Distinguishing Dactyls of Crab Species Using Relational Machine Learning

Mark Goadrich and Jeffrey Agnew
Departments of Computer Science and Geology
Centenary College of Louisiana

{mgoadric, jagnew}@centenary.edu

Abstract

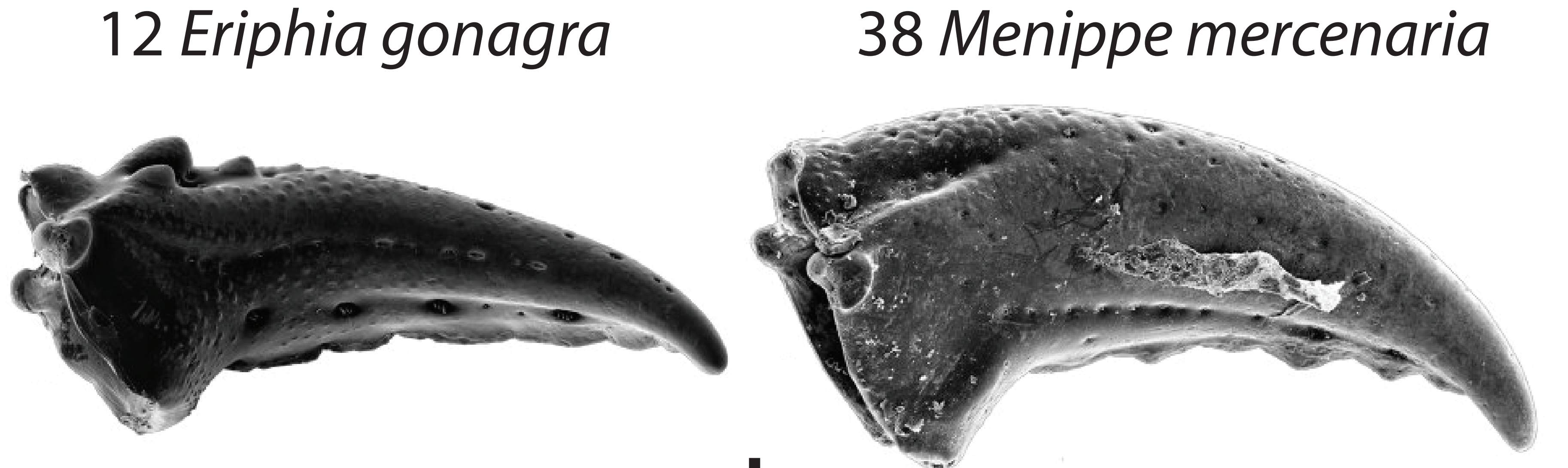
Fossils of decapod crustaceans are more common than published records suggest. Dactyls (movable fingers of claws) are well represented in shell-rich fossil assemblages, but are usually ignored because of the assumption that they can be identified only to high taxonomic levels. Recent studies using outline-based and geometric morphometric methods have demonstrated that closely related species of decapods can be distinguished by their dactyls. Although these techniques allow statistical tests of differences in dactyl morphologies, dactyl shapes must still be described qualitatively.

Our research introduces a new method for distinguishing dactyl shapes by automatically extracting relational features that describe their underlying spatial structure. We first use medial axis techniques, used for shape recognition algorithms in computer vision, to find the shock graph of each dactyl outline. Next, these shock graphs are converted into a first-order logic representation capturing the connections, distances and angles between the nodes in each graph. We then use Aleph, an Inductive Logic Programming algorithm, to find relational classification rules based on the shock graph representations. These relational rules provide a concise and human-understandable way to describe the morphological differences among dactyls of closely related decapods, and can be seen as a first step to creating automatically learned quantitative taxonomic keys.

