PhylOnt: A Domain-Specific Ontology for Phylogeny Analysis

Maryam Panahiazar*‡, Ajith Ranabahu*, Vahid Taslimi*, Hima Yalamanchili*,Arlin Stoltzfus§¶,
Jim Leebens-Mack†, Amit P. Sheth*
* *Ohio Center for Excellence in Knowledge-enabled Computing (kno.e.sis)*
College of Computer Science and Engineering, Wright State University, Dayton, OH, USA
§ *Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850 USA*
¶ *Biochemical Science Division, NIST, 100 Bureau Drive, Gaithersburg, MD, 20899 USA*
† *Department of Plant Biology, University of Georgia, Athens, GA, USA*
‡ *Bioinformatics Institute, University of Georgia, Athens, GA, USA*

***Abstract*—Phylogenetic analyses can resolve historical relationships among genes, organisms or higher taxa. Understanding such relationships can elucidate a wide range of biological phenomena including the role of adaptation as a driver of diversification, the importance of gene and genome duplications in the evolution gene function, or the evolutionary consequences of biogeographic shifts. The variety of methods of analysis and data types typically employed in phylogenetic analyses can pose challenges for semantic reasoning due to significant representational and computational complexity. These challenges could be ameliorated with the development of an ontology designed to capture and organize the variety of concepts used to describe phylogenetic data, methods of analysis and the results of phylogenetic analyses.**

In this paper, we discuss the development of PhylOnt - an ontology for phylogenetic analyses, which establishes a foundation for semantics-based workflows including meta-analyses of phylogentic data and trees. PhylOnt is an extensible ontology, which describes the methods employed to estimate trees given a data matrix, models and programs used for phylogenetic analysis and descriptions of phylogenetic trees including branch-length information and support values. The relational vocabulary included in PhylOnt will facilitate the integration of heterogeneous data types derived from both structured and unstructured sources. To illustrate the utility of PhylOnt, we annotated scientific literature to support semantic search. The semantic annotations can subsequently support workflows that requiring the exchange and integration of heterogeneous phylogenetic information.

***Keywords*-ontology; semantic technology; phylogeny; phylogenetic workflows; data integration; annotation; phylogenetic analysis.**

I. Introduction

A large amount of life science data, ranging from genotype to phenotype, is publicly available now. The focus of life science domains has shifted towards not only acquiring but also meaningfully using these large data sets. However, a number of challenges exist in data integration, intelligent processing and reasoning, knowledge discovery and decision making.

A growing number of biologists are using gene and species phylogenies to address research problems. These research problems range from assessment of shifting biogeographic patterns to the elucidation of amino acid substitutions associated with pathogenicity. Since phylogenies depict historical relationships among genes and species, we believe that phylogeny study can provide the unifying context across the life sciences for investigating diversification of biological form and function from genotype to phenotype. The increased interest in using and reusing phylogenies however, has exposed major limitations in the accessibility of published phylogenetic trees and the data used to estimate these trees. Most phylogenies can only be found in graphical format embedded in printed or electronic versions of research publications. This greatly limits the ability of biologists to use gene trees and species trees. Not only are the published phylogenetic trees typically inaccessible for semantic integration, but the underlying data and methods of analysis are often not adequately described.

The specific objective of this research is to develop and deploy an ontology for a novel ontology-driven semantic problem solving in phylogenetic analyses and downstream use of phylogenetic trees. This is the foundation to allow an integrated platform in phylogenetically-based comparative analyses and data integration. We named this ontology *PhylOnt*. It describes the methods employed with estimate trees given a data matrix, models, programs and provenance data associated with phylogenetic analyses. PhylOnt also supports the Minimum Information About Phylogenetic Analyses (MIAPA) specification [1], [16], [18] by providing a formal vocabulary for that as a reporting standard. PhylOnt has been publicly shared through the BioPortal [17] at the National Center for Biomedical Ontologies (NCBO), which is a web-services based portal universally accessible over the Internet. Thus, the contributions in this paper are the following:

1) We describe the PhylOnt ontology, an extensible ontology targeted towards data integration in life sciences.
2) We describe the systematic process taken in developing PhylOnt.

