

TreeScaper: Software to visualize and understand tree landscapes

Wen Huang¹ Guifang Zhou²
Jeremy Ash³ Jeremy Brown³
Kyle Gallivan² Melissa Marchand²
James Wilgenbusch¹

1, Department of Scientific Computing, FSU,

2, Department of Mathematics, FSU,

3, Department of Biological Sciences, LSU



FLORIDA STATE UNIVERSITY



Shortcomings of Current Approaches

- Consensus tree
 - Discards information concerning competing trees
- Projection into low dimensional Euclidean space (e.g., Hillis et al. 2005)
 - May be difficult to interpret



Alternative Approach

- Apply graph-based methods to understand relationships among:
 - Tree topologies
 - Bipartitions within tree topologies



TreeScaper (Version 1)

- NLDR
 - Optimization Algorithms
 - Linear iteration i.e., nonlinear relaxation
 - Majorization
 - Gauss-Newton
 - Stochastic gradient descent
 - MCMC simulated annealing
 - Cost functions
 - Kruskal-1 stress
 - Normalized stress
 - Sammon stress
 - Curvilinear components analysis
- Dimension estimators
 - Nearest neighbor estimator
 - Correlation dimension
 - Maximum likelihood estimator
- Visualization



TreeScaper (Version 2)

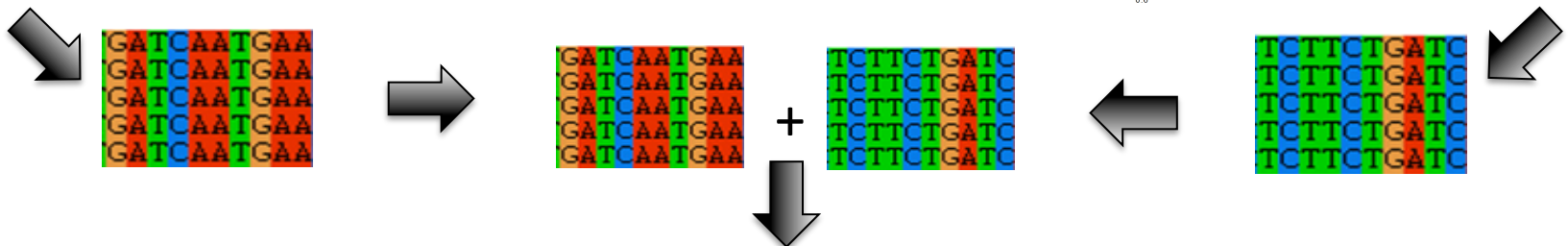
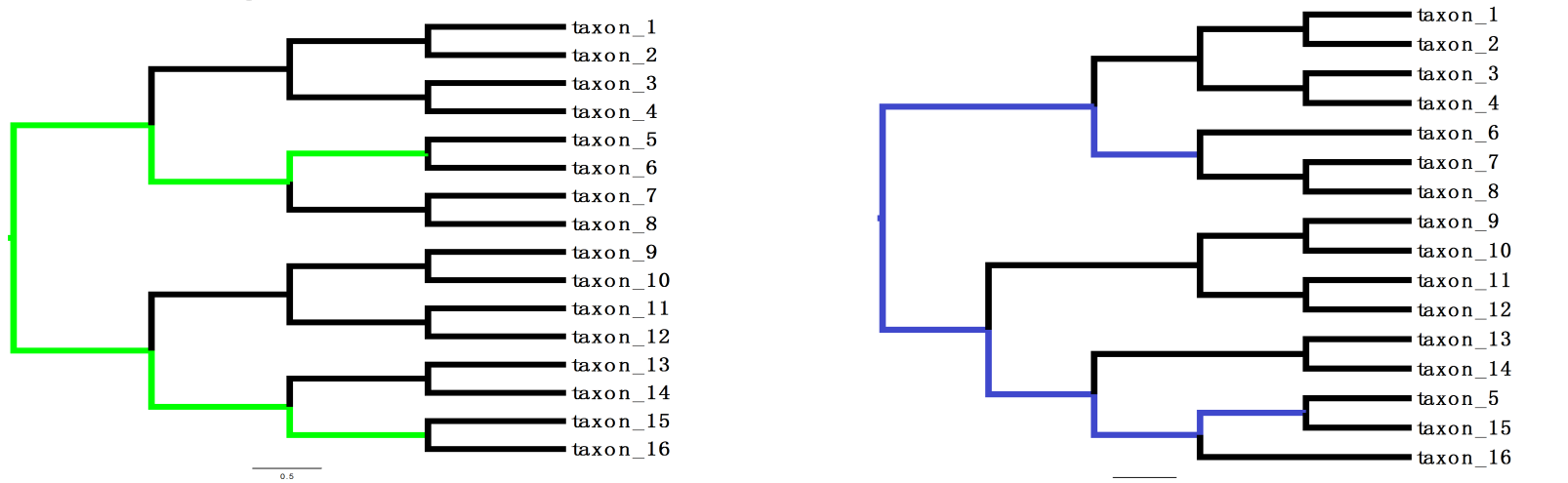
Graph-based extension

