

Using Biological and Paleontological Evidence to Calculate Evolutionary History



Yiteng Dang¹
 Supervisor: Martin R. Smith²
 Department of Earth Sciences, University of Cambridge

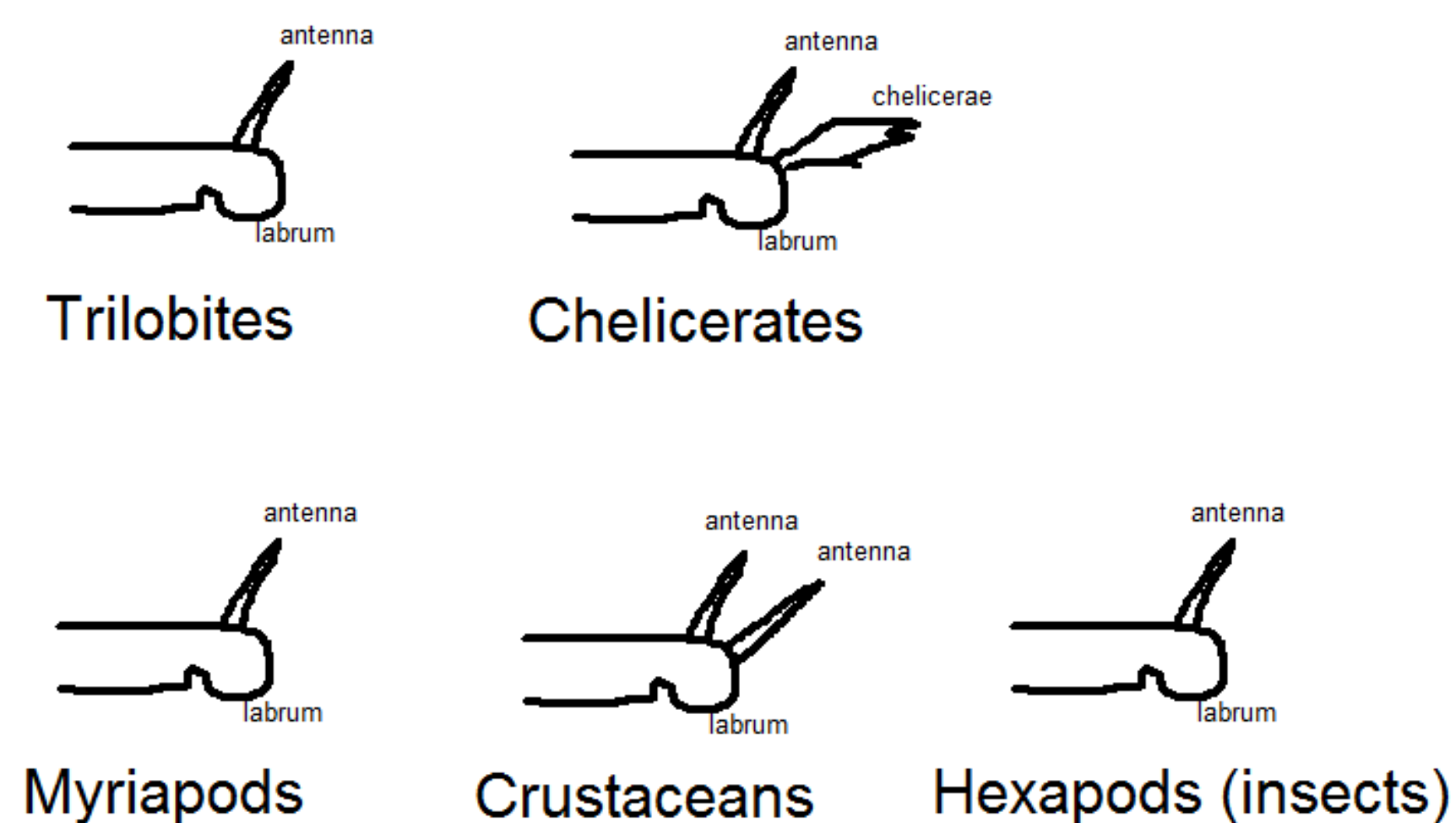
Funded by the Post Master
 Consultancy Scheme

Project Description

Historical biology is primarily concerned with the inference of evolutionary history and relationships. Whereas molecular data can be readily interrogated under a range of models, the reconstruction of phylogenetic trees from morphological data – the mainstay of palaeontologists and conventional taxonomists – has outstanding problems. This project's objective was to adapt existing models of morphological evolution to incorporate peculiarities of morphological data sets.

