Effect of migration, carrying capacity, and fecundity on the formation of clinal patterns during range expansions.

Neha J. Angal
University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd

Part of the Biological and Physical Anthropology Commons, Other Anthropology Commons, and the Other Ecology and Evolutionary Biology Commons

Recommended Citation
https://doi.org/10.18297/etd/2507

This Master's Thesis is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.
EFFECTS OF MIGRATION, CARRYING CAPACITY, AND FECUNDITY ON THE FORMATION OF CLINAL PATTERNS DURING RANGE EXPANSIONS

By

Neha J. Angal
B.A., Western Kentucky University, 2013

A Thesis
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Arts
in Anthropology

Department of Anthropology
University of Louisville
Louisville, Kentucky

August 2016
EFFECTS OF MIGRATION, CARRYING CAPACITY, AND FECUNDITY ON THE FORMATION OF CLINAL PATTERNS DURING RANGE EXPANSIONS

By

Neha J. Angal
B.A., Western Kentucky University, 2013

A Thesis Approved on

August 5, 2016

by the following Thesis Committee:

Christopher Tillquist

Fabian Crespo

Forrest Stevens
ACKNOWLEDGEMENTS

This work was conducted in part using the resources of the University of Louisville’s research computing group and the Cardinal Research Cluster.

I would like to acknowledge Dr. Fabián Crespo (Department of Anthropology, University of Louisville) for his encouragement throughout my graduate career and committee service; Dr. Forrest Stevens (Department of Geography and Geosciences, University of Louisville) for his invaluable assistance with data management in R and committee service; Harrison Simrall (PhD Candidate, Department of Physics and Astronomy, and Senior Academic Consultant at IT support services and CRC research computing group, University of Louisville) for all his help facilitating this work with the Cardinal Research Cluster and Dr. Christopher Tillquist (Department of Anthropology, University of Louisville), for all of the support and encouragement throughout my graduate career, for his insightful comments and guidance through my thesis research, for mentoring me as an academic and teacher for the past three years, and for serving as my committee chair. I would also like to thank the Anthropology Department at UofL for funding two years of my graduate education, in addition to providing funding support for field school and conference opportunities. Lastly, I would like to thank all friends and family who have supported me throughout my graduate career, especially my partner Adam, who has been a wonderful friend and cheerleader.
ABSTRACT

EFFECTS OF MIGRATION, CARRYING CAPACITY, AND FECUNDITY ON THE FORMATION OF CLINAL PATTERNS DURING RANGE EXPANSIONS

Neha J. Angal

August 5, 2016

Range expansions, empirically and in simulations, lead to clinal patterns of genetic diversity. Clines are often used as spatial markers of past migrations. This study investigated the effects of migration, growth, and carrying capacities on clinal patterns during range expansions, using forward-time simulations in Nemo. Initial results show, in the absence of prior population structure, range expansions result in a loss of diversity strongly affected by migration, growth, and carrying capacity. This loss of diversity did not persist to the final generation, corresponding to 10,000 years, indicating clinal patterns are less durable than previously assumed—challenging the utility of clinal patterns as specific markers of past migrations. Further simulations are necessary to evaluate the effects of large demographic collapses, negative selection, and non-equilibrium migration upon clines. While the case study for these experiments is the peopling of Europe, these results are broadly applicable to other human colonization events.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 1—BACKGROUND, THEORY, AND LITERATURE REVIEW</td>
<td>5</td>
</tr>
<tr>
<td>CHAPTER 2—MATERIALS &amp; METHODS</td>
<td>21</td>
</tr>
<tr>
<td>CHAPTER 3—EXPERIMENTAL DESIGN</td>
<td>26</td>
</tr>
<tr>
<td>CHAPTER 4—RESULTS AND DISCUSSION</td>
<td>36</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>72</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>84</td>
</tr>
<tr>
<td>CURRICULUM VITA</td>
<td>162</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLE PAGE
1. Migration Rates ................................................................. 22
2. Growth Rates and Fecundity .................................................. 23
3. Carrying Capacity ............................................................... 24
4. Key Software Attributes ...................................................... 28
5. Parameter Values ............................................................... 37
6. Parameter Results ............................................................... 51
7. Moran’s I for Average Patch Heterozygosity ......................... 53
8. Moran’s I for Allele Frequencies ............................................ 59
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Analysis Pipeline</td>
<td>27</td>
</tr>
<tr>
<td>2. A-D</td>
<td>40</td>
</tr>
<tr>
<td>3. A-D</td>
<td>41</td>
</tr>
<tr>
<td>4. A-D</td>
<td>43</td>
</tr>
<tr>
<td>5. A-D</td>
<td>44</td>
</tr>
<tr>
<td>6. A-B</td>
<td>46</td>
</tr>
<tr>
<td>7. A-D</td>
<td>47</td>
</tr>
<tr>
<td>8. A-D</td>
<td>48</td>
</tr>
<tr>
<td>9. A-D</td>
<td>50</td>
</tr>
<tr>
<td>10. A-D</td>
<td>54</td>
</tr>
<tr>
<td>11. A-D</td>
<td>55</td>
</tr>
<tr>
<td>12. A-D</td>
<td>57</td>
</tr>
<tr>
<td>13. A</td>
<td>58</td>
</tr>
</tbody>
</table>
INTRODUCTION

The developments and dissolutions of spatial signatures of past migration events are challenging to analyze within biological systems, particularly for organisms with long generation times, such as humans; time simply moves too slow for any such study to be feasible. An elegant solution to this problem is to test hypotheses of these processes via simulations, which are virtual reconstructions of real systems that model evolutionary processes by utilizing virtual environments. This study used one of these programs, Nemo (Guillaume and Rougemont 2006), to simulate complex evolutionary processes that occur during the colonization of continental regions.

A central component to this and other studies of the colonization of continental regions involves the study of the processes of organisms moving into a region and colonizing it—or range expansions. The theoretical study of range expansions is still an ongoing and active area of research. The exact contributions of expansion dynamics such as carrying capacity, migration rate, and growth rate/average fecundity to the formation of spatial gradients of genetic diversity within populations, or clines, are still under investigation (Ray, Currat and Excoffier 2003; Klopfstein, Currat and Excoffier 2006; Wegmann, Currat and Excoffier 2006; Travis et al. 2007; Excoffier and Ray 2008; Ray and Excoffier 2009; Arenas et al. 2012; Arenas et al. 2013; Flaxman 2013; Peischl et al. 2013; Peischl and Excoffier 2015). This study addresses this problem through
investigating which range expansion dynamics—carrying capacity, fecundity, and migration rate—are most salient for cline formation. Ultimately, this study demonstrates which expansion dynamics are most important for the development of clinal patterns and comments on potential limitations to utilizing clinal patterns as evidence of historical migration events.

In order to address this problem, literature searches were performed for the fields of molecular ecology, phylogeography, geography, population genetics, population genetics simulations, anthropology, and anthropological genetics. Generally, literature fell into one of three categories: A) Theory papers investigating cline formation in a variety of evolutionary contexts (Slatkin 1973; May, Endler and McMurtrie 1975; Slatkin and Maruyama 1975; Endler 1977; Fix 1996, 1997) B) Studies explicitly investigating the confluence of range expansion dynamics with a focus on the effects of different variables and their interactions (Ray, Currat and Excoffier 2003; Klopstein, Currat and Excoffier 2006; Wegmann, Currat and Excoffier 2006; Excoffier and Ray 2008; Ray and Excoffier 2009; Arenas et al. 2012; Arenas et al. 2013; Flaxman 2013; Peischl et al. 2013; Peischl and Excoffier 2015); or C) Studies explicitly attempting to test specific hypothesized models for the peopling of the European continent—here the focus was on the Paleolithic/post-glacial recolonization versus Neolithic colonization (also known as the demic diffusion versus cultural diffusion) debate, and the results were taken to be indicative of supporting one colonization model over the other (Sokal and Menozzi 1982; Rendine, Piazza and Cavalli-Sforza 1986; Barbujani, Sokal and Oden 1995; Itan et al. 2009; François et al. 2010; Arenas et al. 2013). While papers of type A or B were
sometimes also type C papers, very rarely were papers found that fit all three types. For a
detailed review of this body of literature, refer to Chapter 1.

This study bridges the conceptual, theoretical, methodological, and historical gaps
amongst these varied fields to address how the dynamics of range expansions lead to
cline formation. Additionally, the question of cline durability is particularly relevant for
anthropology and anthropological genetics. The Paleolithic/post-glacial re-colonization
versus Neolithic colonization debate has been prominent within anthropology and
anthropological genetics for more than two decades, and many results obtained and used
to support one model over its alternative(s) were obtained from studies which implicitly
assumed clines were, once established, were indissoluble (Ammerman and Cavalli-Sforza
1984; Rendine, Piazza and Cavalli-Sforza 1986; Cavalli-Sforza, Menozzi and Piazza
1994; Barbujani, Sokal and Oden 1995; François et al. 2010; Arenas et al. 2013); this
study comments on the assumption that has been essential to the value and applicability
of those past works by providing preliminary evidence that clines, once established, can
in fact dissolve. As such, this study calls into question anthropological and
anthropological genetics research from the past two decades in which colonization
models were precipitated upon assumptions of cline persistence.

Chapter one provides a more detailed analysis of this body of literature and the
theoretical underpinnings and implications of this study. Chapter two notes the materials
used and introduces the methods employed, while chapter three treats their application.
Chapter four presents the results and discussion. Lastly, the Appendices provide all input
scripts for Nemo, analysis scripts, and additional figures.
OBJECTIVE AND HYPOTHESES

Ultimately, the purpose of this study was to investigate the contributions of migration, fecundity, and carrying capacity to the development of clinal patterns of diversity. It is expected, from theoretical and empirical studies of range expansions, that clinal distributions of diversity may result. Clines, as previously noted, are gradient distributions of diversity across space. In the context of a range expansion, it is expected that a cline indicates a decline in diversity across space from source to terminus. With regard to diversity statistics, it is generally expected that range expansions result in clinal patterns of heterozygosity and Fixation indices ($F_{st}$) (Excoffier, Foll and Petit 2009). Data were collected as genotypes for 10 biallelic Single Nucleotide Polymorphism (SNP) loci for all adult individuals in all 1,024 demes in the metapopulation. Data were collected at the first generation (generation 60) at which each deme had at least 25 adult individuals, and the final generation (either 500 or 750).

The research question was: How do carrying capacity, fecundity, and migration rates interact to generate clines? The purpose of this study was to establish, via forward-simulations, the contributions of migration, fecundity, and carrying capacity to the development of clinal patterns of diversity. The hypotheses tested were:

1. Clines will form more readily with lower carrying capacities, fecundities, and migration rates than with higher carrying capacities, fecundities, and migration rates.

2. Cline presence will be impacted by later bottleneck events.
CHAPTER 1—BACKGROUND, THEORY, AND LITERATURE REVIEW

The processes by which clines form and dissolve during continental range expansions have not yet been fully elucidated or statistically evaluated within population genetics; despite several promising attempts via studies utilizing simulation software, the contributions of carrying capacities, fecundity, and migration rates to the development of clines are still under investigation, and the dissolution of clines has not yet been addressed (Edmonds, Lillie and Cavalli-Sforza 2004; Klopstein, Currat and Excoffier 2006; Wegmann, Currat and Excoffier 2006; Excoffier, Foll and Petit 2009; Hofer et al. 2009; Itan et al. 2009; Ray and Excoffier 2009; Gerbault et al. 2009, 2011; Peng, Kimmel and Amos 2012; Slatkin and Excoffier 2012; Antoniazza et al. 2014). This thesis studies, via forward simulations, the contributions of migration, fecundity, and carrying capacity to the development of clinal patterns in the context of range expansions.

SIMULATIONS

Simulation programs are collections of scripts, which are code texts (computer instructions written in particular computer languages), and simulations are simply the
execution (‘running’, or processing) of those coded instructions, where the code creates and models data according to its instructions. Simulations allow scholars several advantages over empirical data: there are no gaps or missing data; data can be simulated for any place or time; any type and number of data can be simulated and simulated reproducibly; and finally, all manner of conditions and scenarios regarding human evolutionary history may be modeled (Hoban, Bertorelle and Gaggiotti 2012; Peng, Kimmel and Amos 2012). With regard to the simulation of genetic data and evolutionary processes, there are two main methods of simulation: forward and backward. A more nuanced explanation of the two may be found in the literature, but generally forward simulators evolve data forward in time, while backward simulators move back in time to a coalescent event (the point in time when lineages coalesce or meet at the most recent common ancestor). Backwards simulators provide insights into the age of lineages and their phylogenies and the evolution of lineages, while forward simulators model the evolution of individuals. Forward simulators can be ‘seeded’ or initialized with either starting parameter values or data previously simulated in a backwards simulator. The primary differences between the two in terms of modeling evolutionary processes are that forward simulators can model selection much more effectively than can backwards simulators, forward simulators are slower and track data for individuals at chosen points in time, backward simulators are faster and do not track individuals but are coalescence based and so are useful for answering questions about lineages (Hoban, Bertorelle and Gaggiotti 2012; Peng, Kimmel and Amos 2012). Investigations of complex demographic processes where selection is modeled require the use of forward simulators, as backwards simulators are mostly based in simple coalescence algorithms that have not added
selection (Hoban, Bertorelle and Gaggiotti 2012; Peng, Kimmel and Amos 2012). Simulations have been well vetted as a class of methodologies within population genetics, and even within anthropological genetics they have been used at least since the mid to late 20th century to test hypotheses for which data were limited due to time scales or difficulties in field collection (Rendine, Piazza and Cavalli-Sforza 1986; Barbujani, Sokal and Oden 1995; Fix 1993, 1996, 1997, 1999, 2004; Peng and Kimmel 2005; Klopfstein, Currat and Excoffier 2006; Peng and Amos 2008, 2010; Arenas 2012; Peng, Kimmel and Amos 2012; Arenas et al. 2013). However, not all simulators are designed equally; the language in which software is written can impact the speed and functionality of the simulators. Additionally simulators often are designed for use within a specific study, so not all programs are useful for simulating evolutionary scenarios allowing for control over the size and number of populations, the types of loci, the demographic scenarios possible, the types of selection regimes, and other key components (Hoban, Bertorelle and Gaggiotti 2012). Despite their general use becoming commonplace within population and anthropological genetics, few studies have been conducted utilizing software capable of modeling both complex demography and selection—making the methods used here novel due to the recentness of the software and its revolutionary flexibility for modeling complex human evolution.

Just as simulators may be limited by the language in which they are written and in their respective base algorithms (forward or backward), so too may they be limited by their application. Their validity as a tool of analysis and data generation for the study of complex evolutionary processes in humans and other organisms is not debated; however, their particular applications are. Simulations as a class of methods provide researchers the
opportunity to generate and test large data sets based in/applied to specific models of evolutionary processes and events. Models both enhance and limit our ability to test hypotheses. Models are meant to provide scenarios that are testable, and applicable; the more complex the model, the more applicable it is to a specific context, but the less generalizable it is and less applicable to other contexts. And the inverse can also be said: the more generalizable the model, the less complex it is, and it is less likely to be a realistic and accurate representation of specific systems represented by real data (Fix 1999; Hoban, Bertorelle and Gaggiotti 2012). The balance between generalizability and complexity for models is especially problematic for simulation studies—where it is assumed that the computational progress lying underneath in itself ensures the accuracy of the results. In fact, the results of simulation studies may be heavily biased by parameter choices, such that the parameter choices utilized in the simulations must be accurate or appropriate if the results can be, and by extension—the parameter choices must be appropriate in order to effectively test the hypothesized model. One method scholars have utilized to address the problem of parameter choice making is ABC. ABC, or Approximate Bayesian Computation, is a means of estimating parameters of new data from prior distributions of previous data and statistically evaluating probabilities in a computationally efficient and effective manner. ABC is particularly effective for investigating which parameters are driving complex evolutionary processes. It has been widely applied to coalescence modeling and many population genetics questions regarding selection and complex demography (Beaumont, Zhang and Balding 2002; Beaumont and Rannala 2004; Csilléry et al. 2010). More recently, ABC has been used as a means to model coalescence processes to better answer such questions as the timing of
colonization events. However, ABC applications to the spatial distributions of genetic data have been limited (Antoniazza et al. 2014). Effective and appropriate use of ABC methods require exponentially increased computational time and resources, as well as appropriate integration with simulation software—all of which make ABC methods beyond the scope of many studies, including this current one (Beaumont, Zhang and Balding 2002; Beaumont and Rannala 2004; Csilléry et al. 2010; Antoniazza et al. 2014). Additionally, while ABC methods have increased effectiveness in parameter selection due to the rapid generation and estimation of parameter values that are more likely than others—this still does not address the problem of choosing an appropriate range of parameter values. For choosing the appropriate range of parameter values from which to select actual parameter values—regardless of method of value selection—the best course is to choose values based on actual data from the population(s) of interest. Another option is to choose values used in previous simulation studies to replicate past results or explore null hypotheses. This study reviewed parameter values from past simulation studies and compared those values to values estimated from studies of archaeological remains, equations of population growth patterns, and ethnographic and demographic data from living human populations of varied subsistence and stratifications; from here, a range of possible values was defined and parameter value choices made.

