Spatial Structure in Extinction and Range Expansion: Models and Experiments

Christopher Weiss-Lehman
University of Colorado at Boulder, tpweiss06@gmail.com

Follow this and additional works at: https://scholar.colorado.edu/ebio_gradetds
Part of the Evolution Commons, Natural Resources and Conservation Commons, and the Population Biology Commons

Recommended Citation
https://scholar.colorado.edu/ebio_gradetds/97
SPATIAL STRUCTURE IN EXTINCTION AND RANGE EXPANSION: MODELS AND EXPERIMENTS

by

CHRISTOPHER WEISS-LEHMAN

B.A., Earlham College, 2010

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Ecology and Evolutionary Biology
2017
This thesis entitled:
Spatial structure in extinction and range expansion: models and experiments

Written by Christopher Weiss-Lehman

has been approved for the Department of Ecology and Evolutionary Biology

____________________________________
Dr. Brett Melbourne (Committee Chair)

____________________________________
Dr. Nolan Kane

Date_____________

The final copy of this thesis has been examined by the signatories, and we find that both the content and form meet acceptable presentation standards of scholarly work in the above mentioned discipline
ABSTRACT

We are facing a global biodiversity crisis as extinction risks increase for species around the globe as a result of anthropogenic activities. Some of the most prominent causes of this trend include habitat fragmentation and invasive species. While these threats to biodiversity differ in their ultimate causes, they share a unifying theme: they both threaten biodiversity via mechanisms involving the spatial structure of populations. In the case of habitat fragmentation, a once continuous population is broken into smaller subpopulations, thus potentially increasing extinction risk. For an invading population, on the other hand, spatial structure is created from the expansion front to the core, potentially increasing invasion success. In my dissertation, I consider extinction risk in spatially subdivided populations and expansion speeds of spatially structured invaders. In Chapter 2, I detail the derivation and validation of a hierarchical model to assess extinction risk for poorly known species with little available data. The model accomplishes this by leveraging the spatial population structure characteristic of habitat fragmentation. In Chapters 3 and 4, I consider the context of invasive species and test the role of spatial structure during range expansions in driving the evolution of key traits at the expansion edge. In Chapter 3, I present data from a tightly controlled microcosm experiment to show that this trait evolution not only increases expansion speed on average, but also dramatically increases variability in expansion speeds. In Chapter 4 I present genomic data from these experimental populations to explore and quantify the evolutionary mechanisms underlying the increased speed and variance. Finally, Chapter 5 considers the intersection of habitat
fragmentation and invasive species by considering the role of evolution due to spatial population structure in range expansions through fragmented habitat. I combine empirical data with a theoretical model to demonstrate that habitat heterogeneity reduces overall variance in spread rates, primarily due to the role of dispersal evolution at the expansion edge. Throughout my dissertation, I use a combination of empirical and theoretical approaches which allows me to provide greater mechanistic understanding of the importance of population spatial structure to ecological dynamics.
ACKNOWLEDGEMENTS

Funding for the work presented here was provided through various grants including National Science Foundation Grants DEB 0949595 to Brett A. Melbourne, DEB 0949619 to Ruth A. Hufbauer, DEB 1601333 to Brett A. Melbourne and myself, a University of Colorado Dean’s Grant to me, and several departmental grants awarded to me from the Ecology and Evolutionary Biology department of the University of Colorado, Boulder. Publication of Chapter 3 in *Nature Communications* was partially funded by the University of Colorado Boulder Libraries Open Access Fund. Additionally, I would like to thank Jasmine Vidreo, Tristin Myers, Sean Race, Nick Richter, Frances Janz, McKenzie Gibson, Jake Depompolo, Kabita Ghale, Adam Solon, Naomi Ochwat, Tyler Turk, Min Duc Le and Rabita Ghale for assistance with data collection. I would also like to thank Brad Ochocki, Tom Miller, Stacy Endriss, and Megan Vahsen, for comments on Chapter 3 specifically. Finally, this would not have been possible without the support of my family, friends, and fellow graduate students in the Ecology and Evolutionary Biology department.
TABLE OF CONTENTS

Chapters

1. Introduction ................................................................. 1

2. Estimating extinction risk with minimal data ......................... 7
   2.1 Introduction ............................................................. 8
   2.2 Material and Methods ............................................... 11
      2.2.1 Derivation of the extinction risk model .................. 11
      2.2.2 Model testing: Simulated Data ............................ 14
      2.2.3 Model testing: Microcosm Data ............................ 16
      2.2.4 Applying the model: Habitat fragmentation experiment .. 18
   2.3 Results ................................................................. 19
      2.3.1 Simulated data .................................................. 19
      2.3.2 Microcosm data ................................................ 26
      2.3.3 Habitat fragmentation ........................................ 28
   2.4 Discussion ............................................................. 31

3. Rapid trait evolution drives increased speed and variance in experimental range expansions 36
   3.1 Introduction ............................................................. 37
   3.2 Methods ................................................................. 40
      3.2.1 Experimental system ......................................... 40
      3.2.2 Range expansion experiment ............................... 41
      3.2.3 Testing for spatial trait evolution ......................... 42
      3.2.4 Statistical analyses .......................................... 43
      3.2.5 Expected maximum change in expansion speed ........... 46
   3.3 Results ................................................................. 48
      3.3.1 Speed and variance of range expansion .................... 48
      3.3.2 Trait evolution ................................................. 50
   3.4 Discussion ............................................................. 53
4. Genetic signatures of adaptation and gene surfing in replicated biological range expansions 60

4.1 Introduction .............................................................. 60
4.2 Methods ................................................................. 63
   4.2.1 Range expansion experiment ................................. 63
   4.2.2 DNA pooling, extraction, and sequencing ................ 64
   4.2.3 Analyzing nucleotide diversity ............................. 66
   4.2.4 Assessing the population covariance structure ...... 67
4.3 Results ................................................................. 69
4.4 Discussion ............................................................. 74

5. Habitat heterogeneity modulates the role of evolution in range expansions ............... 79

5.1 Introduction ............................................................. 79
5.2 Methods ................................................................. 83
   5.2.1 Experimental system ........................................... 83
   5.2.2 Individual-based model ....................................... 84
      5.2.2.1 Trait evolution ........................................... 85
      5.2.2.2 Local population dynamics ........................... 87
      5.2.2.3 Dispersal among patches ............................. 88
   5.2.3 Simulation experiment ........................................ 89
5.3 Results ................................................................. 90
5.4 Discussion ............................................................. 98

6. Conclusion .............................................................. 103

Bibliography ............................................................. 107

Appendices
   Appendix A. Derivation of extinction risk model from Chapter 2 ................. 117
   Appendix B. Trimming and alignment statistics for sequence data from Chapter 4 .. 120
   Appendix C. Results for the X chromosome from Chapter 4 ....................... 122
Appendix D. Sensitivity analysis for the model presented in Chapter 5 . . . . . . . . . . 124
TABLES

Table 5.1 Parameter and initial condition values for individual-based model .............. 90
FIGURES

Figure 2.1 Accuracy and precision with missing data .......................... 21
Figure 2.2 Probability of regional extinction with colonization .................. 22
Figure 2.3 Model error due to the amount of data ................................. 24
Figure 2.4 Model precision due to the amount of data ............................. 25
Figure 2.5 Probability of extinction for protist microcosms ......................... 27
Figure 2.6 Posterior distribution of the effect of fragmentation on extinction risk ........ 29
Figure 2.7 Posterior distribution of standard deviations of random effects. ............ 31
Figure 3.1 Distance spread and variance of structured and shuffled expansions ........ 48
Figure 3.2 CV and mean to variance relationship .................................. 50
Figure 3.3 Dispersal propensity in core, edge, and shuffled populations .............. 51
Figure 3.4 Mean growth rates in core, edge, and shuffled populations. ............... 52
Figure 3.5 Density dependent growth rates in core, edge, and shuffled populations .... 53
Figure 3.6 Correlation between growth rate and dispersal in edge populations .......... 56
Figure 4.1 Nucleotide diversity in founding and generation 8 populations ............. 69
Figure 4.2 Differences in nucleotide diversity between founding and generation 8. .... 70
Figure 4.3 Log transformed nucleotide diversity across the genome .................. 71
Figure 4.4 Smoothed nucleotide diversity across the genome ...................... 72
Figure 4.5 Probability of sharing low diversity sites across replicate populations .......... 73
Figure 4.6 Principal component analysis of the population covariance matrix .......... 74
Figure 4.7 Distribution of populations along principal component 1 ................. 75
Figure 5.1 Distance spread of beetle populations through homogenous and heterogeneous landscapes ......................................................... 91
Figure 5.2 Mean distance spread of simulated populations through different landscapes ... 92
Figure 5.3 Difference in simulated mean distance spread between different landscapes. 93
Figure 5.4 CV of distance spread of simulated populations through different landscapes. 94
Figure 5.5 Difference in simulated CV between different landscapes. .......................... 95
Figure 5.6 Dispersal evolution after 7 generations of simulated expansion. ...................... 96
Figure 5.7 Fitness evolution after 7 generations of simulated expansion. .......................... 97
CHAPTER 1

Introduction

The formation and maintenance of spatial structure within populations and communities underlies many areas of ecology from pattern formation to demography (MacArthur & Wilson 1967; Levins 1969; Chesson 2000; Leibold et al. 2004; Kubisch et al. 2013). Spatial structure in populations or communities can be thought of simply as spatial patterns of population abundance or species richness in a landscape. Well known examples of spatial structure span multiple scales of ecology including biodiversity gradients across latitudes (Gaston 2000), species range boundaries (Holt & Keitt 2000), species occupying a fragmented landscape (Haddad et al. 2015), and species invading novel environments (Melbourne et al. 2007). While spatial structure is often viewed as the outcome of multiple underlying processes, such as demographic rates, dispersal, and population and community dynamics, spatial population structure itself is often an important mechanism driving population dynamics. For example, feedbacks between population dynamics and spatial structure form the basis of both metapopulation and metacommunity theory (Levins 1969; Leibold et al. 2004; Shoemaker & Melbourne 2016), which examine population and community dynamics in landscapes with discrete habitat patches and dispersal of individuals between patches.

The relationship between spatial structure and population dynamics is particularly critical for conservation. Many of the most severe threats to biodiversity explicitly involve direct or indirect anthropogenic alteration of species’ spatial structure. For example, Jared Diamond coined the term “evil quartet” to describe the main drivers of modern day extinctions: over exploitation, habitat destruction and fragmentation, invasive species, and chains of extinctions in which one extinction triggers subsequent losses (Diamond 1989). Since Diamond’s original use
of the term, others have argued that climate change is a critical addition to the list (Brook et al. 2008). Of these major drivers of extinction, three explicitly involve changes to species’ spatial population structure: habitat fragmentation, invasive species, and climate change.

Habitat fragmentation not only decreases environmental quantity and quality, but also inherently alters the spatial structure of species’ habitat both by reducing the amount of habitat and the connectivity among remnant habitat patches (Haddad et al. 2015). This change in spatial habitat structure directly impacts species’ spatial population structure by dividing a single population into distinct subpopulations, sometimes connected by dispersal depending on the scale of the fragmentation. This change in a species’ spatial population structure often results in an increased risk of extinction via several possible mechanisms. Extinction risk can be heightened simply due to a reduction in population size resulting from the decrease in available habitat (Haddad et al. 2016), as a result of the reduced connectivity among subpopulations leading to unsustainable metapopulation dynamics (Tilman 1994), or, more likely, due to an interaction between the two (Haddad et al. 2016). Regardless of the specific cause, one of the main challenges in conservation biology is quantifying the extinction risk faced by species for which we have little data beyond simple, spatial metrics like site occupancy (Morais et al. 2013; Bland et al. 2015a). In Chapter 1 of my dissertation, I address this need by creating and testing a hierarchical, statistical model to evaluate extinction risk in fragmented habitats for species with minimal available data. I demonstrate that by using information on the spatial occupancy structure of a population, the model can accurately recover extinction risks even when faced with missing data or small datasets.

Unlike habitat destruction and fragmentation, the conservation threat from invasive species and climate change does not come from anthropogenic alterations to spatial habitat
structure, but rather from changes to the spatial population structure of one or more species’ (i.e. range expansions and shifts). For invasive species, this alteration in spatial population structure typically arises from the removal or reduction of a historical barrier to very long range dispersal and is typically facilitated by human activities (e.g. intercontinental travel and commerce by humans provides a much higher probability of dispersal between continents; Seebens et al. 2016). Similarly, species shifting their ranges in response to climate change are also responding to the anthropogenic alteration of historical range limitations and factors such as temperature and precipitation (Chen et al. 2011). As species move into new areas they often increase the extinction risk of the current inhabitants through competition and predation pressures, creating a conservation concern, especially in the case of invasive species (Elton 1958). It is therefore crucial to understand how changes to population structure that occur via range expansions and shifts cause a feedback that impacts population dynamics.

In range expansions, spatial population structure forms directly as a result of the expansion process. Population density increases to carrying capacity in the range core, but remains low at the expansion edge as it moves through space (Skellam 1951). This characteristic spatial population structure of range expansions is often referred to as an expansion wave due to the consistent gradient in population density from core to edge as the expansion proceeds (Fisher 1937). As a direct consequence of the spatial population structure of range expansions, they are subject to several unique evolutionary mechanisms that can result in evolved changes in dispersal ability (Shine et al. 2011), competitive ability (Burton et al. 2010), and fitness (Peischl et al. 2013). While the effects of these evolutionary mechanisms on evolution of individual traits have been explored previously (Burton et al. 2010; Phillips et al. 2010; Peischl et al. 2013), the overall effect of evolution due to the spatial population structure of range expansions on
expansion dynamics had not been experimentally tested. In Chapter 3, I use tightly controlled experimental microcosms of the red flour beetle (*Tribolium castaneum*) to test the importance of evolution due to spatial population structure in range expansions. I show that the evolution resulting from spatial population structure over just eight generations of expansion is responsible for dramatically increasing variance in expansion speeds compared to control populations without the spatial population structure.

The evolutionary mechanisms resulting from spatial population structure in range expansions, while sharing the same ultimate cause, act via distinct processes at the genomic level. In particular, several mechanisms involve differing selection pressures imposed at the core versus the edge of expansion, resulting in directional, adaptive changes in allele frequencies (Burton et al. 2010; Phillips 2015). In contrast, another key mechanism increases the role of neutral dynamics, resulting in increasingly stochastic changes in allele frequencies (Excoffier et al. 2009). While these mechanisms act in fundamentally different ways (i.e. highly deterministic vs. highly stochastic changes in allele frequencies), in practice they can be difficult to distinguish within a single range expansion. In Chapter 4, I sequence DNA from the beetle populations used in Chapter 3 to quantify the relative roles of these different mechanisms in driving evolution due to the spatial structure of range expansions. By analyzing patterns in genomic evolution from replicated range expansions, I am able to distinguish the signatures of adaptive and neutral evolution, demonstrating that neutral evolution plays a highly significant role in the evolution of edge populations.

As demonstrated in Chapters 2 to 4, even a single mechanism that alters a species’ spatial structure, such as range expansion, can have large conservation implications. However, due to the scale of anthropogenic changes to the planet, multiple mechanisms often work together to
influence spatial structure. Habitat across the globe is being destroyed and fragmented at alarming rates (Haddad et al. 2015) and, due to the increasing numbers of invasive species (Pimentel et al. 2005) and the global nature of climate change (Loarie et al. 2009), more species are experiencing range expansions and shifts. It is therefore critical to understand how multiple mechanisms that alter spatial structure might interact. In Chapter 5, I explore this question by constructing a mechanistic, individual-based model to simulate range expansion through heterogeneous landscapes. Combining this model with experimental data from the *T. castaneum* system, I explore the interplay between spatial evolutionary processes from range expansion and the patchy spatial structure of habitats, characteristic of habitat destruction and fragmentation. As I show in Chapter 5, the combination of evolution due to the spatial population structure of range expansion and habitat fragmentation interact to reduce both expansion speed and variability in fragmented habitats.