Background

A typical problem that involves morphological data is the arthropod head problem. Arthropods form a phylum which includes the hexapods, crustaceans, myriapods, chelicerates and trilobites). Their heads have complex structures which include one or more pairs of antennae. The evolutionary relationships and between these appendages – whether some antennae have the same origin as others – is an ongoing topic of debate among zoologists.



Schematic representation of head structures of five groups of arthropods

References

- [1] Felsenstein, Joseph (2004). *Inferring Phylogenies*.
- [2] Brazeau, Martin D. Problematic character coding methods in morphology and their effect, *Biological Journal of the Linnean Society*, 2011, 103, p. 489-498.
- [3] Swofford, David L. and Maddison, Wayne P. *Reconstructing Ancestral Character States Under Wagner Parsimony*, *Mathematical Biosciences* 87, 199-229 (1987).
- [4] De Laet, Jan E (2005). Parsimony and the problem of inapplicable in sequence data, in: Albert, V.A. (ed.) *Parsimony, phylogeny and genomics*.
- [5] Budd, Graham E. *A paleontological solution to the arthropod head problem*, *Nature*, vol. 417, 16 May 2002, p. 271-275.

Method

The literature contains many algorithms for inferring phylogenies [1], often based on the principle of parsimony: the most likely scenario is the one with the fewest number of evolutionary changes. An example is the Fitch algorithm, suited for non-additive data. However, in a morphological study of arthropod appendages we deal with additive data in the form of a character describing the number of appendages. We would also like to include other data such as appendage colour or shape, but they are dependent on the presence of a certain appendage. This contingency leads to complications involving inapplicable data [2].

We split the problem into two parts: (1) reconstruct the tree with the number of appendages as single character, and (2) apply an algorithm adapted from the Fitch algorithm to the full data set, whereby only evolutionary changes in present states are taken into account. The first part relies on Wagner parsimony, and algorithms to find optimal trees have been proposed decades ago [3]. However, these existing algorithms are either too general or fail to maximize the number of homologies, as desirable by the Hennig-Farris principle [4]. Hence we have tried to find a novel algorithm suitable for dealing with the particular problem studied.