Simulations are a critical class of methods for testing hypotheses within anthropological genetics for which data cannot be collected in the field—and are particularly crucial for studying the colonization history of *Homo sapiens* in continental contexts. Travis et al. (2007) applied colonization theory through expanding our current understanding of allele surfing by demonstrating through spatially explicit, individual-
based simulations that even highly deleterious mutations could surf during an expansion event—findings which have implications for additional studies on the apparently high frequency of rare, deleterious mutations in non-African populations, often cited as evidence of recent range expansions out of Africa (Flaxman 2013; Peischl et al. 2013; Peischl et al. 2015; Peischl and Excoffier 2015). Fix (1996) utilized simple simulations of a linear stepping stone colonization process to challenge previous assumptions regarding selection and the timing of events during range expansions to demonstrate that varied temporal experiences of selection could lead to the development of clines similar to those resulting from drift during expansion. Ray et al. (2005) simulated the peopling of all major continents to test the goodness of fit between empirical data and the multi regional and unique origins models, with their results strongly supporting recent out of Africa origins. Additionally, some early examples of simulations within anthropological genetics were conducted by Barbujani, Sokal and Oden (1995) and Rendine et al. (1986), who used simulations to test hypotheses for the peopling of Europe and found the greatest support for Neolithic colonization models. While many of these papers using simulations to test hypotheses within anthropological genetics focused on the peopling of Europe or the Out of Africa and Multiregional models, recent papers are branching into other regions. Di, Sanchez-Mazas and Currat (2015) used spatially explicit simulations with Approximate Bayesian Computation to test two different hypotheses for the peopling of Asia, concluding that the best supported model was two separate routes of colonization, and Currat, Poloni and Sanchez-Mazas (2010) used simulations of range expansion models to critique the lack to attention paid to the Strait of Gibraltar as a potential route of colonization of Europe and effectively demonstrate its importance.
Studies of spatial distributions of genetic information at continental scales often rely upon colonization theory, a sub-area of population genetics focused on the spatial/geographic and genetic distributions of diversity—differences in the frequency or number of mutations at a locus or set of loci—that result from expansion events whereby a population or groups of populations settles and fills a territory (Hofer et al. 2009; Excoffier, Foll and Petit 2009; Jay et al. 2013; Jobling et al. 2014). Colonization theory provides the theoretical framework for modeling range expansions and analyzing their effects on clinal spatial distributions. The body of work utilizing colonization theory to model these events and processes is ever expanding.

Typically, the expansion events of interest are demographic expansions and range expansions. Demographic expansions are characterized by an increase in a given population’s size, increased heterogeneity, and the ‘freezing’ of diversity due to the reduced ability of drift to operate in larger populations. This increased heterogeneity during a demographic expansion, despite the freezing of diversity, occurs as a function of an increase in rare variants during the expansive population growth, and is often represented in the form of star-like phylogenies, or many branches off a node, usually with shallow roots (Excoffier, Foll and Petit 2009). Range expansions are spatial expansions of the area in which a population mates and lives, and typically result in a loss of diversity across space (Excoffier, Foll and Petit 2009). It should be noted that demographic and range expansions may occur simultaneously. Additionally, it should be understood that these are idealized representations of pure expansions; reality is far more complex and events such as range contractions, demographic collapses and re-expansions (also known as bottlenecks), or even serial bottlenecking (small deme-sized bottlenecking events...
within the context of a range expansion) have been demonstrated to have commonly occurred in human populations throughout human evolution (Fix 1993, 1996, 1997, 1999, 2004; Excoffier and Schneider 1999; Ray, Currat and Excoffier 2003; Moreau et al. 2011; Excoffier, Foll and Petit 2009; Ray and Excoffier 2010; François et al. 2010; Arenas et al. 2013). Expanding populations may encounter heterogeneous landscapes, leading to heterogeneous expansion waves due to ‘resistance’ or ‘friction’ of the landscape preventing the wave from moving forward at a homogeneous rate (Ray et al. 2005; McRae 2006; Ray and Excoffier 2010). Selection may also occur during any of these demographic events.

Critical to any investigation of past human expansion events are two processes which impact the spatial distributions of alleles within the population. First, the ‘wave of advance’ model put forth by Fisher (1937) describes the process driving an allele spatially across a population due to selective pressures favoring the allele and gene flow distributing the allele across continuously distributed demes. In Fisher (1937)’s model, there was no spatial expansion, and selection and gene flow, not drift, is the force moving alleles across. Fisher’s model was adapted by Ammerman and Cavalli-Sforza, early pioneers of anthropological genetics, who altered Fisher’s original model to instead model the movement of an allele across space due not to selection but to the physical movement of people across space—or a ‘demic diffusion’ (Ammerman and Cavalli-Sforza 1984). The demic diffusion model of Ammerman and Cavalli-Sforza (1984) differed from Fisher’s in focusing on large, demographically and spatially expanding populations; genes were, in this model, simply spread by people physically moving from one place to another, not through selection pushing a variant through a static, continuous
population. Fisher’s model allowed for alleles to move across space without massive spatial population shifts, while such shifts were required for Ammerman and Cavalli-Sforza (Fisher 1937; Ammerman and Cavalli-Sforza 1984).

However, Fisher (1937) did not consider the role of drift in moving alleles across space, and Ammerman and Cavalli-Sforza (1984) did not consider the potential for an allele to move across space dynamically within a range expansion event. Within the last two decades, scholars experimentally demonstrated the existence of a process by which alleles were distributed during range expansion events can move across space by drifting dynamically within the range expansion event itself. This ‘allele surfing’ causes alleles to reach higher than expected frequencies during a range expansion event if the mutations are close enough to the wave front, the size and density of the expanding populations are moderate, and the expansion occurs swiftly—in essence, the surfing mutations push through the wave faster than the wave itself because of the enhanced role of drift (Edmonds, Lillie and Cavalli-Sforza 2004; Klopfstein, Currat and Excoffier 2006; Excoffier and Ray 2008; Excoffier, Foll and Petit 2009). Within the context of allele surfing, the ‘wave front’ is referring to the edge of the range expansion—to the populations at the front end of the expansion event (Klopfstein, Currat and Excoffier 2006; Excoffier and Ray 2008; Excoffier, Foll and Petit 2009). In simulations of the surfing phenomenon, surfing occurs regardless of whether the mutations are novel or preexisting, neutral or under selection—with those alleles found closer to the wave front faring better (Klopfstein, Currat and Excoffier 2006). It has been argued that mutations at the wave front are favored simply as a matter of density dependence, where mutations ‘build’ up at the wave’s edge, leading to a homogenization at the wave edge that make it
difficult for alleles found more commonly further from the wave’s edge to advance spatially (Waters, Fraser and Hewitt 2013). This is in essence an argument that the dynamics of the range expansion a) lead to the fixation of mutations at the edge, and b) that fitness must somehow be increased enough at the wave edge to allow for fixation to occur; fitness coefficients conditioned by deme position within the wave or wave front in addition to the number of deleterious mutations has been termed the ‘expansion load’ of the population (Peischl et al. 2015; Peischl and Excoffier 2015)—and many more simulation studies are needed to confirm these seemingly contradictory results regarding the effects of expansion load on fitness, fixation, and general likelihood of allele surfing. Although fitness is typically understood as the effect of selection upon a phenotype, it is modeled in these simulations as the effect of selection upon the individual and the individual’s ability to successfully pass mutations on to offspring; for simulations of quantitative traits modeling fitness is relatively straightforward as phenotypic distributions themselves are modeled, but for non-quantitative traits without phenotypic distributions, selection, and therefore fitness, are directly affecting the mutations (Neuenschwander, Guillaume and Goudet 2008; Flaxman 2013; Peischl et al. 2015; Peischl and Excoffier 2015).

Several studies of selection, surfing, and expansion load during range expansions have been conducted via simulations, and have generally concluded that fitness coefficients are reduced by increased ‘mutational load’ or build up of deleterious mutations at the wave front (which speaks to the power of allelic surfing, as deleterious mutations would not be expected to increase in frequency while simultaneously reducing fitness without this powerful form of drift) (Flaxman 2013; Peischl and Excoffier 2015;
Peischl et al. 2015); while others have noted that fitness coefficients may increase at the wave front if the mutations are beneficial (Peischl et al. 2015); and still further studies have noted that surfing is predictable solely on a mutation’s wave position (Travis et al. 2007), and by deme size (Klopfstein, Currat and Excoffier 2006). In fact, Klopfstein et al. (2006) noted that the likelihood of surfing increases with growth rate and decreases with increased deme size and migration rate. Increased likelihood of surfing has been shown to associate positively with increased likelihood of cline formation during range expansion events (Klopfstein, Currat and Excoffier 2006).

CLINES

Clines are gradations across space of genetic information (Haldane 1948)—often visualized as gradients of allele frequencies, Fst, or heterozygosity. These gradation patterns can be established through time across space by drift or natural selection, or both, increasing the frequency of a variant; they are maintained over time through a balance between those forces and gene flow (Slatkin 1973; Slatkin and Maruyama 1975; Endler 1977). Clines can also change across time and space, either through shifting in shape or location or by forming/dissolving, though the context of their establishment may affect their future shape, location, and presence (Endler 1977). Additionally, as either neutral or selective forces may establish clines, these spatial patterns of diversity are nonspecific indicators of past evolutionary processes and population history (Endler 1977). It is possible to demonstrate a cline as representative of specific evolutionary
processes if there is strong evidence for natural selection, or through the use of simulations and complex models (Endler 1977).

The theory of clines developed from a lengthy debate on speciation and hybridization, where clines were initially conceptualized of as occurring as a result of hybrid zones during hybridization or admixture (Haldane 1948); however, it was noted that cline theory could also be applied to continuous populations (Endler 1977). This idea of a gradation in variation forming within a ‘contact zone’ was rooted in conceptions of barriers. That said, it has been shown both empirically and through simulations that clines form within a species as well (Haldane 1948; Slatkin 1973; Slatkin and Maruyama 1975; (Endler 1977). Cline formation in simulations is directly related to migration models, because clines form and are maintained through a balance between drift and gene flow, and these models are approximations of gene flow (Endler 1977). There are two major types of migration models, discrete and continuous. In discrete migration models, populations are subdivided into breeding units or demes defined by size, and mating occurs as a function of deme membership. In continuous migration models, populations are grouped into neighborhoods defined by size and distance, but mating is not considered to occurs between distinct breeding units, but as a function of distance (Fix 1999).

Under the assumptions of a discrete island model, this is somewhat simple to see: if islands are arranged in a linear fashion, those farthest away from the hypothesized ‘source’ continental region will experience the least amount of gene flow, and hence, will become more differentiated from the islands closest to the continental region (Endler 1977; Fix 1999). Under the assumptions of a continuous isolation by distance model,
subpopulations are not separated by physical barriers other than distance, though there are some complexities: in Wright (1943)’s isolation by distance model, mating distances follow a normal distribution (so mating from far away distances is still possible but uncommon) and are conditioned by neighborhood sizes and distances, and genetic distances are correlated by geographic distances; whereas in Malecot (1973)’s isolation by distance model mating distances follow a negative exponential distribution (mating is almost exclusively occurring within a very narrow distance and declines rapidly, so that mating farther away is possible but extremely rare) and there is an increased probability of increased genetic distance with increased geographic distance (Endler 1977; Fix 1999).

Differentiation, and therefore clines, occur due to a decline in gene flow with distance in isolation by distance models, with the greatest differentiation occurring when those distances are minimal (Endler 1977; Fix 1999). While isolation by distance is a well-known migration model in population genetics, many population genetics simulation studies have relied upon one or two-dimensional stepping stone models or migration matrices. Stepping stone models are similar to the island model with the added constraint that mating does not occur outside of the nearest neighbors. These models are limited by the problem of isotropic migration, which allows these models to be generalizable but is not realistic for living populations—a problem addressed by using migration matrices, which are more specific because they are specific migration rates calculated between individual pairs of demes from field data on human population movements. Migration matrices also suffer from problems, as they are not generalizable or really extensible beyond a few generations (Endler 1977; Fix 1999).
With regard to clines then, a decline in genetic distance with geographic distance in island or isolation by distance models is expected to lead to a decline in gene flow with distance, and hence, the formation of a cline; within two dimensions, such as under the assumptions of a continental model (such as a two dimensional stepping stone model), the impact of a decline in gene flow is expected to be more significant (Endler 1977; Fix 1999). The decline in gene flow across distance has been shown to result in an increase in genetic distance across geographic distance—which has been shown to form the geographic gradient or cline (Endler 1977; Fix 1999).

Clines are identified in data statistically and visually. Many methods for cline identification have relied upon the concept of spatial autocorrelation, or spatially dependent data (Barbujani 1987). The scenario where data spatially close are expected to be more similar, and this correlation declines with distance is described as positive spatial autocorrelation, while a negative correlation between values and distances is referred to as negative autocorrelation, and randomly dispersed data are not spatially autocorrelated (Cliff and Ord 1968; Barbujani 1987). Significant positive spatial autocorrelation indices are considered to indicate the presence of a cline as a correlation of similar values at closer proximity, and declining with distance, mirrors the cline. One of the most common methods for identifying a cline using spatial autocorrelation analysis is through the calculation of a spatial autocorrelation index, typically either Moran’s I (Moran 1950) or Geary’s C (Geary 1954). The two indices both test for positive, negative, or no autocorrelation, and have a range of -1 to +1 and 0 to 2, respectively (Moran 1950; Geary 1954). While either may be used, Moran’s I may be preferable as a global index of spatial autocorrelation as it has been noted to be more accurate at larger scales, while Geary’s C
is preferred at finer scales (Moran 1950; Cliff and Ord 1968; Barbujani 1987). In addition to calculating spatial autocorrelation indices, these indices can be plotted against bins of distances as spatial correlograms (Oden and Sokal 1986) to visually inspect clines. Clines have also been identified quantitatively and visually through the calculation of linear regressions (Griffith 2000; Dray 2011).

CARRYING CAPACITY, MIGRATION, GROWTH, AND BOTTLENECKS

Carrying capacity is a measure of the number of individuals that can be supported in an area due to subsistence strategy and resource availability (Fix 1999). Migration rates are approximations of gene flow modeled as the proportion of the population that is comprised of immigrants (Endler 1977). Growth as modeled here is the rate of population growth (r) per generation, or: (net new population size / net initial population size) / net initial population size, where net size is the size of the population after all births and deaths (Endler 1977; Fix 1999; Guillemette and Rougemont 2006). Bottleneck events are population contractions or reductions in population size (Excoffier, Foll and Petit 2009). In a two dimensional stepping stone model, the impact of gene flow is a function of migration rates among neighboring demes, growth rates of demes, and metapopulation size (Endler 1977; Fix 1999). Clines calculated from allele frequencies were more found to be more likely to occur when carrying capacities were lower as opposed to larger—though allele frequency clines can still form even when carrying capacities are higher, though infrequently (Klopfstein, Currat and Excoffier 2006). This makes sense theoretically, as smaller demes would experience greater drift than larger demes,
particularly during the context of a range expansion; clines can still form even when demes have higher carrying capacities in the context of expansion events due to isolation by distance (Endler 1977; Fix 1999; Excoffier, Foll and Petit 2009; Slatkin and Excoffier 2012). On the basis of theoretical and simulation studies it is expected that smaller populations will experience greater drift, that higher growth rates will result in demographic growth and larger populations (thereby reducing the ability of drift to cause differentiation between demes) more readily than would lower growth rates (Endler 1977; Fix 1997, 1999; Excoffier, Foll and Petit 2009). Additionally, it is expected, and has been shown, that lower rates of migration restricted to close neighbors results in differentiation between demes and promotes cline formation while long distance dispersal maintains diversity across space (Endler 1977; Fix 1999; Ray and Excoffier 2010). As such, these parameters are critical to the formation and dissolution of clines during and after range expansion events (Endler 1977; Fix 1999). Bottleneck events can also lead to dramatic shifts in the numbers of alleles relative to heterozygosity and may affect signals left by demographic and range expansion events (Maruyama and Fuerst 1985).
CHAPTER 2—MATERIALS & METHODS

The simulator environment Nemo (Guillaume and Rougemont 2006) was used to test the hypotheses. A more detailed description of Nemo may be found in chapter three; this chapter details the parameter values chosen to test in simulation models. Additionally, this chapter highlights the diversity indices utilized during analyses. See chapter three for the application of software and the simulation environment. Forward simulations were carried out in Nemo v.2.3.45 (2015) (Guillaume and Rougemont 2006). These data are in the form of SNP genotypes for each adult individual in each occupied patch, for each run of every simulation, in the FSTAT format for neutral mutations. Nemo .ini files were written with some parameterization to allow for batch processing. A bash/R script was used to pre-process the .dat files and convert them from FSTAT-extended to FSTAT. PGDSpider (Lischer and Excoffier 2012) was utilized for converting the FSTAT data files into arlsumstat input files. All analyses were performed in arlsumstat (Excoffier and Lischer 2010), and in R\(^1\) v. 3.0 (R Core Team 2013). Data generation, conversion, and analyses were completed using the Linux-based Cardinal Research Cluster computing network at the University of Louisville, Information Technology Department.

