Using Mapper to Reveal Morphological Relationships in Passiflora Leaves

Sarah Percival, Erik Amézquita, Sourabh Palande, Aman Husbands, Arjun Krishnan, Beronda Montgomery, Elizabeth Munch and Daniel Chitwood

Michigan State University

May 19, 2022

Morphometric analysis of Passiflora leaves

Passiflora

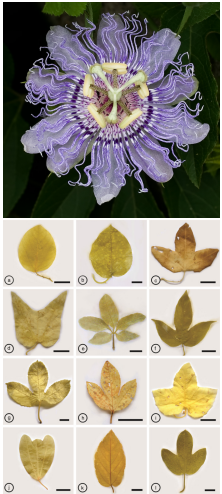
- Over 550 species of Passiflora are found among warm climates in the Americas, Asia, and Oceania, with new species continuing to be identified.
- Most species produce edible fruit.
- Passiflora exhibit some of the most leaf shape diversity among plants, which helps disguise them from butterflies.
-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

Image credit: Quillia (2008), Wosch et al. (2015)

Passiflora

- Over 550 species of Passiflora are found among warm climates in the Americas, Asia, and Oceania, with new species continuing to be identified.
- Most species produce edible fruit.
- 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.

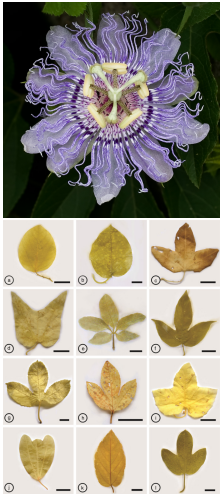
Image credit: Quillia (2008), Wosch et al. (2015)



Passiflora

- Over 550 species of Passiflora are found among warm climates in the Americas, Asia, and Oceania, with new species continuing to be identified.
- Most species produce edible fruit.
- 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.

Image credit: Quillia (2008), Wosch et al. (2015)



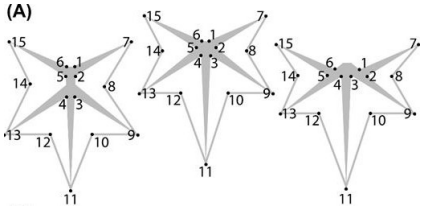
Passiflora

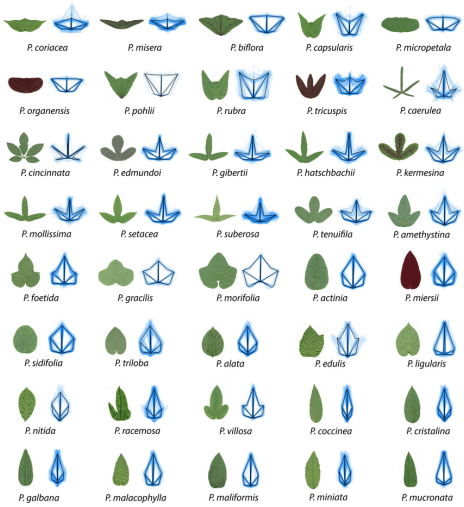
- Over 550 species of Passiflora are found among warm climates in the Americas, Asia, and Oceania, with new species continuing to be identified.
- Most species produce edible fruit.
- Passiflora exhibit some of the most leaf shape diversity among plants, which helps disguise them from butterflies.
-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

Image credit: Quillia (2008), Wosch et al. (2015)

Input Data

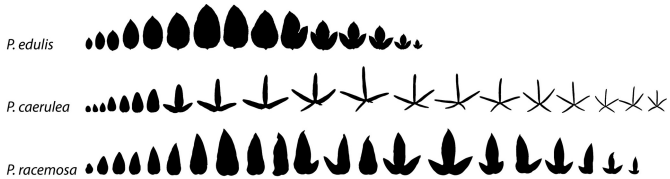
Figure: Vanderhoff (2022)





Heteroblasty

The first leaves to grow on a vine may look very different than the ones that grow later on. *Heteroblasty* refers to the morphological changes of leaves across a vine.