- Community detection methods
 - Configuration Null Model
 - Constant Potts Model
 - Erdos-Renyi Null Model
 - No Null Model
- New input data types
- Interactive tree viewer



Example Data

- Trees generated from two different processes



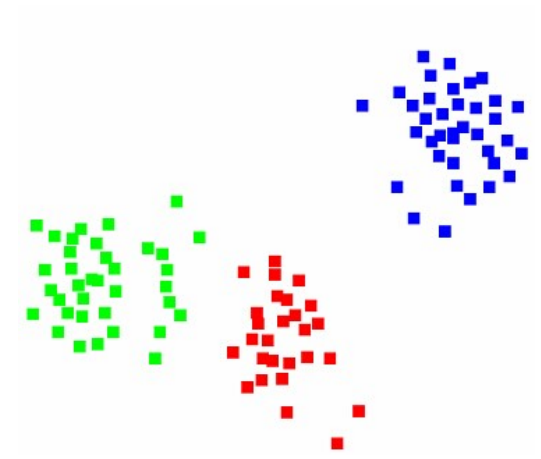
100 Trees



Affinity Matrix Analysis

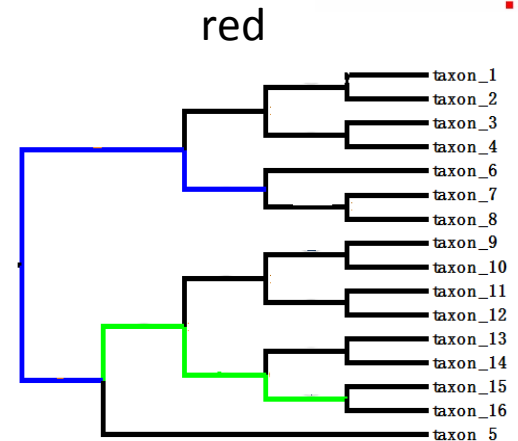
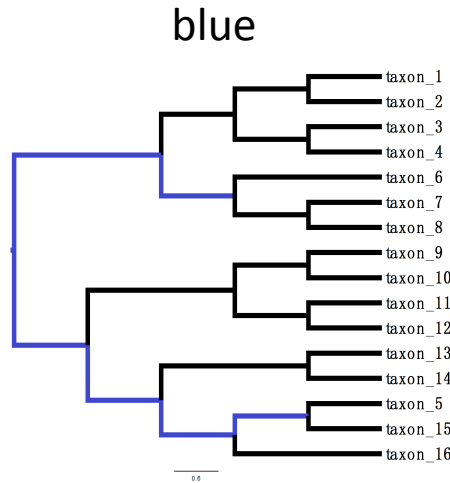
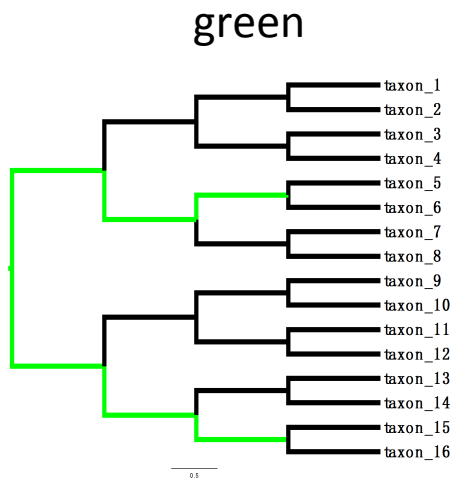
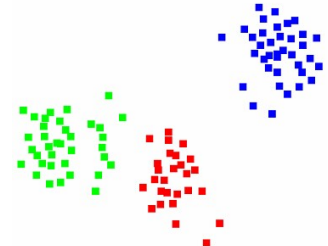
- Affinity matrix
 - Reciprocal of pairwise distances
- Detect communities
 - Discovered 3 communities
- Project data
 - Colors represent each community