By using a combination of hierarchical statistical models, experimental data at both population and genomic scales, and mechanistic simulation modeling, my dissertation examines the importance of spatial structure in population dynamics generally and in some important conservation settings. I demonstrate that information on the spatial occupancy structure of a population enables accurate assessments of the species’ extinction risk, even with small amounts of data. Further, I show that spatial population structure due to range expansion can act as an important causal mechanism in driving patterns of expansion and delve into the specific evolutionary mechanisms responsible. Finally, I explore the combined role of altered population spatial structure due to range expansion and habitat fragmentation in altering expansion dynamics. In the last chapter of my dissertation, I conclude by discussing the consequences of
my findings and future research directions to gain a more complete understanding of the role of population spatial structure in ecology and evolution.
CHAPTER 2

Estimating extinction risk with minimal data

Christopher Weiss-Lehman, Kendi F. Davies, Christopher Clements, and Brett A. Melbourne

Adapted from

Weiss-Lehman, C., Davies, K. F., Clements, C., & Melbourne, B. A. In review. Estimating extinction risk with minimal data. *Biological Conservation*

Abstract

Increasing anthropogenic pressures on the global biome are causing widespread species declines and extinctions. The ability to assess the extinction risk faced by individual species is a critical first step in combating this trend. However, we lack high quality demographic data to do so for the vast majority of plant and animal species. We present an efficient modeling approach to assess extinction risk based on a statistical framework from the mark-recapture literature. We assessed the model’s performance using a combination of simulated data, results from a protist microcosm experiment, and data from a long-term, large-scale habitat fragmentation experiment in southeastern Australia. Simulation experiments showed the model is robust to missing data as well as biological processes not included explicitly in the model’s assumptions. Fitting the model to data from the protist experiment yielded accurate predictions of the regional extinction dynamics observed in the system, even with a relatively low level of replication. Finally, the model was able to accurately predict the observed dynamics in the habitat fragmentation experiment. The model provides a robust and accurate method to evaluate a species’ extinction risk. Since it only requires presence/absence data, applies to a wide range of survey designs, and
allows for observational uncertainty and missing data, it can be applied to many datasets that existing models cannot accommodate. For these reasons, the model should be useful in conservation settings.

**Introduction**

The ability to accurately assess a species' risk of extinction is one of the most important goals of conservation biology (McCarthy et al. 2004). This is not a trivial task, in part due to the role of stochasticity in extinction (Melbourne & Hastings 2008), and to our lack of data for the vast majority of organisms on Earth (Costello et al. 2013). A recent study estimates the current rate of extinction at 1000 times higher than the extinction rate before anthropogenic pressure (Pimm et al. 2014). However the Red List of Threatened Species from the International Union for Conservation of Nature has only assessed a small fraction of known plants and animals, and of those many are classified as data deficient (IUCN 2015), a rapidly growing category (Bland et al. 2016). While much work has focused on assessing extinction risk specifically for these data deficient species (e.g. Morais et al. 2013; Bland et al. 2015a), many more species have yet to be assessed at all. However, the costs associated with thorough surveys to assess extinction risk can be prohibitive (Bland et al. 2015b). Statistical methods are therefore needed both to extract more relevant information from the datasets we do have and to reduce the logistic difficulties associated with data collection by using less data, of types that are easier to collect.

Many different models exist for the estimation of extinction risk but most have data requirements that cannot always be satisfied by even the highest quality datasets available (Coulson et al. 2001). The most commonly used tool to assess a species’ extinction risk is Population Viability Analysis (PVA) (Shaffer 1983; Brook et al. 2000; Lindenmayer & McCarthy 2006). First developed by Shaffer (1983) for grizzly bears in Yellowstone National
Park, PVAs use demographic and environmental data to parameterize stochastic simulations of population dynamics through time. These simulations are then used to assess the probability of the population falling below a quasi-extinction threshold within a given amount of time (Shaffer 1983). Since their development, however, PVAs have proven controversial due to the amount of data required for accurate predictions (Fieberg & Ellner 2000; Coulson et al. 2001; Ellner et al. 2002).

In addition to lacking the amount and duration of data necessary for PVAs, many datasets of threatened and endangered species are subject to uncertainties and errors in the data collection process (Mackenzie et al. 2006). The inherent difficulty of attempting to infer the true state of a population from imperfect observational data has driven much progress in the related fields of occupancy modeling and mark-recapture modeling (Mackenzie et al. 2006; Lindberg 2010). However, to use occupancy modeling to assess the extinction risk of a patchily distributed species usually requires data conforming to Pollock’s robust design (1982) in which data must be stratified into primary periods (e.g. breeding seasons) with repeated, independent data collections within each primary period. This design allows the estimation of both extinction and colonization rates (Royle & Kéry 2007), thus allowing an explicit calculation of a regional extinction probability. While this approach significantly reduces data requirements compared to PVAs, data collection under the robust design can still be costly due to the repeated, independent sampling events required. This is perhaps one reason why studies accounting for imperfect detection are still so rare despite the abundance of methods correcting for it (Kellner and Swihart 2014). Recent studies have attempted to bypass the data requirements of the robust design, using assumed spatial relationships among occupancy for neighboring patches (Eaton et al. 2014) or unique covariates between the processes governing occupancy and observation (Lele et al. 2012).
These techniques can be quite useful, but still require additional data either in the form of information on the dispersal behavior and neighborhood size of the focal species or additional covariates to collect. Here, we propose a new method for assessing extinction risk while accounting for observation errors that does not rely on the robust design. The method we develop combines occupancy models with classic models from the mark-recapture literature (Cormack 1964; Jolly 1965; Seber 1965) to provide an order of magnitude estimate of extinction risk for species with minimal data available. It should be noted, however, that this method is not intended to replace dynamic occupancy models or PVAs for use in extinction risk assessments when the necessary data are available.

By treating habitat patches as analogous to marked animals with the probability of death corresponding to the probability of permanent extinction, we derive a likelihood equation analogous to classic models for mark-recapture studies (Darroch 1959), which can be applied to a range of experimental designs and incorporate system specific features such as spatial structure, community composition, and habitat alteration or degradation. The stochastic model has three basic parameters and may be applied using either a frequentist or a Bayesian approach. We first demonstrate the robustness of this extinction risk model to missing data and to violations of model assumptions using simulated data. We then use simulated data to examine the relationship between the size of a dataset and the accuracy of the model estimates. We also apply the model to two different biological datasets. The first consists of a protist microcosm experiment (Clements et al. 2013), which we use to show the accuracy of the model for replicated biological populations in which the time of extinction is known. The second dataset documents beetle population dynamics in a large-scale, long-term habitat fragmentation
experiment in southeastern Australia (Margules 1992; Davies et al. 2004) and demonstrates the model’s ability to accurately assess extinction risk for species coping with habitat fragmentation.

**Materials and Methods**

*Derivation of the extinction risk model*

We consider observations of a species in multiple locations through time as a function of three probabilities ($n$, $w$, and $f$), which are the basic parameters of the model. This approach requires binary data (observed, not observed) from multiple locations and multiple time points at each location but with only modest replication and is well suited to most monitoring programs. Analogous to the way in which the probability of observing a marked animal through time is modeled as a combination of processes affecting observation and the probability of death or emigration (Cormack 1964; Jolly 1965; Seber 1965), we model observation events (observed/not observed) at a location as a combination of the probabilities mentioned above. The first parameter, $n$, is the probability that the focal species will occupy a given patch at least once during the study period (whether or not it is detected). This parameter is the key differentiator between the current model and similar models from the mark-recapture literature which only consider modeling marked animals after their first appearance in the data. Thus, this parameter allows the inclusion of habitat patches in which the focal species might not have been sighted yet and relates to the natural rarity of the species in the region. This is an important consideration as natural rarity can have important effects on extinction risk for a species (Caughley 1994). The next parameter, $w$, is the conditional probability that a currently or previously occupied patch will transition to permanently unoccupied (i.e. permanent rather than temporary extinction). This parameter, while related to the extinction process, is not equivalent to extinction as it is sometimes conceptualized in metapopulation models in which patches can go temporarily extinct.
and become recolonized later (Levins 1969). While we do not assume this process of temporary extinction and recolonization is absent from the data, we do not attempt to model it explicitly so as to reduce the data requirements from the model and allow the estimation of the parameters of greatest conservation concern (i.e. likelihood of a permanent extinction event absent any management intervention). The final parameter, \( \phi \), is the conditional probability that the focal species is observed in a patch, given the patch has not yet become permanently unoccupied. While this parameter is related to a traditional detection probability in occupancy models, it should not be confused with one as it also encompasses the probability that a patch may be temporarily unoccupied, but not permanently so (i.e. recolonization may still occur).

Consider first the case where all locations have a complete set of observations through time. The likelihood of the data can be described by

\[
\prod_{i=1}^{n} \left[ \sum_{t=1}^{T} y_{i,t} = 0 \right] (1 - \nu) + \nu \left( \prod_{t=1}^{k_i} (1 - \omega) \phi^{y_{i,t}} (1 - \phi)^{1-y_{i,t}} \right) \left[ \sum_{t=k_i}^{T} y_{i,t} = 0 \right] \left( \sum_{t=1}^{T} \left[ \sum_{t=1}^{T} y_{i,t} = 0 \right] \right) \left( \sum_{t=k_i}^{T} (1 - \omega)^{t-k_i} \omega (1 - \phi)^{t-1} + (1 - \omega)^{T-k_i} (1 - \phi)^{T-k_i} \right),
\]

(Eqn 1)

where \( n \) is the number of locations, \( T \) is the number of time points, \( y_{i,t} \) is the binary observation (1 for observed or 0 for not observed) at location \( i \) at time \( t \), \( k_i \) is the time of the last recorded presence for location \( i \) (i.e. \( y_{i,t} = 1 \) for \( t = k_i \), and \( y_{i,t} = 0 \) for \( t > k_i \)), and the parameters are defined as above. Double square brackets denote an indicator function, which evaluates to 1 when the interior condition is true and 0 otherwise. Briefly, we can assume that a species has not gone permanently extinct at a location at least up until the last time, \( k_i \), that it was observed to be present there. This assumption allows the probability of any single observation to be conditioned on data both before and after it in the time series for a location and contrasts with most patch-occupancy models, which instead assume a Markov process. Thus, before \( k_i \), zeros are due to temporary local extinction or failure to detect the species, whereas zeros after \( k_i \) could also be
explained by permanent extinction at any time after $k_i$. A location with all zeroes could be explained by unsuitable habitat (i.e. it cannot be occupied), or any combination of local extinction, failure to detect, or permanent extinction.

The likelihood can be modified to account for missing data, such as might occur if some site visits were missed, data were lost, or visits were irregular. The likelihood then becomes

$$
\prod_{i=1}^n \left[ (\sum_{t=1}^T y_{i,t} = 0) (1 - \nu) + \nu \left( (\prod_{t=1}^{k_i} (1 - \omega)) [\phi^{y_{i,t}} (1 - \phi) (1 - \omega) (1 - \phi)^{T-k_i} (1 - \phi)^{T-k_i}) + \left( \sum_{j=1}^{\tau_{i,j}} (\prod_{t=1}^{r_{i,j}} (1 - \omega)) [\phi^{y_{i,t}} (1 - \phi) (1 - \omega) (1 - \phi)^{T-r_{i,j}} (1 - \phi)^{T-r_{i,j}}) \right] \right) \right],
$$

(Eqn 2)

where $z_i$ is the number of missing data points occurring after $k_i$ (i.e. after the last successful observation) at location $i$ and $\tau_{i,j}$ is the time point of the $j^{th}$ missing data point after $k_i$ at location $i$. Missing data points occurring before $k_i$ (i.e. before the last confirmed observation) do not alter the likelihood as we can infer the species has not gone permanently extinct at the location yet. However, for missing data points after $k_i$, it is possible that one of the missing data points might have been an observation, and thus could have changed the value of $k_i$. Therefore, for missing data points after $k_i$, Eqn. 2 considers both possible realizations (presence or absence) to account for the resulting uncertainty in the true value of $k_i$. Detailed derivations of equations 1-2 are given in Appendix A.

While $\omega$ relates to the probability of local extinction, the probability of regional extinction, $P_{\text{reg}}$, will most often be the quantity of interest. For any integer time $t$, the cumulative probability of regional extinction, $P_{\text{reg}}(t)$, is the probability that the species has either gone extinct or never occurred at all sites, that is

$$
P_{\text{reg}}(t) = \prod_{i=1}^n [1 - \omega + \nu (1 - (1 - \omega)^t)].
$$

(Eqn 3)
In a frequentist framework, point estimates and confidence intervals of $v$ and $\omega$ may be used to generate a probability curve over a range of times and in a Bayesian framework the full posterior distribution of this curve may be calculated directly. For the examples that follow, we use a Bayesian approach, which is convenient for the hierarchical nature of these and many other biological datasets.

**Model testing: Simulated data**

We used simulated data to investigate the accuracy and bias of the extinction risk model when fitted to datasets with missing data, varying rates of temporary extinction and recolonization, and of varying size.

To simulate datasets with missing data, we set values of all three model parameters and used them to simulate observations of a focal species across 20 locations with 10 time points. At the start of each simulation, each site was assigned as either unsuitable (i.e. permanently unoccupied) or not according to a Bernoulli process with parameter $v$. Then, the time point of permanent extinction for each patch was determined using a series of Bernoulli trials with parameter $1 - \omega$ (i.e. the probability of not going permanently extinct). This is also equivalent to a single draw from a negative binomial distribution with $p = 1 - \omega$ and only one allowed “failure” (extinction). After the time of permanent extinction, all data for a patch were set to 0 (non-detection), but before that time, data points were determined by a Bernoulli process with parameter $f$. As noted above, $f$ determines the probability of a detection prior to permanent extinction and can encompass processes related to both observation and temporary extinction and recolonization dynamics. After simulating the full dataset, we randomly removed a number of data points ranging between 0 and 100. For each level of missing data, we simulated 100 datasets as described above, fitted the extinction risk model to them, and compared the posterior mean of
ω to the true value used in the simulations. To examine the effect of missing data on precision of model estimates as well, we calculated the average width of the 95% credible intervals across the 100 replicates for each level of missing data.

To investigate the performance of the extinction risk model when colonization dynamics are actually present but not explicitly included in the extinction risk model, we generated data using a different model that incorporated colonization and fitted the extinction risk model (i.e. lacking colonization dynamics) to those data. Henceforth, we will call the data generating model the true model. For the true model we used the same simulation procedure as above, without removing data points. To include recolonization of sites in the true model, we did not model the time of permanent extinction via a negative binomial distribution as described above, but instead explicitly modeled the state of each patch through time as dependent on its state at the previous time point (i.e. Markovian dynamics). Initially occupied sites (as determined by a positive outcome from the Bernoulli trial with ν) either remained occupied or transitioned to extinct via another Bernoulli process with parameter 1 - ω. If the site became unoccupied, then its subsequent state was determined via a different Bernoulli process with parameter $P_{col}$. $P_{col}$, or the probability of an unoccupied site being recolonized, did not depend on spatial relationships among sites (i.e. we assumed global dispersal) or the number of occupied sites (i.e. we assumed constant propagule pressure), provided at least one site was occupied. Once no occupied sites remained, $P_{col}$ became 0 and regional extinction occurred. We evaluated three levels of $P_{col}$ to examine the effects of no (0), low (0.05), and high (0.1) colonization rates. The case of $P_{col} = 0$ is equivalent to the extinction risk model. For each colonization rate, we simulated 50 datasets, fitted the extinction risk model to the simulated data, and obtained the posterior distribution of $P_{reg}(t)$, the cumulative probability of regional extinction through time. From each posterior, we
calculated the mean for the cumulative probability of regional extinction, $P_{\text{reg}}(t)$, at each time, yielding a distribution of estimated means from the 50 simulated datasets through time. To obtain the expected values for $P_{\text{reg}}(t)$ for the true model, we simulated each parameter combination 10,000 times, recorded the time of regional extinction in each realization, and calculated the proportion of simulations out of 10,000 to have gone regionally extinct through time. To assess the accuracy of the extinction risk model in predicting regional extinction we compared the curves for $P_{\text{reg}}(t)$ estimated from the fit of Eqn. 1 to the data to the expected value of $P_{\text{reg}}(t)$ from the true model.

We explored the relationship between amount of data and model accuracy using datasets simulated from the extinction risk model with varying numbers of locations and time points. We varied the number of locations and time points from 2 to 128 by powers of 2 and simulated data for four different parameter combinations of low and high extinction risk and detection frequency. We chose low and high values of extinction risk by setting values of $\omega$ such that the cumulative probability of regional extinction was 0.95 in 100 time points and 1000 time points respectively. We adjusted detection frequency by simulating data according to low (0.3) and high (0.8) values of $\phi$. For each combination of parameter values, number of locations, and number of time points we simulated 50 datasets and fitted the extinction risk model to those datasets. We then calculated the average error of the estimated $\omega$ compared with the true value used in the simulations and the mean width of the 95% credible intervals for all 50 replicates.

