A population phylogenetics case study with *in silico* ddRADseq and the highly clonal desert moss, *Syntrichia caninervis* (Pottiaceae)

Jenna T. Baughman

IB 200 Final Project

Spring 2016

University and Jepson Herbaria

and

Integrative Biology

University of California, Berkeley

Contents

List of Figures

List of Tables

1 Abstract

Populations of clonal organisms may be "tree-like," in that they can be represented accurately with a bifurcating tree. This study aimed to explore the use of phylogenetic analyses on populations of highly clonal organisms, with individual branches of the desert moss *Syntrichia caninervis* as the OTUs. Clonal organisms (even with some sexual reproduction) have processes analogous to those at higher levels: the process of cloning is analogous to speciation or splitting while sexual reproduction is analogous to introgression or hybridization. This project also used a novel *in silico* double digest restriction site-associated DNA sequencing (ddRADseq) technique that utilizes existing transcriptomic data for comparison with original ddRADseq data. I aimed to test the prediction that relationships among individuals within populations will have good support values in a bifurcating tree, suggesting clonal reproduction and accumulation of mutations. Further, this research aimed to test the hypothesis that the higher elevation population of Sheep Creek Wash (SC) will be nested within the lower elevation population suggesting a younger age and a shift upward in elevation. Using a sequence matrix of 136,160 coding nucleotide characters and 131 Mojave Desert and 1 Chinese Gurbantunggut Desert chimera as an outgroup, parsimony and maximum likelihood (ML) trees were built. These results showed polytomies in a strict consensus parsimony tree and low bootstrap support along the spine of the ML tree, indicating the data is not "tree-like" and is not represented well with a bifurcating tree. Both the parsimony consensus tree and the ML trees placed individuals from the higher elevation population (SCH) closer to the root and sister to the rest of SC, which does not support the hypothesis that SCH would be nested within the lower elevation population. Further, the software `TreeMix` (Pickrell and Pritchard 2012) was used to assess migration history and found support for one migration event to SCL. In sum, population history and relationships are inconclusive for all but the clades at the tips, which support recent clonal relationships. Together these results indicate the potential utility in population phylogenetics for inferring population histories of clonal organisms but the need for robust data, broad sampling, and the proper tools to detect introgression and other deviations from a bifurcating tree. This project represents a novel use for existing data and an *in silico* ddRADseq comparative analysis that can be applied to other, larger data sets.

2 Introduction

2.1 Background

The Mojave Desert moss *Syntrichia caninervis* is the dominant species of the Mojave Desert biological soil crust community (Bowker et al. 2000; Stark, McLetchie, et al. 2005). Mosses are bryophytes (small, non-vascular plants) that lack the complex water transport tissues of vascular plants and whose tissues quickly equilibrate to ambient water content, a trait known as poikilohydry (Mishler 2001). Because of these characteristics, mosses often live in wet or damp environments. Mojave Desert *S. caninervis*, however, is extremely desiccation-tolerant and spends much of its life in a dry, suspended animation state where all its biological functions are limited to infrequent post-rainfall periods, primarily in winter months (Stark 1997; Stark et al. 1998). *Syntrichia caninervis* is dioicous and has distinct male and female individuals, as well as sterile individuals that do not produce gametangia (sex organs). Dioicous mosses and can reproduce both sexually, via water-dependent swimming sperm (Mishler 2001), and asexually via vegetative cloning. Energetic trade-offs may play a role in determining an individual's reproductive strategy at a given time, and male and female life histories may experience different trade-offs. For example, in the dioicous moss *Ceratodon purpureus*, males experience a trade-off between production of vegetative and reproductive tissues (McDaniel 2005). Especially under stressful conditions, individuals may shift allocation of energy into the mode of reproduction that is less energetically costly.

Syntrichia caninervis individuals of both sexes may have the ability to pursue reproductive strategies that maximize their energy usage. Asexual propagation allows mosses to multiply and disperse, within and between populations, without the energetic costs of sexual reproduction (A. E. Newton and Mishler 1994) and females have some asexual fitness advantages over males. Females produce twice the number of protonema (gametophyte precursor) and shoots from detached leaf tissue than males do under both cool conditions and desiccation stress (Stark et al. 1998). Female *S. caninervis* individuals can also abort fertilized sporophytes, presumably to reduce further energy allotment to the sporophyte, and do so at a frequency of 0.64 even in a highly sporophytic population (Stark et al. 2000). The different reproductive costs and strategies for males and females in this species may indicate an evolutionary conflict between the sexes. Since gametophytes are capable of multiple rounds of reproduction, both sexual and asexual, theory predicts a sexual conflict over energy allocation from the female to offspring sporophytes (Haig and Wilczek 2006).

As a perennial plant, *S. caninervis* is not unusual in its ability to reproduce both

sexually and asexually (Richards 1986). However, the influence of moss reproductive modes on population structure and genetic diversity is not well understood, as a growing body of research has produced conflicting results and the dominant mode of reproduction varies among species and even populations (Cronberg et al. 2006; Mishler 1988). One might hypothesize that populations persisting through vegetative clonal growth alone (or intra-gametophytic selfing in hermaphroditic species) would have lower genotypic variation than those with primarily sexual reproduction, since there would be a large number of clones and therefore reduced number of genetic individuals and reduced gene recombination. Furthermore, as genetic individuals clone and accrue mutations, the clonal lineages may form clades with synapomorphies and plesiomorphies as evidence of their shared histories. Yet while mutations may contribute to genetic diversity in clonal populations, the dominant haploidy of mosses may decrease the potential for that diversity as haploid organisms are more susceptible to purifying selection (M. E. Newton 1988). Thus, genetic variation might be expected to be low in dominantly clonal populations. Indeed, several moss population diversity studies have found low genetic diversity in clonal populations (Cronberg 1996; Cronberg 2002; Cronberg 2003; Karlin, Andrus, et al. 2011; Shaw and Srodon 1995). Still, others have found clonal populations to maintain comparable genetic diversity to highly sexual populations (Akiyama 1999; Cronberg 2002; Cronberg 2003; Gunnarsson et al. 2005; Karlin, Hotchkiss, et al. 2012; H. Korpelainen et al. 2013; Paasch et al. 2015; Spagnuolo et al. 2007; Zouhair et al. 2001). Clonal populations with variable modes of reproduction may be able to retain high levels of genetic diversity, such as from previous sexual reproduction, under periods of reduced or absent sexual reproduction.

While the present research focuses on Mojave Desert *S. caninervis*, this species is also known to occur in arid climates around the world (see Figure 1 for worldwide distribution). However, the relationships among populations around the world remain unknown as no comprehensive comparative or phylogenetic analyses of the *S. caninervis* complex has been performed to date. In fact, some moss population studies have found significant gene flow across broad geographic ranges (Helena Korpelainen et al. 2012; McDaniel and Shaw 2005; Shaw, Golinski, et al. 2014; Vanderpoorten et al. 2008). In some cases, asexual propagation, such as from wind-dispersed vegetative fragments, can provide sufficient gene flow to maintain genetic diversity (Cronberg 2002). Still others have found significant genetic structure over relatively small geographic areas (Hutsemékers et al. 2010; Patiño et al. 2010; Pisa et al. 2013; Wang et al. 2012). Previous research on population structure and gene flow has only shown that there is no single pattern for mosses and each system needs to be

studied in its own right to fully understand the history of relationships within it.