\(^1\) See methods for a detailed discussion of parameter choices and defaults for the simulations. For a full listing of all R packages utilized, confer the References.
Previous simulation studies focused on the peopling of Europe, from 1986 through 2013 were surveyed for migration rates. The results of this survey are found in table 1 below, ordered ascending by year. The rates refer to the discrete parameter values unless otherwise explicitly noted as a range.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Migration Model</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rendine, Piazza and Cavalli-Sforza 1986</td>
<td>2DSS*</td>
<td>0.04, 0.25, 0.20</td>
</tr>
<tr>
<td>Barbujani, Sokal and Oden 1995</td>
<td>2DSS</td>
<td>0.065</td>
</tr>
<tr>
<td>Fix 1996</td>
<td>1DSS</td>
<td>0.10</td>
</tr>
<tr>
<td>Ray, Currat and Excoffier 2003</td>
<td>2DSS</td>
<td>0.01, 0.02, 0.05, 0.1, 0.5</td>
</tr>
<tr>
<td>Edmonds, Lillie and Cavalli-Sforza 2004</td>
<td>2DSS</td>
<td>0.05, 0.1, 0.2, 0.4</td>
</tr>
<tr>
<td>Currat and Excoffier 2005</td>
<td>2DSS</td>
<td>0.25</td>
</tr>
<tr>
<td>Wegmann, Currat and Excoffier 2006</td>
<td>2DSS</td>
<td>0.05, 0.2</td>
</tr>
<tr>
<td>Klopfstein, Currat and Excoffier 2006</td>
<td>2DSS</td>
<td>0.1-0.4</td>
</tr>
<tr>
<td>François et al. 2010</td>
<td>2DSS</td>
<td>0.2, 0.5, 0.8</td>
</tr>
<tr>
<td>Arenas et al. 2012</td>
<td>2DSS</td>
<td>0.04, 0.13, 0.2, 0.23</td>
</tr>
<tr>
<td>Arenas et al. 2013</td>
<td>2DSS</td>
<td>0.4</td>
</tr>
</tbody>
</table>

*Two-dimensional stepping stone model

In comparison with these parameter value choices used in simulation studies of farmer and hunter-gatherer expansion events in Europe, are field data, simulations, and comparative studies conducted by Alan Fix (1996, 1999). These empirically derived data demonstrate that a migration rate of 0.2 is more appropriate for human populations (Fix
1996, 1999). In order to model realistic and high (but still realistic) migration rates, we have chosen the parameter values of 0.2 and 0.3.

With the exception of (Edmonds, Lillie and Cavalli-Sforza 2004), all other studies referenced for migration rates were also referenced for growth rates. The results of this survey are found table 2 below, ordered ascending by year. The rates refer to the discrete parameter values unless otherwise explicitly noted as a range.

<table>
<thead>
<tr>
<th>References</th>
<th>Growth Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rendine, Piazza and Cavalli-Sforza 1986</td>
<td>0.25, 0.5</td>
</tr>
<tr>
<td>Barbujani, Sokal and Oden 1995</td>
<td>0.5</td>
</tr>
<tr>
<td>Fix 1996</td>
<td>0.1</td>
</tr>
<tr>
<td>Ray, Currat and Excoffier 2003</td>
<td>0.1</td>
</tr>
<tr>
<td>Currat and Excoffier 2005</td>
<td>0.6, 0.8</td>
</tr>
<tr>
<td>Wegmann, Currat and Excoffier 2006</td>
<td>0.5</td>
</tr>
<tr>
<td>Klopfstein, Currat and Excoffier 2006</td>
<td>0.2-0.8</td>
</tr>
<tr>
<td>François et al. 2010</td>
<td>0.4, 0.5, 1</td>
</tr>
<tr>
<td>Arenas et al. 2012</td>
<td>0.6, 0.8, 1</td>
</tr>
<tr>
<td>Arenas et al. 2013</td>
<td>0.4, 0.5</td>
</tr>
</tbody>
</table>

As with the migration rates surveyed, work conducted by Alan Fix (Fix 1999, 1996) demonstrated that growth rates of 0.5 were unreasonably high for all but industrial and post industrial populations; instead, it was noted that growth rates around 0.1 were more realistic for pre industrial populations. Growth rates for this study were approximately 0.1 and 0.5 and were set using mean fecundity values of three or four offspring with a 30%
population-wide die-off rate per generation plus logistic regulation through carrying capacity. This die-off rate was used to keep growth rates around constants of 0.1 and 0.5 for mean fecundities of 3 and 4, respectively.

The same references were surveyed for carrying capacities as well, also excluding Edmonds, Lillie and Cavalli-Sforza (2004). The results of this survey are found in table 3 below, ordered ascending by year. The rates refer to the discrete parameter values unless otherwise explicitly noted as a range.

<table>
<thead>
<tr>
<th>References</th>
<th>Carrying Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rendine, Piazza and Cavalli-Sforza 1986</td>
<td>300, 8000</td>
</tr>
<tr>
<td>Barbujani, Sokal and Oden 1995</td>
<td>114, 7560</td>
</tr>
<tr>
<td>Fix 1996</td>
<td>300</td>
</tr>
<tr>
<td>Ray, Currat and Excoffier 2003</td>
<td>100, 500, 1000, 5000, 10000</td>
</tr>
<tr>
<td>Currat and Excoffier 2005</td>
<td>800</td>
</tr>
<tr>
<td>Wegmann, Currat and Excoffier 2006</td>
<td>50, 100, 200, 1000</td>
</tr>
<tr>
<td>Klopfstein, Currat and Excoffier 2006</td>
<td>10-500</td>
</tr>
<tr>
<td>François et al. 2010</td>
<td>500, 1000</td>
</tr>
<tr>
<td>Arenas et al. 2012</td>
<td>50, 100, 200</td>
</tr>
<tr>
<td>Arenas et al. 2013</td>
<td>500</td>
</tr>
</tbody>
</table>

For these studies explicitly modeling hunter-gatherer or farming populations, the lower carrying capacities (less than 500) were assumed to be appropriate for modeling hunter gatherer populations, and values up to 10,000 were considered appropriate for modeling farming populations. This assumption that hunter gatherer population carrying
capacities were dramatically lower than early farming population carrying capacities has been challenged (Gamble et al. 2005; Bocquet-Appel et al. 2005). As such, carrying capacities of 200 and 500 were chosen as more realistic carrying capacities for hunter-gatherer/early farmer populations.

SNP genotype data were analyzed through the calculation of allele frequencies in arlsumstat (Excoffier and Lischer 2010). Allele frequencies were calculated for each replicate and then averaged across all replicates. Heterozygosities were calculated from the averaged allele frequencies, averaged across all loci, and then plotted in lattices as heat maps of high to low heterozygosity across space in R. Heat maps were also plotted for heterozygosity calculated at the first locus for each patch. This additional set of heat maps was generated to investigate the implications of analyzing heterozygosity at a single locus versus a set of loci.
CHAPTER 3—EXPERIMENTAL DESIGN

Materials and methods outlined in chapter two were applied using a complex pipeline moving from the simulator environment in Nemo (Guillaume and Rougemont 2006) through the command line, PGDSpider (Lischer and Excoffier 2012), and finally on to arlsumstat (Excoffier and Lischer 2010) and R (R Core Team 2013) for analyses. This chapter describes each step of the analysis pipeline including the methods for applying the software, necessary inputs and outputs, and methods for ensuring data integrity. Scripts used in the pipeline are located in the Appendices.
F1. Analysis Pipeline

The choice of simulator was initially made upon reviewing figures in Hoban, Bertorelle and Gaggiotti (2012) of populations genetics simulation environments. This choice was further confirmed and refined through utilizing the National Cancer Institute: Division of Cancer Control & Population Science’s Genetic Simulation Resources Compare Software Resources tool (Peng et al. 2015), available at popmodels.cancercontrol.cancer.gov/gsr/search/. This tool proved useful in refining the choice of simulator because it provided details for additional attributes not noted by Hoban, Bertorelle and Gaggiotti (2012), such as license, platform, and language. The simulator chosen for the current study was selected according to the following desired
attributes and capabilities. Initial preferences of simulators were simuPOP, Nemo, and quantiNEMO; these simulators were also selected by the National Cancer Institute: Division of Cancer Control & Population Science’s Genetic Simulation Resources Compare Software Resources tool algorithm.

<table>
<thead>
<tr>
<th>Selection Attributes Choice</th>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Simulated Data</td>
<td>Genotype at Genetic Markers, Diploid DNA Sequence</td>
</tr>
<tr>
<td>Variations</td>
<td>Biallelic Marker, Single Nucleotide Variation</td>
</tr>
<tr>
<td>Simulation Method</td>
<td>Forward-time</td>
</tr>
<tr>
<td>Output Data Type</td>
<td>Genotype or sequence, Demographic</td>
</tr>
<tr>
<td>Output File Format</td>
<td>Fstat</td>
</tr>
<tr>
<td>Population Size Changes</td>
<td>Exponential Growth or Decline, Logistic Growth, Bottleneck, Carrying Capacity, User Defined</td>
</tr>
<tr>
<td>Gene Flow</td>
<td>Stepping Stone Models, Influence by Environmental Factors, User-Defined Matrix</td>
</tr>
<tr>
<td>Life Cycle</td>
<td>Discrete Generation Model</td>
</tr>
<tr>
<td>Mating System</td>
<td>Random Mating, Monogamous</td>
</tr>
<tr>
<td>Natural Selection Determinant</td>
<td>Single-Locus</td>
</tr>
<tr>
<td>Natural Selection Models</td>
<td>Directional Selection</td>
</tr>
<tr>
<td>Mutation Models</td>
<td>K-allele Model</td>
</tr>
<tr>
<td>Events Allowed</td>
<td>Population Merge and Split, Varying Demographic Features, Population Events</td>
</tr>
<tr>
<td>Interface</td>
<td>Command-Line, Script Based</td>
</tr>
<tr>
<td>Tested Platforms</td>
<td>Mac OS X, Linux and Unix</td>
</tr>
<tr>
<td>Language</td>
<td>C or C++, R, Python</td>
</tr>
<tr>
<td>License</td>
<td>GNU Public License</td>
</tr>
<tr>
<td>GSR Certification</td>
<td>Accessibility, Documentation, Application, Support</td>
</tr>
</tbody>
</table>
According to the results of the compare software resources tool as presented in table 4, the most flexible and appropriate simulators were simuPOP (Peng and Kimmel 2005; Peng and Amos 2008, 2010; Peng, Kimmel and Amos 2012); Nemo (Guillaume and Rougemont 2006), and quantiNEMO (Neuenschwander, Guillaume and Goudet 2008); in descending order of best fit. The initial final decision lay between quantiNEMO and simuPOP; while quantiNEMO had a more robust user interface and required only configuration input files, it lacked the vast flexibility of simuPOP which used user-written python scripts to control all simulation processes (Peng and Kimmel 2005; Neuenschwander, Guillaume and Goudet 2008; Peng and Amos 2010; Peng, Kimmel and Amos 2012). Additionally, simuPOP had a developed website and full documentation of all classes and functions, and an active mailing list—all serious pros to the con of a sharp learning curve. However, after several months of developing test scripts and testing range expansion models in simuPOP, it became apparent that even though simuPOP is an incredibly flexible platform, the ‘gluing’ together of scripts resulted in a program that sometimes developed deep internal conflicts. One of these deep internal conflicts was revealed in range expansion test scripts; initial development tests showed that it was possible to run a range expansion model in simuPOP with very limited lattice sizes of 4 or 6 demes. However, upon scaling up the models to a full lattice of 1,000 demes, the simulator would repeatedly crash due to a failure of the program to produce enough
mating events in demes. simuPOP controls mating through mating functions, which rely on generators that will introduce an error and halt the simulations if there are fewer than two individuals available for mating, or in any other case where there were not enough males or females for mating; the mating function calls the demographic function (which controls the range expansion and logistic regulation of population growth). Even with high migration and growth rates, and extremely high population densities, simulations were halted unless more than half of the lattice was already populated (Peng and Kimmel 2005; Peng and Amos 2010; Peng, Kimmel and Amos 2012). This meant that it was not possible to run range expansion models in simuPOP at the resolution this study required. Upon this realization, it became apparent that another simulator was necessary.

quantiNEMO was the next choice pursued because it offered many of the same features as simuPOP, a detailed manual, and was user friendly (Neuenschwander, Guillaume and Goudet 2008). The next choice after quantiNEMO was Nemo—an earlier program also developed by the same working group, but more recently revised than quantiNEMO. Nemo was chosen over quantiNEMO because quantiNEMO does not yet allow for modeling of directional selection while Nemo does; while this current study did not investigate natural selection, incorporating selection will be important in future work (Guillaume and Rougemont 2006; Neuenschwander, Guillaume and Goudet 2008). Nemo lacks some of the more flexible components of quantiNEMO, but it contains the core features necessary for this study: the possibility of instituting a range expansion; of controlling population size with carrying capacity and fecundity; and of setting the migration rate and model (as well as custom matrices) (Guillaume and Rougemont 2006; Neuenschwander, Guillaume and Goudet 2008). Additionally, Nemo has a built system
for generating directory structures and data file organization, simplifying data management; for these reasons, Nemo was ultimately chosen as the study simulator environment (Guillaume and Rougemont 2006)

Simulations in Nemo (Guillaume and Rougemont 2006) were broken into two demographic models, plus an additional set of these same models instituting a bottleneck event. The lattice for all models was 32 rows by 32 columns with 1,024 demes. Ten loci with two alleles were simulated at a mutation rate of 0.00001 to model SNPs. Alleles were set at maximal variance at the start of all models. The two demographic models tested were: a control model where the entire lattice was full at the start of the simulation, and a range expansion model where three full rows and four columns along a corner and edges of the 1,024 deme lattice were populated at the start of the simulation, and the expansion moved outwards. Starting deme sizes for the control models were 100, and were 100 and 150 for range expansion models.

Demographic parameters are presented in table 5. Demographic parameters used were carrying capacity, migration rate, fecundity, and die-off probability. We used a two-stage model to investigate the implications of changing range expansion dynamics across time and a one-stage model to investigate the impact of specific parameters on range expansions. The two stage models used carrying capacities of 200 and 500, and ran for 750 generations, where the carrying capacity either increased or decreased to the other value at generation 250. Similarly, migration rates of 0.2 and 0.3 either increased or decreased in the two-stage models at generation 250. Fecundity increased from a mean of 3 to 4 at generation 250, and was regulated through an die-off rate with a constant 100% probability that 30% of the population would be randomly removed, regardless of
carrying capacity. Carrying capacity logistically regulated population growth. For the one-stage models, carrying capacity was either 200 or 500, migration was either 0.2 or 0.3, and fecundity was either a mean of 3 or 4. As with the two-stage models, die-off rate and probability remained constant at 100% that 30% of the population would randomly be killed off regardless of carrying capacity. All one-stage models were run for 500 generations. Twenty independent replicate runs of each simulation were performed for all simulations to account for stochasticity. Additionally, all one-stage models in which carrying capacity was 500 were re-run with an additional late bottleneck event from generation 450 to 455 of either 30 or 60% of carrying capacity before the carrying capacity was allowed to return to its previous level. The carrying capacities for the bottleneck events were chosen by calculating a reduction of 30 or 60% of the metapopulation size to approximate a severe population crash, such as that experienced in Eurasia during the Black Death (Gottfried 2010).

Other parameter settings utilized were held constant across all models. The general structure of the life cycle was: breeding, die-off, dispersal (migration), aging, resizing, and the timing of the calculation of statistics (see supplemental figure S36). The aging parameter removes all existing adults and uses the offspring generation for the new adult generation, and its placement prior to the statistics calls results in Nemo only recording adult data (Guillaume and Rougemont 2006). The breed model was set to monogamy with specified fecundity values at the mean of a Poisson distribution. Population growth was regulated logistically by carrying capacity and the parameter resize_do_regulate, which removed all excess individuals, and die-off of 30% of each deme each generation. The die-off rate of 30% per generation was set to hold the growth
rate constant at either 0.1 or 0.5 for mean fecundities of 3 or 4, respectively. Migration was modeled on a 2D stepping stone lattice with up to eight possible neighbors with reflective boundaries. Refer to the Appendices for all simulation scripts with all simulation inputs, for each set of models.