**Model testing: Microcosm data**

To assess the model’s ability to capture known extinction dynamics in biological data, we applied the model to data from a protist microcosm experiment. This experiment examined the interaction of biotic stress (competition) and abiotic stress (increased temperature) in extinction
events for the protist *Loxocephallus sp.* (Clements et al. 2013). Experimental treatments consisted of two levels of competition (one or two competitors present) and two temperatures: 15°C (normal) and 20°C (stressful). The experiment used three other protist species to provide the biotic stress: *Paramecium caudatum*, *Colpidium striatum*, and *Blepharisma japonicum*. These species were combined with *Loxocephallus* to form three communities of two species and three communities of three species. Each of these six community types were replicated five times at each temperature. Individual microcosms started with 100 individuals of each species assigned to it and were sampled three times a week for 163 days.

To test the model, we divided these data into test and validation datasets. The test dataset consisted of a single community composition at each level of competition: (1) *Loxocephallus* with *P. caudatum* and (2) *Loxocephallus* with *P. caudatum* and *B. japonicum*. We fitted the model to data on *Loxocephallus* from both temperature treatments for the test dataset. As is expected from microcosm data, there was little to no observation error in occupancy state and we therefore introduced artificial observation error to the test data by modeling each recorded occurrence as a Bernoulli random variable with 0.6 probability of a successful observation. To account for fixed effects of treatment as well as random effects associated with individual microcosms, we modeled \( \omega \) as logit linear:

\[
\text{logit}(\omega) = \beta_c + \varepsilon_m, \quad (\text{Eqn 4})
\]

where \( \beta_c \) is the fixed effect associated with treatment \( c \) (temperature and competition levels) and \( \varepsilon_m \) is the random effect associated with each individual microcosm such that \( \varepsilon \sim \text{Normal}(0, \sigma^2) \). \( \sigma \) was assigned an uninformative prior with uniform distribution from 0 to 100 (Gelman 2006).

From the model fits, we calculated the posterior distribution for the cumulative probability of regional extinction through time (\( P_{reg}(t) \)) for each treatment. We then compared
the model’s predictions for *Loxocephallus* extinction dynamics with the validation datasets consisting of *Loxocephallus* in the other two (*Loxocephallus* with *C. striatum*; *Loxocephallus* with *B. japonicum*) and three (*Loxocephallus* with *P. caudatum* and *C. striatum*; *Loxocephallus* with *C. striatum* and *B. japonicum*) species communities at both temperatures. As each community type and temperature combination results in only a single value for the time of regional extinction (the time when *Loxocephallus* goes extinct in the last of the five replicates) this yields two time points to compare against each model generated posterior for $P_{reg}(t)$.

*Applying the model: Habitat fragmentation experiment*

To demonstrate the utility of the model to field data, we used data from the Wog Wog habitat fragmentation experiment in southeastern Australia (Margules 1992; Davies et al. 2004). Briefly, the Wog Wog experiment consists of six experimental replicates, each consisting of small, medium, and large patches of native *Eucalyptus* forest. Four of the replicates are fragmented from the surrounding *Eucalyptus* forest by a plantation of *Pinus radiata* and the remaining two replicates are in continuous *Eucalyptus* forest and serve as controls. Each fragment contains eight sample sites with two permanent pitfall traps for surface-dwelling invertebrates. Habitat fragmentation occurred in 1987 and five years of data were collected following fragmentation. Sampling ended after this initial period but began again in 2010 to assess the long-term effects of habitat fragmentation.

As an example of assessing extinction risk in a situation of conservation concern, we analyze data for 18 beetle species known to occur only in fragments and not in the pine matrix. We used the first five years of data to estimate extinction risk into the future and compared these predictions to data from 2010 (23 years after fragmentation). We fitted a mixed effects model with all three parameters of the extinction risk model further modeled as combinations of fixed
effects (species identity, fragment size, and their interaction) and random effects (nested spatial heterogeneity among experimental replicates, patches, and sample sites) as follows:

\[
\text{logit}(\omega) = S_i^\omega + F_j^\omega + D_{ij}^\omega + \epsilon_r^\omega + \epsilon_p^\omega + \epsilon_s^\omega,
\]

\[
\text{logit}(\phi) = S_i^\phi + F_j^\phi + D_{ij}^\phi + \epsilon_r^\phi + \epsilon_p^\phi + \epsilon_s^\phi,
\]

\[
\text{logit}(\nu) = S_i^\nu + \epsilon_r^\nu + \epsilon_p^\nu + \epsilon_s^\nu,
\]

(Eqn 5)

In the above equations, \(S_i\) and \(F_j\) denote the mean effects of species \(i\) and fragment size \(j\) respectively. \(D_{ij}\) (the interaction term) is the deviation of species \(i\) in fragment size \(j\) from the mean, \(S_i + F_j\). Superscripts indicate the biological parameter (\(\omega, \phi, \nu\)) that is being modelled. The \(\epsilon\) terms are the spatial random effects for each parameter associated with each replicate \(r\), patch \(p\), and site \(s\). The spatial random effects were distributed normally with mean 0 and standard deviations \(\sigma_r, \sigma_p, \text{ and } \sigma_s\) respectively. The standard deviations were given a half-Cauchy prior (Gelman 2006) and the fixed effects were given uninformative normal priors with means of 0 and standard deviations of 100. We compared the extinction risk estimated by the model to the results of the most recent data collection, beginning in 2010.

We used R version 3.2.1 (R Core Team 2015) to simulate data and all models were fitted using the rstan package (Stan Development Team 2014) with post-processing in R.

Results

Simulated data

When accounting for missing data (Eqn. 2), the extinction risk model retained high accuracy with increasing numbers of missing data points (Fig. 2.1a). The model error was relatively small and estimates remained unbiased up to approximately 40 missing data points (20% of the dataset). With between 40 and 60 missing data points out of 200 (i.e. 20% - 30% of the dataset), the model underestimated the true value of \(\omega\) slightly on average, but the
interquartile range of model estimates still included the true value. Once more than 60 data points were missing, the model estimates became notably more biased. Precision was similarly robust to missing data. The mean width of the 95% credible intervals remained constant over the full range of missing data examined (up to 50% missing). However, estimated interval widths
Figure 2.1: Model accuracy and precision for simulated datasets with increasing numbers of missing data points. (a) Accuracy measured as the difference between the model estimate (mean of the posterior distribution) of $\omega$ and the true value used in the simulations. Negative values indicate underestimates by the model. (b) Precision measured as width of the 95% credible interval for $\omega$. For all simulations, datasets were simulated with 20 locations and 10 time points for each location. Parameters for the simulated datasets were as follows: $\omega = 0.1$, $\phi = 0.6$, and $\nu = 0.75$. Data were removed randomly with no bias among locations or through time. Each point is the average of 100 replicates of fitting the model to a simulated dataset. Bars are the interquartile ranges across the 100 replicates.
became more variable once 60 or more data points were missing, the same point at which model estimates became notably biased (Fig. 2.1).

The model was also robust to colonization dynamics, accurately capturing the order of magnitude time frames in which the species was at risk of regional extinction (Fig. 2.2). For lower probabilities of regional extinction, model estimates were biased downward compared to the true values (i.e. extinction risk was underestimated), which is a known issue for these types of CJS models (Kendall et al. 1997).

**Figure 2.2:** The cumulative probability of regional extinction, $P_{reg}(t)$, as estimated by the extinction risk model (which lacks explicit colonization), compared to three true stochastic models, each with different colonization probability ($P_{col}$). The dashed lines are the medians and the dotted lines the means from replicated fits of the extinction risk model to data simulated from the true models (50 replicates). The shaded regions are the interquartile ranges across the 50 replicates. The solid lines are the expected value for each true stochastic model, calculated as the proportion of simulated metapopulations to have gone extinct at each time point (10,000 realizations). The three curves represent three colonization probabilities as labeled on the figure. Aside from $P_{col}$ other parameters for all three true models were: $\omega = 0.35$, $\phi = 0.6$, and $\nu = 0.75$, 20 locations, 10 time points.
Results for the effect of dataset size on model accuracy and precision show five important patterns. First and unsurprisingly, more data and higher detection frequencies gave better model
Figure 2.3: Model error as a function of the size of a dataset. The number of sites and time points for a dataset were varied from 2 to 128 by powers of 2 and data were simulated for four different parameter combinations: low and high detection frequency \((f)\), crossed with low and high extinction risk \((\omega)\). For each combination of detection frequency and extinction risk, number of sites, and number of time points, 50 datasets were simulated and the model fitted to the simulated data. The error between the expected value for \(\omega\) and the true value \((\omega_{\text{est}} - \omega_{\text{true}})\) was then calculated and averaged over the 50 simulations. Values of \(\omega\) and \(f\) were chosen as described in the text to generate low and high extinction risks and detection frequencies respectively. \(\nu\) was 0.75 for all simulations.
predictions both in terms of accuracy ($\omega_{\text{est}} - \omega_{\text{true}}$; Fig. 2.3) and precision (width of the credible interval; Fig. 2.4). Second, a high level of accuracy was achieved with only a small number of locations and time points (e.g. between four and eight locations and time points), after which accuracy was only marginally improved by adding sites or time points (Fig. 2.3). Third, for all
combinations of detection frequency and extinction risk, the smallest datasets led to consistent overestimation of extinction risk by the model, indicating a directional bias in model estimates for low amounts of data. Fourth, there was a similar pattern for precision, although slightly more data were needed for high precision (e.g. eight locations and eight time points; Fig. 2.4). Finally, extinction risk affected model accuracy, but was modulated by detection frequency. When extinction risk was high with low detection frequency, model estimates were more accurate compared to simulations with low extinction risk for the same amount of data (Fig. 2.3).

**Microcosm data**

The posterior distributions for the cumulative probability of regional extinction in the protist microcosm system demonstrated the expected relationship between biotic and abiotic stress and extinction dynamics. Increasing the number of competitors and increasing the environmental stress both led to higher estimated cumulative probabilities of regional extinction over the same time period (Fig. 2.5). Furthermore, the fitted models suggested that environmental stress played a more important role in this system as $P_{reg}(t)$ was more affected by increased
Figure 2.5: Estimated probability of regional extinction, $P_{reg}(t)$, for *Loxocephallus sp.* in protist microcosms compared to experimental extinction times. The solid line is the mean of the posterior distribution for the cumulative probability of regional extinction through time. Shading shows the 95% credible interval. The vertical, black arrows show the observed times of extinction for *Loxocephallus* in the two validation datasets. The columns indicate the number of competitors of *Loxocephallus* and the rows indicate the temperature treatments. For each treatment combination the model was fitted to data from five experimental replicates (“locations”). The observed regional extinctions also represent regional extinction from a set of five replicates.
temperature than by increasing the number of competitors (Fig. 2.5). The credible intervals were relatively wide; however this is to be expected based on the previous analysis relating the amount of data to model precision. This experiment had only five replicates ("locations" for the purposes of the model) per treatment. This number of locations, while sufficient for relatively high accuracy (Fig. 2.3), gives model estimates with relatively wide credible intervals, even from systems with high detection frequency (Fig. 2.4). Despite low precision, the cumulative probabilities of regional extinction predicted by the model were consistent with the observed extinction times of *Loxocephallus* in the validation data (Fig. 2.5).

*Habitat fragmentation*
Using the model to investigate the role of fragmentation in extinction of beetle species in the Wog Wog experiment yielded several important results. First, the posterior distributions for $F^0_{w}$ were essentially identical among the different sized fragments and the control treatment. Additionally, these posteriors all corresponded to extremely low probabilities of extinction, meaning that the fragmentation treatment did not increase extinction risk for these beetles (Fig. 2.6). This was confirmed by the posterior distributions of $P_{\text{reg}}(t)$ for each of the 18 beetle species, which display essentially zero probability of regional extinction for any of the species over at least the next 100 years. The most recent data collection at Wog Wog confirmed this prediction as all 18 species were still present 23 years after the initial fragmentation event. Second, the estimated random effects showed that the scale of spatial heterogeneity in this system could be

**Figure 2.6:** Posterior distributions for the $F^0_{w}$ terms in the Wog Wog experiment. The posteriors are shown on the logit scale, meaning the lower the value, the closer it is to a probability of 0. For context, a value on the logit scale of -200 (near the mean of all four posterior distributions) corresponds to a probability of $1.38 \times 10^{-87}$.
important for beetle movement, observation probability, or both. The posterior distributions of the standard deviations for the random effects reveal potentially important spatial patterns in the model parameters. The $\phi$ parameter (related to movement between sites and/or the probability of observation; see Section 2.1) is most affected by random variation at the site scale, as indicated by the larger standard deviation for these effects (Fig. 2.7). However, $\nu$ is dominated by random variation at the scale of experimental replicates (Fig. 2.7), for which the standard deviation is so large that these effects likely wash out the fixed effects of species on this parameter. Thus, the
processes driving variation in these two parameters seem to be operating at different spatial scales.

**Discussion**

Anthropogenic factors continue to cause declines and extinctions among well-known species for which we have ample data (Dirzo et al. 2014). Recent work has focused on how to assess extinction risk for those species for which we have little to no data (Holmes 2001; Morais

**Figure 2.7:** Posterior distributions for the standard deviations of the random effects on each parameter and each spatial scale for the Wog Wog data. The standard deviations associated with each parameter are plotted together and the parameter corresponding to each graph is indicated at the top. The standard deviations for each spatial scale are indicated by different line types (shown in the legend). The posterior for the standard deviation of replicate level random effects is shown separately in the inset of the right panel due its drastically different scale. The standard deviations were used to model hierarchical random effects as described in the text.
This work targeted at data deficient species has advanced our ability to assess extinction risk and therefore apply effective conservation strategies. The model presented here is another useful tool for species lacking the data necessary for traditional models of extinction risk. It only requires presence/absence data, bypassing the potential cost and difficulty of obtaining abundance or life history data. The model can be used with very simple sampling designs, performs well with missing data and small datasets, and does not require additional information on dispersal behavior (Eaton et al. 2014) or multiple, unique covariates to the occupancy and observation process (Lele et al. 2012). The model could therefore minimize the survey costs associated with collecting new data (Bland et al. 2015b) or allow preliminary conclusions to be drawn from minimal historical sighting records. Finally, the model is easily extended to allow for more complex designs or to include explanatory variables.

The model remains accurate when faced with moderate amounts of missing data (Fig. 2.1) and is robust to temporary extinction and recolonization dynamics (Fig. 2.2), despite the known biases that can occur when CJS models are faced with temporary emigration. In the extinction risk model, temporary extinction at a patch is analogous to temporary emigration in a CJS model and is by definition a Markovian process, with the probability of temporary extinction depending on the previous state of a patch (i.e. an unoccupied patch has a different probability of temporary extinction compared to an occupied patch). Markovian emigration dynamics are known to bias CJS model estimates, and they have a similar effect on estimates of the extinction risk model, lowering estimates of extinction risk when temporary extinction and recolonization are prevalent in a system (Fig. 2.2). However, even for high levels of colonization, the model accurately captures the time scales for which the species is at moderate
to high risk of regional extinction ($\geq 0.4$), which accomplishes the goal of providing a first pass assessment of extinction risk and the relevant time frame.

These features of the model make it ideal for use with species for which we need information but have few resources for data collection. As the model accurately predicts the time frames over which a species will be at a moderate to high risk of regional extinction both with and without colonization dynamics, it can be used even when the importance of dispersal for a species is unknown. This, combined with the model’s ability to cope with missing data, means that it can be used effectively for species with poorly known life histories and imperfect datasets.

Furthermore, the model only requires a small amount of data to make sufficiently accurate predictions, depending on the detection frequency and extinction risk of a species (Fig. 2.3 & 2.4). While the model tends to overestimate $\omega$ with small datasets, this bias disappears with more data. As might be expected, a higher detection frequency will provide more accurate estimates and narrower credible intervals. Interestingly, for species that are difficult to detect (i.e. low detection frequency) the model gives more accurate estimates for species at greater risk of extinction. Thus, the model should perform more accurately for exactly those species that are of highest conservation concern. Considering the relative ease of collecting occupancy data compared to other quantities (e.g. abundance) (Mackenzie et al. 2006), the model allows for efficient sampling designs for species with no preexisting data. The model could be particularly useful for citizen science programs since occupancy data require less expertise and investment to collect, the model is robust to sporadic sampling, and potential observer biases could also be modeled (van Strien et al. 2013).