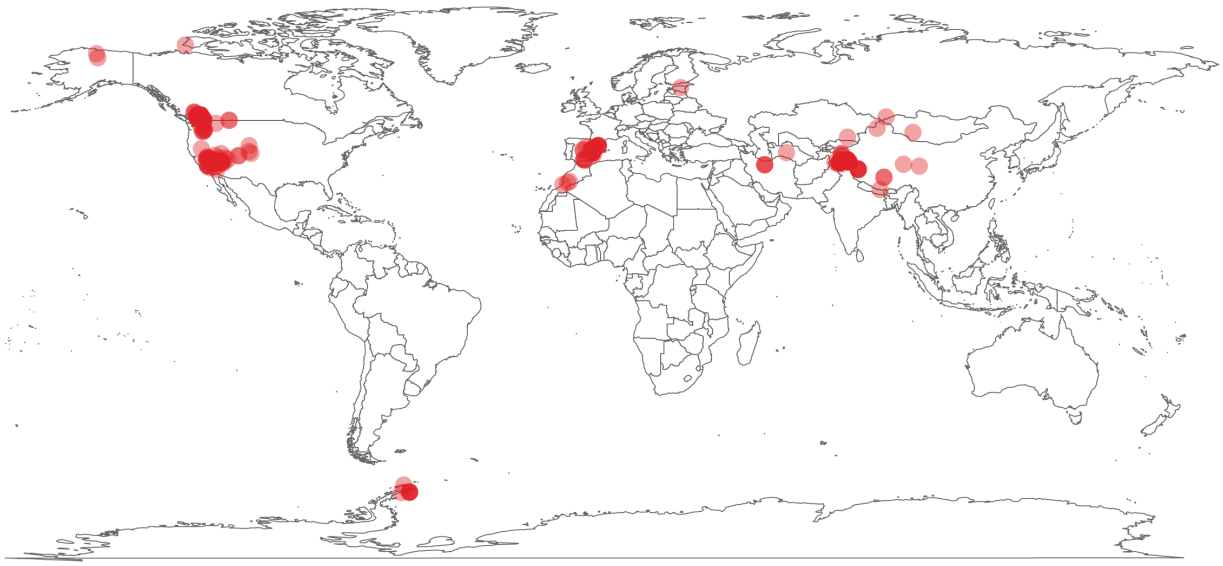

Figure 1: Worldwide occurrence of *S. caninervis* from the Global Biodiversity Information Facility accessed through the `R` (R Core Team 2016) package `rgbif` (Chamberlain et al. 2014) and plotted with the `R` package `ggplot2` (Wickham 2009).

The effect of environment on moss genetic structure and diversity is also unclear (Hutsemékers et al. 2010; Helena Korpelainen et al. 2012; Patiño et al. 2010; Pisa et al. 2013). However, previous research on the reproductive ecology of *S. caninervis* has revealed patterns of gametangial expression and sexual reproduction across gradients of environmental stress. Environmental variation that corresponds with changes in timing and duration of the biologically active period appears to affect overall levels of sex expression in a population. A survey of 890 *S. caninervis* individuals from a 10 hectare elevation gradient found that total percentage of expressing individuals increases with elevation (Bowker et al. 2000), and that male sex expression only occurs at the higher elevations, with lower elevation populations containing few expressing females (Bowker et al. 2000). In parallel with low levels of sex expression, sexual fertilization and production of sporophytes is relatively rare, and established desert S. caninervis populations seem to persist through vegetative cloning (Paasch et al. 2015). Recently, reliable sex-associated markers were used to find that genetic sex ratios in Mojave Desert *S. caninervis* populations are also female biased, with some populations entirely genetically female and sterile (Baughman 2015; manuscript in preparation). Further investigation of moss genetic diversity, particularly in understudied environments such as deserts, over large and small geographic ranges is needed in order to more fully understand the influence of life history and environment on genetic structure and relationships within

and among populations.

2.2 Research Question & Hypotheses

This study aim to explore the use of phylogenetic analyses on populations of clonal organisms, with individual branches as the semaphoront OTUs. In this system, it is impossible to tell the limits of genetic individual as they may occur mixed within a small patch or a single genotype covering a large area. Thus, branches are suitable as known potentially independent units, capable of both sexual and asexual reproduction. Clonal organisms (even with some sexual reproduction) have processes analogous to those at higher levels: the process of cloning is analogous to speciation or splitting while sexual reproduction is analogous to introgression or hybridization. So long as there is significant clonal replication with accumulation of new mutations, phylogenetic techniques may be applicable to the level of individuals within populations. This project also uses a novel *in silico* double digest restriction site-associated DNA sequencing (ddRADseq) technique that utilizes existing transcriptomic data for comparison with original ddRADseq data. Upon successful application of these analyses, this study aims to uncover the population history of two Mojave Desert populations of *S. caninervis*. Specifically, is there a pattern of one population being older than the other? Can I infer a direction of population shifting, if any? These questions lead to two distinct hypotheses:

- Relationships among individuals within populations will have good support values in a bifurcating tree, suggesting clonal reproduction and accumulation of mutations. Conversely, low support on the tree may indicate insufficient data, past sexual reproduction and recombination, or gene flow from other populations.
- The higher elevation population of Sheep Creek Wash will be nested within the lower elevation population suggesting a younger age and a shift upward in elevation.

3 Methods

3.1 Data Acquisition

3.1.1 Sampling

Samples (n = 131) were collected previously from two geographically distinct Mojave Desert sites where each site is considered a 'population': One at a higher elevation with

lower environmental stress and the other at a lower elevation with higher environmental stress. Sheep Creek High (SCH) is located in Sheep Creek Wash (SC) near Wrightwood, CA (34° 22' 33.85" N, 117° 36' 34.59" W), at the west end of the Mojave Desert and the northern base of the San Gabriel Mountains. At 1800 m elevation, this site is the highest known elevation of S. caninervis in the Mojave Desert. SCH is located at the edge of the Angeles National and is thus in a more protected location than the low elevation site. The average high and low annual temperatures are 16.8° C and 1.7° C, respectively, with an average annual precipitation of 49.4 cm (2007-2011, Wrightwood Weather Station, NOAA National Climatic Data Center).

