Population phylogenetics with the highly clonal desert moss, *Syntrichia caninervis*

Jenna T. Baughman

University and Jepson Herbaria and Integrative Biology
University of California, Berkeley

IB 200 Final Project
May 9, 2016

Outline

1 Background
- System
- Hypotheses

2 Methods
- Data
- Analyses

3 Results
- Parsimony
- ML

4 Conclusions
- Next Steps

System

Mojave Desert moss *Syntrichia caninervis*

- Small, nonvascular plant
- Extremely desiccation tolerant
- The dominant plant in some parts of the Mojave
- Also occurs in arid climates around the world

System

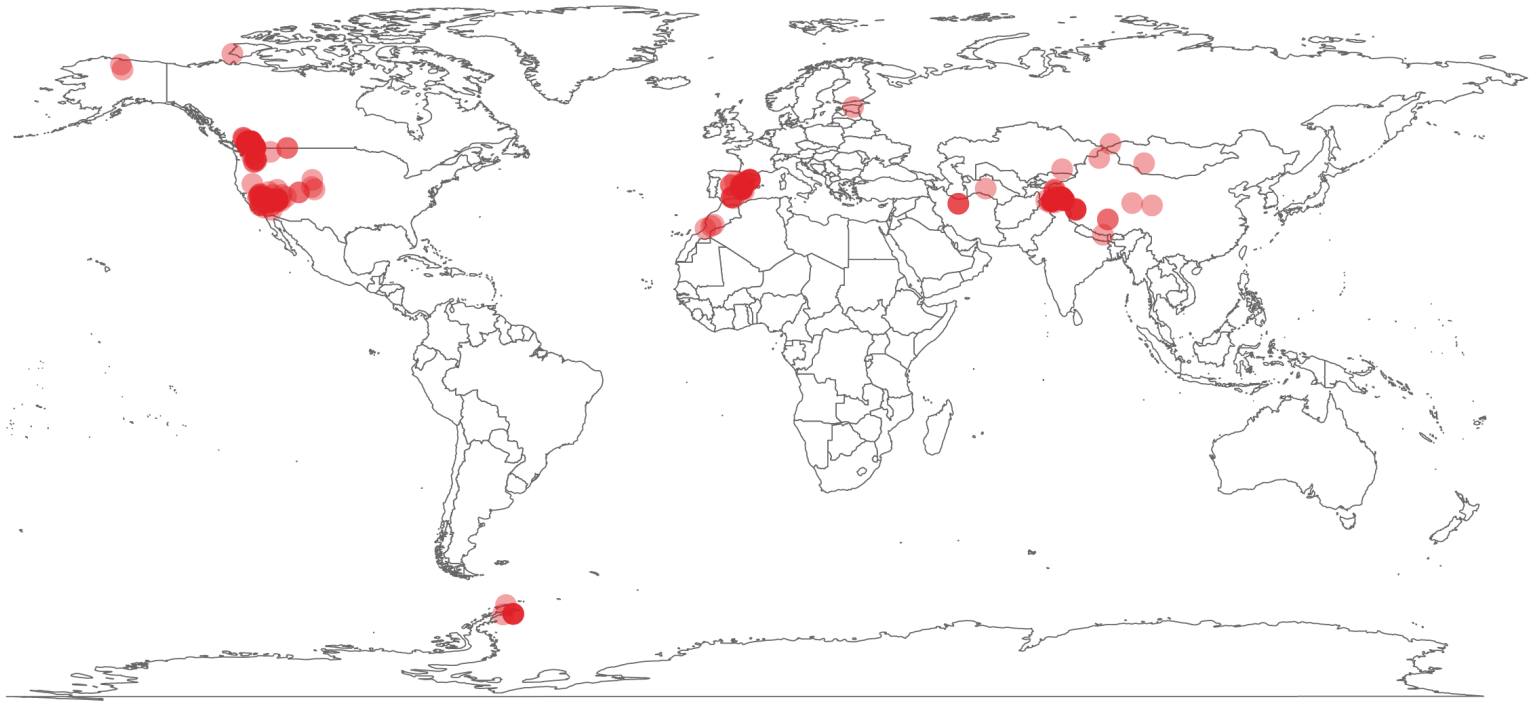


Figure 1: *S. caninervis* GBIF distribution

Research Question

Can phylogenetic methods be used with populations of a clonal organism?

Which population is "older"?

Hypotheses

- Relationships among individuals within populations will have good support values in a bifurcating tree, suggesting clonal reproduction and accumulation of mutations.
  - Low support may indicate:
    - Insufficient data
    - Past sexual reproduction and recombination
    - Gene flow
- The higher elevation population will be younger than and nested within the lower elevation population.

Sequence Data

- 131 sterile samples
  - Double-digest restriction site-associated DNA sequencing (ddRADseq); Illumina HiSeq platform
- Outgroup: *Syntrichia caninervis* transcriptome from a population in China
  - *in silico* ddRADseq

in silico ddRADseq

- Mirrored wet lab ddRADseq
  - "Digested" transcriptome with:
    - EcoRI (cuts G|AATTC)
    - MseI (cuts T|TAA)
  - Size selection; keep only fragments 250-500 bp long
  - "Sequence" (keep) only fragments that start with "AATTC"
  - Remove enzyme cut site and trim to 80 bp
- Want both strands so reverse-complement transcriptome and repeat!