978-1-4673-2560-8/12/$31.00 ©2012 IEEE

3) We provide a comprehensive use case of using PhylOnt to annotate a test set of publications. We used a subset of our Kino annotation Tools[3] which enables faceted search over the annotated publications.

The subsequent sections are organized as follows: Section II reviews the background and related work in phylogeny. Section III presents the Challenges and Opportunities in this field. Section IV explains developing a data set and foundation for ontology development. Section V describes the ontology development process. Section VI describes our annotation use case, Section VII presents our evaluation and section VIII includes the conclusion.

II. Background and Related Work

A. Background

The rapidly increasing number of published gene and species trees creates significant opportunities for addressing a variety of biological questions. Further, this trend is certain to pick up pace with the explosion of data generated by the next generation of sequencing technologies. One of the major challenges in this space, data integration, has been successfully addressed by using ontologies. Ontologies are being used as the core knowledge component in a number of sophisticated, integrated platforms for data analysis and integration.

In phylogenic studies, a standard workflow consists of the following steps [1], [10]:

1) Formulation of hypotheses and questions.
2) Identifying steps for a gene or taxon sampling scheme for that question.
3) Data collection, in both scientific and informatics contexts.
4) Constructing the data matrix.
5) Estimating trees with support values.
6) Publishing the results.

Phylogenetic workflows are more varied and complicated than many other types of analyses that have well-developed ontologies [4]. Because of this complexity, development of an ontology to support phylogeny studies is more challenging. When recreation of workflows is important, this kind of complexity can be even more problematic. Storing data items such as documents, publications, underlying data and workflows in structured, exchangeable and easily retrievable formats would facilitate interoperability among various researchers. Such practices would allow researchers to access, explore and reuse the products of phylogenetic studies including innovative workflows.

With these considerations in mind, domain scientists with an interest in archiving and reuse of phylogenetic data have outlined the requirements of a reporting standard, which is called *Minimum Information About Phylogenetic Analyses* (MIAPA) [1]. The main objective of the proposed MIAPA standard is to enable the interpretation of phylogenetic data by multiple researchers. The need for such a reporting standard is clear, but specification of the standard has been hampered by the absence of controlled vocabularies to describe phylogenetic methodologies and workflows with common concepts.

B. Related Work

There are two ontologies that stand out in the domain of Phylogeny.

1) Comparative Data Analysis Ontology: CDAO [5], [6], is an ontology of comparative data analysis that provides a formal ontology for semantic descriptions of data and transformations commonly found in the domain of phylogenetic analysis. However, there seems to be a major gap in CDAO between what is available and what is needed by a scientific researcher for phylogenetic analysis. CDAO does not cover certain concepts related to phylogenetic analysis like methods, models, and programs which need to be described for the community to estimate trees. For example, CDAO includes concepts like node, edge, branch, and network that explain the structure of a phylogenetic tree/network but not the analysis of the phylogeny study.

2) Embrace Data And Methods: EDAM [7] is an ontology developed for general bioinformatics concepts including operations, topics, types and formats. EDAM includes phylogeny related concepts but some of the concrete terms forming the core of phylogenetic analysis including methods, models and programs have neither been explicitly defined under the correct hierarchy nor reported in EDAM.

III. Challenges and Opportunities

The most significant challenge in phylogeny study is the variety and complexity of data being used in phylogeny reconstruction. Some of the reasons that challenge the reuse of this data are incomplete and non-tractable provenance data, insufficient method descriptions to reproduce the results and the lack of semantic annotations.

The different types of data, used in a typical molecular phylogeny analysis workflow, include [1]:

1) Sample description, including taxonomy, collection locality, DNA/RNA preparation.
2) Raw sequence data, sequencing methods, sequence assemblies, assembly method.
3) Alignments and trees including branch lengths (with units) and support values.
4) Alignment programs and their parameter settings.
5) Phylogeny estimation programs, models of evolution, methods, search algorithms, support assessments, and relevant parameter settings.