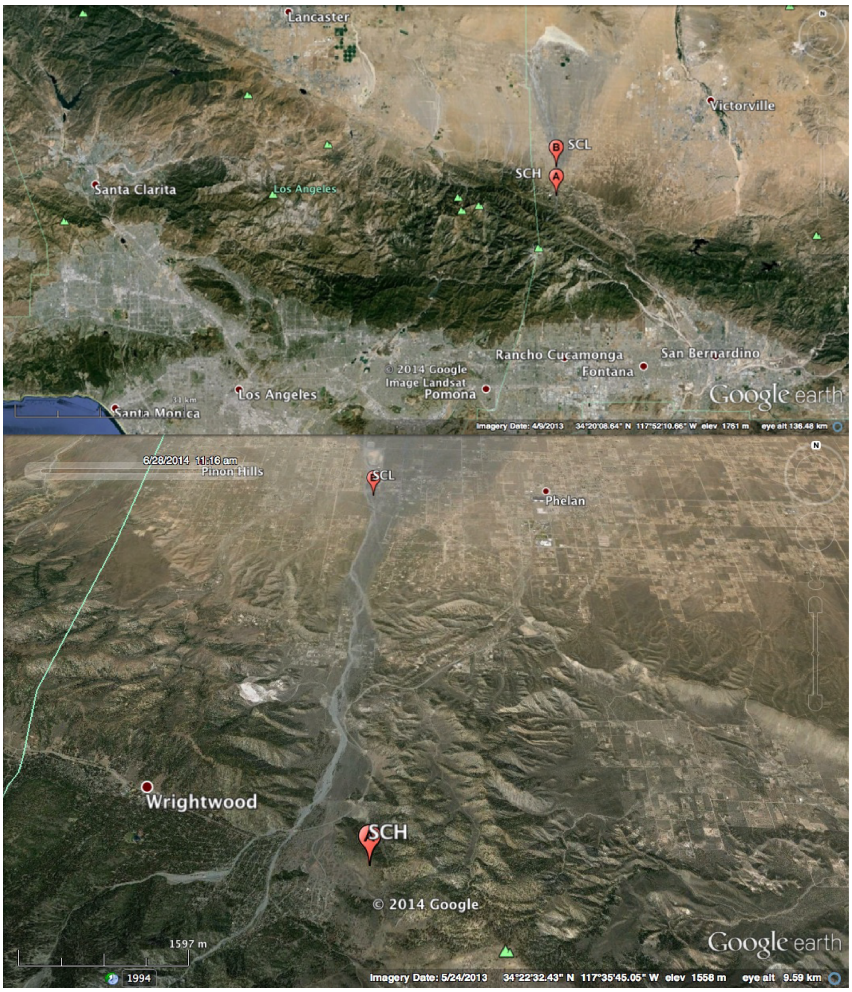

Figure 2: Satellite images of Sheep Creek Wash sites. *Top:* The two geographically distinct populations can be seen along the gray, inverted triangle that is SC Wash in the western Mojave Desert. *Bottom:* Closer view showing locol topography. *Both:* Sheep Creek High, the 1800 m elevation, lower stress site, is indicated with SCH at point A. Sheep Creek Low, the 1257 m elevation, higher stress site, is indicated with SCL at point B.

Sheep Creek Low (SCL) is a site at 1257 m elevation near Phelan, CA (34° 25' 29.80" N, 117° 36' 30.91" W), about nine miles northeast of SCH. This site is highly disturbed by foot traffic and erosion. The average high annual temperature of SCL is 27° C and the average low annual temperature is 10° C, while average annual precipitation is 28.2 cm (2005-2009, Phelan, CA, NOAA National Weather Service). Both sites are located along the Sheep Creek Wash and experience seasonal flow and snow melt from the San Gabriel Mountains (Figure 2).

3.1.2 Sequencing

Double digest restriction site-associated sequencing (ddRADseq) data was obtained from a previous study (Baughman 2015; manuscript in preparation). Sequencing was performed at the University of Florida NextGen Sequencing Core facility. The prepared library of 131 samples was sequenced for 150 bp from the 5' (EcoRI cut site) end in a single Illumina NextSeq lane. EcoRI cut sites are expected to occur approximately every 4,100 bases in the genome but the fragments were also digested with MseI, a more frequent cutter, and were further subjected to 250-500 bp size selection. These steps reduce the total amount of genome sequenced, thereby increasing sequencing depth for the same regions on all individuals. Raw reads were 150 bp long and were multiplexed; identifiable to the sample with unique sequence barcode adaptors. *Syntrichia caninervis* Illumina RNA HiSeq transcriptomic data was provided by the authors of its original publication (Gao et al. 2014).

3.2 Data Cleaning & Processing

3.2.1 Sequence Quality Filtering

All scripts used for this project will be available at the author's github repository: `https://github.com/jennabaughman/IB200`. Original raw ddRADseq data for 131 Mojave Desert samples was cleaned and organized prior to SNP genotyping and phylogenetic analyses. Quality of the sequence data was assessed with `FastQC` of the `FastX-Toolkit` (Gordon and Hannon 2010). A custom Perl script was used to de-multiplex the samples, check for correct enzyme cut site, and retain only reads with a minimum Phred quality score of 20 for at least 90% of the read. Any samples with fewer than 45,000 reads were removed due to insufficient or low quantity data. Finally, custom Perl and Bash scripts were used to remove barcodes, adaptors, and EcoRI cut sites, and to trim all fragments to 100 bp to remove lower quality 3' ends.

3.2.2 *in silico* Double Digest Restriction Site-Associated Sequencing

In order to root a phylogenetic tree of individuals from Mojave Desert populations, transcriptomic data from Chinese *S. caninervis* was used as the outgroup. This transcriptome was assembled with cDNA (derived from mRNA) from an unknown number of branches collected within a 10 m^2 research plot in the Gurbantunggut Desert of Xinjiang Uygur Autonomous Region of China (44° 32' 30" N, 88° 6' 42" E). Thus, although it is potentially made up of different genetic individuals, creating a chimeric sequence, it will still suffice as an outgroup so long as individuals in the Mojave Desert populations are more closely related to each other than they are to any individual from the Chinese chimeric transcriptome. However, gene flow between Chiense Gurbantunggut Desert and Mojave Desert is still possile and gene flow across this large of a distance is not unprecedented (Helena Korpelainen et al. 2012; McDaniel and Shaw 2005; Shaw, Golinski, et al. 2014; Vanderpoorten et al. 2008), so this assumption must be considered in the interpretation of the results of this study.

A high degree of sequence homology confidence was required in order to align and compare 80 bp ddRADseq loci with sequences from the *S. caninervis* transcriptome. One option for finding homologous regions would be to BLAST (Altschul et al. 1990 the ddRADseq fragments to the transcriptome, but this could potentially introduce some error when using such short sequences. Depending on how diverged the loci are, it would be difficult to tell if mismatches are due to genuine mutations or due to misalignment (alignment of non-homologus loci). A second step would be to check if the two enzyme cut sites, EcoRI and MseI, were present immediately up and within 500 bp down stream, respectively, from the alignment region. However this seemed computationally intensive and would still require some *a priori* alignment sequence identity and similarity restrictions. As an alternative method, I chose to perform an *in silico* ddRADseq procedure on the *S. caninervis* transcriptome to limit it to the same genome regions that the 131 samples were limited to. This technique may be better suited for a whole genome sequence, as transcriptomes necessarily reduce the amount of sequence data to work with, but at the time of writing a whole *S. caninervis* genome has not yet been sequenced or made publicly available.