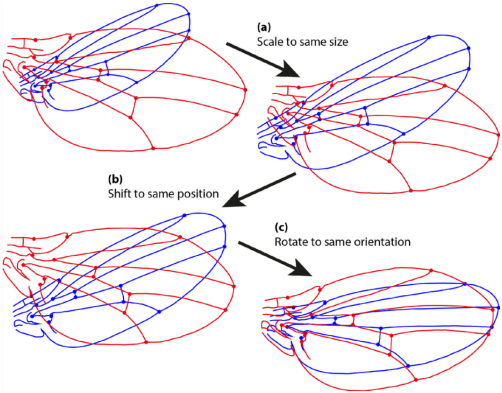


The leaves in each row come from a single vine (Chitwood and Otoni (2017a))

The first leaf on a vine might look more similar to a leaf from a different species than it does to the tenth leaf on the same vine.

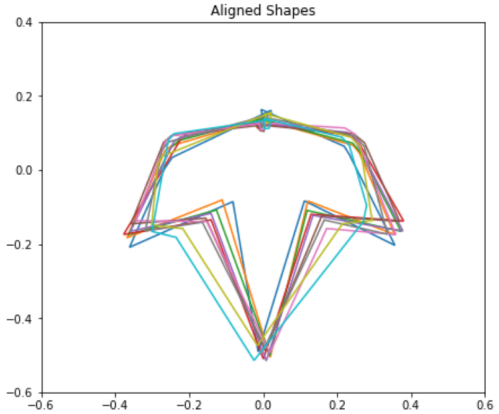
Generalized Procrustes Analysis

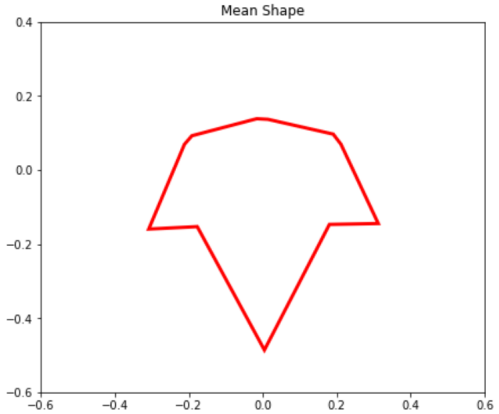
To obtain normalized leaf shape, we use Generalized Procrustes Analysis (Gower (1975)).



Procrustes alignment of fly wings (Klingenberg (2015))

Mean Leaf Shape

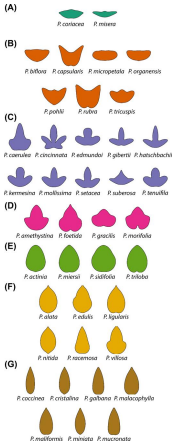




The average leaf shape of each node is computing by averaging coordinates across all leaves in the node.

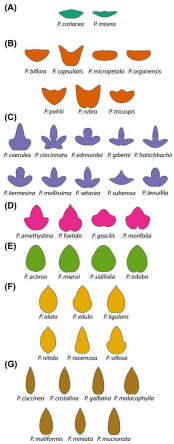
Morphotype

- To make the data manageable, Chitwood and Otoni (2017a) examined 40 *Passiflora* species and grouped each species into one of seven morphotypes.
- The morphotypes were determined based on Principal Component Analysis on leaf landmark data.
- The figure shows the average leaf shape across all leaves in each species.



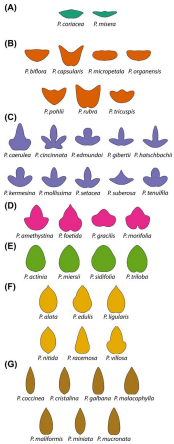
Morphotype

- To make the data manageable, Chitwood and Otoni (2017a) examined 40 *Passiflora* species and grouped each species into one of seven morphotypes.
- The morphotypes were determined based on Principal Component Analysis on leaf landmark data.
- The figure shows the average leaf shape across all leaves in each species.