Our focus in this research is on the last two components: formally characterizing these data types and identifying the relationships between them to develop an ontology for phylogenetic analyses. Developing an ontology and using it to annotate the data and services in workflows can provide a foundation for other semantic technologies, such as concept based searches and comprehensive federated queries on all the data sources.

IV. Data Collection

The first step of ontology development is to understand the domain of study for which it is being developed. This is usually achieved by reviewing and harvesting concepts from exemplary publications and data sets. For this study, we reviewed exemplary papers identified by experts in phylogenetic.

We use the Phyloways [9] environment as a base repository for adding selected papers and results of analysis. In order to perform data extraction, a standard reporting method and formalized methods to extract and classify data from the papers is required. We developed these extraction methods for Phyloways to provide a description of phyloinformatics data and workflows extracted from publications. These descriptions pave the way for classification of concepts associated with phylogenetic data (including provenance information) phylogenetic workflows, and the results of phylogenic analyses.

A. Advantages of data collection with PhyloWays

The PhyloWays data collection has been used as a foundation for making and evaluating diagrams depicting the relationships among concepts that will ultimately evolve into an extensible ontology for phylogenetic analyses. PhyloWays also serves as an archive where users can share comments and link to workflow descriptions of phylogenetic documents. Finally, PhyloWays includes a set of exemplary publications for annotation and validation of the PhylOnt ontology.

B. Candidate cases for analysis

Phylogenetic analyses included in PhyloWays were categorized into protein-based and DNA-based groups. Our initial step has focused on analyses of molecular data rather than other data types such as morphology. For each paper we harvested the following information: publication, data type, alignment method, method of tree estimation, models, programs, parameters, provenance data and additional comments. In Section IV-C, we provide an example of analyzing a publication and writing descriptions for the trees and methods used for phylogeny estimation.

C. Example Analysis

As an example, we provide the detailed steps taken in analyzing the paper from Soltis, et al [19]. The study mentioned in the paper was first converted to a more structured description using commonly used concepts. Compilation of information in PhyloWays provides a foundation to develop workflow diagrams, lists of concepts, and the classification that captures relationships among concepts used in phylogenetic analyses. Some of the categories we collect details under, are as follows.

1) Goal of the selected paper: The main descriptive statement can be the goal of the paper.

2) About the publication:

- *Pub1 has_authors "Soltis DE", "Smith SA"*
- *Pub1 has_citation "Am J Bot 2011:ajb.1000404"*

3) About the phylogeny result:

- *PhylogenyResult1. has_value **
- *PhylogenyResult1. has_method **

The actual values are not included for brevity. The relevant place holders are marked with * symbols. PhylogenyResult1 is an instance of phylogeny results. has_method is object properties and has_value is a data type property. We omit the rest of the details for brevity.

D. Domain and Source

Data Sources in phylogenetic studies can be classified into scientific and meta data categories. Scientific data exists as published data, such as literature with text, images, excel files, and other supplemental materials. Scientific data also refers to methods, models, programs and even parameters used in programs. Meta data includes data about whom, when and where the data was created. This information plays a very important role in the re-usability of valuable resources. For example when researchers want to re-use or repeat any kind of experiment, this kind of knowledge not only helps them to find the data, it also allows them to evaluate the data source, and methods of analysis. The availability of meta-data and provenance information can also aid disambiguation between experimental data entities.

V. A Systematic Approach for Ontology development

A. Data analytical diagrams

Based on discussions with domain experts, literature reviews and the data in PhyloWays, we created initial relational diagrams to describe methods of sequence analysis, models, and widely used phylogenetic programs.

B. Methods in phylogeny analysis

Phylogenetic methods vary considerably in the concepts upon which they are developed and the way they are used to infer relationships. We created a relational diagram that includes several popular methods.

