AN ABSTRACT OF THE DISSERTATION OF

Zhian N. Kamvar for the degree of Doctor of Philosophy in Plant Pathology presented on December 6, 2016.

Title: Development and Application of Tools for Analysis of Clonal Population Genetics

Abstract approved: _____

Niklaus J. Grünwald

This is a LaTeXtemplate derived from the BeamerTeXtemplate found on overleaf (https://www.overleaf.com/latex/templates/oregon-state-university-thesis-and-dissertation/wnvzcdhqshxf) and cloned using `git clone https://git.overleaf.com/5973847rnpkdf` I have made the following changes

- modified `beavtex.cls` from v1.1 to include support for verbatim font styles in R code
- stripped the template of its original content to create a pandoc template
- placed this into a bookdown (https://bookdown.org/) framework
- added packages latexsym, amsmath amssymb,amsthm, longtable, booktabs, setspace, url, hyperref, lmodern, caption, and float

This is still a WIP, but it seems to work for the moment. Zhian N. Kamvar 2016-08-25

Development and Application of Tools for Analysis of Clonal Population Genetics

by

Zhian N. Kamvar

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented December 6, 2016
Commencement June 2017

Doctor of Philosophy dissertation of Zhian N. Kamvar presented on December 6, 2016.

APPROVED:

Major Professor, representing Plant Pathology

Head of the Department of Botany and Plant Pathology

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Zhian N. Kamvar, Author

ACKNOWLEDGEMENTS

I would like to acknowledge... Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas vel eros sed mauris porttitor semper nec a orci. Nullam vestibulum mi nec condimentum posuere. Pellentesque eget diam id sapien aliquet ullamcorper. Pellentesque blandit nec lectus ut mollis. Praesent in facilisis justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Sed eget congue leo, sed consequat libero. In rutrum malesuada nisi. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi sollicitudin tortor ut sem facilisis mollis.

CONTRIBUTION OF AUTHORS

The following people contributed to this dissertation:

Chapter 1

Jane R. Professor assisted in the design, analysis, and editing of the manuscript.

Chapter 2

Lisa Simpson developed the initial concept and experimental design for the study. Ellen Ripley advised and assisted with statistical analysis. Jane R. Professor assisted in the design, analysis, and editing of the manuscript.

Chapter 3

Jane R. Professor assisted in the design, analysis, and editing of the manuscript.

Chapter 4

Jane R. Professor assisted in the design, analysis, and editing of the manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

This is for my mother who paved the way.

Chapter 1: Introduction

- Plant Pathogens
  - Evolution and Adaptation
  - Understanding clonal population dynamics (*Phytophthora* and *Zymoseptoria* examples)
- Tools for analysis
  - Tools for clonal populations always come after sexual populations
  - We wrote our own tools
- Goals of my work
  1. development of computational tools to characterize populations
  2. application of these tools to address empirical and theoretical questions

1.1 Population genetics of clonal organisms

- Population genetics is traditionally centered around HWE
- Clonal populations violate basic statistical inference
  - non-independent samples (clones)
  - excess heterozygosity

  - Meselson effect (Butlin 2000; Welch & Meselson 2000, 2001; Balloux *et al.* 2003)

1.2 The Genus *Phytophthora*

The genus *Phytophthora*, translating to "plant destroyer" in Greek, contains over 100 species (Kroon *et al.* 2012), many of which have significant impact on US agriculture. *P. sojae* is a major problem on soybean, causing $1-2 billion in losses each year (Tyler 2007). *P. ramorum* is changing the landscape of the North American West due to its wide host range (Grünwald *et al.* 2008), which results in devastating losses for the US forestry and nursery industries. And *P. infestans*, which was a root cause of over a million deaths during the Irish Potato Famine and continues to be a problem on tomato and potato crops, resulting in losses exceeding $6 billion, annually (Haas *et al.* 2009). *Phytophthora spp.* are water molds characterized by production of oospores and biflagellate zoospores that place them into the Stramenopiles (Baldauf 2003). They are most closely related to golden brown algae and quite diverged from fungi.

1.2.1 life cycle

1.2.2 Sex and mating types

1.2.3 Heterothallic, clonal: *P. ramorum*

- Sudden Oak Death

- Population genetics in US nurseries (Goss *et al.* 2009)

1.2.4 Homothallic, partially clonal: *P. syringae*

- Abundant in OR Nurseries (found in foliar isolates) (Parke *et al.* 2014)
- Genetic structure uncharacterized

1.3 Tools for analysis of clonal population genetics

Recommendations have been made for analysis (Arnaud-Hanod *et al.* 2007)

- MLG diversity
- Genotype Accumulation Curve
- P_{sex} and P_{gen}

1.3.1 Index of Association

- Standardized Version (Agapow & Burt 2001)
- Previous Simulation Analyses (de Meeûs & Balloux 2004)
- What's missing
  - Sympatric clonal lineages
  - Analysis of significance testing
  - HTS markers

1.3.2 Software Limitations

Plethora of tools, most designed for sexual populations except:

- GenClone
- GenoDive

Problems with file formatting, time, and reproducibility.