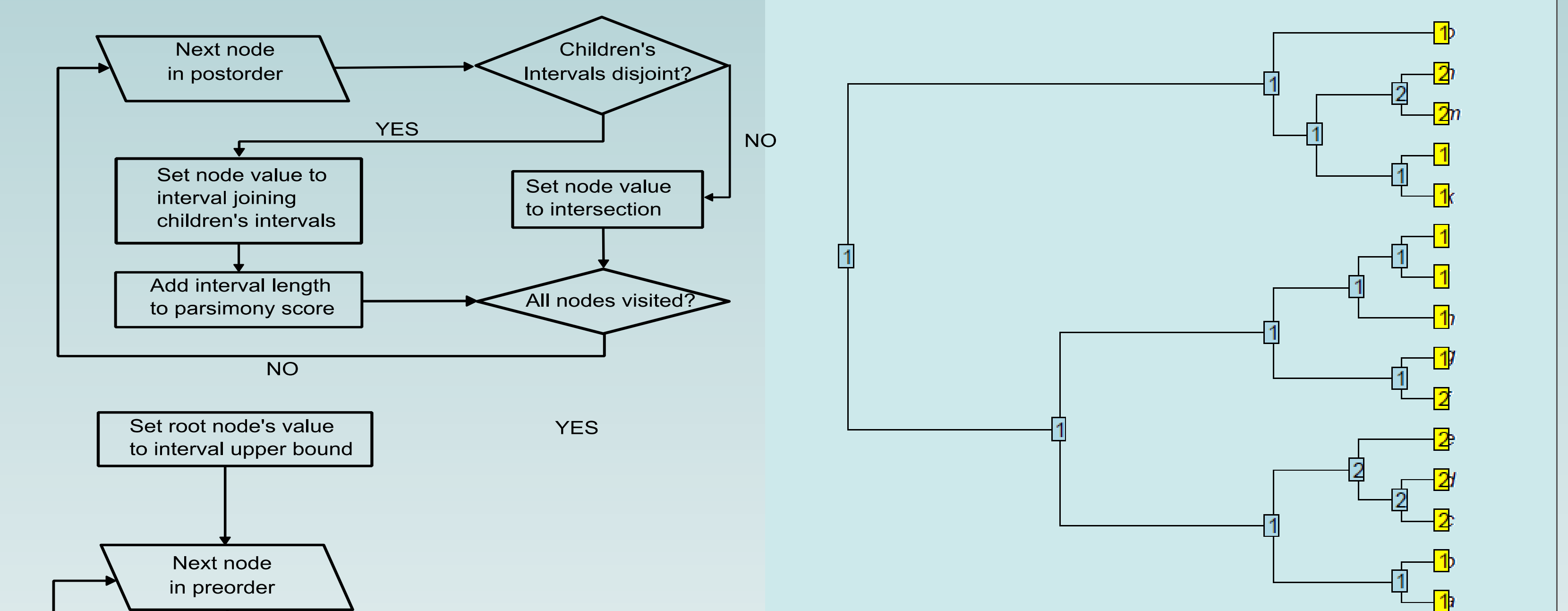
Results

A new algorithm has been proposed for inferring the tree for the first part of our approach. It consists of a downpass from the tips to the root and an uppass from the root upwards assigning definite values to all internal nodes and tips (see flowchart). The algorithm was devised by examining trees locally by looking at an internal node and its immediate ancestor and descendants. By considering the various possibilities for their datasets exhaustively and optimizing the internal node in each of the cases, we obtained the algorithm displayed to the right.

This algorithm has been implemented in R and C code and has been tested for a range of sample trees. The results found all coincide with the best tree obtained by parsimony and the Hennig-Farris principle (example: top-right figure).

Conclusion

The algorithm is a proposed solution for the first part of the problem. Each part of the tree is optimized locally during the uppass, but it remains to prove that the resulting full tree is optimal. Furthermore, ongoing work has resulted in an algorithm for the second part of the problem for single-character trees. An extension to multi-character trees seems possible, but has not yet been realized.



Above: algorithm applied to arbitrary tree

Left: flowchart of the algorithm.

Below: example of phylogenetic tree (source: [5])