One of the main methods in phylogeny is the optimal tree under the maximum parsimony criterion, which is the minimal tree after evaluating different trees. This method gives the order not the branch length. It means this method searches all possible trees to find the best tree. The optimal tree under the maximum likelihood criterion is the method of search for the tree with the highest probability or likelihood. Bayesian inference in phylogeny generates a posterior

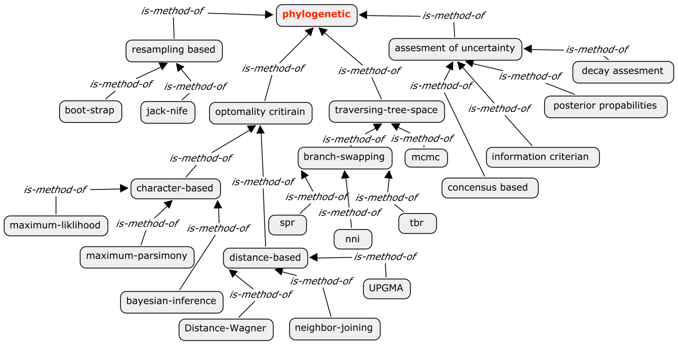


Figure 1. Data analysis diagram for most popular methods in phylogeny

distribution for a parameter, composed of a phylogenetic tree for that parameter and the likelihood of the data, generated by a multiple alignment. All of these methods are character based.The second main group is distance-based methods. Neighbor joining and UPGMA are two different distance-based methods, the former works on unrooted trees while the latter works on rooted trees. Two main validation methods are bootstrap and jack knife resampling. Bootstrap resampling is a method to test how good a dataset fits on an evolutionary model by checking the branch arrangement topology of the tree with bootstrap value.The basic idea behind jack knife resampling is to re-computing the statistic estimate leaving out one observation at a time from the sample set in phylogeny for validation [10], [11]. Figure 1 demonstrates a relational data diagram for the most popular methods in phylogeny study.

C. Models in phylogeny analysis

Model selection is very important and effects most of the stages in phylogenetic inference. The rational development of a phylogenetic method needs a model of evolution as a starting point. Maximum likelihood, Bayesian Inference and most distance-based methods rely on substitution models, but parsimony simply assumes all types of change are equally possible. Substitution models are are classified as DNA model and Model of protein at the first level. JC, HKY, SYM, F81, GTR, and K80 are all models of Nucleotide Substitution Models [12], [1]. Figure 2 is the relational diagram for the most popular models in phylogeny study.

D. Programs in phylogeny analysis

There are approximately 400 phylogeny packages and more than 50 free web servers [8] for such as analysis. PhylOnt currently includes the most commonly used phylogenetic inference programs such as fast-tree, mrbayes, dambe, nona, garli, paup*, raxml, and mega. Programs can be categorized based on the method they used. For example paup* can be used to perform most major methods of analysis such as parsimony, and maximum-likelihood.