Projections of trees



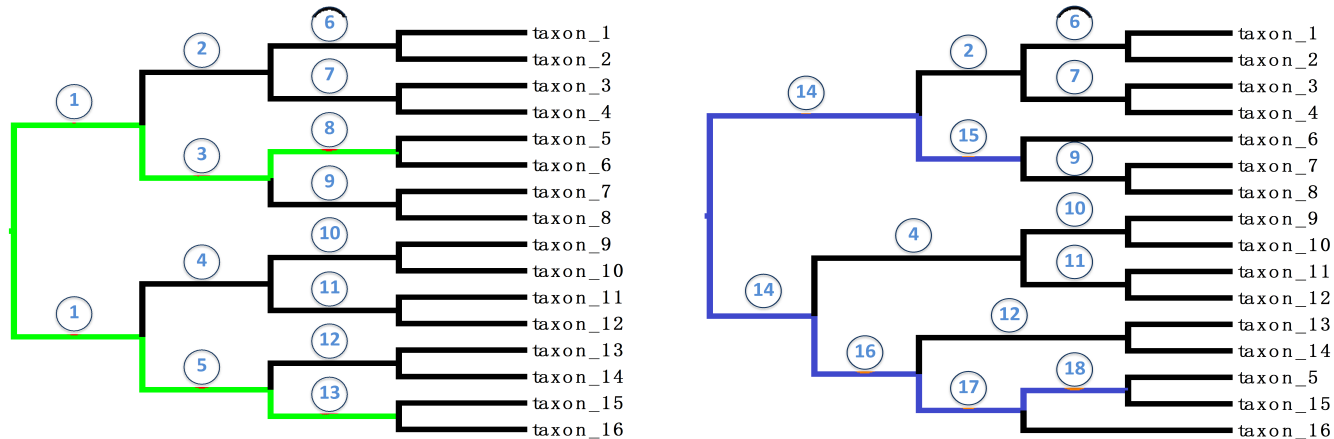
Affinity Matrix Analysis

- Consensus trees for each community
 - Two trees used to generate data are discovered
 - Extra tree is also obtained
 - Should be caused for further considerations



