



PAUL G.
ALLEN
SCHOOL
W

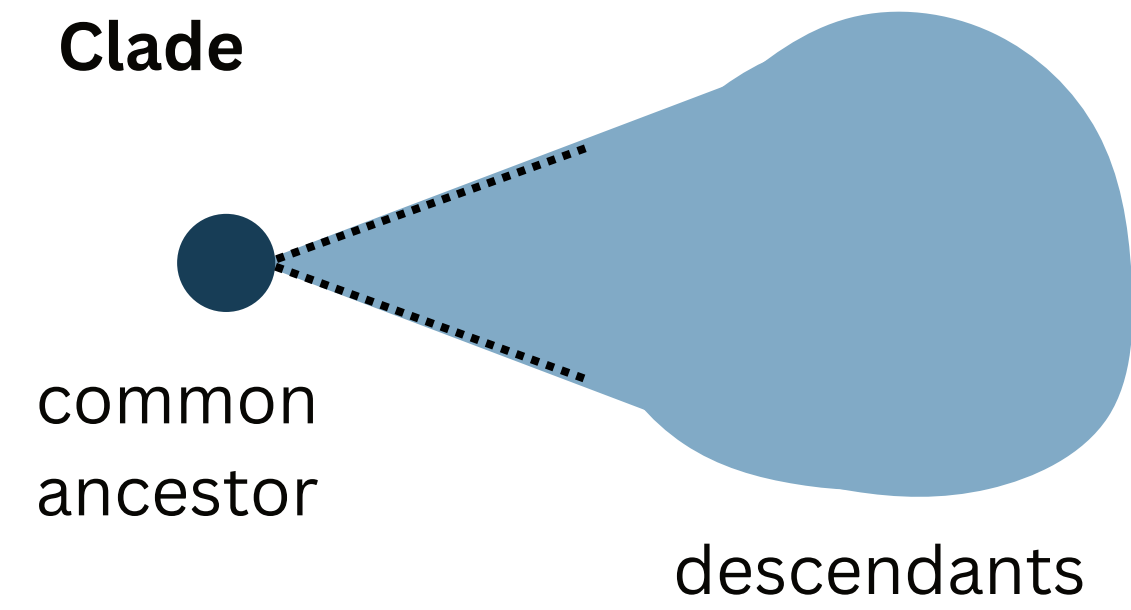
Examining Clade Structures: an ADT Approach

Yuning Hu
Faculty Advisor: Rose Novick
yuninh@uw.edu

W DEPARTMENT OF PHILOSOPHY
UNIVERSITY of WASHINGTON

Introduction

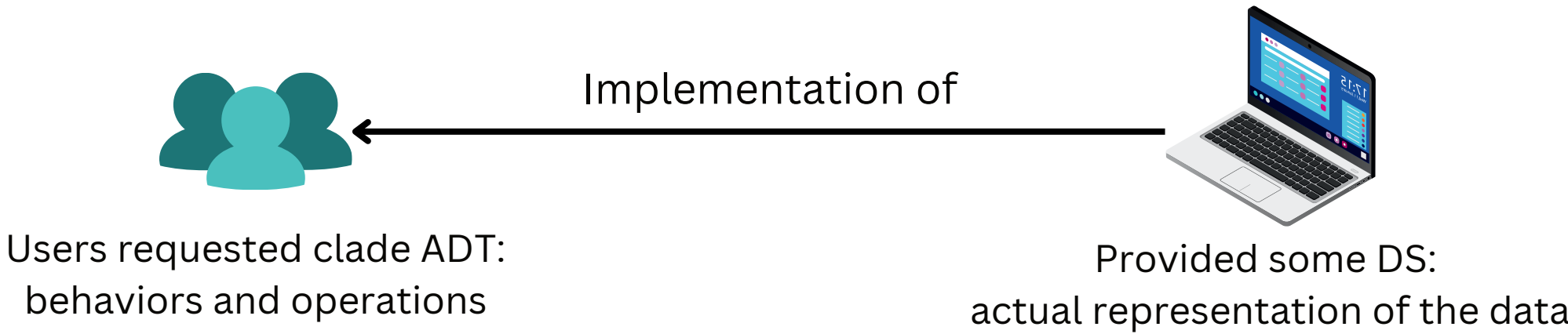
Due to the recognition of limited resources on the earth, it is true that the majority of scientists agree that natural selection plays an important role in shaping the diversity of life on the earth. As outlined by Lewontin (1970, 1980), Godfrey-Smith (2007), and others, evolution by natural selection (ENS) generally should fulfill the two key criterias (or four specified criterias): selection (population, variation) and response to selection (fitness, heritability). If an entity meets the ENS requirements for undergoing a natural selection, we may refer to it as a "Darwinian" entity or a unit of selection. Similarly, we may refer to the biological hierarchy of the unit of selection as a level of selection. The recognition of species as sufficiently "Darwinian" entities for natural selection (Lewontin, 1970) has led to the exploration of clades as alternative units of selection to species in the field of philosophy of biology. Clade refers to the taxonomic group of organisms that are monophyletic. In other words, a clade is composed of a common ancestor and all of its descendants. Previous research has demonstrated that clades meet the criteria of "Darwinian" entities (Doolittle, 2017). Also, the concept of clades aligns with Penny's (2011) belief that Darwin suggests that, as with branches in a living tree, lineages of species competed with and supplanted one another. Therefore, it is intriguing for philosophers to understand and explain natural selection at the clade level. The objective is to scrutinize the inner structure of clades and uncover the interplay between clades and species, a relationship not explicitly implied in their definitions.