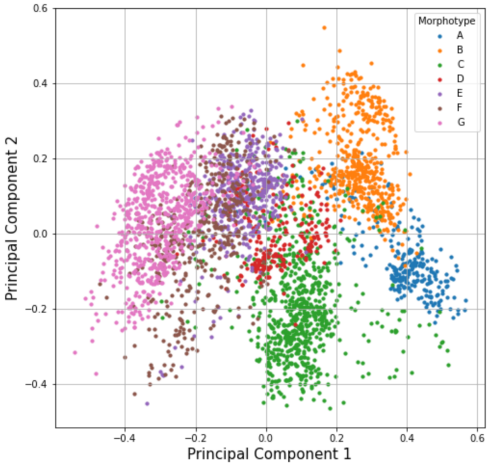


Morphotype

- To make the data manageable, Chitwood and Otoni (2017a) examined 40 *Passiflora* species and grouped each species into one of seven morphotypes.
- The morphotypes were determined based on Principal Component Analysis on leaf landmark data.
- The figure shows the average leaf shape across all leaves in each species.



Principal Component Analysis



PCA shows separation between species classes

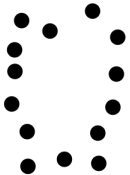
Topological Data Analysis

The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.

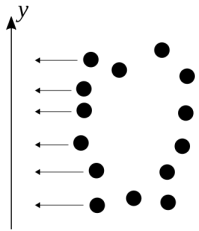
The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.



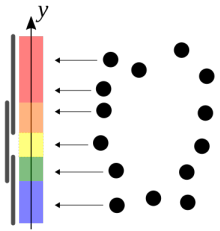
The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.



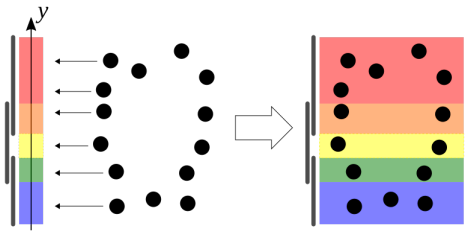
The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.



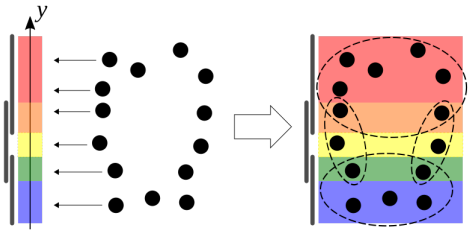
The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.



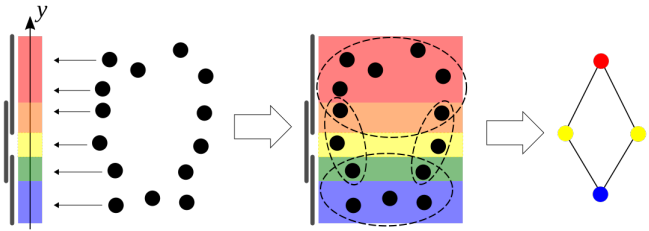
The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.



The Mapper Algorithm (Singh et al. (2007))

The Mapper algorithm produces a one-dimensional set that captures the underlying structure of the input data.



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:** There are often multiple ways to measure distance between points in a dataset. The choice of metric is often informed by what aspect of the data is being studied.
- **Covering:** Can be arbitrary, but usually determined by the number of intervals and by how much they overlap.
- **Clustering algorithm:** Many to choose from, each with their own user-selected parameters.

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:** There are often multiple ways to measure distance between points in a dataset. The choice of metric is often informed by what aspect of the data is being studied.
- **Covering:** Can be arbitrary, but usually determined by the number of intervals and by how much they overlap.
- **Clustering algorithm:** Many to choose from, each with their own user-selected parameters.

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:** There are often multiple ways to measure distance between points in a dataset. The choice of metric is often informed by what aspect of the data is being studied.
- **Covering:** Can be arbitrary, but usually determined by the number of intervals and by how much they overlap.
- **Clustering algorithm:** Many to choose from, each with their own user-selected parameters.

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:** There are often multiple ways to measure distance between points in a dataset. The choice of metric is often informed by what aspect of the data is being studied.
- **Covering:** Can be arbitrary, but usually determined by the number of intervals and by how much they overlap.
- **Clustering algorithm:** Many to choose from, each with their own user-selected parameters.