Beyond the results from computer simulations, we also demonstrate the utility of the model using biological data. Using the protist data, we confirmed that, while the credible
intervals estimated from data with relatively few locations were large, the expected values of the posterior distributions remained accurate. The model predicted cumulative probabilities of regional extinction were consistent with the validation data and the comparison of the curves among treatments demonstrated the expected relationship between biotic and abiotic stress and extinction risk. The analysis of the habitat fragmentation data demonstrated the model’s ability to correctly predict very low extinction probabilities, which is an important quality for a model to be used in conservation settings. The use of the logit transformation to model the parameters as mixtures of fixed and random effects can yield important biological insight into the dynamics of this system as well, as shown for the importance of spatial scale to variability in the $\phi$ and $v$ parameters. The $\phi$ parameter can encompass variability in movement between sites as it is related to the frequency of extinction/recolonization dynamics as well as processes related to the probability of observing the beetle species at each site. The large posterior for the standard deviation of random effects at the site scale for $\phi$ (Fig. 2.7) indicates that heterogeneity among individual sites could play an important role in movement dynamics, observation probability, or both in this system. Further, the large standard deviation estimated for random effects at the replicate scale for the $v$ parameter suggests the processes driving species occurrence in Wog Wog are most important at this larger spatial scale. For example, the geographic arrangement of the replicates or landscape hydrology could play a large role in determining the presence or absence of a beetle species in experimental replicates at Wog Wog. The model specification does not allow us to directly determine the underlying causes of the variation in these parameters, but it does demonstrate the importance of considering processes at multiple spatial scales in future work in this system.
The logistically efficient design allowed by the extinction risk model lessens the burdens on estimating extinction risk for poorly known species. It does not require knowledge of behavior, community interactions, or environmental covariates, although it can incorporate them. The model is appropriate for simple, easily collected data from sighting records, citizen science projects, or planned surveys and lessens the need to estimate higher-level information beyond a binary value for presence or absence. Given its ability to incorporate observation error as well as the robustness of its estimates in the face of missing data (Fig. 2.1) and small datasets (Fig. 2.3 & 2.4), the model should provide a useful framework for evaluating poorly known species for extinction risk and therefore assisting in the determination of conservation priorities and strategies. Thus, the model presented here fills a gap in existing methods for assessing extinction risk by allowing for logistically efficient estimates of extinction risk for species with little data available, the results of which can be used to inform future distribution of conservation resources and effort.
CHAPTER 3

Rapid trait evolution drives increased speed and variance in experimental range expansions

Christopher Weiss-Lehman, Ruth A. Hufbauer, and Brett A. Melbourne

Adapted from

Abstract

Range expansions are central to two ecological issues reshaping patterns of global biodiversity: biological invasions and climate change. Traditional theory considers range expansion as the outcome of the demographic processes of birth, death, and dispersal, while ignoring the evolutionary implications of such processes. Recent research suggests evolution could also play a critical role in determining expansion speed but controlled experiments are lacking. Here we use flour beetles (*Tribolium castaneum*) to show experimentally that mean expansion speed and stochastic variation in speed are both increased by rapid evolution of traits at the expansion edge. We find that higher dispersal ability and lower intrinsic growth rates evolve at the expansion edge compared to spatially non-evolving controls. Furthermore, evolution of these traits is variable, leading to enhanced variance in speed among replicate population expansions. Our results demonstrate that evolutionary processes must be considered alongside demographic ones to better understand and predict range expansions.
Introduction

Evolution is predicted to change the dynamics of range expansions through multiple processes. For example, the expanding population may encounter different biotic and abiotic conditions in the newly colonized habitat that impose novel selection pressures (Urban et al. 2007). Evolution in response to a novel environment is necessarily context dependent and, while important, difficult to generalize across range expansions. However, three other processes by which evolution can change the dynamics of range expansions are direct outcomes of the intrinsic spatial population structure formed during range expansion and should therefore be general to any range expansion. First, spatial sorting, the non-random aggregation of highly successful colonizers at the expansion edge, and subsequent assortative mating among them, can lead to increases in traits related to colonization (such as dispersal ability) (Travis & Dytham 2002; Shine et al. 2011; Perkins et al. 2013). Second, the selection imposed by population density varies across the expanding range, with individuals at the expansion edge typically experiencing lower density than individuals in the core of a species’ range (Melbourne & Hastings 2009; Burton et al. 2010). Traits conferring fitness at high densities, generally referred to as competitive ability, are expected to evolve upwards in the core of the range but not at the edge where densities are low (Burton et al. 2010; Phillips et al. 2010a). In contrast, edge populations are expected to evolve higher fecundity at the expense of competitive ability (Burton et al. 2010), yielding higher intrinsic growth rates when compared to the core. Here, intrinsic growth rate refers to the population growth rate achievable in a given environment unhindered by any negative influence of population density and is typically determined by fitness at low density when competition is absent. A third way that spatial evolution can change the dynamics of range expansions is that allele frequencies can be influenced by the small population sizes and repeated
founder events associated with the edge of the expanding population, resulting in the recently discovered phenomenon of gene surfing (Edmonds et al. 2004; Klopfstein et al. 2006; Hallatschek & Hersen 2007; Travis et al. 2007; Hallatschek & Nelson 2010; Roques et al. 2012; Peischl et al. 2013; Peischl et al. 2015). In gene surfing, serial founding events at the expansion edge (Excoffier et al. 2009) can allow deleterious alleles to increase in frequency and travel with the expansion edge, while advantageous alleles can be lost due to chance events (Travis et al. 2007; Peischl et al. 2013). Increased frequency of deleterious alleles in edge populations is predicted to reduce mean fitness (Travis et al. 2007; Peischl et al. 2013) and thus decrease intrinsic growth rates at the edge, regardless of density. These three spatial evolutionary processes could act alone or together to change the speed of a range expansion over time.

Expansion speed is determined largely by growth rate and dispersal at the edge (Skellam 1951; Hastings et al. 2005), so the effect of evolution on expansion speed should depend on the balance of evolutionary processes that increase or decrease growth rate and/or dispersal at low density. Theory shows that evolved increases in intrinsic growth rate or dispersal ability at the edge should increase expansion speed (Burton et al. 2010; Perkins et al. 2013; Phillips 2015; Williams et al. 2016a). On the other hand, theory shows that gene surfing should decrease fitness at the expansion edge (Travis et al. 2007; Peischl et al. 2013), thus depressing intrinsic growth rates and slowing expansion (Peischl et al. 2015). Given the stochastic nature of evolutionary processes, the evolutionary outcome of any one realization of a range expansion will be partly randomly determined and is thus likely to be unique. Indeed, the stochastic and potentially opposing effects of different spatial evolutionary processes on expansion speed offer a possible explanation for previous experimental work showing that demographic factors alone are not sufficient to explain observed variability among range expansions (Melbourne & Hastings 2009).
A theoretical model combining dispersal evolution with gene surfing predicts that evolution will increase stochasticity in expansion speed among range expansions relative to models of expansion that do not include evolutionary processes (Phillips 2015). However, evolution of increased dispersal is still predicted to dominate, so that evolution is predicted to lead overall to an increase in mean expansion speed along with increased variance (Phillips 2015).

The three spatial evolutionary processes outlined above (spatial sorting, selection by density, and gene surfing) have varying degrees of empirical support. For example, evolved increases in dispersal ability have been observed in natural and experimental range expansions (Phillips et al. 2010b; Berthouly-Salazar et al. 2012; Lombaert et al. 2014; Fronhofer & Altermatt 2015; Williams et al. 2016b) as have signatures of gene surfing (Hallatschek & Hersen 2007; Gracia et al. 2013; Henn et al. 2016). On the other hand, investigation of spatial evolution of competitive ability via selection across the density gradient of a range expansion has been primarily theoretical (Burton et al. 2010; Phillips et al. 2010a). Implicit in each of the three spatial evolutionary processes is the idea that the dynamic development of spatial structure across the species range, from core to edge, drives evolution in range expansions. This has yet to be tested empirically, and controlled experiments are needed to determine the effects of spatial structure on evolutionary dynamics of range expansions. Here, we evaluate the role played by spatial structure in driving evolution in experimental range expansions using laboratory microcosms of the red flour beetle, Tribolium castaneum.

We founded replicate experimental populations from the same large, well-mixed source population, and allowed them to expand from the founding point. To isolate the effects of evolution due to spatial structure on expansion speed and variance, we compared expanding populations subjected to two experimental treatments. In one treatment populations were allowed
to develop natural spatial genetic structure over time (that is, spatial evolution was allowed) and in the other treatment we prevented the development of spatial genetic structure (that is, spatial evolution was prevented) by randomly shuffling the location of individuals without disrupting population density or demographic processes (Methods). After eight generations of range expansion, we compared the effect of the treatments on two key traits that contribute to expansion speed, dispersal ability and intrinsic growth rate, by assaying G1 (first generation) descendants of beetles from structured and shuffled populations (Methods).

The treatment allowing spatial evolution has a significantly higher mean spread rate compared to the shuffled treatment as well as heightened variability in spread rates. The trait experiments suggest that these patterns are driven by evolution of dispersal and growth rate in beetles at the expansion edge of structured populations. Our results demonstrate the importance of spatial evolutionary changes in determining the dynamics of range expansion over short timescales.

Methods

Experimental system

We used Tribolium castaneum, the red flour beetle, in experimental microcosms to test the effects of spatial evolutionary mechanisms on range expansions. Tribolium castaneum populations were kept in 4cm by 4cm by 6cm acrylic containers with 20 g of a standard medium (95% wheat flour and 5% brewer’s yeast). We will henceforth refer to a container complete with medium as a patch. Patches can contain single, isolated populations of T. castaneum (as used in growth assays) or be connected in linear arrays via 2 mm holes drilled in the sides to form landscapes of patches (Melbourne & Hastings 2009) linked by dispersal (as used in the range expansion experiment and dispersal assays). The life cycle of T. castaneum was constrained to
mimic that of a seasonally breeding organism with non-overlapping generations and a discrete dispersal phase (Melbourne & Hastings 2008), a life history found commonly among plants and animals. Adult beetles were placed in individual patches with fresh standard medium for 24 hours to reproduce and start the next generation. At the end of the reproduction phase, adults were removed from the patches and the eggs were left to mature into adults over a 34-day period. In landscapes linked by dispersal via holes drilled in the sides of the patches, thin plastic sheets were placed between patches for 34 days to prevent dispersal. These sheets were removed for 24 hours on day 34 to allow for a discrete dispersal phase before the populations were censused and adult beetles were placed in fresh medium to begin the next generation. While egg deposition was constricted to only this 24-hour period, mating could occur during that 24-hour period and at any point prior, once beetles reached maturity (usually a few days before the dispersal period). Therefore, a single beetle dispersing to an unoccupied patch could give rise to offspring in the next generation if it was a mated female. This experimental protocol yielded a 35-day generation time. Beetles were kept in incubators at 31°C with approximately 80% relative humidity. Three incubators were used and landscapes or single isolated patches were randomized among and within incubators once each week to prevent systematic effects of incubation conditions.

Range expansion experiment

We tested the effects of spatial evolutionary processes on range expansions with 60 experimental landscapes of *T. castaneum* divided between two treatments, which we will call ‘structured’ and ‘shuffled’. In the structured treatment, to begin each generation we returned beetles to the same patch in which we recorded them, thus allowing the formation of spatial genetic structure due to evolutionary processes. In the shuffled treatment, we prevented spatial evolution by randomizing the spatial location of beetles within landscapes at the start of each
generation. To randomize beetles, we first recorded population densities in each patch after dispersal, then all beetles from a landscape were mixed together and redistributed throughout the landscape according to the recorded density of each patch. This procedure disrupted the formation of spatial genetic structure by decoupling an individual’s genetics from its location, but maintained the demographic processes of the range expansion (e.g. density dependent growth or dispersal). Landscapes from each treatment were additionally divided randomly among three temporal blocks.

Following common procedures from other experimental evolution studies (Huey et al. 1991; Gilchrist et al. 1997; Yoshida et al. 2003; Collins et al. 2004; Duffy & Sivars-Becker 2007; Palkovacs et al. 2008; Burke et al. 2010; Tompkins et al. 2011; Egan et al. 2015; Huang et al. 2015), populations were initially founded from the same large, well-mixed population and then randomly assigned to an experimental treatment (structured or shuffled). Landscapes were founded by placing 20 randomly selected adult beetles into the first patch of the landscape for 24 hours to reproduce. Expansions proceeded for eight generations. Both treatments began with 30 replicate landscapes but several replicates were lost due to laboratory mishaps, yielding 28 replicates of the structured treatment, and 29 replicates of the shuffled treatment.

Testing for spatial trait evolution

At the end of the eighth generation of expansion, we tested for evolved differences in dispersal ability and growth rate. For each structured landscape, we drew 20 beetles at random from the range core (i.e. the first patch of the landscape) and took the 20 furthest forward at the range edge (drawing randomly where necessary to make up to 20). For each shuffled landscape, we drew two random samples of 20 beetles each (since random beetles began the generation at the core and edge of shuffled landscapes). Each sample of 20 beetles was placed in an individual
patch to reproduce for 24 hours. We used the first generation offspring (i.e. G1 generation) to perform assays of dispersal ability and growth rate within a common environment to determine if there were evolved differences in these traits (Savolainen et al. 2013). Differences in trait values in a common environment such as this are considered to be due to evolution in those traits, since populations in different experimental treatments were founded from the same stock population (Savolainen et al. 2013).

To conduct growth rate assays, we placed offspring from core, edge, and shuffled populations in individual patches at densities of 5, 10, 15, and 20 to reproduce for 24 hours. This allowed us to compare growth rates among G1 descendants from the core and edge of structured populations and G1 descendants from shuffled populations and to examine how growth rate varied with density among core, edge, and shuffled populations.

To conduct dispersal assays, we placed offspring from core, edge, and shuffled populations at densities of 10 and 40 in the first patch of freshly prepared landscapes. Beetles remained in the first patch of these landscapes for 48 hours to equilibrate movement behavior within patches before the plastic dispersal barriers were removed for a 24-hour dispersal period. After dispersal finished, the number of beetles that dispersed away from the patch of origin was recorded for each replicate. All of the patches and landscapes for all growth and dispersal assays were kept in the same incubator conditions as previously described.

Statistical analyses

For each landscape at each generation, we recorded the furthest distance spread, defined as the furthest patch to have a single beetle present (as a single beetle could produce offspring if it was a mated female). To analyze expansion speed, we used a linear mixed effects model with distance spread as the response variable. Fixed effects were generation (a continuous variable
allowing estimation of the slope of distance on time, which is the mean expansion speed), treatment (a categorical variable with two levels: structured and shuffled), and the interaction of treatment and generation (which represents the difference in mean expansion speed between structured and shuffled populations). To account for the non-independence of repeated data from the same landscapes, we included a random effects term to allow the speed to vary among replicate landscapes (i.e. a random slopes model) (Clark 2007). These random slopes were nested within a second random term, block, to account for the three temporal blocks of the experimental design. To test for the significance of treatment on expansion speed, we used a parametric bootstrap to calculate the p value for the interaction of treatment and generation (Hufbauer et al. 2015). For this and all other bootstrapped p values and confidence intervals described below, we performed 10,000 simulations.

For each landscape at each generation, we calculated the standard deviation and coefficient of variation of distance spread. We used multiple regression to model the linear change of the standard deviation and coefficient of variation through time and to test for a significant effect of treatment or interaction of treatment by generation (likelihood ratio test). This model contained no random effects since each treatment had one summary data point (standard deviation or the coefficient of variation) per generation. Similarly, we used multiple regression to model the relationship of the standard deviation to the mean distance spread and to test for a significant interaction of treatment with the mean (likelihood ratio test). Model predictions for the standard deviation were transformed to variances for display and comparison with experimental results.

Data from the dispersal assay (number of beetles dispersed away from the origin patch out of the number starting in the origin patch) were analyzed using a generalized linear mixed
effects model with a binomial distribution and a logit link function for the probability that a beetle disperses away from the origin patch. Fixed effects included population density starting in the origin patch, location of population origin (core, edge, or shuffled), and the interaction of density and location. Random intercepts were modeled for each landscape nested within temporal block. To test for a significant effect of treatment or density on dispersal probability, we used a parametric bootstrap to calculate the p value for location and the interaction between density and location.

Data from the growth rate assay were analyzed using a linearized Ricker model previously developed for this system (Hufbauer et al. 2015). As described in Hufbauer et al. (2015), the population dynamics of T. castaneum in this system can be modeled with a generalized Ricker model allowing for a potential Allee effect: \( N_{t+1} = RN_t \theta e^{-\alpha N_t} \), where \( N_t \) is population density (number of beetles per patch) in generation \( t \), \( R \) is the finite growth rate, \( \alpha \) is the strength of negative density dependence (egg cannibalism by adult beetles), and \( \theta \) is the strength of the Allee effect with a value of 1 corresponding to no Allee effect. This model can be linearized via a log transformation to yield:

\[
\ln \left( \frac{N_{t+1}}{N_t} \right) = a + bN_t + c\ln(N_t),
\]  
(Eqn 1)

where \( a = \ln(R) \), \( b = -\alpha \), and \( c = \theta - 1 \). We used a mixed effects implementation of this model to fit the data from the growth rate assay. Fixed effects for this model included population density, the natural log of population density, location of population origin, the interaction between location and density (to allow \( \alpha \) to vary among locations), and the interaction between location and log density (to allow \( \theta \) to vary among locations). Random intercepts for landscapes nested within temporal blocks were included as with the dispersal analysis. We used a parametric
bootstrap to test for the significance of location on the intercept of the model \( \ln(R) \) as well as the interaction between location and density to test for evolved differences in \( \alpha \) between offspring from core and edge populations.

