomy ontologies. The resulting knowledge base will contribute to a community

resource for exploring gene expression data. In XSPAN, mappings can be used to express correspondences between tissues in terms of their evolution (*Evolutionary Homology*), development (*Common Lineage Homology*) or function (*Analogy*). Creating a mapping is necessarily a human decision, made complex by the nature of the task and the size of the anatomies. Within XSPAN, COBrA supports acquisition and exploration of these human-specified mappings.

## COBrA

COBrA is an editor and mapping tool for GO and Open Biological Ontologies (OBO) (<http://obo.sourceforge.net>) that allows users to explore two ontologies simultaneously and to make links between them. COBrA provides both a tree-based view and a node-based view of an ontology, where the latter displays the selected term's parents, children and definitional information. The ontology can be edited by direct manipulation of the tree or by calling a term editor. Initial evaluation of the tool over a range of tasks, and user-types, confirms the design choices (Korf, 2003). Figure 1 shows a mapping between *adult epidermis* (Drosophila) and *hypodermis* (C.elegans), with the justification *barrier epithelial cell* denoting the cell type shared by the tissues.

Concepts and relations in Semantic Web languages such as OWL require both a name and a namespace (combined into a URIRef), and COBrA provides visualizations and interfaces to these new (and potentially unfamiliar) elements of the ontology. COBrA also provides a graphical interface to the analysis functions we now describe.

Concept mapping, ontology merging, and verification are problems that COBrA solves through the use of OWL. A mapping is a pointer created to link a concept in one ontology to a concept in another: A mapping is a new term that relates two existing URIRefs. It can be created and saved without modifying the original ontologies. Meta-data such as authorship is associated with the mapping term, and mapping terms can be organized hierarchically, as illustrated on the right-hand side of Figure 1.

Turning to ontology comparison and merger, these can be computed by finding the intersection and union, respectively, of the RDF graphs derived from the OWL representations of two ontologies. These graph-based operations improve on the equivalent operations that might be performed on textual representations of the ontologies (e.g. in CVS), but do not involve verification of the results.

For ontology verification, the semantics of the GO *is\_a* and *part\_of* relations must be defined, hence we use OWL *subClassOf* and *define*

\*To whom correspondence should be addressed.

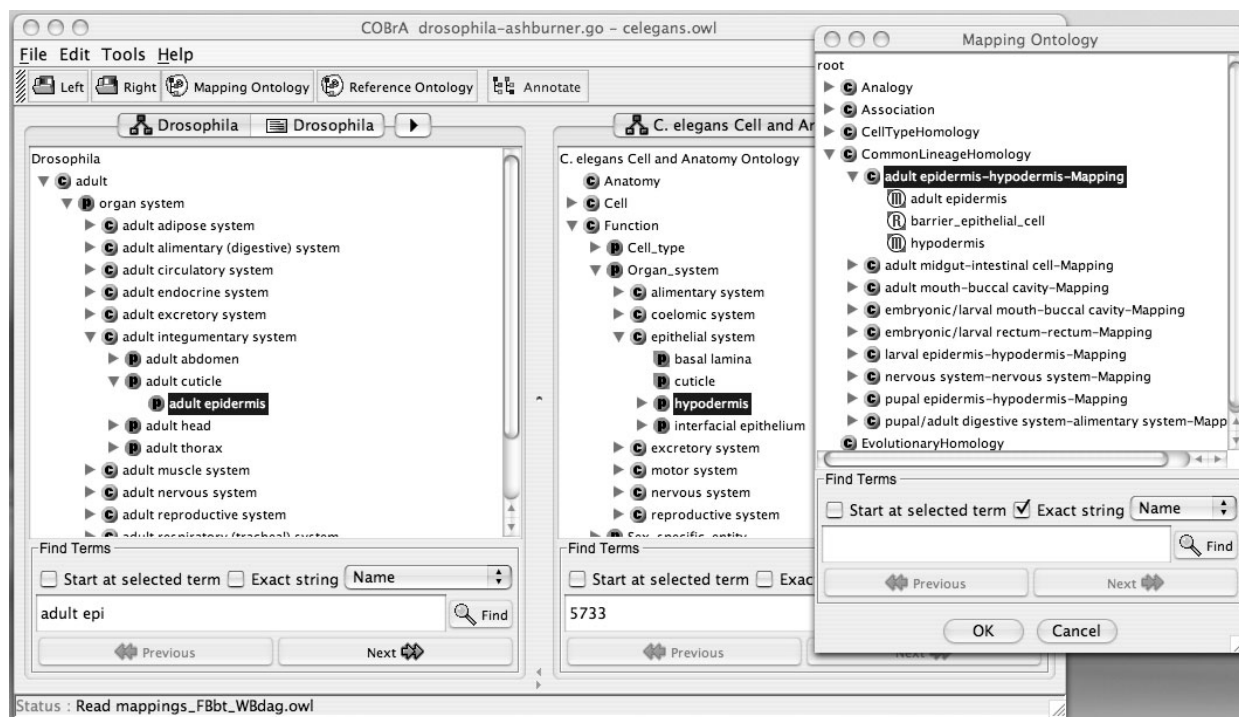


Fig. 1. Two anatomies displayed in COBrA, with a Mapping Ontology dialog inset right.

the interpretation of *partOf* (Aitken *et al.*, 2004). These steps allow verification. An inference mechanism implements rule-based reasoning over the RDF graph, for example, to propagate properties across *partOf* links. COBrA can also perform a more complex ontology analysis that checks for cycles in the graph and in the ontology. Both graph manipulation and inference methods are provided by the Jena Semantic Web toolkit which provides Java methods to read, write and create RDF graphs (<http://www.hpl.hp.com/semweb>).

In addition, COBrA supports the import and export of bio-ontologies in RDF, RDFS and OWL. However, COBrA is not a generic OWL editor. The GO RDF format is that specified by the GO Consortium; the RDFS format is a modification of that where *is\_a* is replaced by *rdfs:subClassOf*. The OWL format is defined by a top-level ontology (Aitken *et al.*, 2004) which specifies a number of classes and relations that are required to state GO-style ontologies in OWL.

Protégé (<http://protege.stanford.edu>), a generic ontology editor, and DAGEdit (<http://www.geneontology.org>) provide comparable editing functions to COBrA. However, neither addresses mapping between ontologies. Protégé would require adaptation to read GO and OBO formats, but is more fully compatible with OWL.

COBrA demonstrates the practical application of Semantic Web techniques in the bioinformatics context by combining familiar ontology-editing functions, and compatibility with existing file formats, with additional features such as mapping, merging and verification that make use of RDF and OWL.

## ACKNOWLEDGEMENTS

This work is supported by BBSRC grant BBSRC 15/BEP 17046.

## REFERENCES

- Aitken, J.S., Webber, B.L. and Bard, J.B.L. (2004) Part-of relations in anatomy ontologies: a proposal for RDFS and OWL formalisations. *Proc. PSB*, **9**, 166–177.
- Bard, J.B.L. and Rhee, S.Y. (2004) Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.*, **5**, 213–222.
- GO Consortium (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
- Korf, R. (2003) COBrA—A concept ontology browser for anatomy. *Informatics Report EDI-INF-IM030022*.
- Smith, B., Williams, J. and Schulze-Kremer, S. (2003) The ontology of the Gene Ontology. *Proc. AMIA Annual Symposium 2003*, 609–613.