To perform an *in silico* ddRADseq procedure, the bench procedure was mirrored in a bioinformatics context. First, the transcriptome was "digested" with the two enzymes that were used in the Mojave Desert samples, EcoRI (cuts G|AATTC) and MseI (cuts T|TAA), using tools from the `Biopieces` package Hansen et al. n.d. Next, the size selection step was also performed using `Biopieces`. As with the bench procedure, only fragments between 250 and 500 bp long were kept. The Illumina sequencer begins sequencing at the primers

that attached to the EcoRI "sticky end." Therefore, `Biopieces` tools were used to only keep fragments that start with the sequence "AATTC." Next, the EcoRI enzyme cut site was removed with `Biopieces` and sequences were trimmed to 80 bp with tools from the `FASTX-Toolkit` (Gordon and Hannon 2010). The whole process was then repeated on the reverse-complement of the transcriptome in order to account for both DNA strands. Finally, the transcriptome ddRAD sequence set was quintupled in order to create an artificial "read depth" of five.

3.2.3 Single Nucleotide Polymorphism Genotyping

The program `STACKS`, version 1.37, (Catchen, Amores, et al. 2011; Catchen, Hohenlohe, et al. 2013) was used to assemble both the actual and *in silico* ddRADseq data *de novo*, or without a reference genome. First all 132 samples were used to create catalog of variable loci. Here, the term locus is used to mean a sequence fragment or read that has passed quality filters and may have other reads that align to it, either perfectly or with some nucleotide differences. The `STACKS denovomap.pl` Perl script parameters were set as follows:

- Each identified catalog locus had a minimum four reads required for each allele (read depth of four). In general, the higher the read depth the more confidence that the SNP or polymorphism is real and not due to sequencing error. Since the study system is a haploid organism, there is no need to require a higher read depth and avoid ambiguity between true homozygosity and apparent homozygosity because only one allele was sequenced. A read depth of four was chosen as a requirement high enough to ensure confidence in sequencing and low enough not to unnecessarily reduce available data.

- Zero mismatches were allowed between loci when processing a single individual in order to require that all loci are "homozygous" to account for the haploid nature of the samples.

- Two mismatches were allowed when aligning secondary reads to primary stacks, allowing for two SNPs per 80 bp locus and, thus, fragments with two or fewer nucleotide differences were considered alleles for the same locus. Increasing this number may allow for the possibility of more highly diverged loci to be counted, but allowing too many mismatches between SNPs increases the likelihood of them being non-homologous or homoplastic.

- One mismatch was allowed between loci when building the catalog of variable loci. This parameter allows for one mismatch (one SNP) between loci that are fixed within but different among populations.

- `STACKS` attempted to remove or break up highly repetitive loci and only export loci that are "biologically plausible" with fewer than three alleles per locus per individual (Catchen, Amores, et al. 2011; Catchen, Hohenlohe, et al. 2013).

The `STACKS POPULATIONS` program (Catchen, Amores, et al. 2011; Catchen, Hohenlohe, et al. 2013 was used to generate "whitelist" of 1,702 80 bp present in at least 75% of individuals from two of the three 'populations' (SCH, SCL, and China), with up to 2 bi-allelic SNPs per locus and a minimum minor allele frequency of 10%. `POPULATIONS` was then run again to produce a the full fasta file for each individual for each locus on the 'whitelist.' One limitation with use of the program `STACKS` on a haploid organism is that all genotyping is processed as if two alleles are possible. Therefore, all SNPs of all 132 haploid individuals are expected to be reported as "homozygous." However, some loci are reported as heterozygous either due to a or sequencing error producing a false SNP or due to paralogous loci with minor nucleotide differences between them. While the construction of a catalog of variable loci was designed to reduce artificially heterozygous loci, few still make it through the restrictions. Therefore, the first allele reported for each loci was selected as the representative for that locus in that individual.

3.2.4 Creating a Sequence Matrix

The `STACKS` output in fasta format has each RADtag or fragment for each individual listed with both the locus number and the sample name in the fasta header. Therefore, before any phylogenetic analyses could be performed the files had to be split according to their catalog locus number. This was performed using a python script, which names each split file by the locus number. Following, bash scripts were used to remove locus numbers from each header in the fasta files and leave only the unique sample names. Finally, these fasta files, one per locus with one entry per individual, were imported into the program `SequenceMatrix` (Vaidya et al. 2011 to create a matrix of 1,702 80 bp loci for each 132 individuals, filling in missing sequences with the 80 '?' symbols for each missing locus to indicate unknown nucleotides. This program was then used to concatenate these 1,702 loci into a single final sequence of length 136,160 bp for each individual for use in phylogenetic analyses. See Table 1 for a summary of the characters in the final sequence matrix.

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

Table 1: Summary of characters used in phylogenetic analyses.

3.3 Phylogenetic Analyses

All phylogenetic analyses were performed using the matrix of 136,160 nucleotide characters (derived from 1,702 80 bp RAD loci with 1-3 SNPs each) and 132 OTUs. Of the 1,702 loci, 60 are represented in both the transcriptomic *S. caninervis* as well as the Mojave Desert samples and only 45 of those are in both populations.

3.3.1 Maximum Likelihood

Maximum likelihood tree-building was performed using `RAxML HPC` with the GTR-GAMMA nucleotide substitution model on Cipres Science Gateway. Several runs were performed to find the tree with the highest likelihood. Additionally, a bootstrap run was performed with 1000 iterations. Final trees were visualized, rooted, and colored in `FIGTree` (Morariu et al. 2008) and further edited in Adobe Illustrator.

3.3.2 Parsimony

Parsimony tree-building was performed using `PaupRat` on Cipres Science Gateway. Trees were built both with and without ratchet. After eight runs of randomized start points, four without ratchet four with 1000 ratchets each, all equally parsimonious trees of the shortest length were imported to the `Paup*` command line program to build a strict consensus tree. Final trees were visualized, rooted, and colored in `FIGTree` (Morariu et al. 2008) and further edited in Adobe Illustrator.

3.3.3 Migration Inference

The program `TreeMix` (Pickrell and Pritchard 2012) was used to infer migration events between SCH and SCL. This program takes a nonstandard or "private" data input format so data had to be converted manually. `STACKS` was used to output variable sites from 1,702 loci in all 132 individuals in the `GENEPOP` format since it was most similar to what `TreeMix`

required. Then, a combination of `TextWrangler` and `Microsoft Excel` was used to format the data as needed. Finally, the program was run both with and without correction for small populations and with up to six migration events allowed.

4 Results

4.1 Phylogenetic Trees

4.1.1 Maximum Likelihood

The likelihoods reported from four `RAxML` runs ranged from -303517.387360 at the lowest to -303389.332110 as the highest. `RAxML` bootstrap tree shown in Figure 3 and highest likelihood topology shown in Figure 4. There is high bootstrap support for placement of the tips but low support throughout the spine of the tree.

4.1.2 Parsimony

Four ratchet and four non-ratchet parsimony tree-building analyses resulted in a range of tree length of 19133 to 19269. The shortest tree length reported was 19133 and was found on three of the runs. One found 72 trees of length 19133, one found 294 trees of this length, and the other reported 187 trees of length 19133. However, it is not known how many of these total 553 trees of length 19133 represent identical topologies. Nonetheless, all trees with length 19133 are summarized in the strict consensus tree shown in Figure 5.