Methods: Modified `Biopieces` scripts for digestion, size selection, and cut site removal. `FASTX-Toolkit` script for trimming. `BioPython` used for data cleaning and preparation. `SequenceMatrix` used to concatenate sequences.

Character Matrix

Item	Quantity
OTUs	132
Characters	136160
Informative SNPs	2492
Uninformative SNPs	133

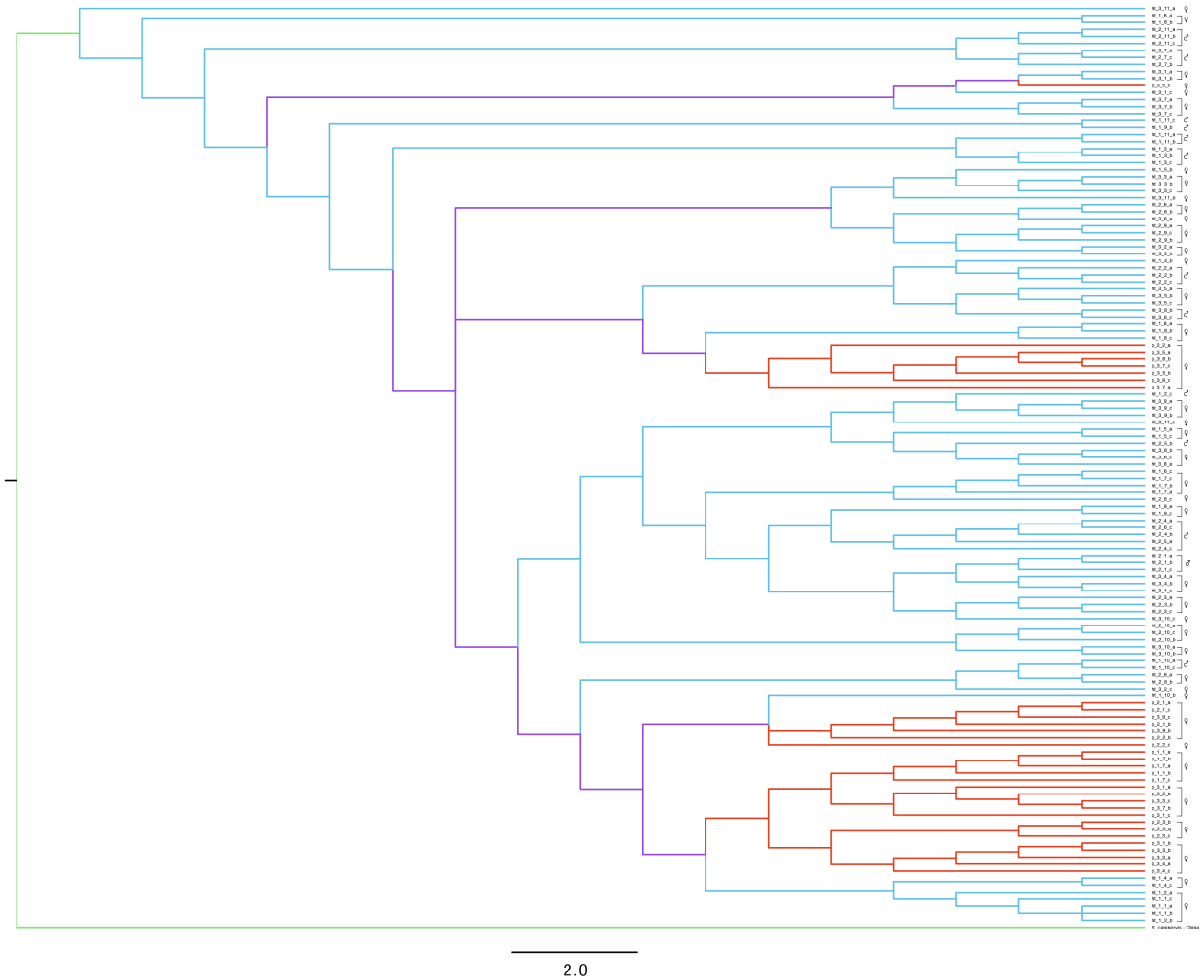
- 1702 80 bp RAD loci with 1-3 SNPs (variable nucleotides) each
  - 60 shared with *S. caninervis* transcriptome outgroup