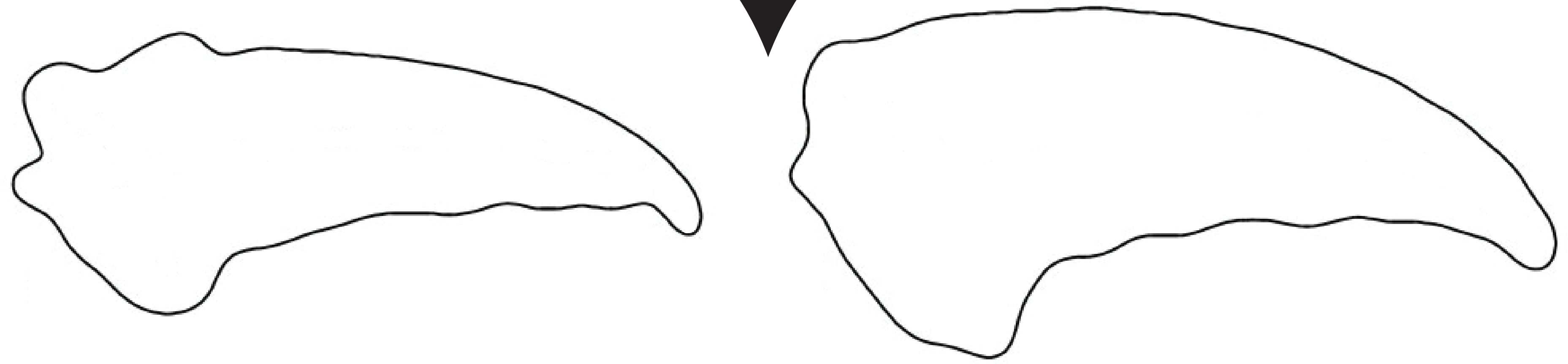
References

1. J. G. Agnew. *Dactyls Reveal Evolutionary Patterns in Decapod Crustaceans*. PhD thesis, Louisiana State University - Baton Rouge, 2008.
2. P. Dimitrov, C. Phillips, and K. Siddiqi. Robust and efficient skeletal graphs. In *Computer Vision and Pattern Recognition*, 2000.
3. D. Macrini. *Indexing and matching for view-based 3-D object recognition using shock graphs*. Master's thesis, University of Toronto, July 2003.
4. A. Srinivasan. The Aleph Manual Version 5. http://web.comlab.ox.ac.uk/oucl/research/areas/ machlearn/Aleph/, 2003.
5. A. B. Williams. *Shrimps, lobsters, and crabs of the Atlantic coast of the eatern United States, Maine to Florida.* Smithsonian Institution Press, Washington, D.C., 1984.

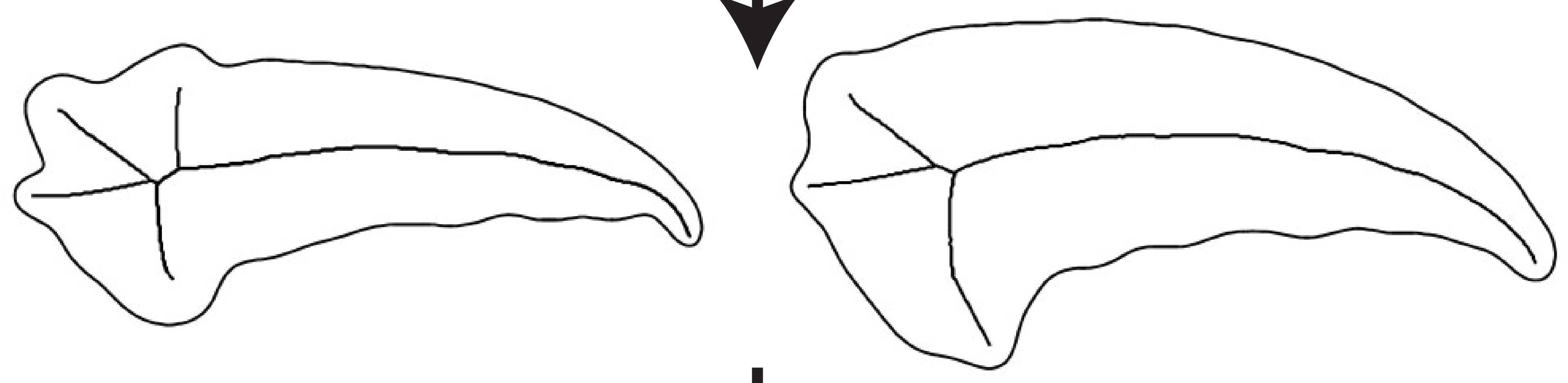
We would like to thank the authors of the software packages Aleph, SHAPE and ShapeMatcher for the availability of their code, and Centenary College for providing travel funds.