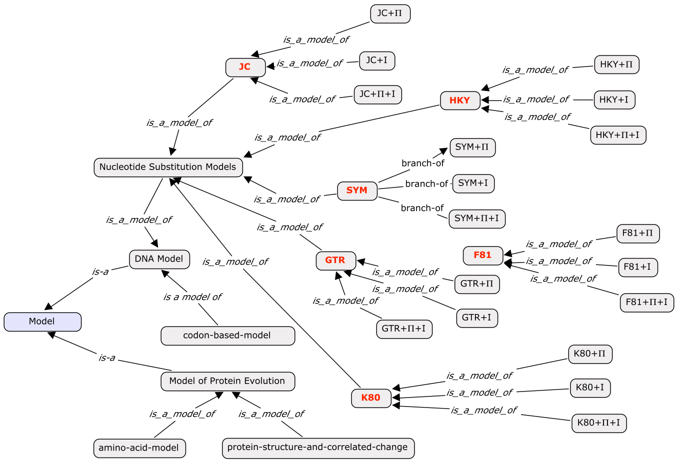


Figure 2. Data analysis diagram for models in Phylogeny

E. Development of the Ontology

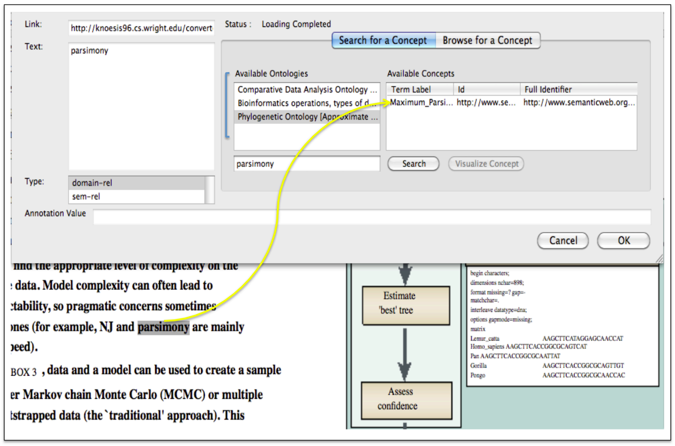
An extensible and broadly applicable ontology for description of phylogenetic analyses can only be developed though synergistic efforts of both the phylogeny community and computer scientists. For this research, we worked closely with phylogeny researchers and computer scientists to develop and validate the ontology.

PhylOnt is being developed in Protege 4.1.0, which supports the Web Ontology Language (OWL). Our ontology includes descriptions of classes, definitions, properties, meta data, usage of classes with an example and relations between them. With the help of NCBO researchers, the PhylOnt Ontology has already been deployed within the BioPortal at NCBO [17]. The BioPortal is a web-services based portal designed to enable universal accessibility over the Internet. The deployment of PhylOnt in the BioPortal maximizes its exposure.

VI. Using Ontology for Annotation - Use case

A fundamental driving principal for the development of ontologies is their utility for data object annotation and management [3]. Therefore, as we developed PhylOnt, we used it to annotate publications from the phylogenetic literature. Here, annotation refers to embedding labels pointing to ontologies from documents. Using accurate annotations pointing to even a single ontology can improve the quality of lookups in a scientific document management system dramatically. From the perspective of database searches, it is very important to have the ability to link from ontology concepts to concepts in publications. Annotating publications with ontology concepts highlights the utility of an ontology in the targeted field of study, and literature searches [3].

One should note, however, that annotation of scientific literature still remains a human-oriented task. Our intention is to provide biologists with a convenient tool to annotate



Link: http://knoesis96.cs.wright.edu/convert

Status : Loading Completed

Text: parsimony

Available Ontologies: Comparative Data Analysis Ontology ... / Bioinformatics operations, types of d... / Phylogenetic Ontology [Approximate ...

Term Label	Id	Full Identifier
Maximum_Parsi...	http://www.se...	http://www.semanticweb.org...

Search: parsimony

Type: domain-rel / sem-rel

Annotation Value

Search for a Concept
Browse for a Concept
Available Concepts
Search
Visualize Concept
Cancel
OK
find the appropriate level of complexity on the
: data. Model complexity can often lead to
:tability, so pragmatic concerns sometimes
ones (for example, NJ and parsimony are mainly
peed).
BOX 3 , data and a model can be used to create a sample
er Markov chain Monte Carlo (MCMC) or multiple
tstrapped data (the `traditional' approach). This
Estimate 'best' tree
Assess confidence
begin characters;
dimensions nchar=898;
format missing=? gap=- matchchar=.
interleave datatype=dna;
options gapmode=missing;
matrix
Lemur_catta AAGCTTCATAGGAGCAACCAT
Homo_sapiens AAGCTTCACCGGCGCAGTCAT
Pan AAGCTTCACCGGCGCAATTAT
Gorilla AAGCTTCACCGGCGCAGTTGT
Pongo AAGCTTCACCGGCGCAACCAC

Figure 3. Annotation of specific concept such as parsimony with kino-phylo tools using PhylOnt

a large volume of documents and retrieve annotated documents for future use.

A. Kino-Phylo tools for annotation

Kino for phylogeny (Kino-phylo) [13] is an integrated suite of tools that enables scientists to annotate phylogeny related web-based documents. Kino-phylo can annotate documents by accessing PhylOnt and other NCBO ontologies. Kino-phylo consists of an NCBO integrated front-end that allows the convenient annotation and submission of web documents through a browser plugin, and an annotation aware back-end, capable of providing faceted search capabilities. These annotations have a variety of uses, ranging from extended search capabilities to advanced data mining.