Lens Function

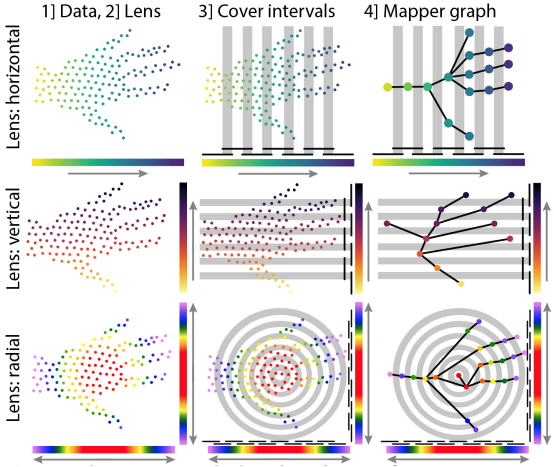


Figure: Three different lenses on the same point cloud data

Building the Mapper Graph

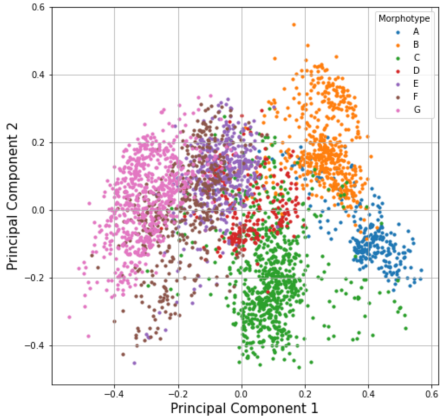
Input data: 3319 leaves from 40 different Passiflora species collected sequentially across the vine collected in Chitwood and Otoni (2017b).

Lens: We assign to each leaf its first principal component value from the PCA analysis.

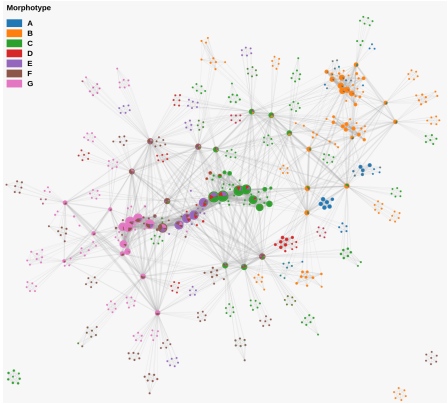
Cover: We experimented with a range of covers, choosing one that resulted in the most interesting graph.

Clustering algorithm: We use the DBSCAN clustering algorithm because it does not require the user to determine the number of clusters a priori, and is robust to outliers.

Comparing Mapper to PCA



2 component PCA

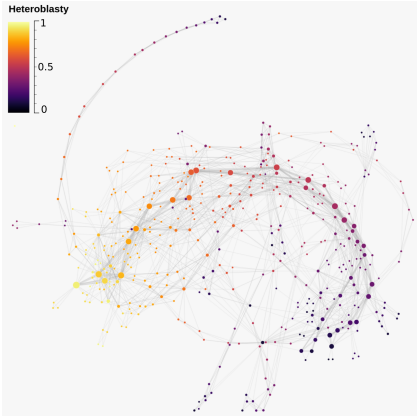


Mapper graph with 1st principal component lens

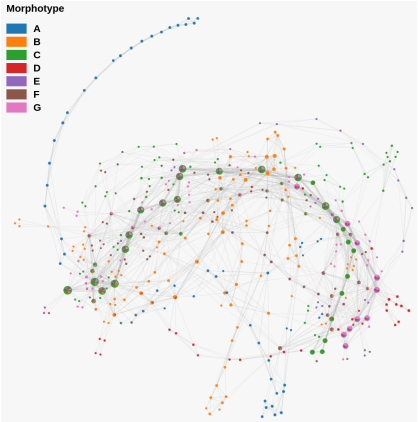
The links in the Mapper graph show additional structure the PCA plot does not.

Lens choice

Here, we assign to each leaf its normalized heteroblasty value (e.g. if a leaf is the fifth leaf out of 14 on the vine, its heteroblasty value is 5/14).