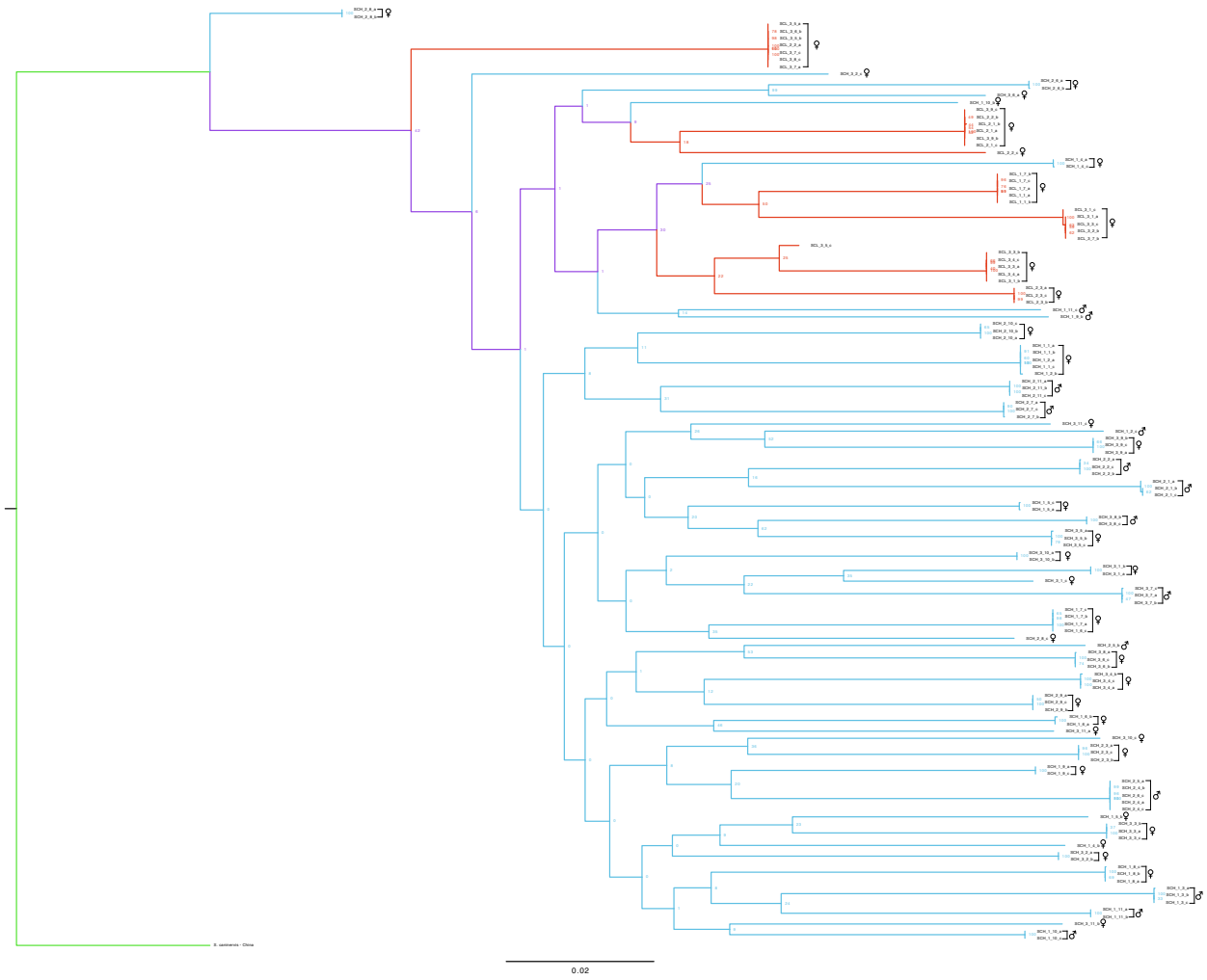

Figure 3: `RAxML` maximum likelihood tree with bootstrap of 131 Mojave Desert samples and *S. caninervis* outgroup. Branches that lead to the higher elevation Sheep Creek Wash population (SCH) are colored in blue and those that lead to the lower elevation population (SCL) are colored red. Branches that lead to mixed-population clades are indicated with purple. Brackets indicate inferred clones and symbols indicate known sex.