All data was generated using the Cardinal Research Cluster at the University of Louisville. Nemo development tests were performed on Mac OS X El Capitan prior to running complete scripts. These development tests (files not included) were performed to test the core functions of Nemo and to verify that all components worked correctly (i.e., that the software was capable of running a range expansion and that surfing could occur). Once development tests were successful, the final scripts were developed. All of the final scripts used are included in the Supplementary Documents. These scripts include all parameters used in the simulations and therefore results obtained are fully reproducible. Final scripts were uploaded to the Cardinal Research Cluster and run via bash job scripts calling the Nemo .ini configuration files and the Nemo executable. Upon the completion of the job scripts, log data files were inspected for runtime errors; if errors in files were detected, configuration files were reviewed, edited, and rerun as necessary. Runtime errors were addressed by increasing the estimations for resource requirements to a higher number of cores and longer run times. Upon completion and clean error log files, data directories were moved to a downloading directory. Log files, .ini files, and .sh job scripts were moved and filed into separate storage directories. Data directories within the downloading directory were compressed using a gnu tar command embedded in a bash job script to reduce directory sizes as a means of conserving space. Directories were
individually uncompressed for analyses and subsequently compressed upon the successful extraction of all necessary files.

Data files were extracted from uncompressed directories using the following methods. First, the generations text results file was downloaded for each range expansion simulation and processed in an R script to obtain the first generation at which the lattice was first filled, where the threshold for a deme being first filled was a deme size of 25 or more individuals. The simulation output files with that corresponding generation number were extracted and downloaded, as well as the corresponding files for the control simulations with the same parameter set. These chosen files were then processed via gawk scripts to convert from the FSTAT extended file format to FSTAT. Cleaned files were converted to arlsumstat project files via a bash script calling PGDSpider, a Java based conversion platform that can take over 20 different kinds of population genetics data files as input and then output them into another format (Lischer and Excoffier 2012). Arlsumstat project files for each simulation were then analyzed in batch in arlsumstat to calculate the average allele frequency in each patch across all replicates for each set of generation files (Excoffier and Lischer 2010). The arlsumstat results file containing the patch specific allele frequencies averaged across replicates for each set of generation files, for each simulation, was then downloaded and text processed and analyzed using R/gawk/awk scripts (R Core Team 2013). This script simply cleaned up the data and then calculated patch average heterozygosity from the ten loci and the patch heterozygosity at a locus, and then plotted these averaged patch heterozygosity in heat maps. Heat maps are a type of visualization tool where data are color coded by bins of values and plotted on a
grid. This makes them particularly useful as a visual analysis tool for spatial data and allows for the identification of regions or clusters of data by low to high values.
CHAPTER 4—RESULTS

This thesis investigated, via forward-simulations, the contributions of migration, fecundity, and carrying capacity to the development of clinal patterns of diversity. On the basis of theoretical and empirical studies of cline formation, it was hypothesized that clines would be most likely to form when range expansion dynamics included lower carrying capacities, lower fecundity, and lower migration rates, rather than higher carrying capacities, fecundity, or migration rates. Additionally, it was hypothesized that the presence of these patterns would be impacted by future bottleneck events. These hypotheses were tested through the analysis of heat maps of average heterozygosity and Moran’s I, where the calculation of patch heterozygosity were averaged across ten loci in a lattice of 32 by 32 or 1,024 demes. A second set of heat maps was developed for each simulation in which heterozygosities were calculated for a single locus.

Heat maps and Moran’s I were analyzed for the first generation at which the lattice filled and the final generation of each simulation. The table 5 contains the parameters and generation-filled information for each set of simulations. Refer back to this table when reviewing the following heat maps.
### Table 5 Parameter Values

Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30%; BN60 - Bottleneck Event of 60%; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

<table>
<thead>
<tr>
<th>Lattice Model</th>
<th>MR 1</th>
<th>MR 2</th>
<th>CC 1</th>
<th>CC 2</th>
<th>G 1</th>
<th>G 2</th>
<th>Gen Fill</th>
<th>Gen Done</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-stage models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range Expansion</td>
<td>0.2</td>
<td>0.3</td>
<td>200</td>
<td>500</td>
<td>3</td>
<td>4</td>
<td>70</td>
<td>750</td>
</tr>
<tr>
<td>RE</td>
<td>0.2</td>
<td>0.3</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>4</td>
<td>60</td>
<td>750</td>
</tr>
<tr>
<td>RE</td>
<td>0.3</td>
<td>0.2</td>
<td>200</td>
<td>500</td>
<td>3</td>
<td>4</td>
<td>60</td>
<td>750</td>
</tr>
<tr>
<td>RE</td>
<td>0.3</td>
<td>0.2</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>4</td>
<td>60</td>
<td>750</td>
</tr>
<tr>
<td>Control</td>
<td>0.2</td>
<td>0.3</td>
<td>200</td>
<td>500</td>
<td>3</td>
<td>4</td>
<td>70</td>
<td>750</td>
</tr>
<tr>
<td>CN</td>
<td>0.2</td>
<td>0.3</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>4</td>
<td>60</td>
<td>750</td>
</tr>
<tr>
<td>CN</td>
<td>0.3</td>
<td>0.2</td>
<td>200</td>
<td>500</td>
<td>3</td>
<td>4</td>
<td>60</td>
<td>750</td>
</tr>
<tr>
<td>CN</td>
<td>0.3</td>
<td>0.2</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>4</td>
<td>60</td>
<td>750</td>
</tr>
<tr>
<td><strong>1-stage models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range Expansion</td>
<td>0.2</td>
<td>200</td>
<td>3</td>
<td></td>
<td>70</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0.2</td>
<td>200</td>
<td>4</td>
<td></td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0.2</td>
<td>500</td>
<td>3</td>
<td></td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td></td>
<td>30</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0.3</td>
<td>500</td>
<td>3</td>
<td></td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0.3</td>
<td>500</td>
<td>4</td>
<td></td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottleneck 30%</td>
<td>0.2</td>
<td>500</td>
<td>350</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN30</td>
<td>0.2</td>
<td>500</td>
<td>350</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN30</td>
<td>0.3</td>
<td>500</td>
<td>350</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN30</td>
<td>0.3</td>
<td>500</td>
<td>350</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottleneck 60%</td>
<td>0.2</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN60</td>
<td>0.2</td>
<td>500</td>
<td>200</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN60</td>
<td>0.3</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN60</td>
<td>0.3</td>
<td>500</td>
<td>200</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.2</td>
<td>200</td>
<td>3</td>
<td></td>
<td>70</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.2</td>
<td>200</td>
<td>4</td>
<td></td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.2</td>
<td>500</td>
<td>3</td>
<td></td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td></td>
<td>50</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.3</td>
<td>200</td>
<td>4</td>
<td></td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.3</td>
<td>500</td>
<td>3</td>
<td></td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.3</td>
<td>500</td>
<td>4</td>
<td></td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottleneck 30%</td>
<td>0.2</td>
<td>500</td>
<td>350</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN30</td>
<td>0.2</td>
<td>500</td>
<td>350</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN30</td>
<td>0.3</td>
<td>500</td>
<td>350</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN30</td>
<td>0.3</td>
<td>500</td>
<td>350</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottleneck 60%</td>
<td>0.2</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN60</td>
<td>0.2</td>
<td>500</td>
<td>200</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN60</td>
<td>0.3</td>
<td>500</td>
<td>200</td>
<td>3</td>
<td>60</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN60</td>
<td>0.3</td>
<td>500</td>
<td>200</td>
<td>4</td>
<td>20</td>
<td>500</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Heat Maps with Discernible Patterns

Generally, it is expected that range expansions will lead to a loss of diversity and that bottleneck events will change any existing spatial patterns of diversity (Maruyama and Fuerst 1985; Excoffier, Foll and Petit 2009). In the context of range expansions, it is also expected that lower migration rates, fecundity, and carrying capacities will result in a greater loss of diversity across space than high migration rates, fecundity, or carry capacities and would thus be more associated with the formation of clinal patterns than would higher parameter values (Endler 1977; Fix 1999; Klopfstein, Currat and Excoffier 2006; Excoffier, Foll and Petit 2009; Ray and Excoffier 2010; Slatkin and Excoffier 2012). Generally, range expansion heat maps had lower diversity than control heat maps in both the average heterozygosity heat maps and the single locus heat maps, though the difference between the range expansion heat maps and the control heat maps was more pronounced in the average heterozygosity heat maps. Controls generally had higher diversity than range expansion heat maps, lacked any identifiable clinal pattern, and typically displayed uniformly high diversity. In several sets of range expansion heat maps a belt-like or hourglass pattern was identified across the central area of the heat map. This pattern was mainly identified in average heterozygosity heat maps and those heat maps with this pattern were markedly different from the generally uniformly high diversity in the controls. We show and describe this pattern for specific experimental conditions in the following section.
F2. A-D GENERATION = 60, CARRYING CAPACITY = 200,
MIGRATION = 0.3, FECUNDITY= 3

In this two-stage model, carrying capacity and fecundity increased at generation 250 while migration decreases at that time. At the first generation in which the lattice filled, generation 60, the simulation parameters were low carrying capacity, high migration, and low fecundity. In the average heterozygosity heat maps, the range expansion heat map F2.B had lower overall diversity, and especially lower diversity at the right and left hips of the central band, and displays a moderate belt-like pattern. The control heat map of single locus heterozygosity F2.C was patchy with mostly high diversity with broadly scattered patches of lower diversity, and the range expansion heat map of single locus heterozygosity F2.D was very similar to the control but had some long regions of lower diversity at the right hip of the central band.
F2. A - CN_LHcc_HLm_LHf_init_060_AvgHets; B - REN_LHcc_HLm_LHf_init_060_AvgHets; C - CN_LHcc_HLm_LHf_init_060_SingHets; D - REN_LHcc_HLm_LHf_init_060_SingHets; Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.
In this next two-stage model, carrying capacity decreased at generation 250 while migration and fecundity increased. At the first generation in which the lattice filled, simulation parameters were high carrying capacity, low migration and low fecundity. In the average heterozygosity heat maps the control heat map F3.A was dominated by demes with high diversity, while the range expansion heat map F3.B had much lower diversity throughout and especially so across the central band, with a cluster of lower diversity patches at the right hip of the central band in the form of a belt-like pattern. The control heat map of single locus heterozygosity F3.C was similar to the control heat map of average heterozygosity F3.A, but was patchier and had more patches with lower diversity. The range expansion single locus heterozygosity heat map F3.D was even patchier than the control heat map of single locus heterozygosity F3.C, and had larger concentrations of lower diversity patches than did the control.
F3. C

F3. D

F3. A-CN_HLcc_LHm_LHf_init_060_AvgHets; B-REN_HLcc_LHm_LHf_init_060_AvgHets; C-CN_HLcc_LHm_LHf_init_060_SingHets; D-REN_HLcc_LHm_LHf_init_060_SingHets;
Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High; change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

F4. A-D GENERATION = 60, CARRYING CAPACITY = 500, MIGRATION = 0.3, FECUNDITY = 3

Carrying capacity and migration decrease at generation 250 in this two-stage model, while fecundity increases. At the first generation in which the lattice filled, the simulation parameters were high carrying capacity, high migration, and low fecundity. With the exception of a few patches, the average heterozygosity control heat map F4.A was entirely high diversity. The range expansion average heterozygosity heat map F4.B had several large clusters of lower diversity across the central band in the form of a belt-like pattern. The single locus heterozygosity control heat map F4.C was mostly high diversity but had more scattered patches of lower diversity than the average heterozygosity control heat map F4.A. The range expansion single locus heterozygosity heat map F4.D was also
patchy, but those patches were of lower diversity compared with the control and occurred in small and large clusters.

F4. A

F4. B

F4. C

F4. D

F4. A-CN_HLcc_HLm_LHf_init_060_AvgHets; B-REN_HLcc_HLm_LHf_init_060_AvgHets; C-CN_HLcc_HLm_LHf_init_060_SingHets; D-REN_HLcc_HLm_LHf_init_060_SingHets;
Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.
F5. A-D GENERATION = 60, CARRYING CAPACITY = 500,
MIGRATION = 0.2, FECUNDITY = 3

All parameter values were held constant for the duration of this one-stage simulation. Migration and fecundity were low and carrying capacity was high. These heat maps are from the first generation in which the lattice was filled. As these heat maps share the same conditions as the two-stage model for this generation (60), they can be considered an independent replication of the models represented in F3.A-D. In the average heterozygosity heat maps the control heat map F5.A was almost completely high diversity, while the range expansion heat map F5.B exhibited a clear belt-like pattern with high diversity along two edges and large regions of low diversity at the left and right hips of the central band. The single locus heterozygosity control heat map F5.C was patchy and mostly high diversity. The range expansion single locus heterozygosity heat map F5.D was also mostly high diversity but had patchy regions with lower diversity and a large region of low diversity on the right hip of the central region.
F5. C

F5. D

F5. A- CN_0.2m_3f_1ini_500cc_060_AvgHets; B- REN_0.2m_3f_1ini_500cc_060_AvgHets; C- CN_0.2m_3f_1ini_500cc_060_SingHets; D- REN_0.2m_3f_1ini_500cc_060_SingHets;

Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

F6. A-B GENERATION = 60, CARRYING CAPACITY = 500, MIGRATION = 0.2, FECUNDITY = 3

This set of simulations also had the same parameter values as F5.A-D the preceding model of low migration, low fecundity, and high carrying capacity, but differed in the addition of a 60% bottleneck event from generations 450 to 455. However, the heat maps were from the first generation filled and were thus an independent replication of the preceding set of heat maps; refer to the controls F5.A and F5.C for the preceding heat maps. Indeed, the average heterozygosity heat map F6.A for the bottleneck model too had a large region of low diversity across the entire central band in the form of a belt-like pattern, while the single locus heterozygosity heat map F6.B again had a much patchier distribution with larger regions of high diversity and some smaller clusters of lower diversity.
**F6. A** - BN60_REN_0.2m_3f_1ini_060_AvgHets; **B** - BN60_REN_0.2m_3f_1ini_060_SingHets;
Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

---

**F7. A-D**

**GENERATION** = 60, **CARRYING CAPACITY** = 500, **MIGRATION** = 0.3, **FECUNDITY** = 3

This one-stage model had parameters values of high migration and carrying capacity with low fecundity. Heat maps were from the first generation the lattice filled. As these heat maps share the same conditions as the two-stage model for this generation (60), they can be considered an independent replication of the models represented in F4.A-D. The average heterozygosity control heat map F7.A, with the exception of a few patches along the edges, was entirely high diversity, while the range expansion heat map F7.B had larger regions of lower diversity across the central band. The single locus heterozygosity control heat map F7.C was similar to the average heterozygosity control heat map F7.A but was slightly patchier, and the range expansion single locus
heterozygosity heat map F7.D was even patchier, with patches forming small clusters of low diversity.

F7. A

F7. B

F7. C

F7. D

F7. A- CN_0.3m_3f_1ini_500cc_060_AvgHets; B- REN_0.3m_3f_1ini_500cc_060_AvgHets; C- CN_0.3m_3f_1ini_500cc_060_SingHets; D- REN_0.3m_3f_1ini_500cc_060_SingHets; Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.
This set of simulations also had parameter values of high migration and high carrying capacity and low fecundity, but differed in the addition of a 60% bottleneck event from generations 450 to 455. The heat maps were from the first-filled generation and were expected to be similar to the preceding set of heat maps F7.A-D. As expected, the control and range expansion average heterozygosity heat maps F8.A-B from this model were almost identical to the non-bottleneck F7.A-B version of this model, with the only really discernible difference as a slightly larger region of low diversity in the right hip of the bottleneck model range expansion heat map. As for the single locus heterozygosity heat maps F8.C-D, these were also highly similar between the regular and bottleneck versions of this model, as expected, with the only noticeable difference to be found in the size of the clusters of low diversity in the range expansion heat map F8.D. These are larger and more numerous in the bottleneck version of the model.
F8. C

A- BN60_CN_0.3m_3f_1ini_060_AvgHets; B- BN60_REN_0.3m_3f_1ini_060_AvgHets; C- BN60_CN_0.3m_3f_1ini_060_SingHets; D- BN60_REN_0.3m_3f_1ini_060_SingHets.

Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

F9. A-D GENERATION = 60, CARRYING CAPACITY = 500, MIGRATION = 0.3, FECUNDITY = 3

As with the previous set of simulations F7-F8, this model also had parameter values of high migration and carrying capacity and low fecundity, but differed in the addition of a 30% bottleneck event from generations 450 to 455. The heat haps were from the first-filled generation and represent an independent replication of the preceding two sets of heat maps. As expected, average heterozygosity heat maps from this model were almost identical to the non-bottleneck and the 60% bottleneck versions of this model in F7-8, but the range expansion heat map F9.B had a much clearer belt-like pattern in this set as compared to the non-bottleneck F7.B and 60% bottleneck F8.B models. As for the single locus heterozygosity heat maps, the only noticeable difference was found in the size of
the clusters of low diversity in the range expansion heat map as compared. These are larger and more numerous in this bottleneck version of the model than in the non-bottleneck model, similar to the 60% bottleneck model.