In Kino-Phylo, Annotated documents are indexed using a faceted indexing and search engine that provides fine grained search capabilities to the scientists [3]. Thus, Kino-phylo is a comprehensive architecture for annotating and indexing phylo oriented documents that should be of great use for the phylogenetics community. Figure 3 illustrates the use of Kino browser plugin to annotate a text segment relevant to phylogeny, using PhylOnt.

VII. Evaluation and discussion

Ontology evaluation is an important task that is needed in many situations. For example, during the process of building of an ontology, ontology evaluation is important to guarantee that what has been built to meet the application requirement. There are different approaches for ontology evaluation, such as evolution-based, rule-based, metric-based and application-based [15]. In this study, we use an annotation-based approach and a metric-based approach to validate quality and quantity of phylOnt.

A. Metric-based approach

These metrics scan through the ontology to gather different types of statistical criteria about the structural knowledge represented in the ontology. In this paper, we followed OntoQA framework that is one of the metric based approaches and used schema metrics [15], [14]. These metrics evaluate ontology design and its potential for rich knowledge representation. In the following, we will list metrics with a brief description and then show the results of our evaluation in Table I. In this table, relationship richness reflects the diversity of the types of relations in the ontology. Attribute richness indicates both quality of ontology design and the amount of information pertaining to instance data. Inheritance richness describes the distribution of information across different levels of the ontologys inheritance tree.

Metric name	Metric formula[1]	Metric value
Relationship Richness	$PR = \frac{\vert P\vert}{\vert H\vert + \vert P\vert}$	0.54
Attribute Richness	$AR = \frac{\vert T\vert}{\vert C\vert}$	0.18
Inheritance Richness	$IR = \frac{\vert H\vert}{\vert C\vert}$	0.94

[1] $|H|$:Number of inheritance relationships, $|P|$:Number of non-inheritance relationships, $|C|$:Number of classes, $|T|$:Number of attributes

Table I
Metric-based approach to evaluate the quantity of PhylOnt

The results of the relationship richness formula show that more than half of the connections between classes are rich relationships compared to all of the possible connections. Inheritance Richness describes our ontology as deep vertical, which indicates that it covers a specific domain in a detailed manner.

Ontology	Precision	Recall	F-measure
PhylOnt	0.64	0.43	0.51
EDAM	0.17	0.013	0.024
CDAO	0.07	0.15	0.095

Table II
Precision, Recall and F-measure results for annotation-based approach

B. Annotation-based approach

In this approach, we tried to annotate the papers selected by experts using PhylOnt. We investigated which concepts are missing in our ontology, in practice by trying to annotate using PhylOnt. The rationale is that we could determine the quality of PhylOnt by counting the relevant concepts encountered in a paper that are not present in PhylOnt, but are present in other relevent ontologies. This approach is used to compute precision, recall, and F-measure [4]. Suppose that $C_{\{P \cap O\}}$ is the set of concepts from the papers which have been annotated using PhylOnt. Then $Precision$ and $Recall$ can be calculated by the following equations.

$$Precision = \frac{|C_{\{P \cap O\}}|}{|C_P|} \quad (1)$$

$$Recall = \frac{|C_{\{P \cap O\}}|}{|C_O|} \qquad (2)$$

C_P and C_O refer to the concepts of the paper and concepts in ontology respectively. The F-measure is the harmonic mean of precision and recall and it is calculated as

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (3)$$

For this experiment, we annotated selected papers using PhylOnt, EDAM and CDAO. As it is shown in Table II the precision of PhylOnt demonstrates that more than half of the concepts in the paper are annotated correctly using PhylOnt. Around 43 percent of the all concepts in the paper that should have been annotated are annotated correctly with PhylOnt. Also the F-measure of PhylOnt is 5 times more than that of CDAO and 20 times more than that of EDAM. The reason may well be that PhylOnt is specifically developed for phylogeny operations, methods, models and programs in phylogeny analysis.

VIII. Conclusion

With the growing importance of using semantic technology in life science, having a well-defined ontology is necessary to make a foundation and facilitate the integrity and accessibility of data and services. To the best of our knowledge and the feedback from Phylogeny community [2], PhylOnt is the first ontology specifically extended for phylogeny analysis operations and related meta data. As of May 2012, the fourth version of phylOnt has been submitted to NCBO and has been used in other projects [17], [16]. Our results show that PhylOnt is a rich ontology, compared to other alternatives.