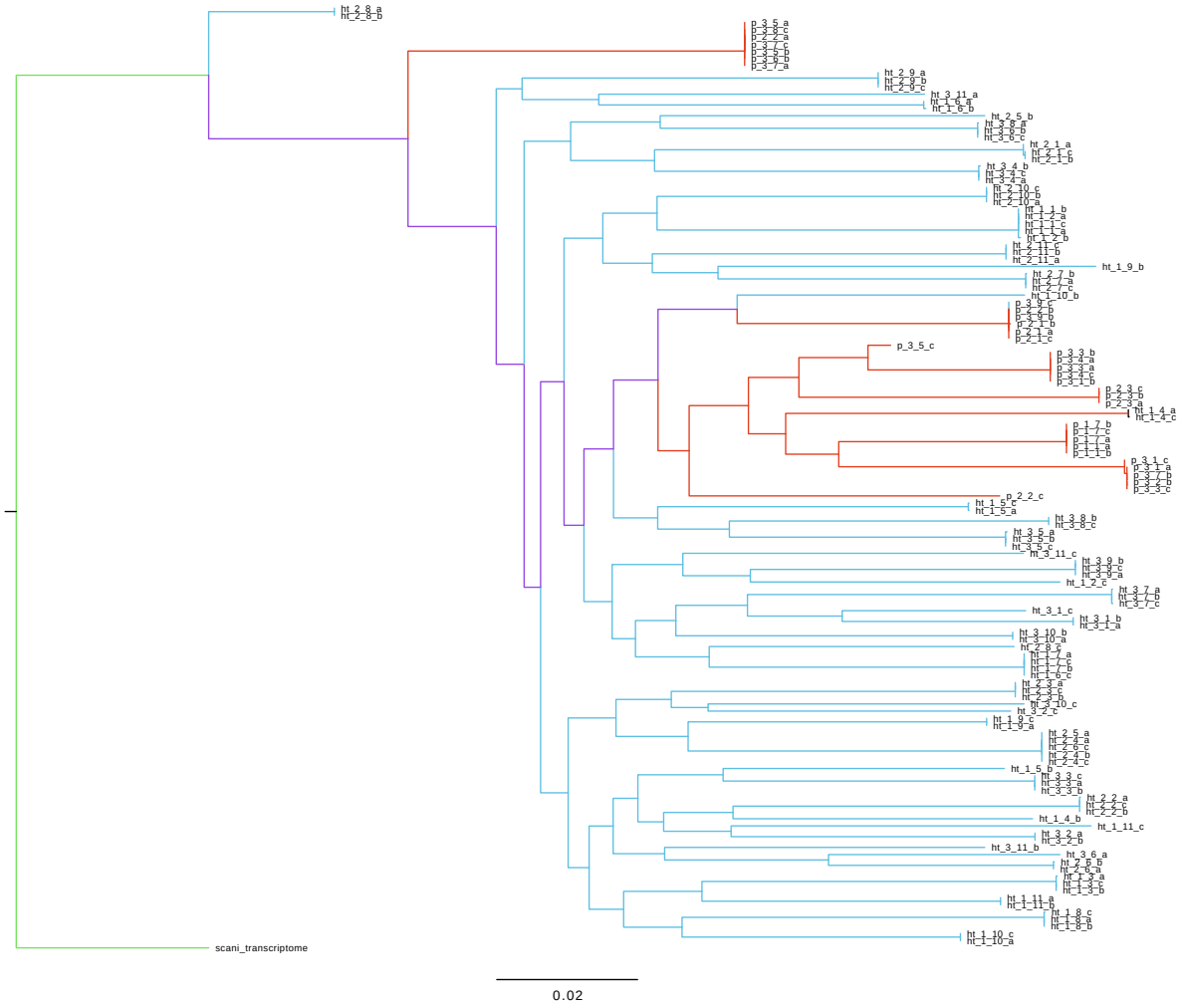

Figure 4: `RAxML` maximum likelihood tree of 131 Mojave Desert samples and *S. caninervis* outgroup. Branches that lead to the higher elevation Sheep Creek Wash population (SCH) are colored in blue and those that lead to the lower elevation population (SCL) are colored red. Branches that lead to mixed-population clades are indicated with purple.

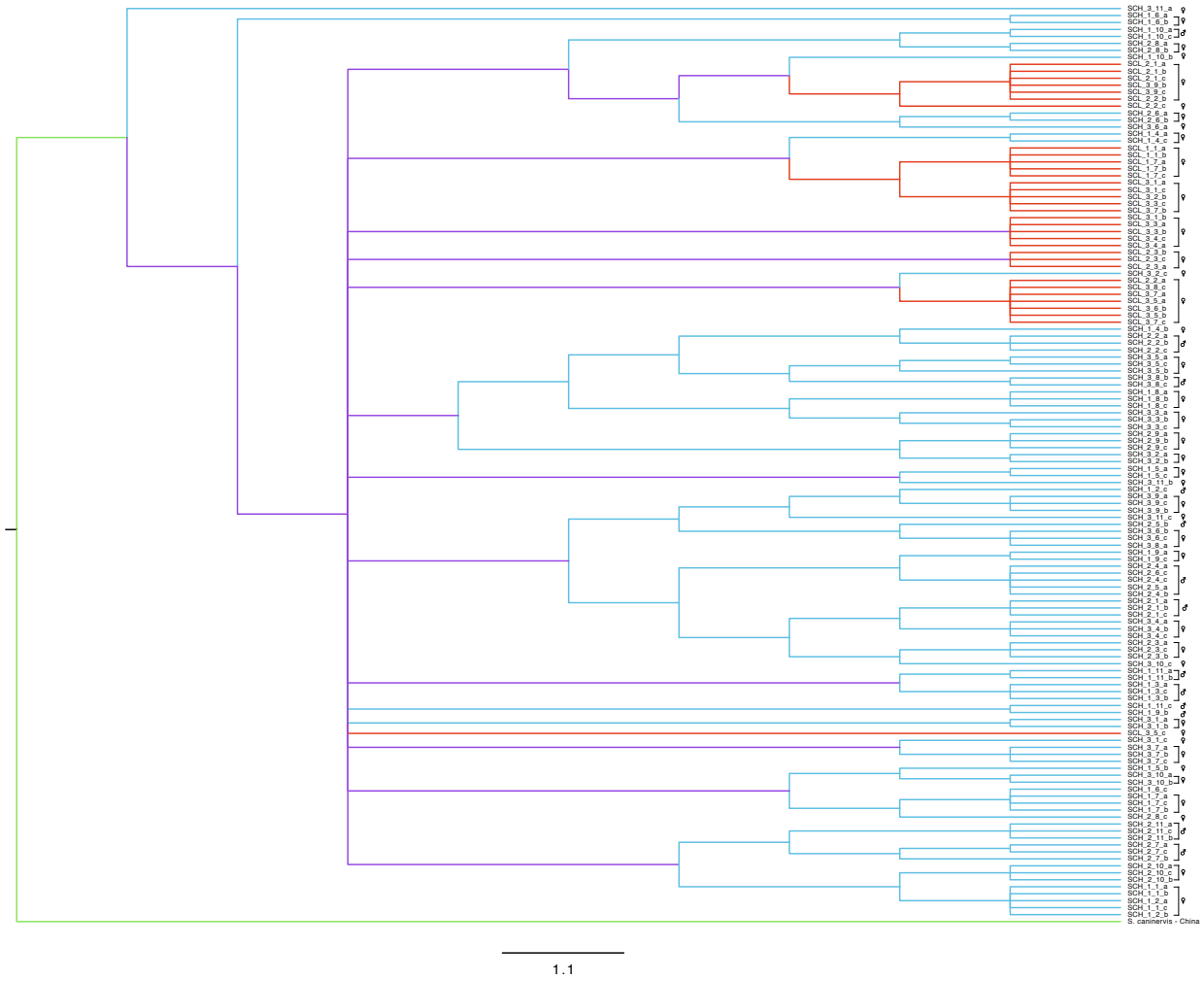

Figure 5: PaupRat parsimony strict consensus of 553 trees of length 19133 for 131 Mojave Desert samples and *S. caninervis* outgroup. Blue represents the higher elevation SCH Sheep Creek Wash population and red represents lower elevation SCL. Ambiguity in population relationships is indicated with purple. Brackets indicate clones inferred from the ML analyses and symbols indicate known sex.

4.1.3 Migration and Population History

One migration event was detected into to SCL from along the branch toward the root of the tree with a low weight or admixture proportion. SCH was found to have more drift than SCL. Results were robust to different parameter values and each produced similar or identical trees and graphs (no significance tests). Representative graph shown in Figure 6.

