Using Mapper to Reveal Morphological Relationships in Passiflora Leaves

Sarah Percival*, Daniel Chitwood[†], and Aman Husbands [‡]
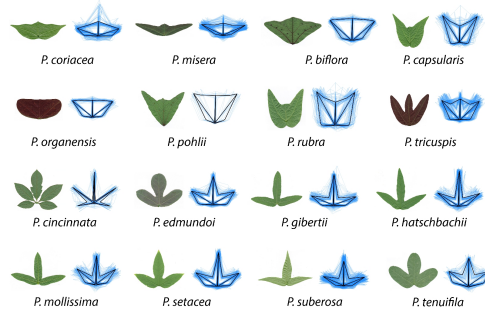
* Department of Biochemistry and Molecular Biology, Michigan State University
[†] Department of Horticulture and Department of Computational Math, Science, and Engineering, Michigan State University
[‡] Department of Biology, University of Pennsylvania

MICHIGAN STATE UNIVERSITY

BMB

Passiflora

Over 550 species of Passiflora are found among warm climates in the Americas, Asia, and Oceania, with new species continuing to be identified. Passiflora exhibit some of the most leaf shape diversity among plants, which helps disguise them from butterflies. In addition to variation across species, leaf shape varies on each individual plant across time.


P. coriacea   *P. misera*   *P. biflora*   *P. capsularis*
P. organensis   *P. pohlii*   *P. rubra*   *P. tricuspis*
P. cincinnata   *P. edmundoi*   *P. gibertii*   *P. hatschbachii*
P. mollissima   *P. setacea*   *P. suberosa*   *P. tenuifila*

The shapes of Passiflora leaves measured using landmarks.

Our dataset consists of 3,319 leaves from 40 different Passiflora species. Each data point consists of 15 Procrustes-aligned (x,y)-coordinate pairs representing 15 landmarks on each leaf. In this project, we use Mapper to visualize the evolutionary and developmental structure of leaf shape.

The Mapper Graph

Given an input set X of point cloud data, the Mapper algorithm works as follows:

1. Choose a lens function $f: X \to \mathbb{R}$ that assigns to each point in X a real number.
2. Cover $\mathbb{R}$ with a set of overlapping intervals $\{U_\alpha\}$.
3. Cluster the points inside each $f^{-1}(U_\alpha)$.
4. Create a node in the Mapper graph for each cluster obtained in Step 3, joining two nodes with an edge if their corresponding clusters share a data point.
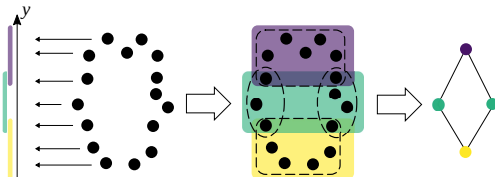

The Mapper algorithm

The Mapper algorithm depends on the choice of lens function, distance metric, covering, and clustering algorithm.

- **Lens function:** A function $f: X \to \mathbb{R}$, often informed by prior knowledge of the data set.
- **Distance metric:** In this project, we use correlation distance.
- **Covering:** Can be arbitrary, but usually determined by the number of intervals and by how much they overlap.
- **Clustering algorithm:** In this project, we use DBSCAN since it does not require the user to determine the number of clusters a priori.

Principal Component Analysis

Traditional dimension reduction algorithms, such as Principal Component Analysis, show separation into distinct morphotypes, but overlapping points make the visualization difficult to interpret.
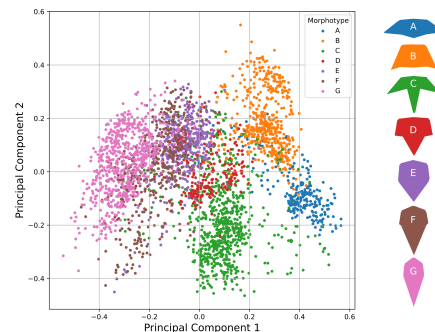
PCA plot with points colored by morphotype

Mapper Reveals Additional Structure

We construct a Mapper graph using a lens function that assigns to each point its first principal component value. The overall structure and relationships resemble the PCA plot, but Mapper shows better separation between morphotypes than PCA alone does. Furthermore, the edges of the Mapper graph show relationships between data clusters missed by the PCA plot.
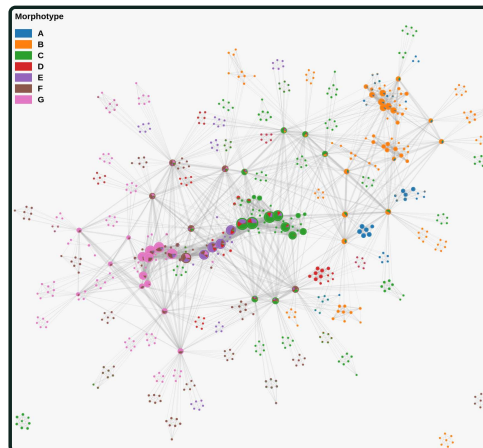
A Mapper graph with the first PCA coordinate as the lens function.

Using Mapper to Study Leaf Development

Heteroblasty refers to the change in leaf shape along a Passiflora vine.

P. edulis
P. caerulea
P. racemosa

Heteroblastic series of three different Passiflora species

Using heteroblasty as a lens reveals how morphometric differences between species relate to the developmental changes that must occur for those shapes to be realized.
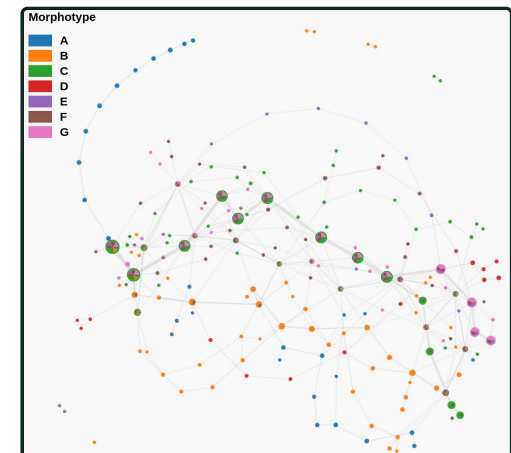
A Mapper graph with heteroblasty as the lens function.

The long branches in the Mapper graph show the linear progression of aging. Many of these branches connect to the central spine, revealing developmental similarities across species.

References

Chitwood, D. H. and Otoni, W. C., "Morphometric analysis of Passiflora leaves: the relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade," *GigaScience* **6** (01 2017).

Singh, G., Memoli, F., & Carlsson, G. "Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition" in [*Eurographics Symposium on Point-Based Graphics*], Botsch, M., Pajarola, R., Chen, B., and Zwicker, M., eds., The Eurographics Association (2007)