Annotating phylogeny documents with ontology is the foundation for other semantic technologies and it is a preliminary step to semantic search, information retrieval, and heterogeneous data integration that can support phylogeny workflows. PhylOnt has been introduced as an important component in our integrated Semphyl platform [13].

Acknowledgment

The authors would like to thank to Dr. Enrico Pontelli, New Mexico State University, Tu Thien Danh and Sreeram Vallabhaneni at Wright State University for their help and discussion during ontology development. The identification of any specific commercial software products is for the purpose of specifying a protocol, and does not imply a recommendation or endorsement by the National Institute of Standards and Technology.

References

[1] Leebens-Mack, J. Vision, T. Brenner, E.Bowers, J.E. Cannon, S.Clement, M.J. Cunningham, C.W. DePamphilis, C. DeSalle, R. Doyle, J.J. and others. Taking the first steps towards a standard for reporting on phylogenies: Minimum Information About a Phylogenetic Analysis (MIAPA). OMICS: A Journal of Integrative Biology. 10(20):231-237. 2006

[2] Panahiazar, M. Vos, R. Pontelli, E. Vision, T. Stoltzfus, A. and Leebens-Mack, J. Building a Foundation to Enable Semantic Technologies for Phylogenetically-Based Comparative Analyses. Talk at iEvoBio 2011, Oklahoma, USA, June 21 - 22. 2011

[3] Ranabahu, A. Parikh, P. Panahiazar, M. Sheth, A. and Logan-Klumpler, F. Kino: A Generic Document Management System for Biologists Using SA-REST and Faceted Search. 2011 Fifth IEEE International Conference. (ICSC 2011), pp.205-208, 18-21 Sept.2011

[4] Cross, V. Stroe, C. Hu, X. Silwal, P. Panahiazar, M. Cruz, I.F. Parikh, P. and Sheth, A. Aligning the Parasite Experiment Ontology and the Ontology for Biomedical Investigations Using AgreementMaker. International Conference on Biomedical Ontology (ICBO 2011), Buffalo, New York, July 26-30, 2011

[5] https://www.nescent.org/wg_evoinfo/CDAO

[6] Chisham, B. Wright, B. Le, T. Son, T. and Pontelli, E. CDAO-Store: ontology-driven data integration for phylogenetic analysis. BMC bioinformatics, 2011

[7] http://purl.bioontology.org/ontology/EDAM

[8] http://evolution.genetics.washington.edu/phylip/software.html

[9] http://www.evoio.org/wiki/MIAPA/PhyloWays

[10] Jill Harrison, C. and Langdale, J.A. A step by step guide to phylogeny reconstruction. The Plant Journal.45:561-572. 2006

[11] Nei, M. and Kumar, S. Molecular evolution and phylogenetics. Oxford University Press, New York, 2000

[12] Posada, D. and Buckley, T.R. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic biology. 53:793–808. 2004

[13] Panahiazar, M. and Leebens-Mac, J. and Ranabah, A. and Sheth, A. Using Semantic Technologies for Phylogeny, Abstract, page 166, 2012 AMIA Summit on Translational Bioinformatics. 2012

[14] Vrandei, D. and Sure, Y. How to design better ontology metrics.The Semantic Web: Research and Applications. 311-325. 2007

[15] Poli, R. Theory and Applications of Ontology: Computr Applications. Val 2. Springer Verlag. 2010

[16] https://www.nescent.org/wg_evoinfo/MIAPA_WhitePaper

[17] http://bioportal.bioontology.org/ontologies/1616

[18] http://wiki.tdwg.org/twiki/bin/view/Phylogenetics/PhylOnt

[19] Soltis, D.E. and Smith, S.A. and Cellinese, N. and Wurdack, K.J. and Tank, D.C. and Brockington, S.F. and Refulio-Rodriguez, N.F. and Walker, J.B. and Moore, M.J. and Carlsward, B.S. and others. Angiosperm phylogeny: 17 genes, 640 taxa. American Journal of Botany. 98:704-730. 2011