To assess correlations between evolved differences in dispersal ability and intrinsic growth rates, we used data from the dispersal and growth rate assays from descendants of the edge populations of structured landscapes. Using landscapes with both a dispersal and growth rate assay, we plotted the proportion of beetles that dispersed away from the origin patch under low-density conditions (from the dispersal assays) against the population growth rate averaged across assay densities (from the growth rate assays). We then calculated Pearson’s \( r \) to assess the correlation between the two.

All statistical analyses were performed using R (version 3.2.3; R Core Team 2015). Linear mixed effects models and generalized linear mixed effects models were fitted using the `lmer` and `glmer` functions in the package `lme4` (version 1.1.11; Bates et al. 2015). Bootstraps were performed using the packages `pbkrtest` (version 0.4.6; Halekoh & Hojsgaard 2014) and `boot` (version 1.3.18; Canty & Ripley 2015). For all analyses, we checked model assumptions using plots of residuals, quantiles, influence, and leverage where appropriate.

*Expected maximum change in expansion speed*

To estimate the expected change in expansion speed between the structured and shuffled landscapes, we used the well-known approximation for expansion speed, \( 2\sqrt{rD} \), in which \( r \) is the intrinsic growth rate and \( D \) is the diffusion coefficient (Hastings et al. 2015). Using data from the phenotypic assays, we calculated the changes in \( r \) and \( D \) between edge populations and shuffled landscapes to determine the expected change in speed. We calculated intrinsic growth rates as \( r \)
\( = \ln(R), \) where \( R \) was estimated from the data as described above. To calculate the diffusion coefficients, we assumed exponentially distributed waiting times for individual beetles to leave a patch. This assumption means that the probability \( p \) of an individual leaving a patch over a time period \( T \) is given by

\[
p = \int_0^T D e^{-Dt} dt. \tag{Eqn 2}
\]

Solving the above integral for \( D \) and setting \( T = 1 \) (to calculate the diffusion coefficient over a single dispersal period) yields \( D = -\ln(1 - p) \), where \( p \) was estimated from the data as described above. The proportional change in speed is then given by \( \sqrt{\Delta_r \Delta_D} \), where \( \Delta_r \) and \( \Delta_D \) are the proportional changes in \( r \) and \( D \) between edge populations and shuffled landscapes.
Results

**Figure 3.1:** Experimental results of range expansions with structured \((n = 28)\) and shuffled \((n = 29)\) populations initiated from a single well-mixed source population. (a) and (b) Distances spread in each treatment. The lines show data for individual replicates while shaded regions show the observed range of distance spread in the other treatment for reference. (c) Mean distance spread through time for each treatment (solid lines) and sample estimated 95% confidence intervals (dashed lines). (d) Model-estimated variances for each treatment with 95% confidence intervals. The observed variances for each treatment are shown as points. In all panels the structured treatment is shown in red and the shuffled treatment in blue. Spatial evolution resulted in a higher mean expansion speed (parametric bootstrap, treatment by generation interaction: \(p = 0.0137\)) and higher variance in expansion speeds (likelihood ratio test, treatment by generation interaction: \(p = 1.93 \times 10^{-5}\)).

**Speed and variance of range expansion**

Spatial evolution led to a 6% (95% CI 1.3 - 11.6%) higher mean expansion speed compared to populations in which individuals (and thus alleles) were shuffled each generation.
(Fig. 3.1c; parametric bootstrap, treatment by generation interaction: $p = 0.0137$). Importantly, variance in distance spread among replicate populations was increased by spatial evolutionary processes (Fig. 3.1d; likelihood ratio test, treatment by generation interaction: $p = 1.93 \times 10^{-5}$), leading to almost doubled variance in the distance spread of structured compared to shuffled landscapes by the eighth generation. Further, to test whether the increased variance in speed was only a result of the increased mean, we analyzed the coefficient of variation (CV) and the variance to mean relationship and confirmed that spatial evolution resulted in higher variation in expansion speeds independent of the differences in mean expansion speed (Fig. 3.2; CV, likelihood ratio test, treatment effect: $p = 0.004$; variance to mean relationship, treatment-mean interaction, $p = 0.025$).

By founding each landscape with randomly selected beetles from the same well-mixed source population, we expected differences between treatments to be negligible at first, before the shuffle treatment began or had an effect, and to develop over time. This expectation was confirmed, as distance spread (Fig. 3.1c; Poisson generalized linear mixed model, generation 1, $p = 0.41$, generation 2, $p = 0.72$) and variance (Fig. 3.1d; F test, generation 1, $p = 0.41$, generation 2, $p = 0.11$) did not differ significantly by treatment in the first two generations. Two of the structured landscapes, however, spread further than any shuffled landscapes in the first generation (Fig. 3.1a). To exclude the possibility of a fortuitous random draw, we redid the
analyses excluding these two landscapes and the results showed the same patterns (expansion speed, treatment by generation interaction: \( p = 0.0138 \); variance, treatment by generation interaction: \( p = 0.0003 \); CV, treatment effect after G1: \( p = 0.002 \); variance to mean relationship, treatment-mean interaction: \( p = 0.030 \)).

**Figure 3.2:** The coefficient of variation (CV) in distance spread through time is shown in (a) and the relationship between the variance in distance spread and the mean distance spread is shown in (b). Solid and dashed lines are means and 95% confidence intervals, respectively, from the regression models described in the methods. Open circles are the observed CV and variance calculated from the replicate populations of the structured (n = 28) and shuffled (n = 29) treatments. The structured treatment is shown in red and the shuffled treatment is shown in blue. Structured populations had a higher coefficient of variation than shuffled populations (likelihood ratio test, treatment effect: \( P = 0.004 \)) and a significantly different relationship between the variance and the mean compared to shuffled populations (likelihood ratio test, treatment by mean interaction: \( P = 0.025 \)).

**Trait evolution**

After eight generations of range expansion, G1 descendants from the expansion edge of structured populations had higher dispersal under low-density conditions compared to G1.
descendants from the range core or shuffled populations (Fig. 3.3; parametric bootstrap, interaction of density and location: \( p = 0.0008 \)). Dispersal was 92\% (CI 42 – 172\%) higher at the edge than the core and 35\% (CI -0.3 – 83\%) higher at the edge than in shuffled populations. Under high-density conditions, the dispersal tendencies of G1 descendants from all populations were similar (Fig. 3.3).

**Figure 3.3:** Open circles are the observed proportion of G1 beetles dispersing out of the natal patch in each replicate of each location and density. Replicates for G1 descendants from the core (n = 28) and edge (n = 28) of structured landscapes and G1 descendants from shuffled landscapes (n = 58) are shown. Density treatments were low = 10 and high = 40. Filled circles and error bars are means and 95\% confidence intervals from the generalized linear mixed effects model described in the methods. Core populations are shown in green, edge populations in fuchsia, and shuffled populations in blue. Spatial evolution led to a higher proportion of beetles dispersing from edge populations at low density (parametric bootstrap, interaction of density and location: \( p = 0.0008 \)).
Intrinsic growth rates were lower among G1 descendants from the expansion edge of structured populations when compared to G1 descendants from range cores and shuffled landscapes (Fig. 3.4; parametric bootstrap, effect of location: p = 0.0012). Intrinsic growth rate was 9.9% (CI 4.7 – 14.9%) lower at the edge compared to shuffled populations, while core and shuffled populations were similar. The degree to which growth rate was lower for edge
populations was similar across densities and not significantly related to density (Fig. 3.5; parametric bootstrap, interaction of density and location: p = 0.4285).

Discussion
By replicating experimental range expansions and comparing them to non-structured controls, we can evaluate rigorously the role of spatial evolutionary processes in range expansions. The congruence of our results with theoretical studies (Perkins et al. 2013; Phillips 2015; Shaw & Kokko 2015) confirms that spatial evolutionary processes can explain previously reported patterns of increased mean and variance in expansion speeds (Phillips et al. 2006; Melbourne & Hastings 2009). Over eight generations, we found a small increase in mean speed and a larger increase in variance in spatially evolving populations compared to populations that were not evolving spatial structure. Furthermore, we observed rapid differential evolution, within eight generations, of two key traits, dispersal ability and intrinsic growth rate, in edge versus core populations of spatially evolving populations. This rapid spatial evolution of traits likely explains the increased speed and variance of spatially evolving populations compared to controls.

Within structured populations, higher low-density dispersal rates and lower intrinsic growth rates evolved at the edge compared to the core and to shuffled populations, leading overall to a slight increase in mean expansion speed relative to shuffled populations. Higher low-density dispersal rates at the edge (where density is low) should lead to higher expansion speeds, whereas lower intrinsic growth rates at the edge should lead to a decrease in expansion speed because fewer colonists are produced. The realized expansion speed is thus a balance of these rates and the increase in mean expansion speed observed suggests that on average the positive effect of dispersal evolution outweighed the negative effect of growth rate evolution. An order of magnitude estimate of expansion speed based on our experimental measurements of dispersal and intrinsic growth rate can be calculated using \(2\sqrt{rD}\), where \(r\) is the instantaneous intrinsic growth rate and \(D\) the diffusion coefficient (Methods). This formula applies across a wide range of expansion models (Hastings et al. 2005), and applying it to our data gives a maximum
increase in speed of 13% for structured compared to shuffled populations, well in line with the observed 6% increase. The observed increase in speed is expected to be lower than the maximum for two reasons. First, beetles from the higher density patches just behind the edge do contribute to spread but should not experience the same boost in dispersal as edge beetles, given the density dependence of the boost (Fig. 3.3). Second, the estimated maximum relies upon dispersal and intrinsic growth measured when they presumably differed the most, at the end of the experiment, and does not account for values of these traits diverging over time.

Our results support the theoretical prediction that stochasticity in the evolution of dispersal ability and intrinsic growth rate at the expanding edge combine to increase variation in expansion speed (Phillips 2015). Although on average individuals from the expanding edge displayed heightened dispersal ability and decreased growth rates compared to individuals from the core and shuffled populations, specific trait values were highly variable (Figs. 3.3 & 3.5) and uncorrelated with each other (Fig. 3.6). Thus, as predicted (Phillips 2015), stochastic combinations of evolving traits at the expansion edge, such as slower dispersers with low
intrinsic growth rates or faster dispersers with higher intrinsic growth rates, would increase variance in expansion speed.

Patterns of traits after eight generations provide some clues to the evolutionary processes that might be dominant across the range expansions. For dispersal, since the low-density dispersal tendencies of G1 descendants in structured populations were within the observed

**Figure 3.6:** The observed proportion of beetles dispersing in low-density conditions plotted against the natural log of growth rate. Dispersal and growth rate data are from the assays of G1 descendants of edge populations in structured landscapes. Low density conditions for dispersal were defined as 10 adult beetles. Growth rate was defined as the ratio of population density at time \( t + 1 \) to population density at time \( t \). Log(growth rate) was averaged across density levels in the growth assay to provide a relative measure of intrinsic growth rate. The correlation between these two traits was weak and not significantly different from zero (Pearson’s correlation \( n = 28 \): \( r = 0.0785, p = 0.6872 \)).
variation of shuffled populations (Fig. 3.3), the difference between edge and core in structured populations was likely due to selection on standing genetic variation of high dispersing genotypes at the edge and low dispersing genotypes in the core, rather than evolution of new genotypes via mutation, consistent with the theory of spatial sorting (Shine et al. 2011). For intrinsic growth rate, spatial sorting is expected to lead to higher growth rate at the expanding edge because producing more offspring increases colonization ability (Phillips 2009). However, we observed lower intrinsic growth rates in edge populations, suggesting that spatial sorting is not the dominant process driving the evolution of growth rates across the landscapes. Similarly, if selection by density dominated, we would expect evidence of an evolved tradeoff across densities between the expansion edge and range core, which we did not find. One possible explanation is that individuals from the leading edge evolved reduced intrinsic growth rates due to a trade-off between dispersal and fecundity, such that highly dispersive individuals at the edge evolved lower fecundity due to greater energy expenditures on movement or metabolism (Burton et al. 2010; Hudson et al. 2015). However, the correlation between dispersal ability and intrinsic growth rate in G1 descendants from edge individuals was low, providing little evidence for a trade-off (Fig. 3.6; Pearson’s r = 0.0785). In contrast, evolution of reduced intrinsic growth rates in individuals from the edge is consistent with predictions of gene surfing in range expansions (Travis et al. 2007; Peischl et al. 2013; Peischl et al. 2015), where the accumulation of deleterious alleles can be expected to reduce intrinsic growth rates, regardless of density. It is feasible that gene surfing of deleterious alleles could have developed quickly, as this process is predicted to be particularly potent in systems with relatively low carrying capacities (about 250 individuals per patch in our system) and high genetic load such as ours (Szűcs et al. 2014) due to the greater influence of genetic drift, founder effects, and inbreeding at the edge in such systems
(Peischl et al. 2013). Future work should focus on the genetic basis of trait evolution to test these hypotheses.

By removing the possibility for spatial evolution in the shuffled landscapes, we obtained an estimate of its relative effects. The differences in traits among core, edge, and shuffled populations (each initiated randomly from the same well-mixed stock populations and thus starting with highly similar trait distributions) measured under common controlled conditions demonstrate that rapid evolution occurred and in what relative direction. Namely, edges evolved higher dispersal and lower intrinsic growth rate compared to shuffled and core. Because we cannot determine the direction of evolution absolutely, a possibility worth considering is that shuffled populations and populations in the core of the structured landscapes both evolved reduced dispersal ability and increased intrinsic growth rate compared to edge populations rather than, or in addition to, edge populations evolving increased dispersal ability or reduced growth rate. Two observations suggest that this direction of evolution is unlikely. First, the observed trait patterns are consistent with proposed mechanisms for evolution at the edge. For dispersal, the pattern of higher dispersal at the edge, lower dispersal in the core, and shuffled populations intermediate is consistent with spatial sorting. Similarly, the pattern of lower intrinsic growth rate at the edge and matching higher intrinsic growth rates in core and shuffled populations is consistent with gene surfing. Second, if evolution were primarily in the core and shuffled treatments, that would imply selection for lower dispersal and higher intrinsic growth rates across the landscape, and the inability of edge populations in structured landscapes to respond to that selection. Mechanisms for a lack of response to selection at the edge are less plausible and would, in any case, need to involve evolutionary processes such as genetic drift in the small edge populations or gene surfing constraining an adaptive response. Thus, the relative shifts in
dispersal and intrinsic growth rate are in accord with theories for evolutionary processes driving these crucial traits along the expanding edge, regardless of the absolute direction of trait evolution.

Rapid evolution of dispersal at the edge provides a direct explanation for increasing expansion speeds observed in range expansions (Phillips et al. 2006). Furthermore, evolution of heightened dispersal ability at the edge compared to the core is countered by opposing evolutionary decreases in intrinsic growth rate of edge individuals compared to the core, and these stochastic processes combine to generate variance in expansion speed. It is thus critical to understand the role of spatial evolutionary processes in range expansions to predict accurately the scope of possible outcomes as a species spreads (Melbourne et al. 2009; Phillips 2015). Building explicit consideration of spatial and temporal evolution into models for the prediction of spatial spread of invasive species or range shifts induced by climate change, most of which currently rely on purely demographic processes, will be a fruitful research area (Urban et al. 2016). In particular, our experimental results suggest that the inclusion of spatial evolutionary processes in such models will provide substantial gains in their predictive accuracy and estimates of uncertainty.
CHAPTER 4

Genetic signatures of adaptation and gene surfing in replicated biological range expansions

Christopher Weiss-Lehman, Silas Tittes, Nolan Kane, Ruth Hufbauer, Brett Melbourne

Abstract

Recent work has demonstrated that evolutionary processes can have dramatic impacts on the outcomes of range expansions. In particular, evolution due to the spatial structure of the expansion process itself can be particularly important. Differences in selection pressures between the core and edge of expansion can lead to different adaptive trajectories in these locations, however neutral processes (i.e. gene surfing) are expected to play a significant role in edge populations as well. The outcomes of adaptation and gene surfing can be difficult to distinguish on a genomic level since both are expected to reduce genetic diversity at the edge by increasing or decreasing particular allele frequencies. Here, we use a highly replicated microcosm experiment to assess the relative roles of deterministic (i.e. selection) versus stochastic (i.e. gene surfing) changes in allele frequencies among replicates. We show that while selection plays a strong role in driving allele frequency changes in edge populations, it cannot fully account for the overall reduction in nucleotide diversity observed at the edge. Therefore, gene surfing must also play a significant role in driving down diversity in edge populations, even over only eight generations. Our findings demonstrate the importance of considering both selective forces and gene surfing in understanding the role of evolution in range expansions.