ENS	General	Clades
Selection	Population	Balance (clade growth, death)
	Variation	
Response to selection	Fitness	Differential persistence
	Heritability	

Method

This study employs an Abstract Data Type (ADT) approach from computer science to model clades. Abstract Data Type (ADT) is a programming paradigm that defines a set of operations and behaviors on a conceptual level. In this context, we conceptualize clades as our ADT. While the specific operations on a clade are not yet defined, we would leave them for the discussion section. For now, our focus lies on the behavioral aspects of a clade ADT, particularly its ability to demonstrate the relationship between ancestors and their descendants. Data structure (DS) serves as a realization or implementation of the abstract concepts defined by the ADT. Given that the ancestor-descendants relationship in clades resembles the node-leaf relationship in trees, we choose to utilize tree-based data structures to implement the clades ADT. In our approach, we would utilize some tree DS to implement the clades ADT both with and without the knowledge of species.



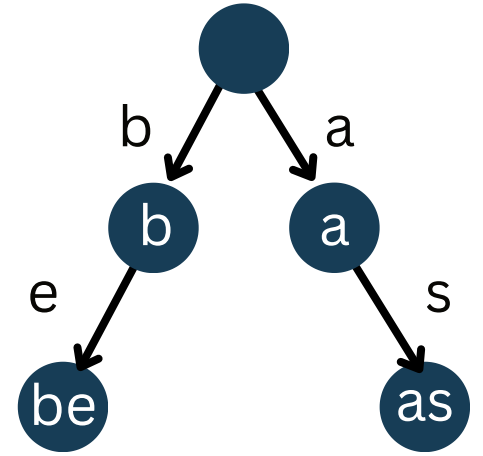
Clade With Species

In both educational and scientific contexts, clade data is typically represented by phylogenetic trees, with each node symbolizing a common ancestor branching into its descendants. Then in terms of programming, we would choose trie (prefix tree) data structure for its resemblance to this hierarchical representation.

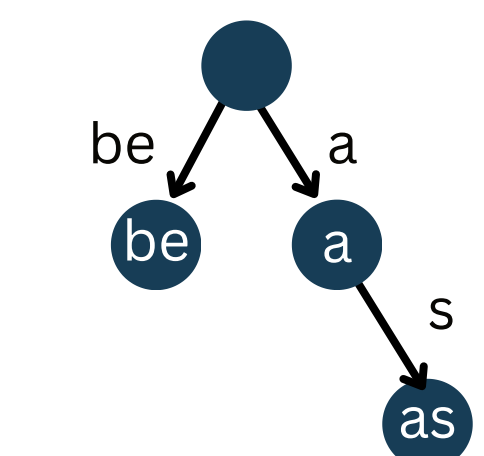
- node Tries as dictionaries
- key (terminate retrieval when reaching a complete English word)
- value