F9. A

F9. B

F9. C

F9. D

F9. A-BN30_CN_0.3m_3f_in_060_AvgHets; B-BN30_REN_0.3m_3f_in_060_AvgHets; C-BN30_CN_0.3m_3f_in_060_SingHets; D-BN30_REN_0.3m_3f_in_060_SingHets;
Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.
In general, range expansion heat maps display lower diversity than controls for both the average heterozygosity heat maps and the single locus heterozygosity heat maps when compared with controls, as expected. Several of the average heterozygosity range expansion heat maps displayed a belt-like pattern, while the single locus heterozygosity range expansion heat maps generally displayed dispersed clusters of low diversity. Of the eight sets of heat maps that displayed the belt-like pattern, three can be considered independent simulations. Of these three independent sets of simulations, the following configurations of parameter values was noted in table 6:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrying Capacity</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Fecundity</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Migration</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>First Fill</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models: CC</th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Fecundity, particularly low fecundity, appears to be critical to the formation of these belt-like patterns. Migration rates and carrying capacity did not appear to be either associated or not associated with the formation of these patterns. Additionally, it should
be noted that all heat maps displaying this pattern were of the first generation at which the lattice filled, and these differences were only noted in the first generation heat maps; in fact, in the heat maps from the final generations of all simulations (not just those heat maps with belt-like patterns), no real differences were observed between range expansion and control heat maps—regardless of parameter values, the presence of a later bottleneck event, changing parameter values across time, or calculating heterozygosity as patch locus averages or investigating patterns at a single locus. As such, not only did no effect of a later bottleneck event appear in the final generation, but also no clinal patterns or overall loss of diversity were visible in the final generation. The only exception was the final generation of the two stage model in which carrying capacity increased, migration decreased, and fecundity increased over time; this simulation showed a clear cline in the average heterozygosity heat map for the final generation but this cline did not follow the belt-like pattern. This pattern also held for those heat maps, which did not display this pattern but still displayed lower diversity in the range expansion heat maps relative to the controls.

MORAN’S I

In addition to calculating average patch heterozygosity and single locus patch heterozygosity and plotting these as heat maps, global Moran’s I was calculated using the gearymoran function in the ade4 package in R (Dray and Dufour 2007; R Core Team 2013). The function calculates the value of Moran’s I (global) and gives p-values, along with the expected value as computed through 999 permutations of Monte-Carlo simulations (Dray and Dufour 2007). Table 7 lists the Moran’s I values for those
simulation generation results where the average patch heterozygosity were tested and had a Moran’s I of 0.095 or greater. In all cases listed, estimates of Moran’s I were statistically significant, where statistical significance was set at \( p \leq 0.001 \).

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Moran’s I</th>
<th>Figures</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>REN_LHcc_HLm_LHf_1init_060</td>
<td>0.104 B*</td>
<td>F2</td>
<td>0.001</td>
</tr>
<tr>
<td>REN_HLcc_LHm_LHf_1init_060</td>
<td>0.113 B</td>
<td>F3</td>
<td>0.001</td>
</tr>
<tr>
<td>REN_HLcc_HLm_LHf_1init_060</td>
<td>0.098 B</td>
<td>F4</td>
<td>0.001</td>
</tr>
<tr>
<td>REN_0.2m_3f_1ini_500cc_060</td>
<td>0.126 B</td>
<td>F5</td>
<td>0.001</td>
</tr>
<tr>
<td>BN60_REN 0.2m_3f_1ini_060</td>
<td>0.102 B</td>
<td>F6</td>
<td>0.001</td>
</tr>
<tr>
<td>REN 0.3m_3f_1ini_500cc_060</td>
<td>0.097 B</td>
<td>F7</td>
<td>0.001</td>
</tr>
<tr>
<td>BN60_REN 0.3m_3f_1ini_060</td>
<td>0.160 B</td>
<td>F8</td>
<td>0.001</td>
</tr>
<tr>
<td>BN30_REN 0.3m_3f_1ini_060</td>
<td>0.135 B</td>
<td>F9</td>
<td>0.001</td>
</tr>
<tr>
<td>REN_LHcc_HLm_LHf_1init_750</td>
<td>0.144 B</td>
<td>F10</td>
<td>0.001</td>
</tr>
<tr>
<td>CN_LHcc_LHm_LHf_1init_750</td>
<td>0.110 B</td>
<td>F11</td>
<td>0.001</td>
</tr>
<tr>
<td>CN 0.3m_4f_1ini_200cc_500</td>
<td>0.104 B</td>
<td>F12</td>
<td>0.001</td>
</tr>
<tr>
<td>BN60_CN 0.3m_4f_1ini_500</td>
<td>0.107 B</td>
<td>F13</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Moran’s I global statistic reported for values of 0.095 and above. *B indicates that simulation generation’s heat map displayed a belt-like pattern. Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

**F10. A-D GENERATION = 750, CARRYING CAPACITY = 500 > 200, MIGRATION = 0.2 > 0.3, FECUNDITY = 3 > 4**

All average heterozygosity heat maps with a belt-like pattern had Moran’s I values above 0.095, but not all sets of tested data with Moran’s I values above 0.095 displayed this pattern. For example, this two-stage model shows a clear cline in the average heterozygosity heat map F10.B for the experimental range expansion model that is noticeably different from the control F10.A, and yet does not display a belt-like pattern.

The conditions for this simulation were a carrying capacity that decreased over time
while migration and fecundity increased, and the heat maps are representing the final generation. The single locus heterozygosity heat maps F10.C-D were patchier with overall lower diversity than the average heterozygosity heat maps F10.A-B, and the control and range expansion heat maps F10.C-D were very similar.

**F10. A**

![Heat Map F10. A](image)

**F10. B**

![Heat Map F10. B](image)

**F10. C**

![Heat Map F10. C](image)

**F10. D**

![Heat Map F10. D](image)

F10. A-CN_LHcc_HLm_LHf_1init_750_AvgHets; B-REN_LHcc_HLm_LHf_1init_750_AvgHets; C-CN_LHcc_HLm_LHf_1init_750_SingHets; D-REN_LHcc_HLm_LHf_1init_750_SingHets;
Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.
F11. A-D GENERATION = 750, CARRYING CAPACITY = 200 > 500, MIGRATION = 0.2 > 0.3, FECUNDITY = 3 > 4

The average heterozygosity heat map F11.A of the control for this simulation had a Moran’s I value above 0.095. In this two-stage model, carrying capacity, migration, and fecundity all increased over time. The heat maps below are from the final generation. The clinal signal of the control heat map detected by Moran’s I is generated by patches of lower diversity in the lower two corners in the context of nearly uniformly high diversity found throughout the remaining lattice.
F11. C

F11. D

F11. A-CN_LHcc_LHm_LHf_1init_750_AvgHets; B-REN_LHcc_LHm_LHf_1init_750_AvgHets; C-CN_LHcc_LHm_LHf_1init_750_SingHets; D-REN_LHcc_LHm_LHf_1init_750_SingHets; Simulation name codes: REN - Range Expansion; CN - Control; BN30 - Bottleneck Event of 30% added; BN60 - Bottleneck Event of 60% added; HL/LH - High to Low or Low to High change in parameter values in a 2-stage simulation; AvgHets/SingHets – Average or Single Heterozygosity Heat Map; m - Migration Rate; cc - Carrying Capacity; f - Fecundity.

F12. A-D GENERATION = 500, CARRYING CAPACITY = 200, MIGRATION = 0.3, FECUNDITY = 4

This one-stage model also had a Moran’s I value above 0.095 for the average heterozygosity control heat map. This model’s parameters were high migration, high fecundity, and low carrying capacity. Heat maps were from the final generation. The control heat map F12.A has a broad region of higher diversity across the center and lower diversity along the corners and edges—the exact opposite of the belt-like pattern observed for most other heat maps with Moran’s I above 0.095, but still clinal.
Lastly, the control heat map F13.A for this one-stage model had a Moran’s I above 0.095. This one stage model had high migration, high fecundity, and high carrying capacity, with the addition of a 60% bottleneck event, with heat maps from the final generation. There was no clear pattern of a decline in diversity across space, but the
upper right corner had some scattered patches of lower diversity, driving the relatively high value of Moran’s I.

F13. A

In general, heat maps with a Moran’s I above 0.095 displayed a belt-like pattern of a decline in diversity across space, though there were a few noted exceptions. In two of these exceptions a clinal pattern was visible in a range expansion and a control heat map, but neither of these patterns was belt-like. The remaining exceptions had patches of lower diversity in corners that appeared to drive the Moran’s I value in the context of nearly uniformly high diversity in the remaining portions of the heat maps. In addition to calculating Moran’s I for average patch heterozygosity, Moran’s I was calculated for the allele frequencies at each locus. Table 8 below presents those simulations for which at least one locus had a Moran’s I of 0.1 or higher, with the proportion out of ten loci of the number of loci meeting this condition. This proportion, when compared with the average patch heterozygosity Moran’s I scores, was not informative of average patch heterozygosity Moran’s I score.
**Table 8 Moran’s I for Allele Frequencies**

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Prop. Loci at I 0.1+</th>
</tr>
</thead>
<tbody>
<tr>
<td>REN LHcc LHm LHf 1init 750</td>
<td>3/10</td>
</tr>
<tr>
<td>REN LHcc LHm LHf 1init 070</td>
<td>4/10</td>
</tr>
<tr>
<td>REN LHcc HLm LHf 1init 750</td>
<td>6/10</td>
</tr>
<tr>
<td>REN LHcc HLm LHf 1init 060</td>
<td>4/10</td>
</tr>
<tr>
<td>REN HLcc LHm LHf 1init 750</td>
<td>6/10</td>
</tr>
<tr>
<td>REN HLcc LHm LHf 1init 060</td>
<td>3/10</td>
</tr>
<tr>
<td>REN HLcc HLm LHf 1init 750</td>
<td>4/10</td>
</tr>
<tr>
<td>REN HLcc HLm LHf 1init 060</td>
<td>3/10</td>
</tr>
<tr>
<td>REN 0.3m 4f 1ini 200cc 500</td>
<td>2/10</td>
</tr>
<tr>
<td>REN 0.3m 3f 1ini 500cc 060</td>
<td>9/10</td>
</tr>
<tr>
<td>REN 0.3m 3f 1ini 500cc 500</td>
<td>5/10</td>
</tr>
<tr>
<td>REN 0.3m 3f 1ini 200cc 500</td>
<td>5/10</td>
</tr>
<tr>
<td>REN 0.3m 3f 1ini 200cc 050</td>
<td>5/10</td>
</tr>
<tr>
<td>REN 0.2m 4f 1ini 500cc 500</td>
<td>3/10</td>
</tr>
<tr>
<td>REN 0.2m 4f 1ini 500cc 020</td>
<td>1/10</td>
</tr>
<tr>
<td>REN 0.2m 4f 1ini 200cc 500</td>
<td>3/10</td>
</tr>
<tr>
<td>REN 0.2m 3f 1ini 500cc 500</td>
<td>4/10</td>
</tr>
<tr>
<td>REN 0.2m 3f 1ini 500cc 060</td>
<td>3/10</td>
</tr>
<tr>
<td>REN 0.2m 3f 1ini 200cc 500</td>
<td>2/10</td>
</tr>
<tr>
<td>REN 0.2m 3f 1ini 200cc 070</td>
<td>1/10</td>
</tr>
<tr>
<td>CN LHcc LHm LHf 1init 750</td>
<td>5/10</td>
</tr>
<tr>
<td>CN LHcc HLm LHf 1init 750</td>
<td>5/10</td>
</tr>
<tr>
<td>CN HLcc LHm LHf 1init 750</td>
<td>3/10</td>
</tr>
<tr>
<td>CN 0.3m 4f 1ini 500cc 500</td>
<td>5/10</td>
</tr>
<tr>
<td>CN 0.3m 4f 1ini 200cc 500</td>
<td>6/10</td>
</tr>
<tr>
<td>CN 0.3m 3f 1ini 500cc 500</td>
<td>1/10</td>
</tr>
<tr>
<td>CN 0.3m 3f 1ini 500cc 060</td>
<td>1/10</td>
</tr>
<tr>
<td>CN 0.3m 3f 1ini 200cc 500</td>
<td>5/10</td>
</tr>
<tr>
<td>CN 0.2m 4f 1ini 500cc 500</td>
<td>5/10</td>
</tr>
<tr>
<td>CN 0.2m 4f 1ini 200cc 500</td>
<td>2/10</td>
</tr>
<tr>
<td>CN 0.2m 3f 1ini 500cc 500</td>
<td>7/10</td>
</tr>
<tr>
<td>CN 0.2m 3f 1ini 500cc 060</td>
<td>2/10</td>
</tr>
<tr>
<td>CN 0.2m 3f 1ini 200cc 500</td>
<td>3/10</td>
</tr>
<tr>
<td>BN60 REN 0.3m 3f 1ini 500</td>
<td>5/10</td>
</tr>
<tr>
<td>BN60 REN 0.3m 3f 1ini 060</td>
<td>7/10</td>
</tr>
<tr>
<td>BN60 REN 0.2m 4f 1ini 500</td>
<td>4/10</td>
</tr>
<tr>
<td>BN60 REN 0.2m 3f 1ini 500</td>
<td>4/10</td>
</tr>
<tr>
<td>BN60 REN 0.2m 3f 1ini 060</td>
<td>1/10</td>
</tr>
<tr>
<td>BN60 CN 0.3m 4f 1ini 500</td>
<td>3/10</td>
</tr>
<tr>
<td>BN60 CN 0.3m 3f 1ini 500</td>
<td>3/10</td>
</tr>
<tr>
<td>Simulation</td>
<td>Moran’s I Global Statistic</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>BN60 CN 0.2m 4f 1ini 500</td>
<td>4/10</td>
</tr>
<tr>
<td>BN30 REN 0.3m 4f 1ini 020</td>
<td>3/10</td>
</tr>
<tr>
<td>BN30 REN 0.3m 3f 1ini 500</td>
<td>4/10</td>
</tr>
<tr>
<td>BN30 REN 0.3m 3f 1ini 060</td>
<td>6/10</td>
</tr>
<tr>
<td>BN30 CN 0.3m 4f 1ini 500</td>
<td>3/10</td>
</tr>
<tr>
<td>BN30 CN 0.3m 3f 1ini 500</td>
<td>5/10</td>
</tr>
<tr>
<td>BN30 CN 0.3m 3f 1ini 060</td>
<td>1/10</td>
</tr>
<tr>
<td>BN30 CN 0.2m 4f 1ini 500</td>
<td>6/10</td>
</tr>
<tr>
<td>BN30 CN 0.2m 3f 1ini 500</td>
<td>1/10</td>
</tr>
</tbody>
</table>

Moran’s I global statistic reported for all simulations in which raw allele frequency data were above 0.1 at one or more loci, listed as a proportion out of ten of the number of loci with allele frequency’s that gave a Moran’s I of 0.1 or greater.

These results showed a decline in diversity associated with range expansions when considering average patch heterozygosity, where this decline in diversity tended to follow a belt-like pattern in which the middle third of the lattice had lower diversity than the upper and lower thirds—with the lowest diversity region almost entirely occurring to the right hip of the central band. These patterns were, generally, associated with a Moran’s I above 0.095 and low fecundity, and were found only in generations in which the lattice was first filled. Two models with Moran’s I above 0.095 but lacking a belt-like pattern still displayed clinal patterns and were from the final generation.

While the association of lower fecundity with clinal patterns was expected, the lack of association between migration rates and carrying capacities and clinal patterns was unexpected (Endler 1977; Fix 1999; Klopfstein, Currat and Excoffier 2006; Excoffier, Foll and Petit 2009; Ray and Excoffier 2010; Slatkin and Excoffier 2012). It follows from theoretical and experimental studies that lower fecundities are associated with the development of clines, but it is generally expected from these studies that lower carrying capacities and migration rates would also be associated with clinal patterns (Endler 1977; Fix 1997, 1999; Klopfstein, Currat and Excoffier 2006; Excoffier, Foll and Petit 2009; Ray and Excoffier 2010). However, Klopfstein, Currat and Excoffier (2006) did show
that clines can form even when carrying capacities are high—it is just less likely than when carrying capacities are lower. These results are still surprising then, as neither higher nor lower carrying capacities were more associated with clinal patterns. It was also surprising that migration rates did not appear to be important for the formation of clinal patterns and neither lower nor higher values were more associated with clinal patterns. Overall, we found that low fecundity—with either migration rate or carrying capacity—was associated with the formation of clinal patterns. The carrying capacities appear to have been balanced by the low fecundity. As for the migration rates, it is possible that these were less important for cline formation because the size of the lattice and restrictions against long distance dispersal would have led to isolation by distance, and hence spatial structuring.

The addition of a bottleneck event for five generations from generation 450 to 455 did not have an apparent impact on the heat maps of the final generation for any of the models tested. Results have been presented for several bottleneck models for generations prior to the bottleneck event. It is interesting to note that these models are identical to non-bottleneck models prior to generation 450, and that the range expansion heat maps for the generations first-filled for both the bottleneck and non-bottleneck models displayed similar clinal patterns.