Introduction

The process of range expansion, or the spatial spread of a species into novel habitat, has long been studied from an ecological perspective (Fisher 1937; Skellam 1951; Hastings et al. 2005), largely due to its importance in understanding the dynamics of invasive species (Elton
1958). Recently, however, research has demonstrated that evolutionary changes can significantly alter the dynamics of range expansions, even over only a handful of generations (Fronhofer & Altermatt 2015; Williams et al. 2016b; Weiss-Lehman et al. 2017; Ochocki & Miller 2017). In particular, rapid evolutionary changes at the expansion edge can lead to dramatic increases in the intrinsic variability of expansion speeds (Weiss-Lehman et al. 2017; Ochocki & Miller 2017). Given the importance of range expansions to conservation efforts, both due to invasive species (Pimentel et al. 2011) and species responding to climate change (Chen et al. 2011), it is important to understand the evolutionary mechanisms at play during range expansions and their relative contributions in driving changes in expansion speed and variance. However, due to the complexities of distinguishing the signatures of different mechanisms of evolutionary change (Hoban et al. 2016), this is not a trivial task.

Evolution during range expansions can occur via several distinct, yet potentially interacting mechanisms. One mechanism, not unique to range expansions, is the adaptive evolution that can occur as a result of encountering novel biotic or abiotic constraints in a new habitat (Szűcs et al. 2017). Three other evolutionary mechanisms are unique to range expansions as they are a direct result of changes in population spatial structure that occur during expansion. The first of these, spatial sorting, occurs via the aggregation of highly dispersive individuals at the expansion edge (Shine et al. 2011). If traits corresponding to high rates of spread, such as fecundity or dispersal ability, are heritable, then the non-random mating of individuals with these traits at the expansion edge will lead to evolution of increased trait values in edge populations over several generations (Phillips 2009; Phillips et al. 2010b). The second mechanism acts via the gradient in population densities that typically occurs in expanding populations, with high
population densities in the range core and relatively low densities at the edge. This can cause a subsequent gradient in selection pressure on density-dependent traits, such as competitive ability (Phillips et al. 2010a; Burton et al. 2010), leading to differential evolutionary trajectories between core and edge populations. These first two mechanisms are both the result of adaptive changes to the unique selection pressures at the edge. However the third mechanism, gene surfing, leads to a greater role for neutral dynamics at the expansion edge. Gene surfing results from the repeated founding events and small population sizes characteristic of the expansion edge (Edmonds et al. 2004; Hallatschek & Nelson 2008; Excoffier et al. 2009), thereby causing increasingly neutral fluctuations in allele frequencies in edge populations through time (Peischl et al. 2013). As a result, gene surfing can lead to a reduction of genetic diversity at the edge (Hallatschek & Nelson 2008) and an accumulation of deleterious alleles, with the potential to slow expansion rates (Peischl et al. 2015).

While there is evidence of all three of these spatially induced evolutionary mechanisms in range expansions, the relative contributions of individual mechanisms are difficult to distinguish in practice. Patterns of trait evolution in controlled laboratory experiments can provide some intuition into the relative roles of the mechanisms in driving changes in dispersal ability or fitness (Weiss-Lehman et al. 2017; Ochocki and Miller 2017), however linking changes in allele frequencies to one or more of these mechanisms is difficult (Hoban et al. 2016). For example, it can be challenging to determine if changes in allele frequency in expanding populations are due to adaptation versus neutral drift since both adaptive (e.g. spatial sorting) and neutral (gene surfing) mechanisms are expected to be present (Lotterhos & Whitlock 2015). However, if data are available for multiple, replicated expansions, it is possible to determine the relative roles of gene surfing and adaptive evolutionary mechanisms by comparing allele frequency changes.
across populations. Due to the stochastic nature of gene surfing, the same loci are not expected to display the same patterns in allele frequency shifts in replicated expansions. Adaptive mechanisms, in contrast, would be expected to produce very similar shifts in allele frequencies across the same genomic regions in replicated expansions through identical environmental conditions.

Here, we leverage this distinction to quantify the relative roles of gene surfing and adaptive evolutionary mechanisms in replicated range expansions of the red flour beetle (*Tribolium castaneum*) through laboratory microcosms (Weiss-Lehman et al. 2017). By using pooled genomic data and a contrast between treatments with and without spatial genetic structure, we quantify changes in allele frequencies specifically due to the spatial structure of range expansions. We further take advantage of the replicated nature of the experiment to assess changes due to adaptive mechanisms and gene surfing.

**Methods**

*Range expansion experiment*

The full details of the range expansion experiment are provided in detail elsewhere (Weiss-Lehman et al. 2017; Chapter 4), but we present a brief summary here. Experimental landscapes consisted of 4 cm x 4 cm x 6 cm acrylic boxes placed in a linear arrangement with holes drilled in the sides that could be easily blocked to allow controlled dispersal events. Each box was filled with a standard medium (95% wheat flour and 5% brewer’s yeast) to serve as both habitat and a food source. *T. castaneum* life cycles were manipulated to produce non-overlapping generations of 35 days each, consisting of discrete growth, dispersal, and reproductive phases.

To begin the experiment, 20 beetles from a large, well-mixed source population were placed in the first patches of 60 empty landscapes. These landscapes were then randomly divided
between two treatments designed to test the role of spatial evolution in range expansions. The first treatment, which we refer to as shuffled, prevented spatial evolution by randomizing the location of individual beetles within the landscape once per generation, while keeping local population abundances constant. In effect, the shuffled treatment decoupled an individual’s genotype from its location. In the second treatment, referred to as structured, beetles remained in the location to which they dispersed, thus allowing the formation of genetic spatial structure necessary for spatial evolution during range expansions. All 60 replicates were further divided into three temporal blocks. Range expansions proceeded for 8 generations, after which phenotypic assays were performed to test for evolution of dispersal, fitness, and competitive ability among beetles from the edge and core of structured treatments and the shuffled treatment, confirming that both dispersal and fitness had evolved in response to the spatial structure imposed by range expansion (Weiss-Lehman et al. 2017).

For each replicate, all 20 founding beetles were collected and stored at -80°C after the first generation. For structured replicates, 20 beetles from the range core (defined as patch 1 of the landscape) and the 20 furthest spreading beetles (defining the expansion edge) were collected and stored at -80°C after the 8th generation of expansion. Similarly, 20 beetles from the shuffled treatment were also stored at -80°C after the 8th generation of the experiment. The beetles collected from shuffled landscapes were randomly selected since spatial location has no biological meaning in that treatment and any group of 20 individuals should be representative of allele frequencies across the population.

DNA pooling, extraction, and sequencing

To pool the DNA for each population, we combined all the beetles from that population with 180 µL Phosphate Buffered Saline (PBS) and homogenized the mixture with a disposable
pestle. We then extracted total genomic DNA from these pooled samples using a DNeasy Blood and Tissue Kit (Qiagen). We used a Qubit fluorometer and a Thermo Scientific NanoDrop Spectrophotometer to assess each extracted DNA sample for both quantity and quality of extracted DNA.

We randomly selected 22 landscapes from the structured treatment and 15 landscapes from the shuffled treatment for sequencing after first excluding 6 structured and 9 shuffled landscapes for which there were potential problems during the DNA extraction process. Each structured landscape provided three populations for analysis: founders (generation 0) and core and edge populations (generation 8). Each shuffled landscape provided two populations: founders and the 20 randomly selected beetles from generation 8. This yielded 96 total populations for sequencing.

Paired end Illumina libraries were prepared for each pool using the Nextera DNA Sample Preparation Kit (Illumina) by the Next-Generation Sequencing Facility at the Biofrontiers Institute of the University of Colorado, Boulder and subsequently sequenced on an Illumina NextSeq V2 machine. Sequencing reads were trimmed using Trimmomatic (v0.36) (Bolger et al. 2014) and aligned to the reference genome for *T. castaneum* (Richards et al. 2008) using bwa mem (v0.7.5a-r405) (Li & Durbin 2009), with default settings for both. After the initial bwa alignment, we used the CleanSam utility from Picard (http://broadinstitute.github.io/picard) to soft-clip alignments beyond the end of reference sequences and to set the mapping quality to 0 for unmapped reads. We used the RealignerTargetCreator and IndelRealigner utilities from the Genome Analysis Toolkit (v3.7-0-gcfedb67) to target and realign reads around indels (McKenna et al. 2010). We used samtools (v0.1.19-96b5f2294a) view and mpileup under default settings to exclude aligned reads with a mapping quality below 20 and to extract multi-sequence pileups,
respectively (Li et al. 2009). Quality scores, read lengths, and average sequencing depth are reported in Appendix B.

**Analyzing nucleotide diversity**

We used Popoolation (Kofler et al. 2011) to calculate \( \pi \) (nucleotide diversity) across the *T. castaneum* genome using a sliding window approach with a window size of 10,000 bp and a pool size of 20. We restricted Popoolation to only consider bi-allelic sites with coverage between 4 and 22 (chosen according to the distribution of average depths across populations; Fig. B1) and a quality score of at least 20. To assess overall trends in \( \pi \), we calculated the mean value across autosomes and the X chromosome separately for each population and calculated the change in means between founding and generation 8 populations for each landscape. We analyzed these changes with a linear model assessing the role of location (core, edge, or shuffled (i.e. no location)) in the change in mean \( \pi \) from founding populations to the 8th generation.

Additionally, to examine trends in \( \pi \) across the genome, we computed average \( \pi \) values for each window across the genome in core, edge, and shuffled populations from generation eight. To assess trends in \( \pi \) across different chromosomes, we further fit a non-parametric, locally weighted polynomial regression (loess smoother; Cleveland & Devlin 1988) to the \( \pi \) values for each chromosome in core, edge, and shuffled populations using the loess function in R (v3.2.3) (R Core Team) with default settings.

To quantify the roles of selection and gene surfing in altering allele frequencies over eight generations of spread, we used a binomial model to determine the probability of a given 10,000 bp window sharing a consistently low \( \pi \) value across replicate populations. If gene surfing caused reductions in \( \pi \), then we would not expect the same windows to be affected across replicate populations due to the stochastic nature of gene surfing, and we would therefore see a
low probability of windows sharing consistently low $\pi$ values. If selection was responsible for reductions in $\pi$, however, we would expect the same windows to be affected across replicate populations and, consequently, we would expect a higher probability of populations sharing low $\pi$ values at the same windows. For this analysis, we first defined the criteria for a window to share consistently low $\pi$ values across populations. To define a “low” value of $\pi$, we used a series of quantiles (0.05, 0.1, 0.15, 0.2, and 0.25) such that the value of $\pi$ for a given window was defined as low if it fell beneath the given quantile (defined on the distribution of $\pi$ values for each population). By defining the quantile according to the distribution of $\pi$ values for each population individually, we focus on relative rather than absolute reductions in $\pi$, and by using a series of quantiles we can consider patterns of shared windows over a gradient of selection strength. We then defined a window as sharing low values of $\pi$ across replicate populations if at least 75% of the replicate populations had a low value of $\pi$ (as defined by the quantile value) for that window. Finally, we fit a generalized linear model (GLM) with a binomial distribution and a logit link function to these binary values (shared or not shared). In the model, we initially included the effects of location, chromosome, quantile, and all their interactions. After discarding non-significant model terms, as determined by analysis of the deviance table for the full model, we used a model with effects of location, chromosome, quantile, and a quantile by chromosome interaction to generate model estimates and 95% confidence intervals for the probability of sharing a window with low $\pi$ values across chromosomes for each of the quantile levels and each population type from generation eight (core, edge, and shuffled).

Assessing the population covariance structure

As an independent assessment of the effect of eight generations of range expansion on the genetic structure of the 96 populations we used the program BayPass (Gautier 2015) to estimate
the standardized allele frequencies and population covariance matrix for autosomal loci. To generate read count data used as input for BayPass, we used a custom Python script to filter and genotype loci reported in the Popoolation sync file. We required loci to have allele counts between 4 and 22 within each population and mean allele counts between 4 and 22 across populations. These cutoffs were chosen based on the distribution of average depth across populations generated using the samtools depth utility (Figure B1). Additionally, we required the minor allele to be counted a total of at least 3 times, with a frequency no less than 0.002. This resulted in 109,885 loci with an average distance of 1,309 (+/- 4,712 s.d.) base pairs between loci, and an average of 12,206 loci per chromosome. We additionally set the pool size to twice our population sizes (40) as recommended in the documentation for BayPass. We then performed a principal component analysis on the population covariance matrix generated by BayPass to visualize differences in genetic structure among populations.
Results

On average, autosomal $\pi$ values were similar for founders of both treatments (Fig. 4.1), confirming that differences arising over the 8 generations of range expansion were due to experimental treatments rather than initial genetic conditions. In all populations from generation eight, mean autosomal $\pi$ values were reduced compared to the founding populations (Fig. 4.1). Analyzing the differences in mean autosomal $\pi$ values with a linear model confirmed that, while all populations had reduced values compared to the founders, this reduction was especially pronounced in edge populations (Fig. 4.2; $p = 0.03$). These results were similar for the X chromosome as well (Appendix C; $p = 2.8 \times 10^{-6}$). Comparing $\pi$ values across chromosomes revealed distinct trends in nucleotide diversity among chromosomes. In particular, the X
chromosome seemed to have distinctly lower values of $\pi$, while chromosome 5 had relatively high values of $\pi$ compared to the rest of the genome (Fig. 4.3). The loess smoother fits to $\pi$ values across the genome for core, edge, and shuffled populations further showed that while edge populations tended to have lower $\pi$ on average, this pattern was much more pronounced in certain chromosomes. For example, chromosomes 2 and 3 tended to display distinctly lower $\pi$ values in edge populations compared to core and shuffled, while the X chromosome and chromosomes 6 and 8 displayed much less differentiation in $\pi$ values among core, edge, and shuffled populations (Fig. 4.4).
The binomial GLM demonstrated that, while the probability of a window sharing low \( \pi \) values across replicate populations varies substantially with chromosome, on average the probability was consistently lower for edge populations (Fig. 4.5). This pattern held across a gradient of quantile values used to define a low diversity window, though the model estimated probability increased with increasing quantile value. While all chromosomes exhibit some variance in the probability of sharing a window, of particular note is chromosome 5, for which

**Figure 4.3:** Nucleotide diversity for sliding windows of 10,000 bp across the genome for populations from generation 8. Data are log transformed to better visualize the spread and chromosomes are colored in alternating black and grey.
there was a particularly low probability of sharing a low diversity window across replicate populations for core, edge, and shuffled populations and across all five quantile values (Fig. 4.5), potentially indicating a strong role for neutral dynamics in driving allele frequencies on chromosome 5.

Principal component analysis of the population covariance matrix generated by BayPass recovered the temporal and spatial structure inherent in the populations. Using the first two principal components (which collectively explained 95.54% of the variance) revealed distinct clustering of populations with all generation eight populations shifted away from the founding populations and edge populations having shifted the furthest (Fig. 4.6). The first principal
Figure 4.5: Model estimated probabilities of sharing a low diversity window in each chromosome. The different quantiles used to define a low diversity window are shown to the left of each plot. Core, edge, and shuffled populations are colored according to the figure legend. Error bars are 95% confidence intervals.

component (which explained 94.56% of the variance) revealed a distinct pattern among populations that seems to mirror the pattern in $\pi$ values (Fig. 4.7), which could indicate that
nucleotide diversity covaries with the first principal component of the population covariance matrix. Additionally, the distribution of edge populations along the first principal component axis is significantly more variable than the distribution of other populations (pair-wise F-test among population categories; $p = 0.02$, $p = 0.04$, and $p = 0.001$ for comparisons between edge and core, shuffled, and founding populations respectively).

**Discussion**

Both neutral and adaptive processes are known to be important in range expansions, however, it can be difficult to fully distinguish the relative roles of each (Hoban et al. 2016). Here, we leveraged the controlled and replicated nature of a laboratory experiment testing the
role of evolution in range expansions to assess the relative contributions of adaptive processes, such as spatial sorting and density-driven evolution, and neutral processes, like gene surfing. By comparing patterns of allele frequency changes among replicate populations, we assessed the role of gene surfing in evolution of edge populations, particularly in light of its effect on nucleotide diversity across the genome.

