Reasoning over anatomical homology in the Phenoscape KB

Paula M. Mabee
Department of Biology
University of South Dakota
Vermillion, SD, USA

James P. Balhoff
Renaissance Computing Institute
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Wasila M. Dahdul
Department of Biology
University of South Dakota
Vermillion, SD, USA

Hilmar Lapp
Center for Genomic and Computational Biology
Duke University
Durham, NC, USA

Christopher J. Mungall
Environmental Genomics and Systems Biology
Lawrence Berkeley National Laboratory
Berkeley, CA, USA

Todd J. Vision
Department of Biology
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

***Abstract*— The Phenoscape project (www.phenoscape.org) has semantically annotated the features of species from the comparative literature, enabling links between candidate genes and novel species phenotypes for which they might be responsible. To enable discovery of homologous phenotypes and associated genes, we incorporated machine-reasoning with knowledge about homology into the Phenoscape Knowledgebase (KB). We show that with homology reasoning enabled, the results of database queries can be expanded to incorporate shared evolutionary history. We discuss the challenges in developing a logical model of homology assertions and implications for database queries, as well as theoretical entailment and practical performance tradeoffs between alternative models.**

Keywords—homology; anatomy ontology; phenotypes; reasoning; evolution

I. INTRODUCTION

The enormous volume of biological data that has become available to researchers has brought with it a rapidly expanding taxonomic range represented by the data. Because different taxa can possess similar features due to shared ancestry, the incorporation of homology in connecting, aggregating, and analyzing data has become increasingly important. For example, without explicit incorporation of homology, the results of user queries for species phenotypes or candidate genes may be incomplete. The Phenoscape Project [1, 2] has semantically annotated the features of species from the comparative literature, enabling links between novel species phenotypes and candidate genes that may underlie them. To enable discovery of homologous phenotypes and associated genes, we incorporated homology reasoning in the Phenoscape Knowledgebase (KB) (kb.phenoscape.org). One of the difficulties in rendering homology knowledge amenable to reasoning is that statements of homology are hypotheses, and in some cases homology assertions regarding the same anatomical structures can be in conflict. Thus, we represent homology assertions separate from a core anatomy ontology as annotations in spreadsheet form with evidence and attribution. The annotations are transformed into OWL axioms according to a model with the desired entailments, and a user can choose whether or not to include hypotheses of homology in reasoning. We explore the ramifications of different logical models of homology and use a series of competency questions to evaluate the performance of each model.

II. ANNOTATION OF HOMOLOGY ASSERTIONS

Homology assertions for both historical and serial homology of vertebrate skeletal elements were extracted from the comparative literature for teleost fishes and early sarcopterygians [3], and from the developmental genetic literature. We constructed these assertions using anatomy terms from the Uberon anatomy ontology [4] and taxon terms from the Vertebrate Taxonomy Ontology [5], resulting in a total of 98 homology assertions pertaining to skeletal anatomy. Attribution for each homology statement was recorded, and the type of evidence (e.g., positional, developmental) provided by the author supporting or contradicting the homology assertion was annotated with terms from the Evidence and Conclusion Ontology [6]. The most common type of evidence for or against homology cited by authors in the collection of

---

Funding provided by National Science Foundation ABI Innovation collaborative grants (1661529, 1661356, 1661456, 1661516) and an ABI Development grant (1062542).

homology assertions were based on development (27 statements), followed by morphological similarity (26 statements), position (20 statements), and gene expression (14 homology statements). Some author statements (5) cited evidence traceable to a different publication, whereas some (6 statements) did not cite traceable evidence. The collection of homology assertions was incorporated in the Phenoscape KB, which currently contains over 600,000 annotated phenotypes for vertebrate taxa from 185 comparative morphological studies.

III. HOMOLOGY REASONING MODELS

We have taken an exploratory approach toward resolving the most effective way to enable machine reasoning on historical and serial homology across anatomical structures. Specifically, we explored the reasoning ramifications of two OWL models of homology that we have developed. In the first model, classes of homologous entities are represented using reciprocal existential property restrictions. In the second model, an OWL individual is introduced that represents the ancestral structure from which all instances of two classes of homologous structures are descended. Using the collection of homology assertions and a sample of fin/limb phenotypes from the KB, we evaluated each model against the expected outcomes for a set of queries formulated as competency questions relevant to research in developmental biology, comparative anatomy, and evolution. We discuss these results and the implementation of homology reasoning in the KB.

IV. CONCLUSIONS

We have enabled homology reasoning in the Phenoscape KB, where it allows discovery of homologous structures. With homology reasoning incorporated, computational tools can now access the results of reasoning across evolutionary history. Although the model we select and implement in the KB satisfies basic reasoning, we expect that it can and will be optimized for different purposes, and as computational methods to represent uncertainty evolve.

ACKNOWLEDGMENT

We are grateful to the many collaborators who have contributed data and expertise to the Phenoscape Project.

REFERENCES

[1] R. C. Edmunds, B. Su, J. P. Balhoff, B. F. Eames, W M. Dahdul, H. Lapp, et al. (2016) Phenoscape: Identifying Candidate Genes for Evolutionary Phenotypes. Mol Biol Evol. 33: 13–24.

[2] P. M. Mabee, J. P. Balhoff, W. M. Dahdul, H. Lapp, P. E. Midford, T. J. Vision, et al. (2012). 500,000 fish phenotypes: The new informatics landscape of evolutionary and developmental skeletal biology. J Applied Ichthy 28(3):300-305.

[3] T. A. Dececchi, J. P. Balhoff, H. Lapp, P. M. Mabee. (2015). Toward synthesizing our knowledge of morphology: using ontologies and machine reasoning to extract presence/absence evolutionary phenotypes across studies. Syst Biol 64(6): 936-952.

[4] M. A. Haendel, J. P. Balhoff, F. B. Bastian, D. C. Blackburn, J. A. Blake, A. Comte, et al. (2014). Uberon: Unification of multi-species vertebrate anatomy ontologies for comparative biology. J Biomed Semantics. 5:21

[5] P. E. Midford, T. A. Dececchi, J. P. Balhoff, W. M. Dahdul, N. Ibrahim, H. Lapp, et al. (2013). The Vertebrate Taxonomy Ontology: A framework for reasoning across model organism and species phenotypes. J Biomed Semantics. 4:34.

[6] M. C. Chibucos, C. J. Mungall, R. Balakrishnan, K. R. Christie, R. P. Huntley, O. White, et al. (2014). Standardized description of scientific evidence using the Evidence Ontology (ECO). Database. 2014:bau075