Generally, these clinal patterns were associated with Moran’s I above 0.095 and these clinal patterns were predominantly found in generations at which the lattice first filled. The only two exceptions were the two-stage model with carrying capacity increasing over time, migration decreasing, and fecundity increasing. In this model, the average heterozygosity range expansion heat map showed a clear cline that differed markedly
from the control heat map. The other exception was the one-stage model where migration and fecundity were high while carrying capacity was low. In this model, the average heterozygosity control heat map showed a clinal pattern in the exact opposite configuration of the belt-like pattern commonly identified previously with a region of high diversity across the central band and lower diversity in the upper and lower thirds. Both of these exceptions had Moran’s I values above 0.095 but neither fit the belt-like pattern and both were from final generations. The second of these exceptions showed this pattern in the control but not the range expansion heat map, so this pattern can be identified as stochastically arising despite a lack of range expansion. The first of these exceptions shows a clear cline in a pattern that is markedly different from all other heat maps with clinal patterns. While this study used Moran’s I global statistic and heat maps as evidence of clines, it would not be a good methodology to use Moran’s I without heat maps or some other visual or statistical representation to indicate presence of a cline.

It is also important to consider that Moran’s I lighting up for a simulation does not constitute an indication of an exact pattern or shape of the data, and it does not provide the biological or evolutionary context for the observed patterns; that will come from additional analyses. While this is intriguing, there is still a general trend of clinal patterns associating with the generation at which the lattice is first filled but not the final generation. In fact, in all but these two exceptions noted above, the first generations at which the lattice filled and the final generations showed no discernible differences. This trend held with or without the addition of a bottleneck event—even when that bottleneck event caused a reduction of 60% of the metapopulation’s carrying capacity. This is an extremely interesting result, as it throws into question both the consequences of
bottleneck events upon clines and the long-term durability of clines across time and space.

DISCUSSION

The decline in diversity associated with range expansion models was expected, though the belt-like pattern was counterintuitive; given the starting populated regions of the range expansion model lattice were two connecting sides in the form of an ‘L’ in the left upper corner, it would have been logical to expect a decline in diversity from that corner to the opposite in the lower right. Another expected result based upon the starting configuration would have been regions of high diversity along the left and top sides of the lattice. The common belt-like pattern was unexpected precisely because it showed regions of lower diversity along the left side, which should have had higher diversity as it was a source region. At the present a good explanation for why this belt-like pattern developed despite the starting configuration is lacking, though this pattern may fit with the literature on sectoring of diversity as a consequence of range expansions—where diversity sectors out from a starting expansion region into large stretches of space with one variant or degree of diversity, and these stretches are bordered by other regions with differing patterns, but in a pin-wheel like shape rather than a traditional clinal gradient (Hallatschek et al. 2007; Excoffier and Ray 2008). Future work is needed to test the impact of this starting configuration on clinal patterns.

These results are interesting within the context of the history of the study of clines in European populations. While the application of the wave of advance to range expansion processes is not inaccurate given recent works on surfing, the application of
the demic diffusion model to the study of allele frequency clines in Europe has been exceedingly problematic. Generally, this body of research has identified a general north-south cline, and the presence of this cline has been explained with the demic diffusion model (a demographic expansion out of the Near East after the advent of agriculture spreads both people and agriculture into Europe and results in the assimilation or replacement of all hunter-gatherers); or with the cultural assimilation model (agriculture is spread by a few long-distance migrants such as traders, and hence there was no major demographic expansion out of the Near East or at least no major replacement of hunter-gatherer populations) (Menozzi, Piazza and Cavalli-Sforza 1978; Barbujani and Sokal 1990; Cavalli-Sforza, Menozzi and Piazza 1994; Fix 1996; Excoffier and Schneider 1999; Belle, Landry and Barbujani 2006; Seldin et al. 2006; Itan et al. 2009; Gerbault et al. 2009, 2011; François et al. 2010; Arenas 2012; Arenas et al. 2013; Jobling et al. 2014). Much of this work relied heavily upon principal components analysis (PCA), a methodology whereby data are summarized by Eigen values (Cavalli-Sforza, Menozzi and Piazza 1994; Price et al. 2006; Novembre and Stephens 2008; Paschou et al. 2008). While the usefulness of this method as a means of data dimension reduction and clustering has been well documented, some scholars have noted that it has been used inappropriately to infer colonization history. Specifically, many researchers have devoted enormous attention to when and how farming spread into Europe, with particular regard to whether modern genetic diversity in European populations was due to farming or hunter-gatherer populations. These studies took genetic data such as allele frequencies and SNPs and ran PCAs, and then results of these PCAs were interpolated (a geographical-statistical method of ‘smoothing’ or filling in the gaps caused by missing
data) and projected onto maps of Europe in the form of ‘synthetic maps’. These maps showed gradients in the interpolated data, and these gradients were then inferred to be clinal patterns reflecting a demographic and range expansion of farming populations moving out of the Near East into Europe (Menozzi, Piazza and Cavalli-Sforza 1978; Sokal and Menozzi 1982; Rendine, Piazza and Cavalli-Sforza 1986; Barbujani and Sokal 1990; Sokal et al. 1990; Sokal 1991; Cavalli-Sforza, Menozzi and Piazza 1994; Sokal, Oden and Thomson 1997; Rendine et al. 1999; Paschou et al. 2008; Itan et al. 2009; Gerbault et al. 2009, 2011; François et al. 2010; Jay et al. 2013).

Later work demonstrated that many different evolutionary scenarios could result in the same synthetic maps, as multiple evolutionary scenarios may often have similar effects upon allele frequencies, and, certain patterns in synthetic maps will appear often as a mathematical artifact of the transformation of PCAs (Novembre and Stephens 2008). As such, it is neither wise nor possible to make or support definitive statements regarding the meaning of gradient patterns observed in synthetic maps constructed from interpolated PCA eigenvalues (Fix 1996; Sokal, Oden and Thomson 1999; Novembre and Stephens 2008). This means any previous publications that relied upon PCA without sufficient effort to corroborate those results via other methods, or took those synthetic maps to be indicative of particular past evolutionary and demographic events, are suspect. The demic diffusion model has been misapplied with regard to clines and the peopling of Europe in several studies where researchers concluded that a correlation between the direction of historical migration events and plots of spatial autocorrelation statistics or synthetic maps represented the evolutionary history of European populations (Menozzi, Piazza and Cavalli-Sforza 1978; Sokal and Menozzi 1982; Rendine, Piazza and Cavalli-
Sforza 1986; Cavalli-Sforza, Menozzi and Piazza 1993; Cavalli-Sforza, Menozzi and Piazza 1994; Barbujani, Sokal and Oden 1995); it was assumed that this correlation was caused by these historical events—despite evidence that other processes could result in the same plots of spatial autocorrelations, and the serious flaws in synthetic maps (Fix 1996; Novembre and Stephens 2008).

It was also found that the axis of an expansion event may not even correspond to the axis of the first principal component, but instead may actually lie orthogonal. According to these papers, younger expansions would be expected to lie parallel, and older expansions orthogonal, to the axis of expansion (Hofer et al. 2009; François et al. 2010; Jay et al. 2013). If true, these results could indicate that a first principal component indicating an axis of southeast-northwest would actually represent an expansion axis centered around Iberia, which then raises the possibility that PCA results were misinterpreted in the past and actually indicate Paleolithic or Mesolithic expansions. However, another study has recently emerged in which this particular result has been strongly challenged (DeGiorgio and Rosenberg 2013). Whether or not the first principal component is orthogonal or parallel to the axis of expansion depends not on the age of the expansion event, but on the sampling scheme of the researchers. Sampling from differently shaped quadrants resulted in first principal components that were either orthogonal or parallel to the true axis of expansion, demonstrating once again that caution with regard to PCA and colonization theory is more than warranted (DeGiorgio and Rosenberg 2013).

It wasn’t until a few decades passed that several more advanced simulation studies with increased spatial, demographic, and temporal complexity were able to
provide data demonstrating the effect of model parameters and expansion dynamics on spatial patterns of diversity in European populations—and these studies provide evidence that the development and direction of clines in Europe were more likely the result of Paleolithic expansions than of Neolithic expansions, given the timings of these events and assumed expansion dynamics (Currat and Excoffier 2005; Klopfstein, Currat and Excoffier 2006; François et al. 2010; Arenas et al. 2012, 2013; Arenas 2012).

As such, it is evident that neither the cultural nor the demic diffusion models in their purest form are sufficient. Additionally, there were numerous methodological errors in the construction and testing of models in these past studies of range expansions in Europe. Furthermore, these studies focused on cline formation and paid little to no study to cline durability. This is not entirely unexpected—the history of research into the development of clinal patterns in European populations, whether due to Paleolithic or Neolithic expansion events, holds the same trend: clines, once established, are assumed to only be disrupted via complete population replacement or significant admixture between very sparse and very dense populations (Menozzi, Piazza and Cavalli-Sforza 1978; Sokal and Menozzi 1982; Rendine, Piazza and Cavalli-Sforza 1986; Cavalli-Sforza, Menozzi and Piazza 1993; Cavalli-Sforza, Menozzi and Piazza 1994; Barbujani, Sokal and Oden 1995; Currat and Excoffier 2005; Klopfstein, Currat and Excoffier 2006; François et al. 2010; Arenas et al. 2012, 2013; Arenas 2012). In fact, perhaps the only hard evidence of durability in clines comes from Endler (1977), who demonstrated that cline presence, shape, and location are strongly affected by how they are established and the first several generations of their persistence—with some shifting around possible but dissolution unlikely after several to several hundred generations (Endler 1977). However, Endler
(1977)’s findings do not eliminate the possibility of cline dissolution due to negative selection or bottleneck events. Often when modeling complex processes, especially via computer simulations, choices must be made regarding the specificity versus the generality of models (Hoban, Bertorelle and Gaggiotti 2012). While the assumption that clines are never dissolved is convenient for the sake of modeling expansion processes at the continental scale, it is an assumption—and one that was challenged this study. The challenge to cline permanence lies in two keys results. The general finding that clinal patterns observed in the first-filled generations did not persist in the final generation poses a direct challenge by demonstrating that clines can dissolve. This finding was further supported by the subsequent results from the bottleneck models, where even the addition of a 30% or 60% bottleneck event did not maintain or form a cline that was present in the final generation.

FUTURE DIRECTIONS

Future work will focus on the interactions of drift and natural selection during range expansions, and how the interactions of these forces and expansion dynamics affect cline durability. It has been well documented that natural selection may establish and maintain a clinal pattern of genetic diversity in the absence of drift or asymmetrical migration (Fix 1996; Jobling et al. 2014). It has also been demonstrated both theoretically, and through simulations, that the allele surfing phenomenon can also establish clinal patterns of genetic diversity (Fisher 1937; Ray, Currat and Excoffier 2003; Edmonds, Lillie and Cavalli-Sforza 2004; Klopfstein, Currat and Excoffier 2006; Excoffier and Ray 2008;
Itan et al. 2009; Hofer et al. 2009; Excoffier, Foll and Petit 2009; Gerbault et al. 2009, 2011; Moreau et al. 2011; Antoniazza et al. 2014). Klopstein (et al. 2006) later demonstrated in simulations that even new mutations could surf, and that a fairly low minimum allele frequency was sufficient to establish a cline; it is not apparent from their or other studies how durable these clines are. Several simulation studies have investigated that natural selection may be important for the development of clines during range expansions, and that even mutations under negative selection may surf (Travis et al. 2007; Gerbault et al. 2009, 2011; Itan et al. 2009; Antoniazza et al. 2014; Peischl and Excoffier 2015). In yet another study it was noted that large continental differences in allele frequencies might be due more to allele surfing than natural selection, as ascertained by a suit of statistics including the direction of change in frequencies when comparing ancestral and derived alleles (Hofer et al. 2009). While Hofer et al. (2009) were able to demonstrate that allelic surfing may be sufficient to establish continental clines, they did not make use of simulation studies, and the relative contributions of natural selection and allelic surfing during range expansions were not tested. Future studies could expand upon this body of literature by testing for the interactions of drift and natural selection during range expansions upon clinal development and maintenance, and by explicitly testing for cline durability.

Cline durability must be explored and tested now that it has been thrown into question. Future tests for cline durability would need to test for cline presence explicitly and would benefit from another series of simulations with a smaller lattice and narrower range of parameters with more constants and absorbing migration, and initialization with clinal patterns of allele frequencies. It would also benefit from investigating the effects of
different populated region starting configurations for the range expansion models, such as an L shape, one side, opposite sides, three sides, all four sides, and a populated central region. Such a study would also need to track more generations at regular steps for heat maps and Moran’s I. Other analyses might include: using variograms to look at patterns across space with more refinement than simple global statistics, taking the average allele frequency to calculate average locus heterozygosity in addition to averaging individual locus heterozygosity, overlaying heterozygosity or allele frequency at each locus in the same variogram plot, or even plotting the allele frequencies themselves as heat maps.

CONCLUSIONS

Simulations of range expansion models using variable migration rates, fecundity, and carrying capacities all resulted in loss of diversity when compared with controls, in the first generation at which the lattice was filled. It was hypothesized that lower carrying capacity, fecundity, and migration rates would be associated with clinal patterns, and that bottleneck events would impact clinal patterns. Parameter choice did a strong role in the overall results, such that lower fecundity (f=3) resulted in positive Moran’s I values and lower diversity. Bottleneck events were not found to have had any impact on clinal patterns. Simulation models with positive Moran’s I values displayed belt-like clinal patterns. However, Moran’s I was also found to indicate positive spatial autocorrelation in some control heat maps. This means that Moran’s I global statistic is insufficient evidence of a cline and must be backed with some other analysis, such as the heat maps used in this current study. Generally, these patterns followed the theoretical expectations.
of a model of range expansion. Results from our simulations are in accord with empirical
data, supporting a strong role of drift in generating human population structure within the
context of continental colonization events. These results also highlight the potential
problems with assuming clines are indissoluble. Future planned simulations will work to
tease out the interactions between the three parameters in the context of natural selection,
as well as explicitly testing for cline durability.
REFERENCES


Hofer, Tamara, Nicolas Ray, D Wegmann, and L Excoffier. 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics* 73 (1): 95-108.


Maruyama, Takeo and Paul A Fuerst. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111 (3): 675-689.


Peischl, Stephan, Mark Kirkpatrick, and Laurent Excoffier. 2015. Expansion Load and the Evolutionary Dynamics of a Species Range. *The


Annual Review of Anthropology 20:119-140.


version 2.17.0. http://CRAN.R-project.org/package=gplots


———. 2014. tidyr: Easily Tidy Data with spread() and gather() Functions. R package version 0.2.0 http://CRAN.R-project.org/package=tidyr


S1. Generation Chooser R Script

## Load Workspace ##
###################################################

## Load Packages ##
library(foreign)
  ## This package is necessary to load the list.files function
library(tidyr)
  ## These packages provide the functions for reshaping the data
library(dplyr)
  ## Un-load plyr before re-running the function below
library(magrittr)

## Load Files ##
list_of_files <- list.files(path="path/to/files", pattern = "*.txt")
my_files <- lapply(list_of_files, read.csv)
names(my_files) <- gsub("\\.txt$", ", list_of_files)

## The above code calls the list.files function to create a list of all the files in the specified
directory matching a given pattern; this list is then assigned so it can be stored in the global environment. Then, `lapply` is called, which creates a loop that moves over the list and reads in all of the files using the chosen read function, and then assigns each as a dataframe to a new list stored in the global environment; the structure call confirms that this worked.

```r
## Function Definition ##
tidy_up <- function(x, threshold){
  tidier <- x %>%
    select(replicate, generation, extrate, contains("adlt")) %>%
    gather(group, pop, -replicate, -generation, -extrate)
  %>%
    separate(group, into = c("group", "sex", "patch"), sep = "\\.") %>%
    select(replicate, generation, sex, patch, pop) %>%
    group_by(replicate, generation, patch, sex) %>%
    summarise(N=sum(pop)) %>%
    spread(sex, N) %>%
    ungroup() %>%
    mutate(size=fem+mal)
  #head(tidier)
  threshold <- threshold
  tidiest <- tidier %>%
    mutate(sizeT=as.numeric(size > threshold)) %>%
    select(replicate, generation, size, sizeT) %>%
    group_by(replicate, generation) %>%
```
```r
summarise(allT=min(sizeT)) %>%
  filter(allT == 1) %>%
  group_by(replicate) %>%
  summarise(gen_min=min(generation))
return(tidiest)
}

########################################################

## Function Calls, Saving Files ##
first_fill <- lapply(my_files, tidy_up, threshold=25)
df.first_fill <- data.frame(lapply(first_fill))
write.table(df.first_fill, "first_fill.txt")

## The plan is to use the fact that the data frames are read in and stored in a list to our advantage: by keeping everything in a list, we can use the lapply function to iterate through each dataframe, extract the info we want, and save it off in a 'results' file. How? the lapply function allows user-defined functions, so it seems rational then that we could use the tidier code inside a function called by lapply, and store the results in a text file to be written out.
```
S2. Gawk programs to convert FSTAT extended files to FSTAT

```bash
  gawk -i inplace 'NR<=8{print $0}; NR>8{print $1 " $2 " $3 " $4 " $5 " $6 " $7 " $8 " $9 " $10}' *.dat
```