Range expansions, particularly in the contexts of invasive species, are typically characterized by an initial population bottleneck, followed by population growth and spread. This initial bottleneck, combined with any novel selection pressures in the new environment, can have substantial effects on genetic diversity, even after multiple generations of growth and expansion (Szűcs et al. 2017). We demonstrate here that the effects of bottlenecks on genetic
diversity are particularly pronounced in edge populations during expansion (Fig. 4.1 & 4.2), because both neutral and adaptive processes are amplified at the edge. The edge is formed during expansion via repeated bottleneck events, leading to gene surfing and an increased role for neutral processes (Peischl et al. 2013). Similarly, edge populations are subject to additional selection pressures due to the spatial structure of range expansions (e.g. spatial sorting (Shine et al. 2011)). This combination of factors makes it highly likely for edge populations to evolve reduced genetic diversity across the genome (Fig. 4.3 & 4.4).

In natural populations, it is difficult to determine the role of adaptive and neutral processes in the reduction of genetic diversity at the edge as both are expected to contribute and any expansion in nature is only one realization of the stochastic process. Replicated experiments are necessary to disentangle the interacting roles of selection and gene surfing in driving this reduction of genetic diversity in edge populations. Using a binomial GLM to determine the probability of the same genomic window displaying low diversity across replicate populations leverages the replicated nature of the experiment to disentangle adaptive and neutral contributions to patterns in nucleotide diversity. The binomial GLM revealed that edge populations had, on average, lower probabilities of sharing the same low diversity windows across both chromosomes and different quantile values compared to core and shuffled populations (Fig. 5). This stands in contrast to the observation of reduced $\pi$ values in edge populations (Fig. 4.1). If only adaptive processes were responsible for the reduction in nucleotide diversity, then we would expect edge populations to share at least as many, if not more, low diversity windows across populations as observed in core and shuffled populations due to the directional nature of adaptation due to selection. However, the combination of reductions in nucleotide diversity beyond what is seen in core and shuffled populations and lower probabilities
of low diversity windows being shared among populations across the genome implies a statistically significant role for neutral mechanisms of genomic evolution in edge populations.

Additionally, when examining the covariance structure among all populations, edge populations display both the greatest segregation from the founding populations and the greatest variation within a population category (Fig. 4.6 & 4.7). The segregation of population categories seen in Figure 4.6 suggests that the population covariance matrix is capturing both spatial and temporal structure among the replicate populations. While there is some overlap among founders and core and shuffled generation 8 populations (Fig 4.7), the patterns within individual landscapes clearly capture patterns of spatial and temporal evolution. Generation 8 populations are always at lower values (on PC1) compared to their founding population and edge populations are similarly always at lower values compared to the core population from the same landscape. This suggests that the variation among edge populations in the population covariance matrix reflects a greater degree of stochasticity in evolutionary trajectories in those populations compared to core and shuffled populations. Given the tightly controlled nature of the experiment, this again suggests a role for neutral mechanisms (i.e. gene surfing) in the evolution of edge populations.

Both analysis of nucleotide diversity in edge populations and the population covariance matrix created by BayPass (Gautier 2015) imply an important role for gene surfing in the evolution of edge populations. This is striking for two reasons: the time scale of the experiment and the implications for the role of stochastic mechanisms in range expansions. While gene surfing has been observed in range expansions previously, both empirical (Gracia et al. 2013; Henn et al. 2016) and theoretical (Peischl et al. 2013; Peischl et al 2015; Gilbert et al. 2017) research has dealt with much longer time scales than presented here (hundreds of generations or
more). Given that we observed evidence for gene surfing over only 8 generations, it now seems
that it could be important even in the early stages of range expansion, for example with recently
introduced invasive species. The effect of gene surfing over short time scales could help to
explain observed increases in the intrinsic variance of expansion speeds of experimental
landscapes with spatial evolutionary processes compared to those without (Ochocki & Miller
2017; Weiss-Lehman et al. 2017; Chapter 3). The stochastic nature of gene surfing could
potentially lead to a greater role of stochasticity in the evolutionary trajectories of edge
populations, as discussed above. As the speed of range expansion is governed by the traits of
individuals at the expansion edge (Hastings et al. 2005), this could lead to the observed increases
in intrinsic variance of expansion speeds (Ochocki & Miller 2017; Weiss-Lehman et al. 2017;
Chapter 3). Hence, gene surfing seems to play a key role in creating a fundamental limit to the
uncertainty associated with predictions of spread for range expansions (Melbourne & Hastings
2009) through increasing the role of stochasticity in evolutionary trajectories of edge
populations.
CHAPTER 5

Habitat heterogeneity modulates the role of evolution in range expansions

Christopher Weiss-Lehman, Alan Hastings, Sam Flaxman, Brett Melbourne

Abstract

The expansion of a species’ range is a fundamental biological process with important implications for invasive species and those species responding to climate change. While range expansions have historically been considered purely demographic processes, recent experimental and theoretical evidence demonstrates that evolution can significantly alter the dynamics of range expansion, particularly by increasing variance in expansion speed. However, other ecological factors, such as landscape structure, are likely to play a role in determining variance in range expansions as well. Here we combine experimental evidence from microcosms of the red flour beetle, Tribolium castaneum, with an individual-based model to explore the interaction of habitat heterogeneity and evolutionary dynamics in range expansions. Results from the T. castaneum system demonstrate that, counterintuitively, habitat heterogeneity reduces variation in expansion speed compared to replicated expansions through homogeneous landscapes. The individual-based model suggests that this effect could primarily be due to the differing roles of dispersal evolution in the two landscape types. Heterogeneous landscapes impose an additional cost on dispersal (i.e. the risk of dispersing into unfavorable habitat), which can alter the relationship between core and edge populations in heterogeneous compared to homogenous landscapes. These results demonstrate the importance of considering both landscape structure and trait evolution in determining the intrinsic uncertainty of range expansions in nature.

Introduction
Range expansion of a species into new territory is a fundamentally important process in ecology and evolution, with the potential to cause local extinctions of native species (Pimentel et al. 2005), alter the composition of communities (Dornelas et al. 2014), or facilitate allopatric speciation (Bocxlaer et al. 2010). Historically, range expansions resulted from major geologic shifts or rare, long-distance dispersal events and often led to the evolution of new species over long periods of time (Bocxlaer et al. 2010). Currently, however, anthropogenic activities have made range expansions much more common and problematic. Expansions of invasive species into new habitats are occurring at unprecedented rates (Pimentel et al. 2005) and climate change is already causing many terrestrial species to alter their ranges, often upward in latitude and elevation (Chen et al. 2011). These contemporary range expansions are leading to increasingly homogenized species assemblages across the globe as endemic species are lost and replaced by introduced species (Dornelas et al. 2014). From a conservation perspective, it is important to understand the factors underlying these range expansions, in some cases to promote species’ tracking a changing climate and in others, to deter the spread of introduced species into new habitats.

While much research has elucidated the role of demographic traits in range expansions (Skellam 1951; Hastings et al. 2005), recent work has demonstrated the importance of rapid evolutionary changes in the dynamics of range expansion (Travis & Dytham 2002; Phillips et al. 2010b; Kubisch et al. 2013; Perkins et al. 2013; Williams et al. 2016b; Ochocki & Miller 2017; Weiss-Lehman et al. 2017). Evolution can affect range expansions via two broad mechanisms: (1) exposure to novel selection pressures in new environments leading to adaptive evolutionary changes and (2) selection imposed by the spatial population structure characteristic of range expansions leading to spatially divergent evolution between the core and edge of the expanding
range. The second of these mechanisms can be broken down into at least two processes: spatial sorting and gene surfing. The first, spatial sorting, operates via assortative mating at the range edge between individuals with phenotypes corresponding to high rates of expansion (e.g. dispersal ability or fecundity). If these phenotypes have a heritable component, then edge populations are expected to evolve increases in these phenotypes relative to core populations (Shine et al. 2011). The second process, gene surfing, can occur due to the small population sizes and repeated founder events characteristic of the expansion edge. These founder events can lead to neutral or near neutral fluctuations in gene frequencies such that beneficial alleles can be lost or deleterious alleles can become fixed at the edge. Gene surfing is expected to both reduce genetic diversity and potentially lead to a buildup of deleterious alleles at the expansion edge, thus reducing fitness and slowing expansion (Peischl et al. 2013; Peischl et al. 2015; Weiss-Lehman et al. 2017; Chapter 4).

Recent research has begun to explore the effect of this second broad mechanism of evolutionary change in range expansions (i.e. spatial evolutionary change), encompassing spatial sorting (Shine et al. 2011; Phillips et al. 2010b; Perkins et al. 2013; Lombaert et al. 2014; Fronhofer & Altermatt 2015; Shaw & Kokko 2015; Ochocki & Miller 2017; Weiss-Lehman et al. 2017) and gene surfing (Edmonds et al. 2004; Hallatschek & Nelson 2010; Graciá et al. 2013; Peischl et al. 2013; Peischl et al. 2015; Chapter 4). However, the role of the first mechanism, differing selection imposed by different environments encountered during range expansions, has received comparably less attention. Given the increasing scope of habitat destruction and fragmentation across the globe (Haddad et al. 2015), it is necessary to consider both of these mechanisms together to understand how evolution will impact range expansion dynamics in the increasingly fragmented global landscape.
Habitat heterogeneity is known to be important in evolution of stationary populations and is likely to interact with evolution in expanding populations. Dispersal evolution, for example, has long been known to be impacted by spatial heterogeneity in stationary populations (McPeek & Holt 1992). Recent empirical and theoretical work has also shown that spatial heterogeneity has important consequences for genetic variation of non-expanding populations. Spatial heterogeneity was shown to increase genetic variance in fitness related traits in experimental populations of Drosophila melanogaster over spatially homogenous treatments (Huang et al. 2015). It has also been shown to affect the stability of the G matrix (genetic variance and covariance matrix among quantitative traits) in theoretical models (Bjorklund & Gustafsson 2015). Recently, researchers have begun to consider the interaction between spatial heterogeneity and evolution due to the spatial population structure of range expansions. One study found that environmental gradients reduce expansion load (i.e. the buildup of deleterious alleles) at the expansion edge (Gilbert et al. 2017). Another demonstrated that large gaps in suitable habitat can lead to the evolution of increased competitive ability at the edge due to the resultant pause in expansion speed (Williams et al. 2016a). Taken together, these findings indicate that gene surfing may play a reduced role in range expansions through heterogeneous habitat. Given the role of gene surfing in driving increased variability in experimental range expansions through homogeneous habitat (Weiss-Lehman et al. 2017; Chapter 4), this suggests that evolution may interact with landscape structure to influence variability in range expansions differently in homogeneous and heterogeneous landscapes. Indeed, a recent experimental study demonstrated that evolution interacted with landscape heterogeneity to impact mean expansion speed (Williams et al. 2016b), implying there is likely an effect on variability of expansion speed as well.
We set out to test the interacting roles of habitat heterogeneity and evolution in range expansions by combining experimental data with a theoretical model. Using microcosms of the red flour beetle (*Tribolium castaneum*), we compared range expansion dynamics through homogeneous and heterogeneous landscapes under tightly controlled laboratory conditions. We then compared results from this experiment to a separately parameterized individual-based model allowing both individual fitness and dispersal ability to evolve. Using the model, we simulated range expansions through homogeneous and heterogeneous landscapes, matching the experimental design for the *T. castaneum* microcosms, and evaluated the contribution of evolution in both traits to mean and variance in distance spread.

**Methods**

*Experimental system*

Details of the *T. castaneum* experimental system are described in detail elsewhere (Melbourne & Hastings 2008; Melbourne & Hastings 2009; Szücs et al. 2017; Weiss-Lehman et al. 2017), but we reiterate them briefly here. Populations were housed in artificial, single dimensional landscapes consisting of 4 cm x 4 cm x 6 cm acrylic boxes with 2 mm holes drilled in the sides to allow dispersal. Each box was filled with 20 g of a carbohydrate food source, the nutritional quality of which could be altered to adjust patch quality. All landscapes were kept in incubators at a constant temperature (31°C). The life cycle of *T. castaneum* was defined by non-overlapping, 35-day generations consisting of discrete reproductive, growth, and dispersal phases. In the reproductive phase, adult beetles were placed in habitat patches with fresh carbohydrate medium for 24 hours to reproduce and lay eggs. This was followed by a 33-day growth phase over which the eggs hatched and matured into adult beetles. Finally, the dispersal barriers in place for most of the life cycle were removed to allow dispersal for 24 hours before
dispersal was halted and populations were censused. The adults were then placed in new landscapes with fresh medium to start the next generation.

To test the effect of habitat heterogeneity, 60 experimental landscapes were founded with 20 individuals from a large, well-mixed population placed in the first patch of the landscape. Landscapes were randomly divided evenly between two treatments: heterogeneous and homogeneous. Heterogeneous landscapes were constructed with an alternating pattern of patch quality between high (95% wheat flour and 5% brewer’s yeast) and low (100% wheat) quality habitat. Homogeneous landscapes used all the same type of medium, representing an intermediate habitat quality (97.5% wheat flour and 2.5% brewer’s yeast). Landscapes were divided randomly among five temporal blocks and range expansions proceeded for seven generations.

Distance spread through time (defined as the furthest occupied patch at the census stage of each generation) in each treatment was analyzed with a linear mixed effects model. Treatment, generation, and their interaction were fixed effects in the model and individual landscapes nested within temporal blocks were treated as random effects. As all landscapes started expansion from the first patch (effectively fixing the intercept of the model) random effects were incorporated for the slope only. Standard deviation of distance spread within each treatment through time was also calculated and analyzed with a standard linear model incorporating treatment, generation, and their interaction.

**Individual-based model**

To directly investigate the role of evolution in range expansions through heterogeneous habitat, we developed an individual-based, quantitative-genetic model to match the life cycle and spatial structure of the *T. castaneum* experimental system. The model simulates range expansion
of a population through a single dimensional, integer lattice representing habitat patches. Populations are defined by a life cycle consisting of non-overlapping generations with discrete reproductive, growth, and dispersal phases. Local population dynamics are governed by a stochastic implementation of the classic Ricker model (Ricker 1954, Melbourne & Hastings 2008) and local patches are coupled via dispersal. Both dispersal and individual fitness are modeled as quantitative genetic traits to allow for evolution during range expansion. Individuals in the model are defined by their location, generation, sex, breeding values, and phenotype values for each trait.

Trait evolution

We used a quantitative genetics framework to model trait evolution (Roughgarden 1979; Lynch & Walsh 1998). Each trait was defined via an individual’s breeding value \( h_{i,T} \) for individual \( i \) and trait \( T \) which was then translated to a phenotypic value \( z_{i,T} \). Inheritance of traits was dictated by a heritability parameter \( h^2_T \) for each trait \( T \), and traits were initialized with a value for the total phenotypic variance present in that trait at the start of simulations \( V_{p,T} \). Using the heritability and total phenotypic variance, a value for the environmental variance parameter for each trait \( V_{e,T} \) and the initial value of the additive genetic variance variable \( V_{a,T} \) were both calculated from (Lynch & Walsh 1998):

\[
V_{p,T} = V_{a,T} + V_{e,T}, \quad (\text{Eqn 1})
\]

\[
V_{a,T} = h^2_T V_{p,T} \quad (\text{Eqn 2}).
\]

We assumed that the additive genetic variance of local subpopulations evolves through time, while the environmental variance is constant for each trait (i.e. an assumption of minimal temporal variance in environmental conditions on the generational scale). This assumption
allows the examination of spatial environmental variation (habitat heterogeneity) without the confounding influence of temporal variation.

Using the initial value of additive genetic variance and assuming a large, randomly mating population, initial individual breeding values were distributed as (Lynch & Walsh 1998):

\[ b_{i,T} \sim Normal(\bar{b}_T, \sqrt{V_{a,T}}) \]  

(Eqn 3),

where \( \bar{b}_T \) is the average breeding value of the initial population for trait \( T \) and was set at the beginning of each simulation. Breeding values then evolved through time via offspring inheriting them from parents with no dominance or epistasis, which yields the following distribution of breeding values (Roughgarden 1979):

\[ b_{o,T} \sim Normal\left(\frac{b_{iT} + b_{jT}}{2}, \sqrt{\frac{V_{a,T,x}}{2}}\right) \]  

(Eqn 4).