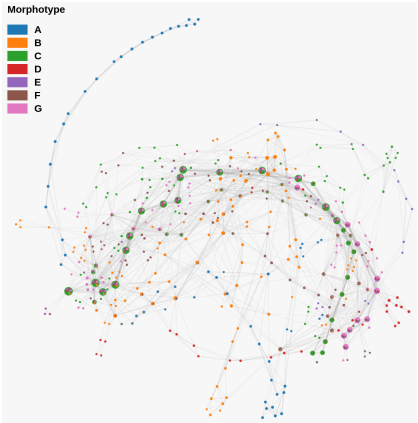


Nodes colored by heteroblasty

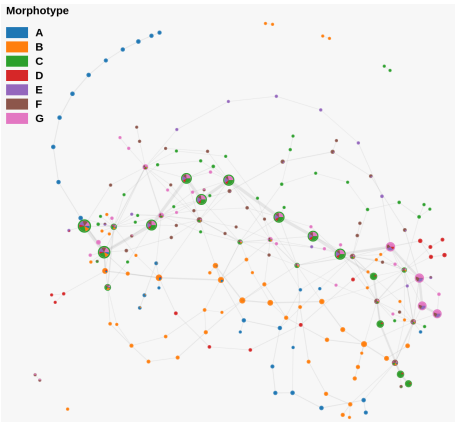


Nodes colored by morphotype

Parameter selection



26 intervals with 75% overlap



13 intervals with 50% overlap

Using Mapper

A New Mapper Visualization

`https://sperciva.github.io/giotto-mapper-pie-nodes/`

Mapper Software

- Giotto-TDA (Python) (Tauzin et al. (2020))
- Kepler Mapper (Python) (van Veen et al. (2019, 2020))
- TDAmapper (R) (Pearson et al. (2015))

Using Mapper

A New Mapper Visualization

`https://sperciva.github.io/giotto-mapper-pie-nodes/`

Mapper Software

- Giotto-TDA (Python) (Tauzin et al. (2020))
- Kepler Mapper (Python) (van Veen et al. (2019, 2020))
- TDAmapper (R) (Pearson et al. (2015))

Using Mapper

A New Mapper Visualization

`https://sperciva.github.io/giotto-mapper-pie-nodes/`

Mapper Software

- Giotto-TDA (Python) (Tauzin et al. (2020))
- Kepler Mapper (Python) (van Veen et al. (2019, 2020))
- TDAmapper (R) (Pearson et al. (2015))

Using Mapper

A New Mapper Visualization

`https://sperciva.github.io/giotto-mapper-pie-nodes/`

Mapper Software

- Giotto-TDA (Python) (Tauzin et al. (2020))
- Kepler Mapper (Python) (van Veen et al. (2019, 2020))
- TDAmapper (R) (Pearson et al. (2015))

Using Mapper

A New Mapper Visualization

`https://sperciva.github.io/giotto-mapper-pie-nodes/`

Mapper Software

- Giotto-TDA (Python) (Tauzin et al. (2020))
- Kepler Mapper (Python) (van Veen et al. (2019, 2020))
- TDAmapper (R) (Pearson et al. (2015))

Chitwood, D. H. and Otoni, W. C. (2017a). Divergent leaf shapes among passiflora species arise from a shared juvenile morphology. *Plant Direct*, 1(5).

Chitwood, D. H. and Otoni, W. C. (2017b). Morphometric analysis of Passiflora leaves: the relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade. *GigaScience*, 6(1).

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1):33–51.

Klingenberg, C. P. (2015). Analyzing fluctuating asymmetry with geometric morphometrics: Concepts, methods, and applications. *Symmetry*, 7(2):843–934.

Pearson, P., Muellner, D., and Singh, G. (2015). *TDAmapper: Analyze High-Dimensional Data Using Discrete Morse Theory*. R package version 1.0.

Quillia, O. P. (2008). Oq passion flower.jpg.

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

Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A., Dassatti, A., and Hess, K. (2020). giotto-tda: A topological data analysis toolkit for machine learning and data exploration.

van Veen, H. J., Saul, N., Eargle, D., and Mangham, S. W. (2019). Kepler mapper: A flexible python implementation of the mapper algorithm. *Journal of Open Source Software*, 4(42):1315.

van Veen, H. J., Saul, N., Eargle, D., and Mangham, S. W. (2020). Kepler Mapper: A flexible Python implementation of the Mapper algorithm.

Vanderhoff, R. (2022). Passiflora caerulea.

Wosch, L., Imig, D. C., Cervi, A. C., Moura, B. B., Budel, J. M., and de Moraes Santos, C. A. (2015). Comparative study of passiflora taxa leaves: I. a morpho-anatomic profile. *Revista Brasileira de Farmacognosia*, 25(4):328–343.