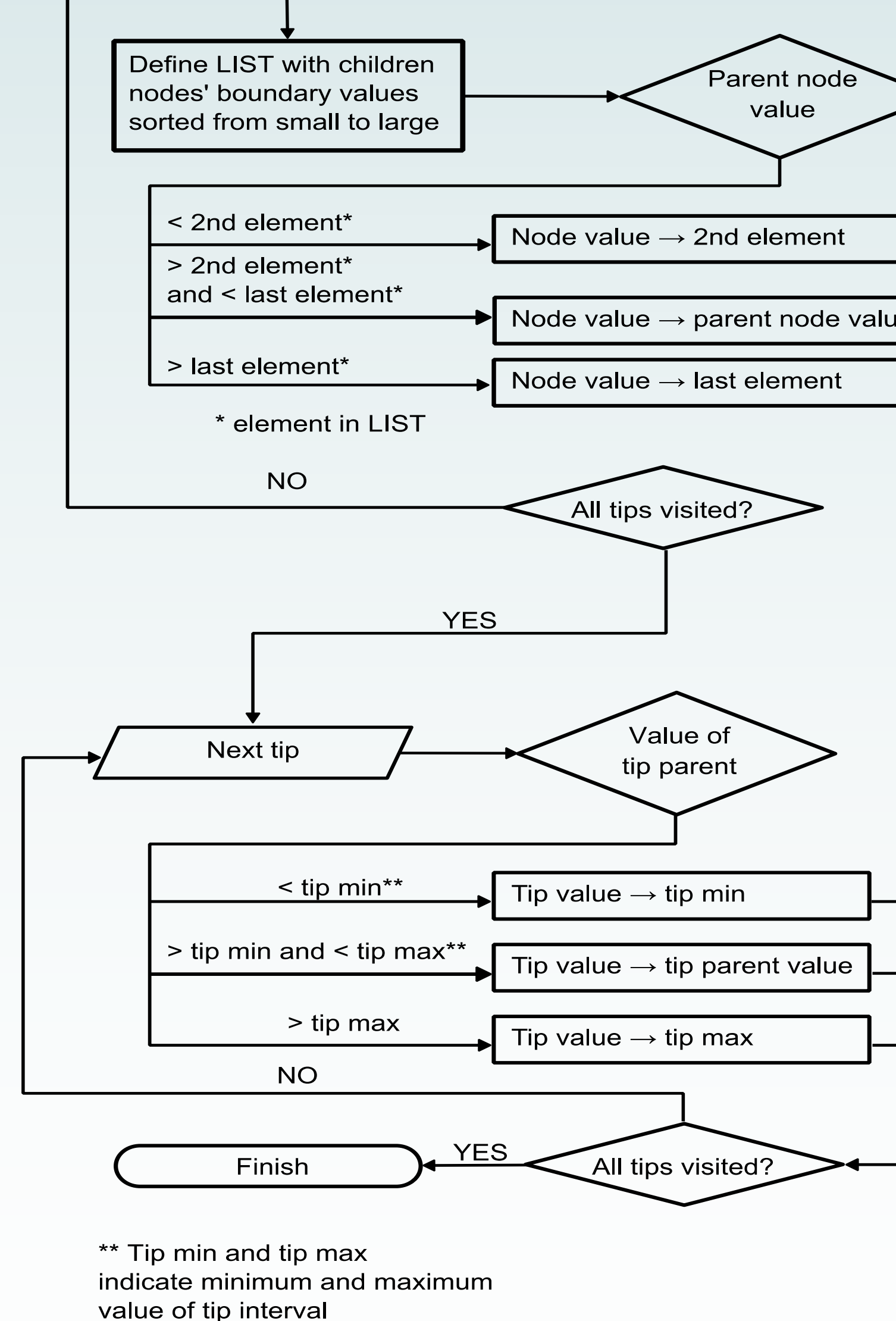
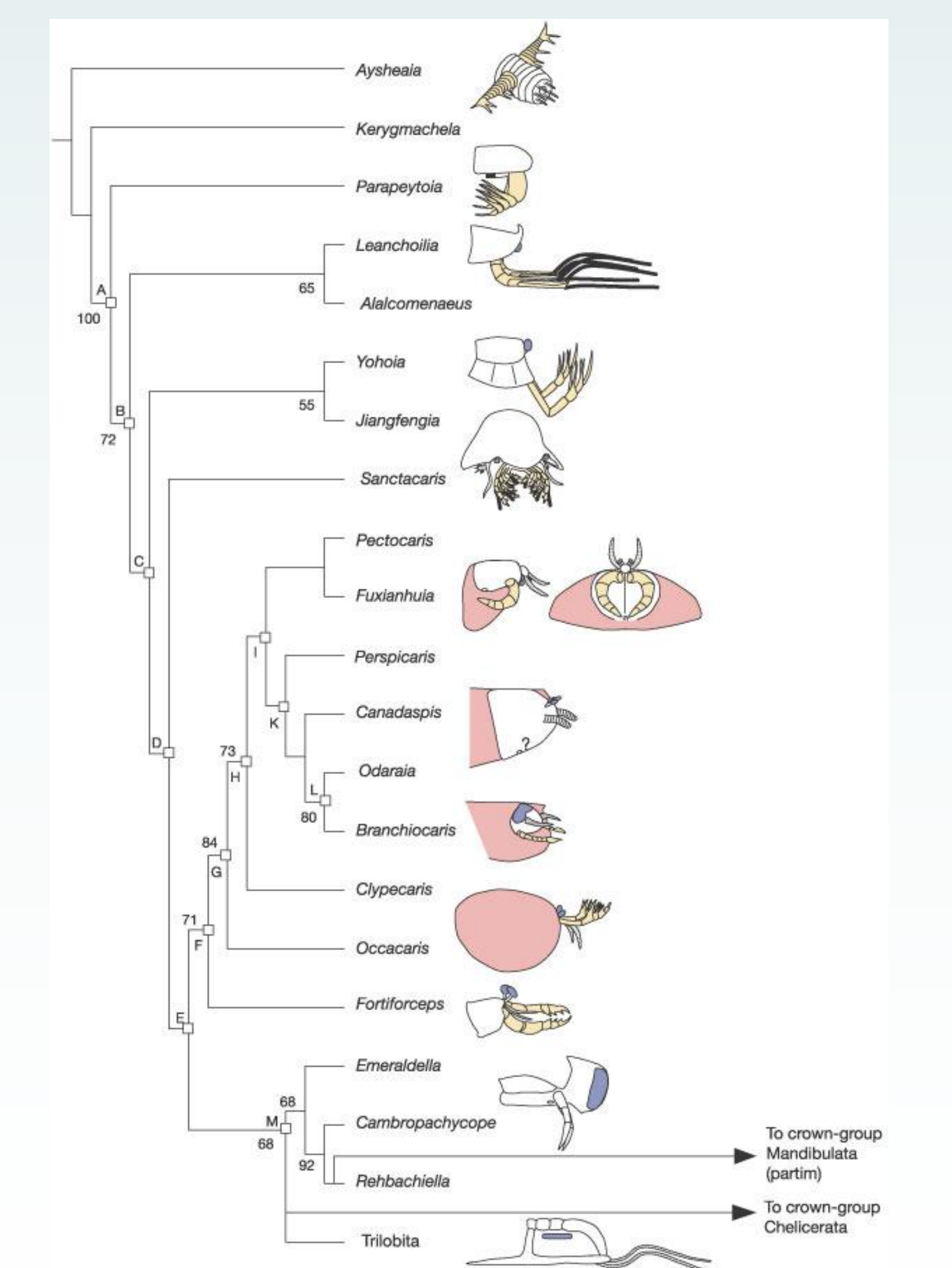


Figure 1: Schematic implementation of the algorithm



Contact information: (1) yiteng@live.nl, (2) ms609@cam.ac.uk