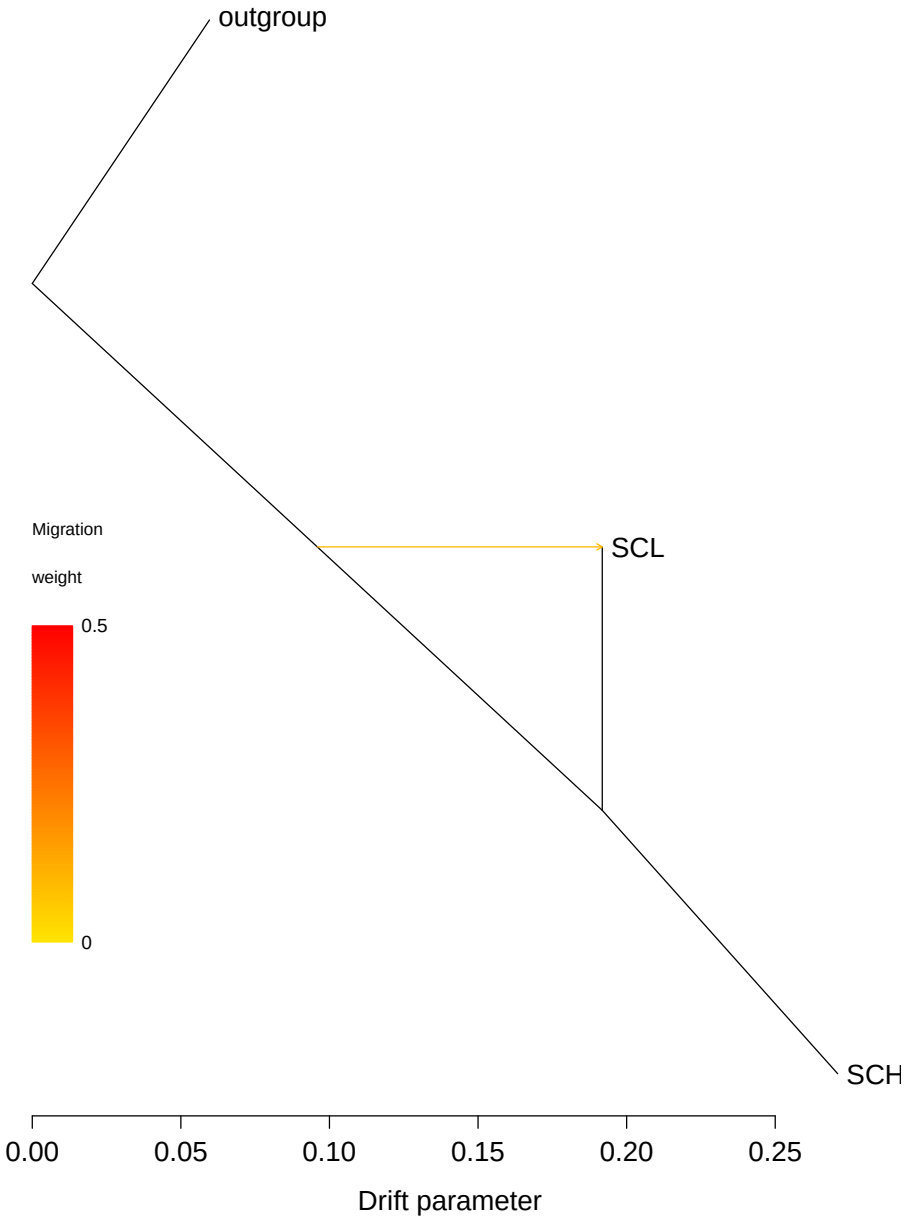

Figure 6: `TreeMix` tree and graph with one inferred migration event. Orange arrow connecting the lower elevation Mojave Desert indicates a single migration event to SCL. The x-axis describes the inferred amount of genetic drift in each population.

5 Discussion

This study showed that phylogenetics approaches may be applied to the population level as a rudimentary analysis of how "tree-like" the data is and therefore how clonal it is. While low bootstrap support on the spine of the tree may suggest past recombination, the lack of support for a bifurcating tree does necessarily mean sexual reproduction; it may just be insufficient data. Further tests need to be be done to understand the population history in these sites. The software `TreeMix` was applied as an exercise in inferring migration and population history, but the results are difficult to interpret and would be more appropriate with more populations (see Figure 6). The single migration even identified suggests and admixed ancestry in SCL while SCH appears to not have experienced any migration into or out of it. The weight of the migration event is the proportion of alleles in SCL that originated in each parental population. Future research in this area would be improved with more populations to allow for more accurate inferences. Another option that can be explored is be to run a `TreeMix` analysis with clones identified as 'populations,' but that would require *a priori* knowledge of what individuals are clones of one another. Future studies can perform a clone analysis to determine the appropriate genetic distance "allowed" between clones, assign individuals to clones, and repeat this with clonal clades as populations.

Interpreting the maximum likelihood trees as is, however, may provide some insight into the history of these two populations. While the highest likelihood and the bootstrap ML trees have similar topologies, they are not identical. In the bootstrap tree (Figure 3), lineages from SCL are placed sister to lineages from SCH more often than they are in the highest likelihood tree, shown in Figure 4. However, both trees group the same OTUs into clades, which likely represent clones, and both place the same SCH sister pair SCH 2.8 a and b as sister to all other SC individuals. Interestingly, the parsimony analysis also placed a SCH lineage as sister to the rest of SC, but it was not the same one as the ML analyses. As shown in Figure 5, sample SCH 3.11 a was found to be the most distantly related individual among the SC samples. Yet, the clade that the ML trees placed as sister to the rest of SC (SCH 2.8 a and b) is also among the more distantly related individuals, according to the parsimony analysis, as it is placed sister to *everyone else* after SCH 3.11 a.

If we can trust the topology and relationships within these trees, particularly that first branch after the outgroup, the hypothesis that SCH would be nested within SCL is not supported–the opposite is true. However, the great deal of ambiguity in these deeper relationships, as indicated by very low bootstrap support and large polytomies or combs,

I am unable to make any robust conclusion about the histories of individuals within these populations. What I can conclude is that the high bootstrap support in ML analyses and consistent placement in close relationships in both ML and parsimony of the tips into clades likely indicates recent asexual reproduction and clonal relationships. The lower bootstrap support and comb topology deeper in the tree may be due to past sexual reproduction but more data is needed to fully test that hypothesis. There also appears to be no pattern to the relationships of known sexes beyond what was already known about the sex ratios in each population. In summary, this research identified a novel approach for use of existing data sets. However, this approach may be more suitable for more distantly related clonal populations and should be followed up with more rigorous testing for migration and sexual reproduction.

6 References

Akiyama, H. (1999). "Genetic variation of the asexually reproducing moss Takakia lepidozioides." In: *Journal of Bryology* 21.3, pp. 177–182.

Altschul, Stephen F. et al. (1990). "Basic local alignment search tool." In: *Journal of molecular biology* 215.3, pp. 403–410.

Baughman, Jenna T (2015). "Sex or survival? The genetic impacts of environment and energetic trade-offs for the Mojave desert moss Syntrichia caninervis (Pottiaceae)." MA thesis. California State University, Los Angeles.

Bowker, Matthew A et al. (2000). "Sex expression, skewed sex ratios, and microhabitat distribution in the dioecious desert moss Syntrichia caninervis (Pottiaceae)." In: *American Journal of Botany* 87.4, pp. 517–526.

Catchen, J., A. Amores, et al. (2011). "Stacks: building and genotyping loci de novo from short-read sequences." In: *G3: Genes, Genomes, Genetics* 1, pp. 171–182.

Catchen, J., P. Hohenlohe, et al. (2013). "Stacks: an analysis tool set for population genomics." In: *Molecular Ecology* 22.11, pp. 3124–3140.

Chamberlain, Scott et al. (2014). "rgbif: Interface to the Global Biodiversity Information Facility API." In: *R package version 0.7* 7.

Cronberg, Nils (1996). "Clonal structure and fertility in a sympatric population of the peat mosses, Sphagnum rubellum and S. capillifolium." In: *Canadian Journal of Botany* 74.9, pp. 1375–1385.

— (2002). "Colonization dynamics of the clonal moss Hylocomium splendens on islands in a Baltic land uplift area: reproduction, genet distribution and genetic variation." In: *Journal of Ecology* 90, pp. 925–935.

— (2003). "Clonal distribution, fertility and sex ratios of the moss Plagiomnium affine (Bland.) T. Kop. in forests of contrasting age." In: *Journal of Bryology* 25.3, pp. 155–162.