Bipartition Analysis

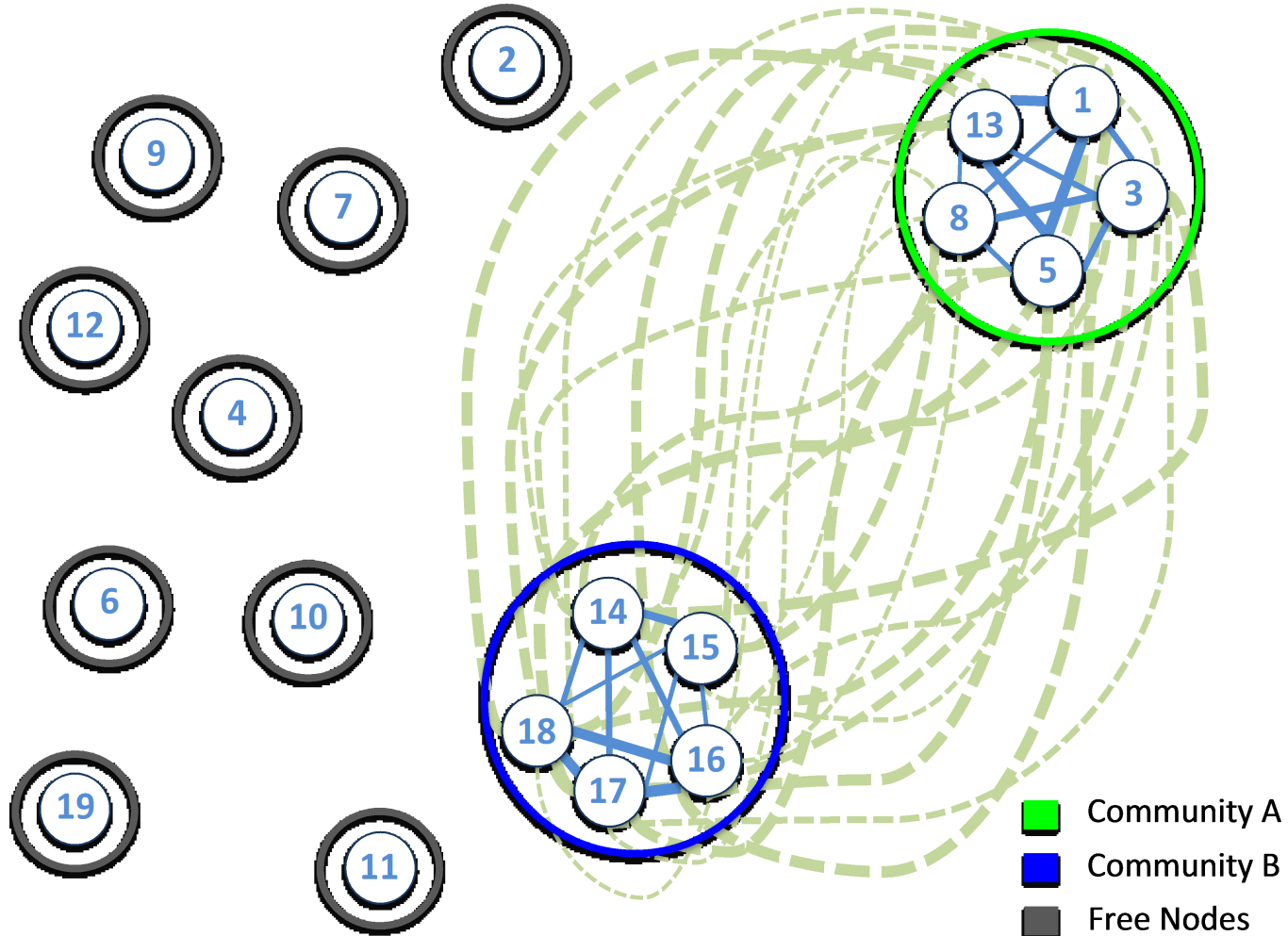
- Same data set generated by the same two input trees



- Covariance matrix based on presence or absence of bipartitions in the bootstrap trees