Depending on implementation, tries may represent termination of retrieval differently

Traditional trie (prefix tree)

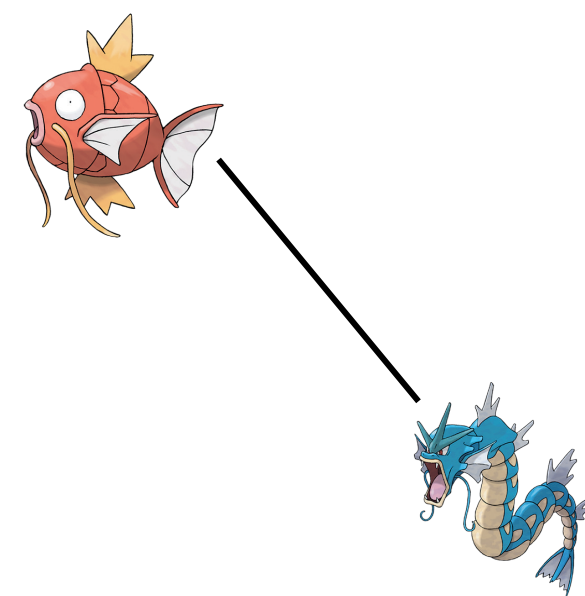


Radix trie (compact prefix tree)

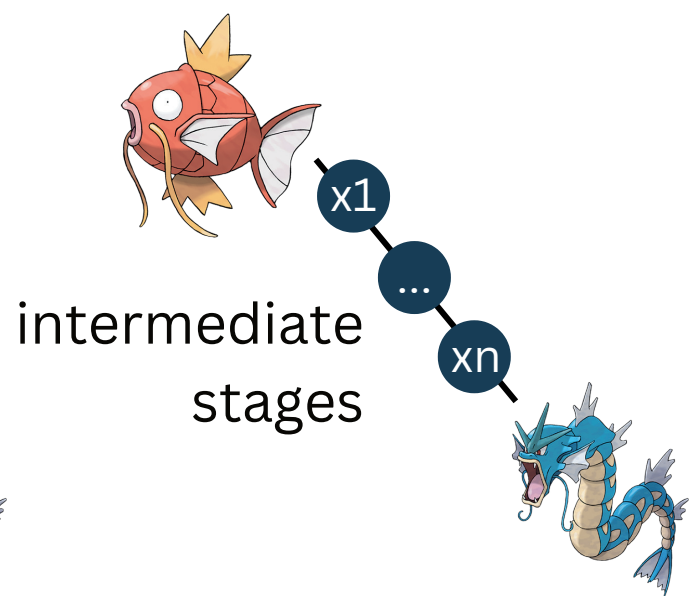


If we use the data extracted from the existing phylogenetic tree...

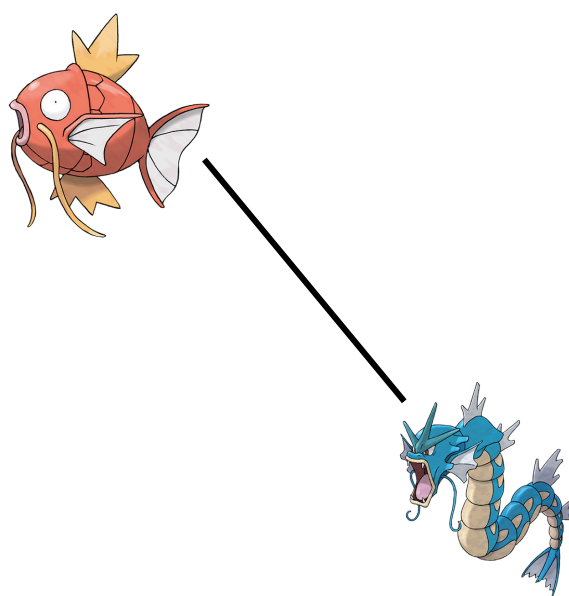
Existing phylogenetic tree



Traditional trie (prefix tree)