##The above gawk program reads in and accepts the first eight lines of the text file. After these eight lines, the program will only print columns 1 through 10, and then will save the text file by overwriting the original contents. This program removes the extraneous columns produced by NEMO; when the simulation genotype files are exported, the FSTAT extended format outputs a text file with the number of loci and alleles at each locus, and the number of demes, as well as the raw genotypes at each locus for every individual in each deme, and four additional columns (age, sex, natal patch, patch at data collection). These four additional columns provide information that is not of use to the current study, increase the amount of necessary storage, and, when converting these files to arelquin project files via PGDSpider, result in corrupted arlsumstat
files. These corrupted arlsumstat files result in faulty analyses in arlsumstat.

```bash
for file in *.dat; do gawk -i inplace 'NR==1{0=gensub(/14/, "10",1)};7' ${file}; done
```

## While the first gawk program removed the four extraneous columns, it did not result in a fully converted extended FSTAT to FSTAT file. As noted in the previous comments, all FSTAT files have a header that contains the number of loci, among other key words describing the content of the FSTAT file. This information is critical, and mismatches between the number of loci as indicated by the header and the number of columns in the file again result in corrupted arlsumstat project files during the FSTAT to arlsumstat conversion process and then faulty analyses.
S3. Gawk programs to convert .del files to FSTAT

for i in *.del; do mv -- "i" "i%.del.dat"; done
    # Loop through all of the .del files in the current working directory and change the file extension to .dat

for file in *.dat; do gawk -i inplace -F " " -v OFS=" " "BEGIN{a[2];a[2];a[3];a[4];a[5];a[7];a[8];a[9];a[10];a[11]} {for(x in a)gsub("1","2",$x)}7" ${file};done
    # Loop through each .dat file and replace every 1 with 2 in every column except the first

for file in *.dat; do gawk -i inplace -F " " -v OFS=" " "BEGIN{a[2];a[2];a[3];a[4];a[5];a[7];a[8];a[9];a[10];a[11]} {for(x in a)gsub("0","1",$x)}7" ${file};done
    # Loop through each .dat file and replace every 0 with 1 in every column except the first

for file in *.dat; do gawk -i inplace -F " " -v OFS=" " "BEGIN{a[2];a[2];a[3];a[4];a[5];a[7];a[8];a[9];a[10];a[11]} {for(x in a)gsub("1","01",$x)}7" ${file};done
    # Loop through each .dat file and replace every 1 with 01 in every column except the first

for file in *.dat; do gawk -i inplace -F " " -v OFS=" " "BEGIN{a[2];a[2];a[3];a[4];a[5];a[7];a[8];a[9];a[10];a[11]} {for(x in a)gsub("2","02",$x)}7" ${file};done
    # Loop through each .dat file and replace every 2 with 02 in every column except the first

for file in *.dat; do gawk -i inplace 'NR<=1{print $0}; NR>1{print $1 "$2" "$3" "$4" "$5" "$6" "$7" "$8" "$9" "$10" "$11}" ${file}; done
    # Loop through each .dat file and print only the first line and the first 11 columns out to file

for file in *.dat; do gawk '{gsub("/pop loc02 loc03 loc04 loc05 loc06 loc07 loc08 loc09 loc0201 age sex ped origin ID/", "1024 10 2 1");print}' ${file}; done
    # Loop through each .dat file and replace the left text with
the right as the header

for file in *.dat; do gawk -i inplace '*/1024 10 2 1/{print $0 RS "loc1" RS "loc2" RS "loc3" RS "loc4" RS "loc5" RS "loc6" RS "loc7" RS "loc8" RS "loc9" RS "loc10";next}1' ${file};done
# Loop through each .dat and print the locus tags after the	header, each on a new line

## This script takes the output of Nemo simulations of deleterious mutations in the
form of a native file format (.del) and converts file format to FSTAT. First, the files are
read in and the extension is changed. Next, the alleles must be recoded; .del files code
alleles as 0 for wild-type and 1 for the mutation, FSTAT expects alleles to be coded as
01 or 02, respectively. In all columns except for the first (to preserve population information), all
alleles were recoded 1 were recoded to 2, all coded to 0 recoded to 1, and then all 1's and 2's were recoded to 01 and 02, respectively. Finally, the excess columns were removed and the header corrected to be consistent with FSTAT.
S4. R tutorial - calculating heterozygosity and plotting average patch heterozygosity in heat maps

library("RColorBrewer")
library("gplots")
library("dplyr")
library("magrittr")

Locus1 <- c(0.9, 0.9, 0.1, 0.4, 0.7, 0.3, 0.5, 0.7, 0.5)
Locus2 <- c(1, 0.4, 0.2, 0.1, 0.9, 0.2, 0.8, 0.6, 0.5)
Locus3 <- c(0.8, 0.3, 0.6, 1, 0.3, 0.1, 0.8, 0.6, 0.5)

q_df <- cbind(Locus1, Locus2, Locus3)
q_df <- as.data.frame(q_df)
str(q_df)

q_df_trans <- q_df %>% transmute(
  hets=(
    2*(Locus1*(1-Locus1)) +
    2*(Locus2*(1-Locus2)) +
    2*(Locus3*(1-Locus3))
  )/3
)

q_df_trans <- matrix(as.numeric(unlist(q_df_trans)), nrow=3, ncol=3, byrow=T)

display.brewer.all()
my_pal <- brewer.pal(7, "YlGnBu")
my_pal

heatmap.2(q_df_trans, breaks=8, Rowv=NA, Colv=NA,
  dendrogram="none", symm=T, trace="none", col=my_pal,
  margins=c(3,13), keysize=1.85,
  density.info="none",
  

key.xlab="Size", srtCol=0, offsetCol= -25, offsetRow=-16)
S5. R Analysis Script- arlsumstat processing, heterozygosity calculations, and heatmap plotting

#CRT April 2016: File processing and import
#CRT and NJA April 2016: Heterozygosity calculations
#NJA April-June 2016: Heat map plotting

## Code to import, re-shape, and plot frequency data in heat maps

### Basic Setup

```r
library(ggplot2)  # For generating the faceted heatmaps
library(gplots)    # For graphing goodness
library(dplyr)     # To massage raw data into correct format
library(readr)     # A better package by HWickham for importing data
library(ezknitr)   # A way to manage directory paths for input, output, and figure directories
library(DataLoader) # For importing multiple files and storing them as a list
library(tidyr)    # Package by HWickham for apply functions
library(plyr)      # Package to split list, apply function, and return results in a list
library(RColorBrewer)
```
library(magrittr)
library(Cairo)
library(ade4)
```

### Configuring knitting
#### Verify and list files in current working directory
```
getwd()
list.files()
```

#### Set up default directories for knitr output
```
ezknit(file = "Range_expansion_heat_maps.Rmd", out_dir = "RangeMapReports", fig_dir = "RangeMapFigures")
```

## Importing raw frequency data
### Cleaning up the frequency text files prior to import.
All of the frequency text files report both alleles giving the frequency for "Locus_1 01" on the first row, and that of "Locus_1 02" on the second row. We only need information on a single allele per locus, and so we can use AWK to retain only those rows with the frequencies of allele 01. The following AWK script does this. First put the set of commands into a variable that then is used in a system call. Remember that system calls are independent. This means that there can't be one system call that changes the directory and then another call that runs the AWK script; both commands must be in the same call. @rTip
```
getwd()  # Just checking ;)
scr.pullAWK <- "cd ./RangeData/; gawk -i inplace '$2=="01" {'print}' *.txt"
system(scr.pullAWK)
```

## Importing the data
The goal is pull in multiple data files, where each data file is average allele frequencies across patches from a single replicate run. Ultimately, we want to be able to calculate heterozygosity for each patch, and then plot this estimate of diversity as faceted heatmaps by replicate runs, and then faceted heatmaps for average heterozygosity by conditions of the simulations.

#### Import Strategy

We need to ensure that each of the files to be imported have the same number of rows (i.e. alleles), because when we create an array of data.frames coerced in matrices, there can be no missing values. This is because if there are any missing values, R will recycle---that would be trouble. Here's is some code that will accomplish iterating over the set of files in the directory counting the number of lines, and then output the count to a file. CAVEAT: The `wc` command is counting the number of newline characters in the file, so if there is ever a case where an additional newline character is accidentally added, it will serve as a proxy for a record. This would be bad.

```r
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
# If we want to remove the resulting file_row_count.output
file_row_count.output
  # scr.removeROWCOUNT <- "cd ./RangeData/; rm
dfile_row_count.output"
  # system(scr.removeCount)
```

```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
It will be much better for us if we allow it to auto-increment for the patches rather than the alleles. The solution is to transpose the rows and columns before using *DataLoader*. The following script accomplishes this:

```bash
# We can use Vim and the plugin 'salsifis/vim-transpose' to transpose our tab delimited files in one go.
# Open vim
# In command line: args **/*.txt
# In command line: bufdo TransposeTab|w <- Will transpose.
# Magic ensues
# In command line: bufdo 1027d|w <- Will strip out the last line of tabs.
# In command line: bufdo 2d|w <- Will strip out second line of "01"..."01"
# In command line: q
```

#### Using *DataLoader* to pull in the data frames

It may be possible to import the long lists of allele frequency text files into R using the package *DataLoader*. From the manual, "importTab function loads various text files which uses a tab delimiter in a selected directory to separate data frames, and stores them as a list. Data frames can be accessed as list elements by using 'listname$filename' or 'listname[]'."

- NOTE: This package threw an error related to the java virtual machine. I found this solution on StackOverflow. In the terminal, type `sudo R CMD javareconf`. Then, in the terminal, type `sudo ln -s $(/usr/libexec/java_home)/jre/lib/server/libjvm.dylib /usr/local/lib`. It should be possible to load *DataLoader* now. See comment by [Jack Tanner](http://conjugateprior.org/2014/12/r-java8-osx/#comment-312349).
```r
allAllele.dflist <- importTab(path = NULL)
allAllele.dflist.cp <- allAllele.dflist
allAllele.dflist.cp2 <- allAllele.dflist.cp
```

### Processing in R to get average heterozygosity

#### Use `dplyr::transmute` in the context of `tidy`

The method here is to process each dataframe individually. This is done by indexing the df from the list of files created via DataLoader, with the following: `df[[i]]`. Then a `transmute` function is used to calculate heterozygosity for each locus in each patch; these values are then averaged for each patch, and the results stored in a new dataframe. This new dataframe should take the original file name, plus the addition of an informative lable, i.e. '_trans'. Next, the resulting dataframe is converted to a numerical matrix of the correct lattice size and shape.

#### Function

```r
tothet_func <- function(alleledf){
  all <- alleledf %>% transmute(
    hets=(
      2*(Locus_1*(1-Locus_1)) + 
      2*(Locus_2*(1-Locus_2)) + 
      2*(Locus_3*(1-Locus_3)) + 
      2*(Locus_4*(1-Locus_4)) + 
      2*(Locus_5*(1-Locus_5)) + 
      2*(Locus_6*(1-Locus_6)) + 
      2*(Locus_7*(1-Locus_7)) + 
    )
  )
  ...
}
```

98
\[ 2 \times (\text{Locus}_8 \times (1 - \text{Locus}_8)) + 2 \times (\text{Locus}_9 \times (1 - \text{Locus}_9)) + 2 \times (\text{Locus}_{10} \times (1 - \text{Locus}_{10})) \] / 10

return(all)

singhet_func <- function(alleledf2){
  all2 <- alleledf2 %>% transmute(
    hets=(
      2 \times (\text{Locus}_1 \times (1 - \text{Locus}_1))
    )
  )
  return(all2)
}

make_mat <- function(trans){
  j <- matrix(as.numeric(unlist(trans)), nrow=32, ncol=32, byrow=T)
  return(j)
}

## These are for Moran's I
xcds <- rep(1:32, each = 32)
ycds <- rep(1:32, 32)
coords <- cbind(xcds, ycds)

# library("ape")
coords.dist <- as.matrix(dist(coords))
coords.dist.inv <- 1/coords.dist
diag(coords.dist.inv) <- 0

###Lapply calls

allAllele.dflist.cp <- lapply(allAllele.dflist.cp, tothet_func)
allAllele.dflist.cpmat <- lapply(allAllele.dflist.cp, make_mat)
allAllele.dflist.cp2 <- lapply(allAllele.dflist.cp2, singhet_func)
allAllele.dflist.cp2mat <- lapply(allAllele.dflist.cp2, make_mat)

###

## Moran's I ###
morans_tot <- lapply(allAllele.dflist.cp, geary.moran, bilis = coords.dist.inv)
names(morans_tot) <- names(allAllele.dflist.cp)
capture.output(morans_tot, file = "morans_totnmtest.txt")
morans_sing <- lapply(allAllele.dflist.cp2, geary.moran, bilis = coords.dist.inv)
names(morans_sing) <- names(allAllele.dflist.cp2)
capture.output(morans_sing, file="morans_sing.txt")

...

## Plotting heat maps
Plotting the heat maps is done using the RColorBrewer package for the color palette, and with the heatmap.2 function from gplots (basically a fancier and more user-friendly version of base). A custom palette is defined and used to define the bins plotted in specific colors. Be sure to select the correct number of shades for the palette! This number should be one less than the number of breaks in the plotting function. See help files for all arguments related to graphical display, labels, etc.

```r, include=TRUE```
system("cd ..")
system("cd ./RangeHeatMaps")
getwd()
my_pal <- brewer.pal(4, "YlGnBu")

for (i in names(allAllele.dflist.cp2mat)) {
  png(paste(i, "HeatmapSingHets_", ".png", sep=""),
      width=6.5, height=6.5,
      res=360, units="in", bg="NA")
  par(bg="transparent", pin=c(4.5, 4.5))

  heatmap.2(allAllele.dflist.cp2mat[[i]],
            breaks=c(0.48,
                    0.485, 0.49, 0.495, 0.5), Rowv=NA, Colv=NA,
            dendrogram="none", symm=T, trace="none", col=my_pal,
            key=F, density.info="none", xlab=NULL, ylab=NULL,
            labRow=" ", labCol=" ")
  dev.off()
}

...
S6. Bash PGDSpider loop

#!/bin/bash
date
hostname
echo $PBS_NODEFILE

source /path/to/bin/bash
module purge
module load java-1.7.0_03

cd /path/to/directory/files
for f in ls ./directory/*.dat; do java -Xmx16g -Xms512m -jar PGDSpider2-cli.jar -inputfile "$f" -outputfile "${f%.dat}.dat.arp" -inputformat FSTAT -outputformat ARLEQUIN -spid thesi_spid.spid; done

#This script calls bash and Java7RE. The job script moves the current directory to the home of the files to be converted. A simple looping construction is used to call the java based PGDSpider on each individual FSTAT file. The code should be read as such: for each file in the list of .dat extension (FSTAT) files in the current directory, use PGDSpider to convert each file from FSTAT to arlsumstat, allowing for up to 5GB of memory for each conversion. Conversions are performed using the data input and output types specified in the .spid file.
S7. PGDSpider .spid file for FSTAT to ARLEQUIN conversions


# FSTAT Parser questions
  PARSER_FORMAT=FSTAT

  # How are Microsat alleles coded?
  FSTAT_PARSER_MICROSAT_CODING_QUESTION=REPEATS

  # Select the type of the data:
  FSTAT_PARSER_DATA_TYPE_QUESTION=SNP

  # Open label file
  FSTAT_PARSER_LABEL_FILE_QUESTION=

  # Do you want to include a label file (listing the name of the populations)?
  FSTAT_PARSER_INCLUDE_LABELS_QUESTION=false

  # Enter the size of the repeated motif (same for all loci: one number; different: comma separated list (e.g.: 2,2,3,2)):
  FSTAT_PARSER_REPEAT_SIZE_QUESTION=

# Arlequin Writer questions
  WRITER_FORMAT=ARLEQUIN

  # Specify which data type should be included in the Arlequin file (Arlequin can only analyze one data type per file):
  ARLEQUIN_WRITER_DATA_TYPE_QUESTION=SNP

  # Specify the DNA locus you want to write to the Arlequin file or write "CONCAT" for concatenation:
  ARLEQUIN_WRITER_CONCATENATE_QUESTION=

  # Specify the locus/locus combination you want to write to the Arlequin file:
  ARLEQUIN_WRITER_LOCUS_COMBINATION_QUESTION=
S8. Sample Nemo Script with Comments

```
## SIMULATION ##
run_mode overwrite  # set to overwrite outputs each time; helps prevent too much space taken up during
root_dir DIRECTORY  # name of the directory where
random_seed 111111  # seed
filename FILENAME   # root filename for all output files
replicates 10       # number of replicate simulations to run
generations 10      # number of generations to run each simulation