Bipartition Analysis



FLORIDA STATE UNIVERSITY

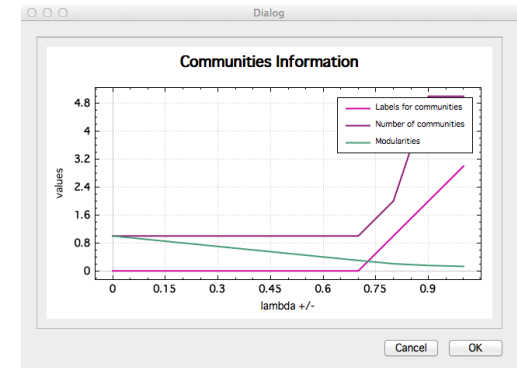
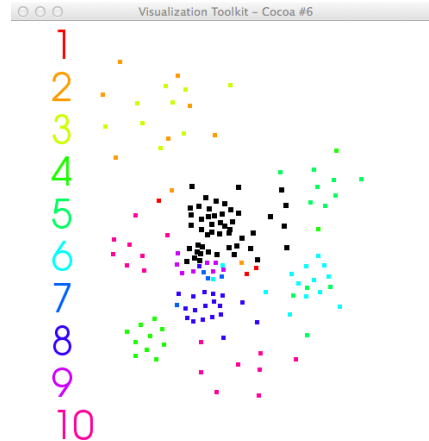
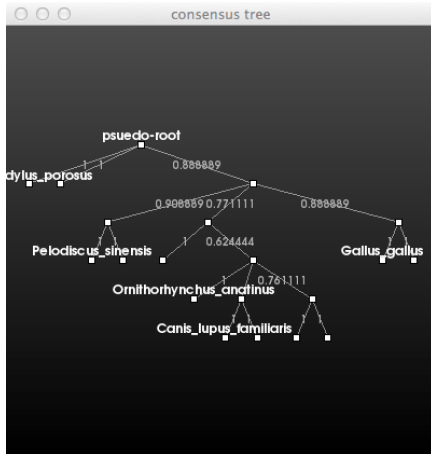
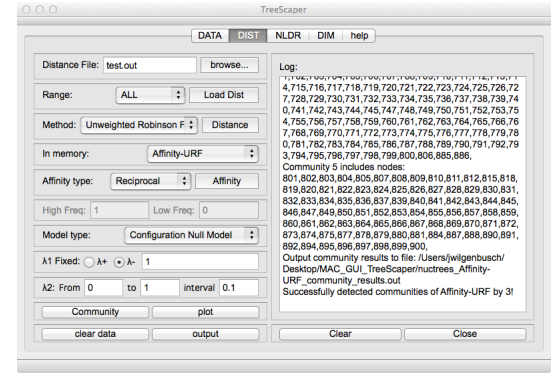
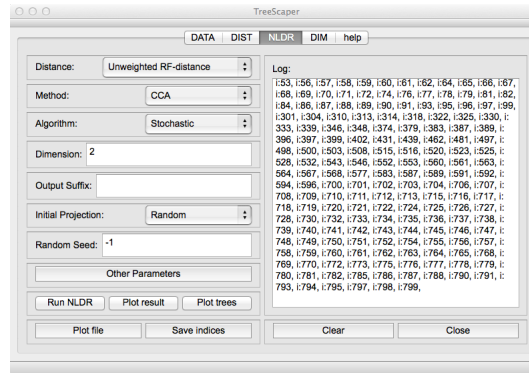
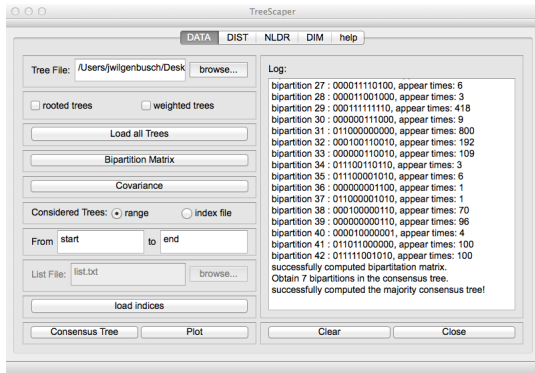


TreeScaper Functionality

- NLDR
- Dimensionality estimation
- *Distance/Affinity matrix*
- *Covariance matrix*
- *Community Detection methods*
- *Interactive visualization interface*



TreeScaper



FLORIDA STATE UNIVERSITY



Software

- Software developed in C++
- Visualization produced using Qt, CLapack and the Visualization Tool Kit (VTK)
- Available under GPL
 - ➔ <https://sourceforge.net/projects/treescaper/>



Support

- FSU's Shared High Performance facility for compute cycles and technical support
- The National Science Foundation for funding to support some of this work (ABI-1262476)



FLORIDA STATE UNIVERSITY