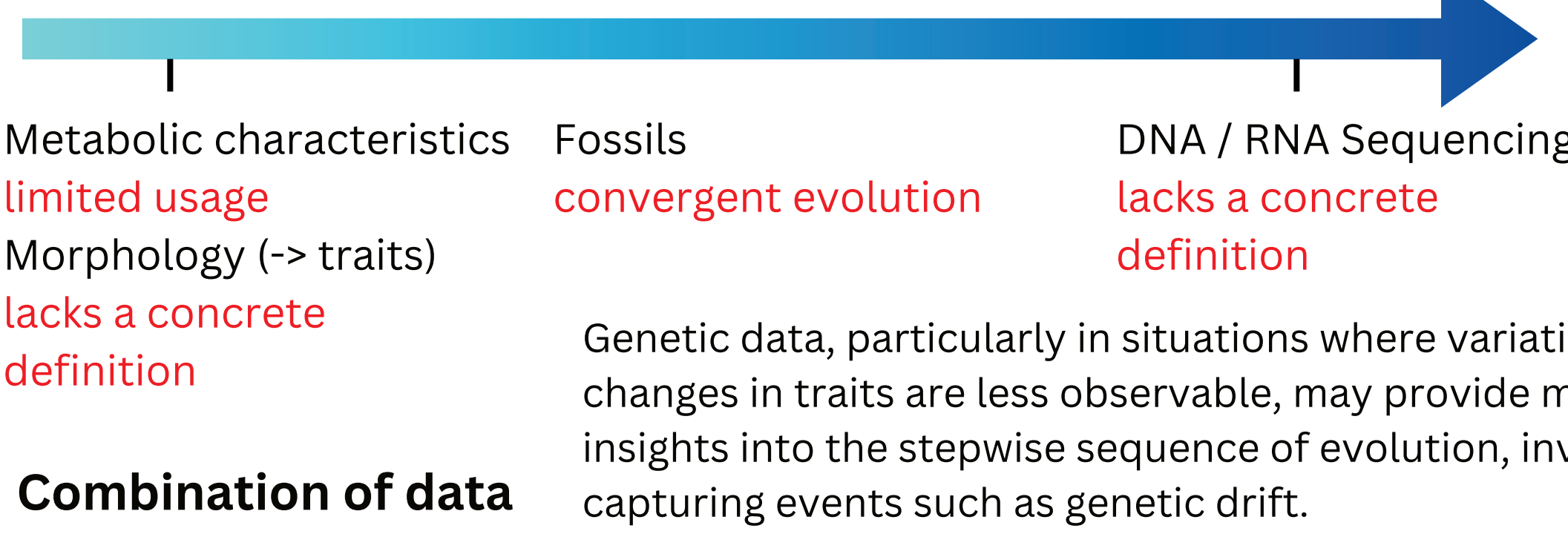


Radix trie (compact prefix tree)