## POPULATION ##
patch_number 10     # number of demes
patch_capacity 10   # patch carrying capacity at start

## RESIZE ##
resize_at_generation 10  # generation at which to introduce the new carrying capacity
resize_patch_capacity 10 20  # new carrying capacity; example of sequential parameter
resize_do_regulate 1  # regulate to carrying
```
capacity

## ORDER OF OPS ##

breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##

mating_system 2 #monogamy
mean_fecundity 2 #mean of poison
distribution;

time

## DISPERSAL ##

dispersal_model 2 #lattice model
dispersal_rate 0.2 #migration rate, changes over
time

dispersal_lattice_range 2 #up to x adjacent
patches

dispersal_border_model 2 #reflective boundaries

## HARVESTING ##

extinction_rate 1 #means this will occur w/100%
0.20

## OUTPUT ##

stat pop.patch #outputs all pop-
specific
stats
#stats output every 10
gens
stat_dir data
#folder to store stat
files
stat_output_CSV 1
#'.txt' file will be formatted as comma delimited

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
#alleles maximally polymorphic at start
ntrl_loci 10
#number of neutral loci
ntrl_all 2
#number of alleles at each locus
ntrl_mutation_rate 0.00001
#mutation rate
ntrl_mutation_model 2
#K-alleles model
ntrl_save_genotype FSTAT
#outputs data in FSTAT format
ntrl_output_logtime 10
#outputs at this frequency of generations
S10. CONTROL 1 stage

## SIMULATION ##
run_mode overwrite
root_dir CN_1S
random_seed 2131455645
filename CN_%1m_%2f_%3ini_%4cc
replicates 20
generations 500

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation 2
resize_patch_capacity 200 500
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity 3 4

## DISPERSAL ##
dispersal_model 4
dispersal_rate 0.3 0.2
dispersal_lattice_range 2
dispersal_border_model 2
## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S11. CONTROL 1 stage; BN 30%

## SIMULATION ##
run_mode overwrite
root_dir BN30_CN_1S
random_seed 21315645
filename BN30_CN_%1m_%2f_%3ini
replicates 20
generations 500

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation {{2, 450, 455}}
resize_patch_capacity (@g0 500, @g450 350, @g455 500)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity 3 4

## DISPERSAL ##
dispersal_model 4
dispersal_rate 0.3 0.2
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S12. CONTROL 1 stage; BN 60%

## SIMULATION ##
run_mode overwrite
root_dir BN60_CN_1S
random_seed 2131455645
filename BN60_CN_%1m_%2f_%3ini
replicates 20
generations 500

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation {{2, 450, 455}}
resize_patch_capacity (@g0 500, @g450 200, @g455 500)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity 3 4

## DISPERSAL ##
dispersal_model 4
dispersal_rate 0.3 0.2
dispersal_lattice_range 2
dispersal_border_model 2
## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S13. RE 1 stage

```bash
## SIMULATION ##
run_mode overwrite
root_dir REN_1S
random_seed 2131455645
filename REN_%1m_%2f_%3ini_%4cc
replicates 20
generations 500

## POPULATION ##
patch_number 1024
patch_capacity @PatchSizesSideFill.txt

## RESIZE ##
resize_at_generation 2
resize_patch_capacity 200 500
resize_do_regulate 1

## ORDER OF OPS ##
resize 1
breed 2
extinction 3
disperse 4
aging 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity 3 4

## DISPERSAL ##
dispersal_model 4
dispersal_rate 0.3 0.2
dispersal_lattice_range 2
```
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S14. RE 1 stage; BN 30%

## SIMULATION ##
run_mode overwrite
root_dir BN30_REN_1S
random_seed 2131455645
filename BN30_REN_%1m_%2f_%3ini
replicates 100
generations 500

## POPULATION ##
patch_number 1024
patch_capacity @PatchSizes.txt

## RESIZE ##
resize_at_generation {{2, 450, 455}}
resize_patch_capacity (@g0 500, @g450 350, @g455 500)
resize_do_regulate 1

## ORDER OF OPS ##
resize 1
breed 2
extinction 3
disperse 4
aging 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity 3 4

## DISPERSAL ##
dispersal_model 4
dispersal_rate 0.3 0.2
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 3
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S15. RE 1 stage; BN 60%

## SIMULATION ##
run_mode overwrite
root_dir BN60_REN_1S
random_seed 2131455645
filename BN60_REN_%1m_%2f_%3ini
replicates 100
generations 500

## POPULATION ##
patch_number 1024
patch_capacity @PatchSizesSideFill.txt

## RESIZE ##
resize_at_generation {{2, 450, 455}}
resize_patch_capacity (@g0 500, @g450 200, @g455 500)
resize_do_regulate 1

## ORDER OF OPS ##
resize 1
breed 2
extinction 3
disperse 4
aging 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity 3 4

## DISPERSAL ##
dispersal_model 4
dispersal_rate 0.3 0.2
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 3
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S16. CONTROL 2 stage
#low cc to high cc
#low to high migr
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir CN_LHcc_LHm_LHf
random_seed 213145246
filename CN_LHcc_LHm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 200, @g250 500)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)

## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.2, @g250 0.3)
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
S17. CONTROL 2 stage
#low cc to high cc
#high to low migr
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir CN_LHcc_HLm_LHf
random_seed 213145245
filename CN_LHcc_HLm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 200, @g250 500)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)

## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.3, @g250 0.2)
dispersal_lattice_range 2
dispersal_border_model 2

### HARVESTING ###
extinction_rate 1
extinction_proportion 0.30

### OUTPUT ###
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

### NEUTRAL MARKERS ###
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S18. CONTROL 2 stage
#high to low cc
#low to high m
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir CN_HLcc_LHm_LHf
random_seed 2131452897
filename CN_HLcc_LHm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 500, @g250 200)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)

## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.2, @g250 0.3)
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S19. CONTROL 2 stage

# high to low cc
# high to low m
# low to high fec

## SIMULATION ##
run_mode overwrite
root_dir CN_HLcc_HLm_LHf
random_seed 2131452895
filename CN_HLcc_HLm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity 100

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 500, @g250 200)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)

## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.3, @g250 0.2)
dispersal_lattice_range 2
dispersal_border_model 2

### HARVESTING ###
extinction_rate 1
extinction_proportion 0.30

### OUTPUT ###
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

### NEUTRAL MARKERS ###
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S20. RE 2 stage

#low cc to high cc
#low to high migr
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir REN_LHcc_LHm_LHf
random_seed 213145246
filename REN_LHcc_LHm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity @PatchSizesSideFill.txt

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 200, @g250 500)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)
## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.2, @g250 0.3)
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S21. RE 2 stage
#low cc to high cc
#high to low migr
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir REN_LHcc_HLm_LHf
random_seed 213145245
filename REN_LHcc_HLm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity @PatchSizesSideFill.txt

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 200, @g250 500)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)
## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.3, @g250 0.2)
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
S22. RE 2 stage
#high to low cc
#low to high m
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir REN_HLcc_LHm_LHf
random_seed 2131452897
filename REN_HLcc_LHm_LHf_%1init
replicates 20
generations 750

## POPULATION ##
patch_number 1024
patch_capacity @PatchSizesSideFill.txt

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 500, @g250 200)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)

## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.2, @g250 0.3)
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop.patch
stat_log_time 10
stat_dir data
stat_output_CSV 1

## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10

6S23. RE 2 stage
#high to low cc
#high to low m
#low to high fec

## SIMULATION ##
run_mode overwrite
root_dir REN_HLcc_HLm_LHf
random_seed 2131452895
filename REN_HLcc_HLm_LHf_%1init
replicates 20
generations 750
## POPULATION ##
patch_number 1024
patch_capacity @PatchSizesSideFill.txt

## RESIZE ##
resize_at_generation {{2, 250}}
resize_patch_capacity (@g0 500, @g250 200)
resize_do_regulate 1

## ORDER OF OPS ##
breed 1
extinction 2
disperse 3
aging 4
resize 5
save_stats 6
save_files 7

## MATING SYSTEM ##
mating_system 3
mean_fecundity (@g0 3, @g250 4)

## DISPERSAL ##
dispersal_model 4
dispersal_rate (@g0 0.3, @g250 0.2)
dispersal_lattice_range 2
dispersal_border_model 2

## HARVESTING ##
extinction_rate 1
extinction_proportion 0.30

## OUTPUT ##
stat pop_patch
stat_log_time 10
stat_dir data
stat_output_CSV 1
## NEUTRAL MARKERS ##
ntrl_init_model 0 1
ntrl_loci 10
ntrl_all 2
ntrl_mutation_rate 0.00001
ntrl_mutation_model 2
ntrl_save_genotype FSTAT
ntrl_output_logtime 10
Supplementary Figures

S24. Control Model Lattice

S25. Range Expansion Model Lattice
S26. Order of Lifecycle Events in NEMO simulations

- Breed
- Extinction
- Dispersal
- Aging
- Resize
- Save Stats and Files
S27. NEMO simulation algorithm
S28. ‘Simulating the effects of migration rates on Neolithic range expansion’. Poster presented at the 2016 American Association of Physical Anthropology annual meetings.
S29. A

S29. B

S29. C

S29. D
S29. A-BN30 CN 0.3m 3f 1ini 500 AvgHets; B-BN30 REN 0.3m 3f 1ini 500 AvgHets; C-BN30 CN 0.3m 3f 1ini 500 SingHets; D-BN30 REN 0.3m 3f 1ini 500 SingHets
S30. A  

S30. B  

S30. C  

S30. D  

S30. A-CN_LHcc_LHm_LHf_1init_070_AvgHets; B-REN_LHcc_LHm_LHf_1init_070_AvgHets;  
C-CN_LHcc_LHm_LHf_1init_070_SingHets; D-REN_LHcc_LHm_LHf_1init_070_SingHets
S31. A

S31. B

S31. C

S31. D

S31. A-CN_HLcc_LHm_LHf_lmit_750_AvgHets; B-REN_HLcc_LHm_LHf_lmit_750_AvgHets;
C-CN_HLcc_LHm_LHf_lmit_750_SingHets; D-REN_HLcc_LHm_LHf_lmit_750_SingHets
S32. A

S32. B

S32. C

S32. D

S. A-CN HLcc HLm LHF linit 750 AvgHets; B-REN HLcc HLm LHF linit 750 AvgHets;
C-CN HLcc HLm LHF linit 750 SingHets; D-REN HLcc HLm LHF linit 750 SingHets
S. A-CN_0.3m_4f_lini_200cc_020_AvgHets; B-REN_0.3m_4f_lini_200cc_020_AvgHets;
C-CN_0.3m_4f_lini_200cc_020_SingHets; D-REN_0.3m_4f_lini_200cc_020_SingHets
S34. A  

S34. B  

S34. C  

S34. D  

S34. A-CN_0.3m_3f_1ini_500cc_500_AvgHets; B-REN_0.3m_3f_1ini_500cc_500_AvgHets;  
C-CN_0.3m_3f_1ini_500cc_500_SingHets; D-REN_0.3m_3f_1ini_500cc_500_SingHets
S35. A  

S35. B  

S35. C  

S35. D  

S35. A-CN_0.3m_3f_1ini_200cc_500_AvgHets; B-REN_0.3m_3f_1ini_200cc_500_AvgHets;  
C-CN_0.3m_3f_1ini_200cc_500_SingHets; D-REN_0.3m_3f_1ini_200cc_500_SingHets
S36. A
S36. B
S36. C
S36. D

S36. A-CN_0.3m_3f_lini_200cc_050_AvgHets; B-REN_0.3m_3f_lini_200cc_050_AvgHets;
C-CN_0.3m_3f_lini_200cc_050_SingHets; D-REN_0.3m_3f_lini_200cc_050_SingHets
S37. A-CN 0.2m 4f lini 500cc 500 AvgHets; B-REN 0.2m 4f lini 500cc 500 AvgHets; C-CN 0.2m 4f lini 500cc_500 SingHets; D-REN 0.2m 4f lini 500cc_500 SingHets
S38. A - CN_0.2m_4f_1ini_500cc_020_AvgHets; B - REN_0.2m_4f_1ini_500cc_020_AvgHets;
C - CN_0.2m_4f_1ini_500cc_020_SingHets; D - REN_0.2m_4f_1ini_500cc_020_SingHets
S39. A

S39. B

S39. C

S39. D

S39. A-CN_0.2m_4f_1ini_200cc_500_AvgHets; B-REN_0.2m_4f_1ini_200cc_500_AvgHets;
C-CN_0.2m_4f_1ini_200cc_500_SingHets; D-REN_0.2m_4f_1ini_200cc_500_SingHets
S40. A

S40. B

S40. C

S40. D

S40. A-CN_0.2m_4f_1ini_200cc_020_AvgHets; B-REN_0.2m_4f_1ini_200cc_020_AvgHets; C-CN_0.2m_4f_1ini_200cc_020_SingHets; D-REN_0.2m_4f_1ini_200cc_020_SingHets
S41. A

S41. B

S41. C

S41. D

S41. A-CN_0.2m_3f_1ini_500cc_500_AvgHets; B-REN_0.2m_3f_1ini_500cc_500_AvgHets; C-CN_0.2m_3f_1ini_500cc_500_SingHets; D-REN_0.2m_3f_1ini_500cc_500_SingHets
S42. A

S42. B

S42. C

S42. D

S42. A-CN 0.2m 3f lini 200cc 500 AvgHets; B-REN 0.2m 3f lini 200cc 500 AvgHets; C-CN_0.2m_3f_lini_200cc_500_SingHets; D-REN_0.2m_3f_lini_200cc_500_SingHets
S43. A

S43. B

S43. C

S43. D

S43. A-CN_0.2m_3f_1ini_200cc_070_AvgHets; B-REN_0.2m_3f_1ini_200cc_070_AvgHets; C-CN_0.2m_3f_1ini_200cc_070_SingHets; D-REN_0.2m_3f_1ini_200cc_070_SingHets
S44. A

S44. B

S44. C

S44. D

S44. A-BN60_CN_0.3m_3f_1ini_500_AvgHets; B-BN60_REN_0.3m_3f_1ini_500_AvgHets;
C-BN60_CN_0.3m_3f_1ini_500_SingHets; D-BN60_REN_0.3m_3f_1ini_500_SingHets
S45. A

S45. B

S45. C

S45. D

S45. A-BN60_CN_0.2m_4f_1ini_500_AvgHets; B-BN60_REN_0.2m_4f_1ini_500_AvgHets; C-BN60_CN_0.2m_4f_1ini_500_SingHets; D-BN60_REN_0.2m_4f_1ini_500_SingHets
S46. A - BN60_CN_0.2m_4f_1ini_020_AvgHets; B - BN60_REN_0.2m_4f_1ini_020_AvgHets; C - BN60_CN_0.2m_4f_1ini_020_SingHets; D - BN60_REN_0.2m_4f_1ini_020_SingHets
S47. A

S47. B

S47. A-BN60_REN_0.2m_3f_lini_500_AvgHets; B-BN60_REN_0.2m_3f_lini_500_SingHets
CURRICULUM VITAE

NEHA ANGAL
Cell: 502.994.7559
Email: neha.angal@louisville.edu
Email: neha.jay.angal@gmail.com

EDUCATION

2016  University of Louisville
      Master of Arts in Anthropology (biological)

2013  Western Kentucky University
      Bachelor of Arts, Cum Laude
      Major: Anthropology (biological)
      Minor: Political Science

AWARDS

2016  Departmental Award- Graduate Student Merit
      Awarded by the Anthropology Department,
      University of Louisville

2016  Graduate Student Council Graduate Student
      Representative of the Year

2015-6  Departmental Graduate Research and Teaching
        Assistantship

2014-5  Departmental Graduate Research Assistantship

2013  Outstanding Undergraduate in the Biological
      Anthropology Concentration
2010-3  Dean’s List (3+ terms)

PROFESSIONAL SOCIETIES
2015-6  American Association of Anthropological Genetics
        Student Member

        American Association of Physical Anthropology
        Student Member

        Representative for the Anthropology Department to the
        College of Arts and Sciences Graduate Student Union
        (Fall only)

        Representative for the Anthropology Department to the
        Graduate Student Council

        Information Officer/Director of Communications for the
        Graduate Student Council

        First Annual Graduate Student Council Graduate Student
        Regional Research Conference Conference Planning
        Committee Member and Conference Co-chair

        Student Government Association Graduate Student
        Senator

2014-5  AAAG Student Member

        AAPA Student Member
Representative for the Anthropology Department to the College of Arts and Sciences Graduate Student Union

Divisional Representative for the Social Sciences to the Graduate Student Union

Representative for the Anthropology Department to the Graduate Student Council

2013-4 Representative for the Anthropology Department to the College of Arts and Sciences Graduate Student Union

PRESENTATIONS

2016 Simulating the effects of migration rates on Neolithic range expansion. Neha J. Angal and Christopher R. Tillquist. Poster presented April 16th, 2016, at the 85th Annual Meetings of the American Association of Physical Anthropology in Atlanta, GA.


Variation in Mandibular Morphology Related to Speech Production in Hominids. Neha J. Angal, Mentor and Contributor: Dr. Darlene Applegate. Paper presented in March at the 43rd Annual Western Kentucky University Student Research Conference.


ADDITIONAL ACTIVITIES

2015-6 Farmers Market Research Project Team Member
University of Louisville, Department of Anthropology
P.I. and Supervisor: Dr. Lisa Markowitz

Mexican Biological Variation Project
Student Member/Field School Attendee
Chiapas, Mexico, Summer 2015
University of Michigan
P.I. and Field School Director: Dr. Abigail Bigham

2014-5 Farmers Market Research Project Team Member
University of Louisville, Department of Anthropology
P.I. and Supervisor: Dr. Lisa Markowitz