1.3.3 poppr

- R
- poppr

1.4 Applications of novel tools for analysis of partially-clonal populations

- Simulation analysis of $\bar{r}_d$
- SSR analysis of *P. ramorum*
- GBS analysis of *P. syringae*

1.5 Conclusion

- Open Source Scientific Software Development of Poppr
  - Related tools
- Simulation Analysis

- Pop gen info for two *Phytophthoras*
- Major results of your work in 2-4 sentences . . .

Chapter 2: *Poppr*: an R Package For Genetic Analysis of Populations With Clonal, Partially Clonal, and/or Sexual Reproduction

2.1 Abstract

Many microbial, fungal, or oomcyete populations violate assumptions for population genetic analysis because these populations are clonal, admixed, partially clonal, and/or sexual. Furthermore, few tools exist that are specifically designed for analyzing data from clonal populations, making analysis difficult and haphazard. We developed the R package poppr providing unique tools for analysis of data from admixed, clonal, mixed, and/or sexual populations. Currently, poppr can be used for dominant/codominant and haploid/diploid genetic data. Data can be imported from several formats including GenAlEx formatted text files and can be analyzed on a user-defined hierarchy that includes unlimited levels of subpopulation structure and clone censoring. New functions include calculation of Bruvo's distance for microsatellites, batch-analysis of the index of association with several indices of genotypic diversity, and graphing including dendrograms with bootstrap support and minimum spanning networks. While functions for genotypic diversity and clone censoring are specific for clonal populations, several functions found in poppr are also valuable to analysis of any populations. A manual with documentation and examples is provided. Poppr is open source and major releases are available on CRAN: http://cran.r-project.org/package=poppr.

More supporting documentation and tutorials can be found under ‘resources’ at: http://grunwaldlab.cgrb.oregonstate.edu/.

Chapter 3: Spatial and Temporal Analysis of Populations of the Sudden Oak Death Pathogen in Oregon Forests

3.1 Abstract

Sudden oak death caused by the oomycete *Phytophthora ramorum* was first discovered in California toward the end of the 20th century and subsequently emerged on tanoak forests in Oregon before its first detection in 2001 by aerial surveys. The Oregon Department of Forestry has since monitored the epidemic and sampled symptomatic tanoak trees from 2001 to the present. Populations sampled over this period were genotyped using microsatellites and studied to infer the population genetic history. To date, only the NA1 clonal lineage is established in this region, although three lineages exist on the North American west coast. The original introduction into the Joe Hall area eventually spread to several regions: mostly north but also east and southwest. A new introduction into Hunter Creek appears to correspond to a second introduction not clustering with the early introduction. Our data are best explained by both introductions originating from nursery populations in California or Oregon and resulting from two distinct introduction events. Continued vigilance and eradication of nursery populations of *P. ramorum* are important to avoid further emergence and potential introduction of other clonal lineages.

Chapter 4: Novel R Tools For Analysis of Genome-Wide Population Genetic Data With Emphasis on Clonality

4.1 Abstract

To gain a detailed understanding of how plant microbes evolve and adapt to hosts, pesticides, and other factors, knowledge of the population dynamics and evolutionary history of populations is crucial. Plant pathogen populations are often clonal or partially clonal which requires different analytical tools. With the advent of high throughput sequencing technologies, obtaining genome-wide population genetic data has become easier than ever before. We previously contributed the R package *poppr* specifically addressing issues with analysis of clonal populations. In this paper we provide several significant extensions to *poppr* with a focus on large, genome-wide SNP data. Specifically, we provide several new functionalities including the new function `mlg.filter` to define clone boundaries allowing for inspection and definition of what is a clonal lineage, minimum spanning networks with reticulation, a sliding-window analysis of the index of association, modular bootstrapping of any genetic distance, and analyses across any level of hierarchies.

Chapter 5: [Tentative Title] Population Dynamics of the Plant Pathogen *Phytophthora syringae* in Oregon Nurseries

5.1 Abstract

5.2 Introduction

Phytophthora syringae is the most important species affecting ornamentals produced in the Pacific Northwest. Recent nursery sampling efforts, aimed at characterizing the diversity of *Phytophthoras* within Oregon nurseries, have revealed the species *P. syringae* to be among the most abundant taxa found in the nurseries surveyed (Parke *et al.* 2014). *P. syringae* is adapted to cold weather and grows best in the cool, wet fall, winter and spring and is least active in summer (Erwin *et al.* 1996). Like *P. ramorum*, it has a wide host range including *Rhododendron*, *Camellia*, *Malus*, and many other taxa. It has the capability for outcrossing, self-fertilizing, and reproducing clonally. This pathogen has been found globally since 1881 and is problematic on woody ornamentals such as crabapple (*Malus spp.*), as it causes unsightly cankers that make the plant unsellable (Erwin *et al.* 1996). While the ecology of this pathogen has been studied to some degree, very little is known about the demographic history and population structure on a local and global scale.

Chapter 6: [Tentative Title] The Effect of Population Dynamics, Sample Size, and Marker Choice on the Index of Association

6.1 Abstract

TBD...