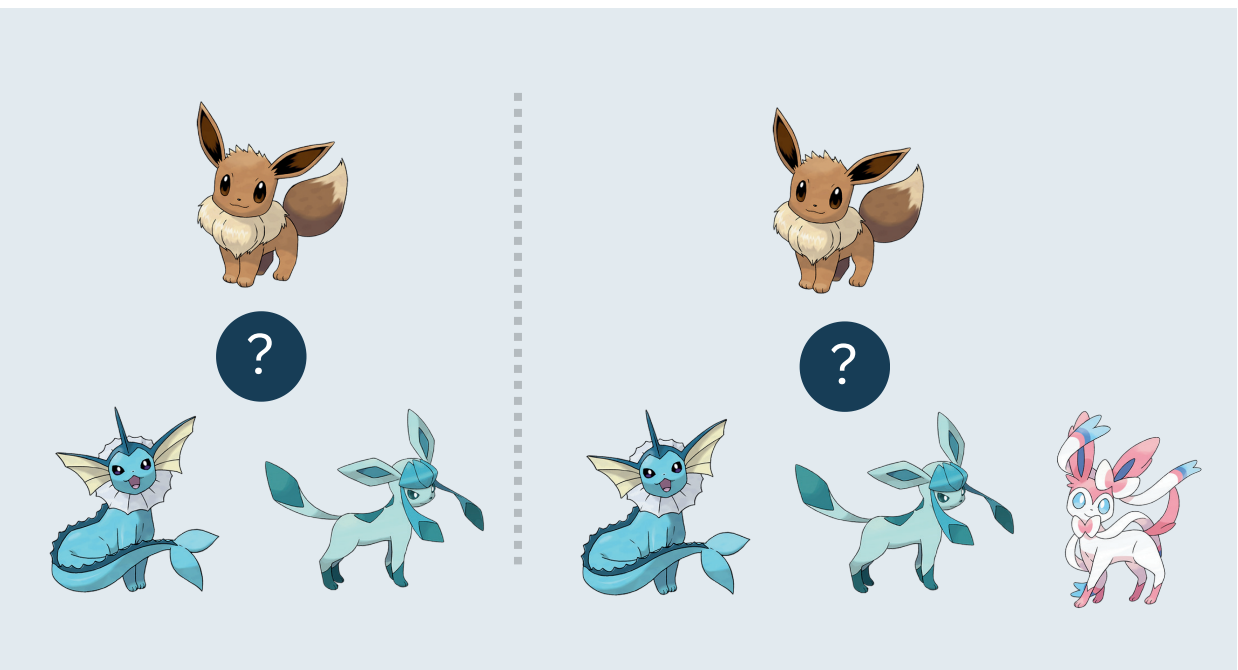


However, we may not directly insert into our tries the data from existing phylogenetic trees, as the data may not accurately represent the true evolutionary path.