Tree-Building

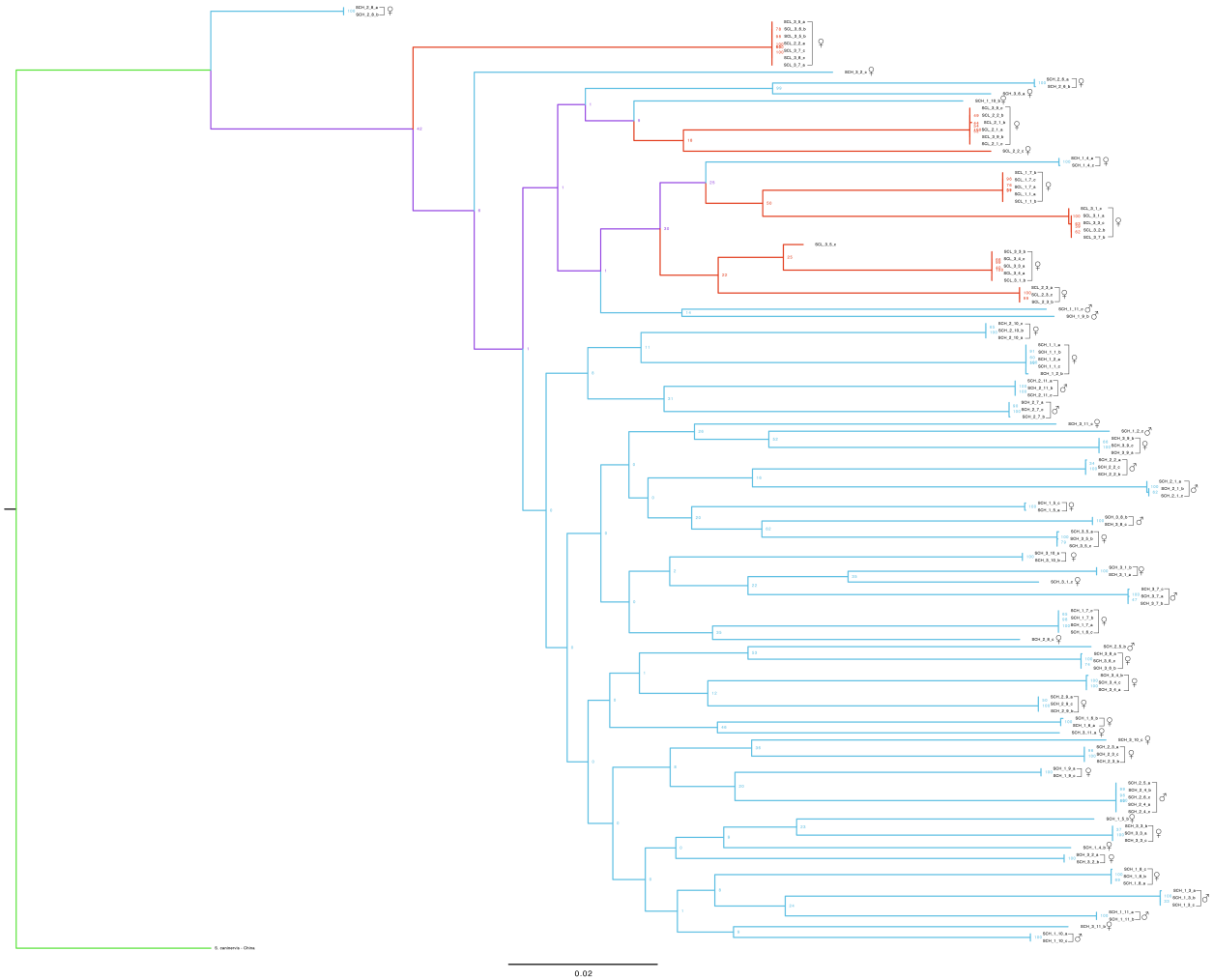
- Parsimony tree-building using Paup on Cipres Science Gateway
  - Strict consensus tree built with Paup*
- Maximum likelihood tree built with RAxML on Cipres
  - 1000 bootstraps
  - GTR GAMMA nucleotide substitution model

Parsimony Strict Consensus Tree

149 equally parsimonious trees



Maximum Likelihood Tree



IB 200

Jenna T. Baughman

Background
System Hypotheses
Methods
Data Analyses
Results
Parsimony
ML
Conclusions
Next Steps

Conclusions

- High support for clades at tips; likely clones
- Lower support for relationships deeper in tree
  - May be due to past sexual reproduction
- This approach may be more suitable for more distantly related clonal populations

Next Steps

- Find out what genes the 60 expressed loci are within.
  - Are the SNPs "silent" mutations?
- `Structure`: a population genetics analysis to determine structure among populations
- `TreeMix`: a phylogenetics approach to determine the historical relationships among populations
- `Migrate`: a coalescence model to determine presence and direction of gene flow between populations

Acknowledgements

My sincere thank you to the course graduate student instructor Will Freyman for your help in the labs and this project. Thanks to my lab mates Javi Juaregui, Caleb Caswell-Levy, Andrew Thornhill, and Sonia Nostratinia for camaraderie and assistance. Last but not least, thank you to the course professors Kip Will, David Ackerly, and my very own advisor Brent Mishler for a great class!

References


THE UNIVERSITY OF CALIFORNIA BERKELEY
LET THERE BE LIGHT
1868