Cronberg, Nils, K. Rydgren, and R. H. Økland (2006). "Clonal structure and genet-level sex ratios suggest different roles of vegetative and sexual reproduction in the clonal moss Hylocomium splendens." In: *Ecography* 29, pp. 95–103.

Gao, Bei et al. (2014). "De novo assembly and characterization of the transcriptome in the desiccation-tolerant moss Syntrichia caninervis." In: *BMC research notes* 7.1, p. 490.

Gordon, A and GJ Hannon (2010). "Fastx-toolkit." In: *Computer program distributed by the author, website http://hannonlab. cshl. edu/fastx_ toolkit/index. html [accessed 2014–2015]*.

Gunnarsson, Urban, Kristian Hassel, and Lars Söderström (2005). "Genetic structure of the endangered peat moss Sphagnum angermanicum in Sweden: A result of historical or contemporary processes?" In: *The Bryologist* 108.2, pp. 194–203.

Haig, David and Amity Wilczek (2006). "Sexual conflict and the alternation of haploid and diploid generations." In: *Philosophical Transactions of The Royal Society* 361, pp. 335–343.

Hansen, Martin A et al. *Biopieces: a bioinformatics toolset and framework*.

Hutsemékers, V. et al. (2010). "Macroecological patterns of genetic structure and diversity in the aquatic moss Platyhipnidium riparioides." In: *New Phytologist* 185, pp. 852–864.

Karlin, Eric F., R. E. Andrus, et al. (2011). "One haploid parent contributes 100% of the gene pool for a widespread species in northwest North America." In: *Molecular Ecology* 20, pp. 753–767.

Karlin, Eric F., Sara C. Hotchkiss, et al. (2012). "High genetic diversity in a remote island population system: sans sex." In: *New Phytologist* 193, pp. 1088–1097.

Korpelainen, Helena, Anikka K. Jägerbrand, and Maria von Cräutlein (2012). "Genetic structure of mosses Pleurozium schreberi (Willd. ex Brid.) Mitt. and Racomitrium lanuginosum (Hedw.) Brid. along altitude gradients in Hokkaido, Japan." In: *Journal of Bryology* 34.4, pp. 309–312.

Korpelainen, H. et al. (2013). "Spatial genetic structure of aquatic bryophytes in a connected lake system." In: *Plant Biology* 15.3, pp. 514–521.

McDaniel, Stuart F. (2005). "Genetic correlations do not constrain the evolution of sexual dimporphism in the moss Ceratodon purpureus." In: *Evolution* 59.11, pp. 2353–2361.

McDaniel, Stuart F. and A. Jonathan Shaw (2005). "Selective sweeps and intercontinental migration in the cosmopolitan moss Ceratodon purpureus (Hedw.) Brid." In: *Molecular Ecology* 14, pp. 1121–1132.

Mishler, Brent D. (1988). "Reproductive ecology of bryophytes." In: ed. by J. Lovett Doust and L. Lovett Doust. Plant Reproductive Ecology. USA: Oxford University Press, pp. 285–306.

— (2001). "The biology of bryophytes—bryophytes aren't just small tracheophytes." In: *American Journal of Botany* 88.11, pp. 2129–2131.

Morariu, Vlad I. et al. (2008). "Automatic online tuning for fast Gaussian summation." In: *Advances in Neural Information Processing Systems (NIPS)*.

Newton, Angela E. and Brent D. Mishler (1994). "The evolutionary significance of asexual reproduction in mosses." In: *Journal of the Hattori Botanical Laboratory* 76, pp. 127–145.

Newton, M. E. (1988). "Chromosomes as indicators of bryophyte reproductive performance." In: *Botanical Journal of the Linnaean Society* 98, pp. 269–275.

Paasch, Amber Elizabeth et al. (2015). "Decoupling of sexual reproduction and genetic diversity in the female-biased Mojave Desert moss Syntrichia caninervis (Pottiaceae)." In: *International Journal of Plant Sciences* 176.8, pp. 751–761.

Patiño, Jairo, Olaf Werner, and Juana-María González-Mancebo (2010). "The impact of forest disturbance on the genetic diversity and population structure of late-successional moss." In: *Journal of Bryology* 32, pp. 220–231.

Pickrell, Joseph K and Jonathan K Pritchard (2012). "Inference of population splits and mixtures from genome-wide allele frequency data." In: *PLoS Genet* 8.11, e1002967.

Pisa, Sergio et al. (2013). "Elevational patterns of genetic variation in the cosmopolitan moss Bryum argenteum (Bryaceae)." In: *Journal of American Botany* 100.10, pp. 2000–2008.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Richards, A. J. (1986). *Plant Breeding Systems*. 2nd. London: Chapman & Hall.

Shaw, A. Jonathan, G. Karen Golinski, et al. (2014). "Intercontinental genetic structure in the amphi-Pacific peatmoss Sphagnum miyabeanum (Bryophyta: Sphagnaceae)." In: *Biological Journal of the Linnean Society* 111, pp. 17–37.

Shaw, A. Jonathan and Monica Srodon (1995). "Clonal Diversity in Sphagnum rubellum Wils." In: *The Bryologist* 98.2, pp. 261–264.

Spagnuolo, Valeria et al. (2007). "Ubiquitous genetic diversity in ISSR markers between and within populations of the asexually producing moss Pleurochaete squarrosa." In: *Plant Ecology* 188, pp. 91–101.

Stark, Lloyd R. (1997). "Phenology and reproductive biology of Syntrichia inermis (Bryopsida, Pottiaceae) in the Mojave Desert." In: *Bryologist*, pp. 13–27.

Stark, Lloyd R., D. Nicholas McLetchie, and Brent D. Mishler (2005). "Sex expression, plant size, and spatial segregation of the sexes across a stress gradient in the desert moss Syntrichia caninervis." In: *The Bryologist* 108.2, pp. 183–193.

Stark, Lloyd R., Brent D. Mishler, and D. Nicholas McLetchie (1998). "Sex expression and growth rates in natural populations of the desert soil crustal moss Syntrichia caninervis." In: *Journal of Arid Environments* 40.4, pp. 401–416.

— (2000). "The cost of realized sexual reproduction: assessing patterns of reproductive allocation and sporophyte abortion in a desert moss." In: *American Journal of Botany* 87.11, pp. 1599–1608.

Vaidya, Gaurav, David J. Lohman, and Rudolf Meier (2011). "SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information." In: *Cladistics* 27.2, pp. 171–180.

Vanderpoorten, Alain et al. (2008). "The barriers to oceanic island radiation in bryophytes: insights from the phylogeography of the moss Grimmia montana." In: *Journal of Biogeography* 35.4, pp. 654–663.

Wang, Yingying, Yongqing Zhu, and Youfang Wang (2012). "Differences in spatial genetic structure and diversity in two mosses with different dispersal strategies in a fragmented landscape." In: *Journal of Bryology* 34.1, pp. 9–16.

Wickham, Hadley (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-0-387-98140-6. URL: `http://ggplot2.org`.

Zouhair, R. et al. (2001). "Growth morphology and genetic diversity of Polytrichum poluations." In: *Journal of Bryology* 23.2, pp. 109–117.

7 Acknowledgments

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