12 *Eriphia gonagra*
38 *Menippe mercenaria*

Calculate Outlines with SHAPE



Find Shock Graph with ShapeMatcher



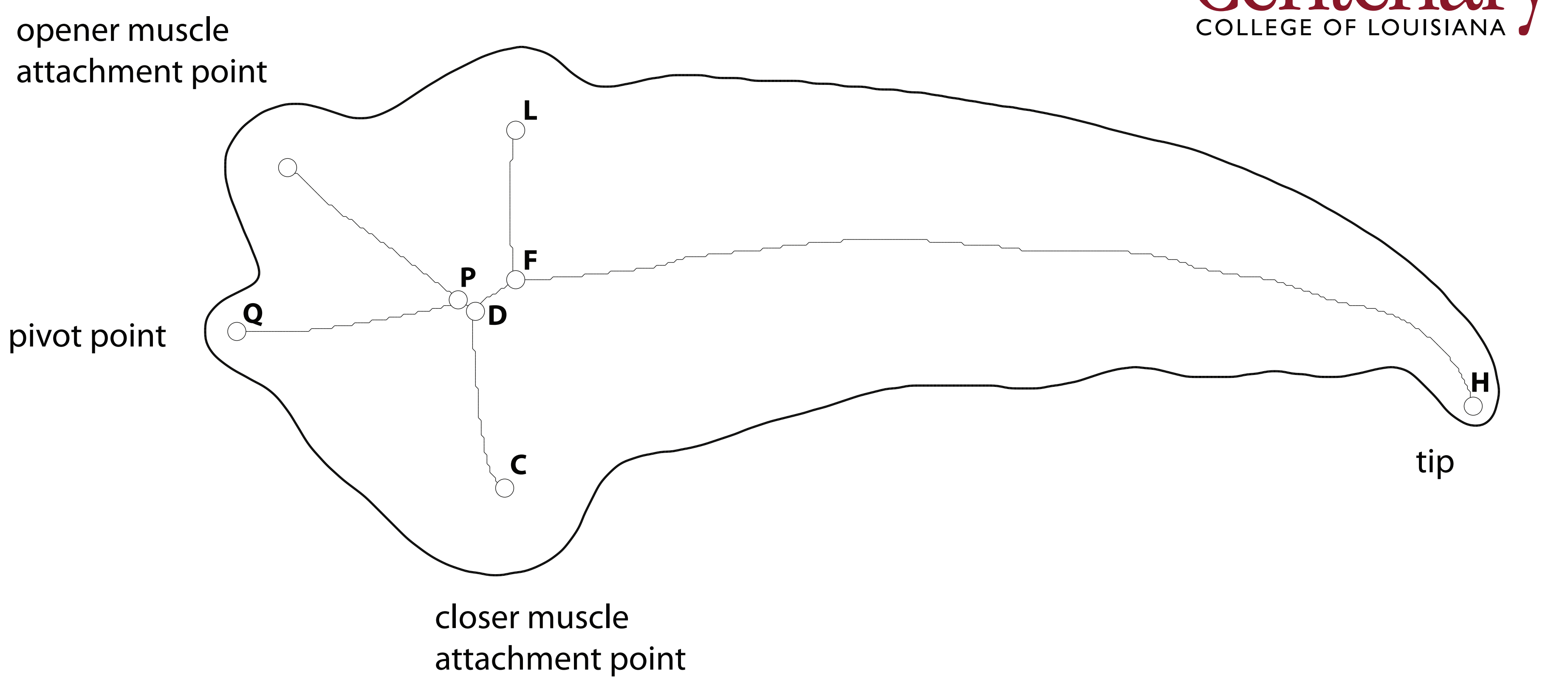
Convert Shock Graphs into First Order Logic

```
example(ex0). node(ex0n0). hasNode(ex0, ex0n0).
nodePoint(ex0n0, 163, 85). node(ex0n5). hasNode(ex0, ex0n5).
edge(ex0e0). hasEdge(ex0, ex0e0, ex0n0, ex0n8). ......
```

Learn Logical Rules with Inductive Logic Programming

```
eriphia(A):-
  hasEdge(A,B,C,D), hasEdge(A,E,D,F), hasEdge(A,G,F,H),
  distance(D,H,I), distance(F,C,J), J<I, hasEdge(A,K,F,L),
  distance(L,H,M), J<M, distance(L,C,N), J<N, hasEdge(A,O,P,Q),
  distance(Q,H,R), M<R, distance(L,Q,S), J<S.
```

Centenary
COLLEGE OF LOUISIANA



Sample rule learned with Aleph, showing one possible match of the nodes when applied to an Eriphia gonagra example.

Our dataset for this study consists of 38 dactyl images, 12 belonging to *Eriphia gonagra* and 26 belonging to *Menippe mercenaria*. We divided the data of 38 examples repeatedly into five training and testing sets, distributing the positive and negative examples separately to ensure a distribution comparable to the complete dataset. We were able to achieve an average of 89.4% accuracy in separating the two species.

Inductive Logic Programming is the process of learning in first-order logic to correctly categorize training set data with hopefully high generalization on the testing set. We generate rules by picking a random example and heuristically searching through the space of possible rules until the best rule is found or time runs out. We then repeat, learning on unexplained examples until enough rules are learned to explain (almost) all of the positive training examples.

The sample rule shown captures all of the positive *Eriphia gonagra* examples and none of the negative examples, in both the training set and testing set. It includes a sequence of connected nodes, C to D to F to H, where the distance between nodes C and F, called J, is less than other calculated distances in this rule. When this rule is applied to the positive examples, as seen in Figure 2, node C frequently corresponds to the closer muscle insertion point and H to the tip point.

Our current dataset is quite small, with test folds having only 2 or 3 positive examples. We plan to further investigate this approach with a larger dataset consisting of 970 major and minor dactyls from nine xanthoid crab species. This dataset will allow us to evaluate whether this method can be used to distinguish dactyls of several closely related species. Also, because many of the dactyls of these species change shape with growth, we can quantify those allometric transformations and identify dactyl sizes where species level differences emerge.