6.2 Introduction

- Population Genetics of partially clonal organisms
  - This has been studied in the past (Orive 1993; Smith *et al.* 1993; Balloux *et al.* 2003; de Meeûs & Balloux 2004)
  - The index of association (Brown *et al.* 1980; Smith *et al.* 1993; Agapow & Burt 2001)
  - In de Meeûs & Balloux (2004), it was shown that $\bar{r}_d$ has a high variance in clonal populations and Smith *et al.* (1993) showed that it's affected by population structure.
- Methods for assessing level of clonal reproduction (CloNcaSe) (Ali *et al.* 2016)
  - This only works for populations with discrete generations with an observable sexual stage.

- Limitations of previous studies
  - No HTS markers
  - Significance tests are routinely performed via permutation analysis, but could not be performed due to software limitations (Burt *et al.* 1996; de Meeûs & Balloux 2004)
- Objectives
  1. Analyze mixtures of clonal populations to assess effect of sampling multiple clonal populations
  2. Re-analyze rates of sexual reproduction to confirm previous study
  3. Assess significance tests for $\bar{r}_d$

6.3 Methods

All simulations were performed with the python package simuPOP version 1.1.7 in python version 3.4. For each scenario, 100 simulations with 10 replicates were created with a census size of 10,000 individuals evolved over 10,000 generations. From each replicate, 10, 25, 50, and 100 individuals were sampled without replacement for downstream analysis.

6.3.1 Microsatellite Simulation

Each population was simulated with 20 co-dominant loci containing 6 to 10 alleles with frequencies drawn from a uniform distribution and normalized. Before mating,

mutations occurred at each locus at a rate of 1e-5 mutations/generation.

6.3.2 Microsatellite Analysis

The standardized index of association ($\bar{r}_d$) was calculated in R version 3.2 with the package *poppr* version 2.2.1 using the function `ia()` within custom scripts (supplementary information). Tests for significance were performed by randomly permuting the alleles at each locus independently and then assessing $\bar{r}_d$. This was done 999 times for each replicate population. The p-values reflect the proportion of observations greater than the observed statistic.

6.3.3 GBS Simulations

Simulations of 10,000 binary loci at intervals of 1 Mbp over 100 chromosomal fragments were simulated with a mutation rate of 1e-5 mutations per generation and a recombination rate of 1e-5 for sexually recombining populations.

References

Agapow P-M, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.

Ali S, Soubeyrand S, Gladieux P *et al.* (2016) Cloncase: Estimation of sex frequency and effective population size by clonemate resampling in partially clonal organisms. *Molecular Ecology Resources*.

Arnaud-Hanod S, Duarte CM, Alberto F, Serrão EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.

Baldauf S (2003) The deep roots of eukaryotes. *Science*, **300**, 1703–1706.

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.

Brown A, Feldman M, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, **96**, 523–536.

Burt A, Carter DA, Koenig GL, White TJ, Taylor JW (1996) Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proceedings of the National*

Academy of Sciences, **93**, 770–773.

Butlin RK (2000) Virgin rotifers. *Trends in Ecology & Evolution*, **15**, 389–390.

de Meeûs T, Balloux F (2004) Clonal reproduction and linkage disequilibrium in diploids: A simulation study. *Infection, genetics and evolution*, **4**, 345–351.

Erwin DC, Ribeiro OK, others (1996) *Phytophthora diseases worldwide*. American Phytopathological Society (APS Press), St. Paul, Minnesota, USA.

Goss EM, Larsen M, Chastagner GA, Givens DR, Grünwald NJ (2009) Population genetic analysis infers migration pathways of *Phytophthora ramorum* in US nurseries. *PLoS Pathogens*, **5**, e1000583.

Grünwald NJ, Goss EM, Press CM (2008) *Phytophthora ramorum*: a pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Molecular Plant Pathology*, **9**, 729–740.

Grünwald NJ, Martin FN, Larsen MM *et al.* (2011) Phytophthora-ID.org: a sequence-based *Phytophthora* identification tool. *Plant Disease*, **95**, 337–342.

Haas BJ, Kamoun S, Zody MC *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*, **461**, 393–398.

Kroon LP, Brouwer H, Cock AW de, Govers F (2012) The Genus *Phytophthora*. *Phytopathology*, **102**, 348–364.

Orive ME (1993) Effective population size in organisms with complex life-histories.

Theoretical population biology, **44**, 316–340.

Parke JL, Knaus BJ, Fieland VJ, Lewis C, Grünwald NJ (2014) Phytophthora community structure analyses in Oregon nurseries inform systems approaches to disease management. *Phytopathology*, **104**, 1052–1062.

Smith JM, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences*, **90**, 4384–4388.

Tyler BM (2007) *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Molecular Plant Pathology*, **8**, 1–8.

Welch DBM, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, **288**, 1211–1215.

Welch DBM, Meselson MS (2001) Rates of nucleotide substitution in sexual and anciently asexual rotifers. *Proceedings of the National Academy of Sciences*, **98**, 6720–6724.