Development history of the phylogenetic tree



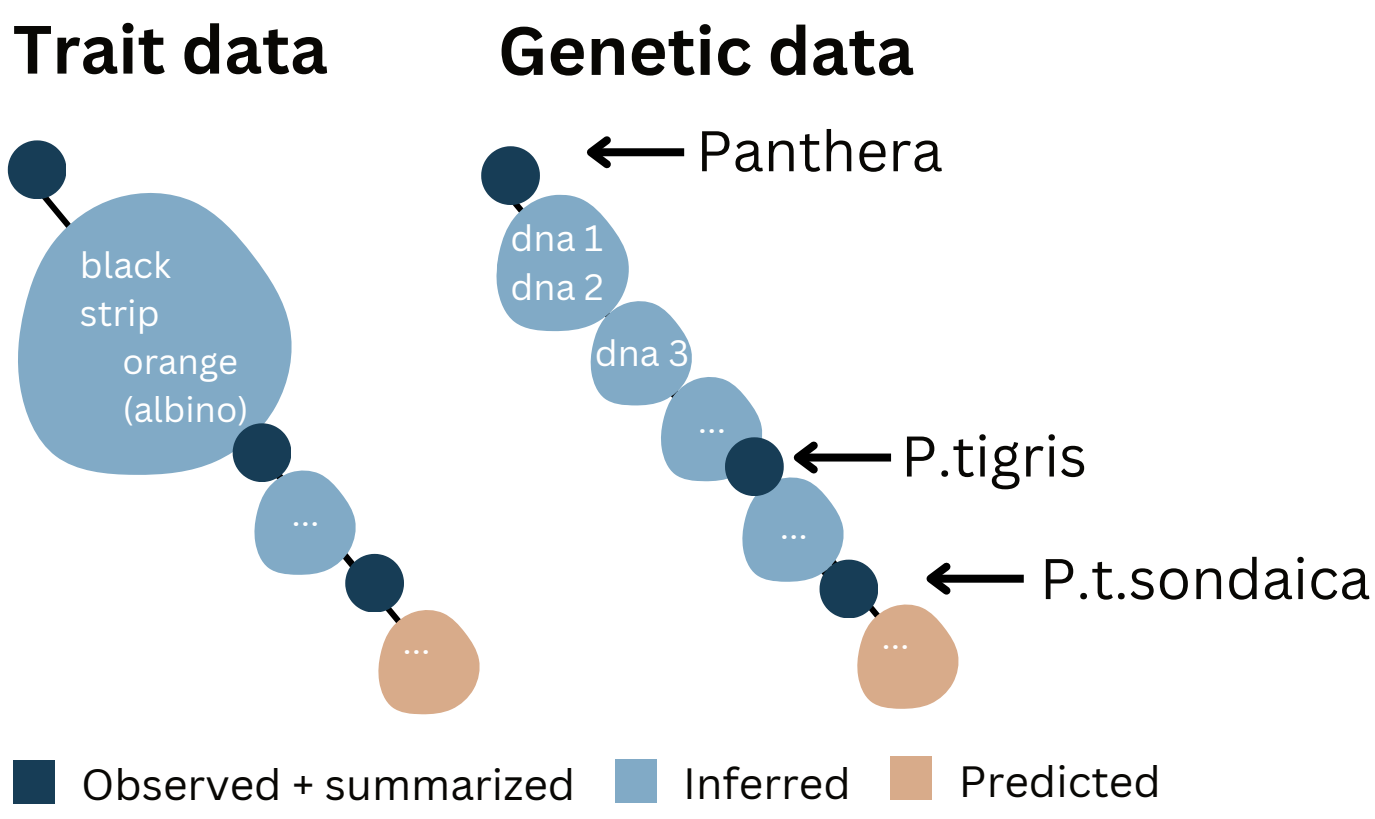
Combination of data



Top down: define speciation (when split) and furcation (what they were split into). **Bottom up:** define aggregation (when stopped splitting) and joining (what was merged into)

Clade Without Species

Methodologically, the data, from both observations and fossil records, used in constructing a foundation clades tree model, are often identified as **static, terminal nodes** or, more directly, **species**. Our understanding of the world is predominantly **species-based**, making it challenging to move away from species-level concepts.



Discussion & Further

- Standardization of alterations in clade data resulting from hybridization and horizontal gene transfer.
- Inclusion of sub-species taxonomy, e.g. breeds.
- Proposal of possible operations on a clade ADT, e.g. incorporating new discoveries.

Conclusion

Findings confirm the technical feasibility of implementing clades ADT with a trie data structure incorporating species knowledge. However, the distinct features of each level of clades remain ambiguous: (1) The complex nature of the data used to construct the data structure representing clades results in a holistic yet disorganized demonstration of evolution. (2) The absence of a consensus on principles for classifying organisms into a clade structure poses a challenge of inconsistency. On the other hand, a clade-tree data structure without species could be constructed with genetic data but still encounters challenges: (1) The empirical nature of genetic data prioritizes understanding at the species level, potentially undermining the species-free intention of the data structure. (2) Interpretation. The revelation of the inconsistency and the dependency on species introduces a compelling reason for reassessing the efficacy of clades in terms of explaining evolution. This study highlights the complexities surrounding the use of clades as units of selection and invites further refinement of biological taxonomy.

References

Crafa, S. (2015). Modelling the Evolution of Programming Languages. ArXiv, abs/1510.04440.
Doolittle, W. F. (2017). Making the Most of Clade Selection. Philosophy of Science, 84(2), 275–295.
Lewontin, R. (1970). "The Units of Selection." Annual Review of Ecology and Systematics 1: 1-18.
Lewontin, R. (1980). "Adaptation." Reprinted in R. Levins and R. Lewontin, The Dialectical Biologist. Cambridge: Harvard University Press, pp. 65-84.
Godfrey-Smith, P. (2007). Conditions for Evolution by Natural Selection. Journal of Philosophy 104 (10):489-516.
Liu, Y. C., Sun, X., Driscoll, C., Miquelle, D. G., Xu, X., Martelli, P., Uphyrkina, O., Smith, J. L. D., O'Brien, S. J., & Luo, S. J. (2018). Genome-Wide Evolutionary Analysis of Natural History and Adaptation in the World's Tigers. Current biology : CB, 28(23), 3840–3849.e6. <https://doi.org/10.1016/j.cub.2018.09.019>
Penny, D. (2011). "Darwin's Theory of Descent with Modification, versus the Biblical Tree of Life". PLOS Biology. 9 (7): e1001096.
Pokedex. Pokemons. pokedex.pokemons.com/us/legal/