Here, \( b_{o,T} \) is the breeding value inherited for offspring \( o \) from parents \( i \) and \( j \). \( V_{a,T,x} \) is the additive genetic variance of trait \( T \) in mating population \( x \), since individuals, and hence potential mating pairs, become segregated in space once range expansion begins. Thus, the model only considers evolution based on standing genetic variation and the effects of gene flow and selection, but it does not consider mutation. As our aim is to consider short time periods (7 generations) to match experimental data, this assumption should have little impact on the results. Individual phenotype values \( z_{i,T} \) are then distributed as (Lynch & Walsh 1998)

\[ z_{i,T} \sim Normal(b_{i,T}, \sqrt{V_{e,T}}) \]  

(Eqn 5)

We assume sexual reproduction, with individuals born as male or female according to a Bernoulli distribution with an even sex ratio. Females randomly select one local (within the same patch) male to mate with either before or after dispersal and then give birth to offspring after dispersal. This creates a polygynous mating system in which each female mates once but males
may mate with multiple females. This also means that the additive genetic variance of a mating population for a given female ($V_{a,T,x}$) is defined as the combined genetic variance of both the current and former location of each female. We made this assumption about mating populations to match the life cycle of *T. castaneum*.

**Local population dynamics**

Local (within patch) population growth is determined via a stochastic, individual-based Ricker model (Ricker 1954; Melbourne & Hastings 2008). This model assumes that density dependence occurs via adult cannibalism of new offspring and has been shown to accurately capture the population dynamics of the *T. castaneum* system (Melbourne & Hastings 2008). The number of offspring produced by each female is distributed as,

$$B_{i,x} \sim \text{Poisson}(R_{i,x} e^{-\alpha N_{t,x}})$$

where $N_{t+1,x} = \sum_{i=1}^{F_{t,x}} B_{i,x}$

(Eqn 6).

$B_{i,x}$ is the number of individuals produced by female $i$ in patch $x$, $R_{i,x}$ is the intrinsic fitness (mean number of surviving offspring at low density) of female $i$ in patch $x$, $\alpha$ is a parameter defining the search rate of adults cannibalizing new offspring, $N_{t,x}$ is the population size at time $t$ in patch $x$, and $F_{t,x}$ is the number of females at time $t$ in patch $x$. The intrinsic fitness is a trait subject to evolution. The intrinsic fitness of each female is determined by its phenotype ($z_{i,R}$) according to

$$R_{i,x} = R_{\max}(H_x)e^{-k(2z_{i,R}^+(-1)^{[H_x=0]}\theta)^2}$$

(Eqn 7)

where $k$ is a parameter defining the strength of stabilizing selection, $\theta$ is a parameter representing the offset of phenotypic optimums in each habitat type, $H_x$ is the habitat type of patch $x$, the double brackets indicate an indicator function evaluating to 1 when the interior condition is true and 0 otherwise, and $R_{\max}$ is a piecewise function evaluating to a different value for each habitat type. In heterogeneous landscapes, this creates disruptive selection between different phenotypic optimums for different habitat types with the strength of disruptive selection governed by the $k$
and $\theta$ parameters. In homogeneous landscapes, $\theta$ is set to 0 and $R_{\text{max}}$ is a single value representing the single phenotypic optimum in a homogeneous landscape.

**Dispersal among patches**

Local populations in this model are linked via dispersal of individuals which occurs after maturation of the new offspring and before birth of the next generation. Dispersal is modelled as a discrete-space diffusion process (Melbourne & Hastings 2009). Dispersal events (defined as a single individual moving one patch forward or backward) occur via a Poisson process in which the event times for each individual $i$ are defined by

$$\tau_i = \delta / D_i n_i \quad \text{where} \quad \delta \sim \exp(\lambda) \quad \text{(Eqn 8)}.$$  

In the above equation, $\tau_i$ is the waiting time for individual $i$, $\delta$ is an exponentially distributed random variable with rate parameter $\lambda$, $D_i$ is the diffusion coefficient for individual $i$ as defined by its trait value (see Equation 9 below), and $n_i$ is the number of directions (forward or back), available to individual $i$, equal to one for individuals in the first patch (because they can only move forward) and two for all other individuals (Melbourne & Hastings 2009). The Poisson process occurs over a length of time equal to one dispersal phase of the life cycle, so the dispersal parameters, in particular $D_i$, are interpreted on that time scale.

The diffusion coefficient, $D_i$, is a trait subject to evolution and is related to an individual’s dispersal phenotype ($z_{i,D}$) as follows

$$D_i = \eta e^{\gamma z_{i,D}} \quad \text{(Eqn 9)}.$$  

This equation is similar to the inverse logit function, but with added parameters $\eta$ and $\gamma$. $\eta$ is the upper limit on the diffusion parameter and $\gamma$ controls the rate at which the function transitions between upper and lower limits with small values of $\gamma$ corresponding to a slow transition rate.
Simulation experiment

To test the role of evolution in range expansions through heterogeneous landscapes, we performed a simulation experiment mirroring the empirical experiment with *T. castaneum*. First, we used published data on range expansion in the *T. castaneum* system (Melbourne & Hastings 2009) to find approximate parameter values for the model to mimic population dynamics in this system. We then simulated range expansions through heterogeneous landscapes of alternating high and low quality habitat and through homogeneous landscapes of uniform, medium quality habitat. Simulated range expansions were founded with 20 beetles in patch 1 and proceeded for 7 generations. For each landscape type, we simulated expansions under a range of heritability values for each trait (from 0 to 1 by values of 0.1), thus systematically altering the strength of evolution in each trait. For each combination of heritability values and landscape structure we simulated 1,000 range expansions, yielding a total of 242,000 simulations. See Table 5.1 for the parameter values and initial conditions used in our simulations. While most of these simulations resulted in successfully expanding populations, several resulted in early extinctions due to demographic fluctuations. These only represented a small number of the original simulations (<1%) and were excluded from subsequent analyses to focus on the dynamics of successfully expanding populations. To fully match the *T. castaneum* experiment in which stock populations were reared in good quality (95% wheat flour and 5% brewer’s yeast) habitat before the start of the experiment, initial breeding values for the fitness trait were set to the optimum trait value for the good quality environment.

Finally, while we used previously published data to approximate parameter values for our model, the possibility exists that the effects we observed in the model are a result of fortuitous choices of those parameter values. Therefore, we performed a sensitivity analysis on our model,
systematically varying those parameters and initial conditions for which we had the least ability
to accurately measure from our parameterization data: $\eta$, $k$, $\gamma$, $V_{p,R}$, and $V_{p,D}$. Details and results
of the sensitivity analysis are presented in Appendix D.

**Table 5.1**: Default parameter values and initial conditions used for each simulation unless
explicitly specified otherwise.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Initial Condition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.0065</td>
<td>$V_{p,R}$</td>
<td>1</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (High quality habitat)</td>
<td>20</td>
<td>$V_{p,D}$</td>
<td>10</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (Medium quality habitat)</td>
<td>15</td>
<td>$\bar{b}_R$</td>
<td>1</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (Low quality habitat)</td>
<td>10</td>
<td>$\bar{b}_d$</td>
<td>0</td>
</tr>
<tr>
<td>$k$</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results**

Experimental populations of *T. castaneum* expanding through homogeneous and
heterogeneous landscapes displayed little change (7% (CI -7 to 19%) in mean expansion speed
from heterogeneous to homogeneous landscapes; parametric bootstrap, treatment by generation
interaction; $p = 0.32$; Fig. 5.1). However, there was a pronounced difference in the variance in
distance spread through time (likelihood ratio test, treatment by generation interaction; $p =
0.0001$; Fig. 5.1d). For integer valued data such as these, effects on the variance can be driven by
effects on the mean, but since the mean distance spread did not differ between treatments (Fig.
5.1c), the increase in variance in homogeneous landscapes is strictly due to differences between
the treatments.

In the simulation model, trait evolution in both individual fitness and dispersal ability
contributed to increased spread rates through both homogeneous and heterogeneous landscapes
(Fig. 5.2). In heterogeneous landscapes, dispersal evolution appeared to play a greater role than
fitness evolution in driving increased spread rates, while evolution of both traits contributed relatively equally to increases in distance spread in homogeneous landscapes. However, homogeneous landscapes displayed a greater range of mean spread rates overall, with both the highest and lowest observed value from the simulations. In comparing distances spread over the 7 generations of the simulation, fitness evolution seemed to determine whether expansions proceeded faster in homogeneous or heterogeneous landscapes with low fitness evolution leading
To greater distance spread in heterogeneous landscapes and vice versa at high levels of fitness evolution (Fig. 5.3).

To examine variability in distance spread in the model simulations, we used the coefficient of variation (CV) to account for the fact that mean spread rate varied across landscape types and parameter space. In homogeneous landscapes, dispersal evolution had a strong effect of increasing the CV for distance spread, while fitness evolution had the opposite effect (Fig. 5.4). Heterogeneous landscapes displayed a similar, but less pronounced pattern. This meant that, unlike for mean distance spread, the difference in CV of distance spread between landscape types was influenced by both fitness and dispersal evolution (Fig. 5.5). When dispersal evolution
was greater than fitness evolution, distance spread was less variable in heterogeneous landscapes and vice versa when fitness evolution dominated. (Fig. 5.5).

To understand the driving forces of these patterns, we examined the trait values in core and edge populations after seven generations of expansion across a range of heritability values. For these analyses, core populations were defined as the point of introduction in the landscape (patch 1), representing the longest occupied area in the landscapes. Edge populations were defined by the patches necessary to contain the 20 furthest forward individuals (Weiss-Lehman et al. 2017). For example, if the last occupied patch in the landscape had 20 or more individuals,
then the edge population was defined as the population occupying only that last patch. On the other hand, if the last patch had less than 20 individuals, then the edge extended to include the second to last patch (and so on if there were still less than 20 individuals). To isolate the dynamics of a particular trait over a range of heritability values, and to make direct comparisons between core and edge in both landscape types, we set the heritability of the other trait to a single, intermediate value (0.5). We then converted the mean breeding values in core and edge populations to values for the intrinsic fitness and diffusion coefficient using equations 7 and 9 respectively.

For the diffusion coefficient, edge populations in both homogeneous and heterogeneous landscapes evolved to higher values with increasing heritability, approaching the maximum achievable value at high levels of dispersal heritability (Fig. 5.6). On the other hand, evolution of the diffusion coefficient diverged in core populations from homogeneous and heterogeneous landscapes. In heterogeneous landscapes, core populations evolved reduced diffusion
coefficients on average from the starting conditions, while core populations in homogeneous landscapes evolved slightly increased diffusion coefficients (Fig. 5.6). Genetic variance in dispersal was similarly reduced in edge populations of both homogeneous and heterogeneous landscapes. Genetic variance in core populations was substantially higher in both landscape types, but core populations in homogenous landscapes displayed slightly lower variance than their heterogeneous counterparts at high levels of dispersal heritability (Fig. 5.6).

**Figure 5.5**: Difference in the coefficient of variation (CV) of distance spread over 7 generations of range expansion between homogeneous and heterogeneous landscapes (positive values indicate greater variation in distance spread in heterogeneous landscapes). The difference in spread rates is indicated by color according to the color key to the right of the figure. Values for heritability of dispersal and individual fitness are shown on the x and y axes respectively. Results are only shown for successful expansions (i.e. no extinctions).
Analysis of the fitness trait revealed that both core and edge populations in homogeneous landscapes evolved towards the fitness optimum for those landscapes (Fig. 5.7), however edge populations appear to lag slightly behind core populations at intermediate values of fitness heritability. Populations in heterogeneous landscapes, on the other hand, started at the fitness optimum for the high quality habitat in those landscapes and core populations in those landscapes remained quite close to the starting value (Fig. 5.7). Edge populations in heterogeneous landscapes, however, evolved slightly away from the optimum trait value with increasing heritability of fitness. Genetic variance in fitness displayed a similar pattern to genetic variance of dispersal with variance reduced in edge populations regardless of landscape type. However, unlike dispersal, edge populations from heterogeneous landscapes evolved a higher genetic variance in fitness compared with edge populations in homogeneous landscapes (Fig. 5.7).
However, the absolute difference in genetic variance between core and edge populations is similar in both landscape types.

Results from the sensitivity analysis demonstrate that the results presented above are robust to choice of parameters and initial conditions (Appendix D). None of the parameters had a substantial effect on the mean distance spread over seven generations, with the exception of $\eta$. Increasing $\eta$ resulted in increases in mean spread rate, which makes sense as $\eta$ defines the upper bound of individual diffusion coefficients (Eqn 9). Importantly, though, the relative relationship between mean distance spread in each landscape type did not vary over the range of any of the parameters explored. For the coefficient of variation, only $k$ seemed to have an effect, specifically increasing the CV with increasing values of $k$. Since $k$ defines the width of the fitness peak around the optimum value (Eqn 7), this is likely due to added variation from

**Figure 7:** Average percent of optimum possible fitness achieved in edge and core populations after 7 generations of spread through homogeneous and heterogeneous landscapes (left panel). Values of $R$ were calculated from breeding values according to Equation X and compared to the $R_{\text{max}}$ values for homogeneous landscapes and the good quality habitat of heterogeneous landscapes. Variance (right panel) of breeding values for individual fitness in core and edge populations after 7 generations of spread through homogeneous and heterogeneous landscapes. The x axis in both panels shows a gradient of fitness heritability. For these results, heritability of the dispersal trait was fixed at 0.5. Solid lines indicate homogeneous landscapes while dashed lines show heterogeneous landscapes. Red and blue lines correspond to edge and core populations as indicated on the figure.
variability in realized individual fitness values as environmental stochasticity or drift pushes some individuals further from the optimum than occurs for lower values of $k$. As with the mean distance spread, however, the essential relationship between homogeneous and heterogeneous landscapes did not significantly change with parameter value. If any change occurred, in fact, it was to slightly increase the difference in CV values between homogeneous and heterogeneous landscapes (Appendix D).

**Discussion**

Evolution has been shown to play a large role in governing the dynamics of range expansions (Phillips et al. 2010b; Perkins et al. 2013; Phillips 2015; Ochocki & Miller 2017; Weiss-Lehman et al. 2017), but research is just beginning to investigate the interaction between evolution (both spatial and adaptive) and other factors that influence expansion dynamics such as landscape structure and habitat quality (Williams et al. 2016a; Williams et al. 2016b; Gilbert et al. 2017). Here, we showed that landscape heterogeneity results in decreased variation in expansion speed compared to homogeneous landscapes in experimental landscapes. We further demonstrated the potential role of trait evolution in this result using an individual-based model tailored to our experimental system. In particular, our model results suggest that the relative roles of dispersal and fitness evolution could be a key factor driving patterns of variation in expansion speed.

Comparing results from the empirical and theoretical components of our study provides several key insights. First, given the role of fitness evolution in driving differences in mean distance spread in the individual-based model (Fig. 5.3), the equal mean expansion speeds observed in *T. castaneum* suggest that heritability values for fitness in this system may be intermediate to low (e.g. between 0.2 and 0.3). This range is reasonable as it also produces the
observed pattern of higher variation in expansion rates through homogeneous landscapes over a range of heritability values for dispersal (Fig. 5.5). This congruence of the model with the experimental results allows us to use the simulation results to further tease apart the mechanistic underpinnings of these patterns in expansion dynamics.

One key result from the model is the importance of dispersal evolution in driving patterns of expansion across landscape types. Dispersal evolution increases expansion rates in both homogeneous and heterogeneous landscapes but appears to be the dominant process responsible for the increase in heterogeneous landscapes (Fig. 5.2). One reason for this pattern could be that populations in homogeneous landscapes are also undergoing adaptive evolution to a fitness trait optimum (Fig. 5.7). This means that fitness evolution in homogeneous landscapes leads to higher individual fitness, thus increasing expansion speed (Skellam 1951; Hastings et al. 2005), while populations in heterogeneous landscapes are already at the fitness optimum for the good quality habitat. Furthermore, dispersal evolution seems to be primarily responsible for increasing variation in expansion speed in both homogenous and heterogeneous landscapes (Fig. 5.4). However, variation in expansion speed is increased less in heterogeneous landscapes leading to a reduction in variation compared to homogenous landscapes for those parameter values (Fig. 5.5). This indicates that landscape structure interacts with dispersal evolution to shape range expansions. In particular, core populations in heterogeneous landscapes evolved lower values of the dispersal trait on average while edge populations displayed largely equivalent patterns among landscape types. This makes sense as heterogeneous landscapes impose an additional cost of dispersal (i.e. the possibility of dispersing into a patch to which an individual is not adapted) not present in homogeneous landscapes. This cost of dispersal leads to a greater deviation in diffusion coefficients between edge and core populations in heterogeneous landscapes (Fig. 5.6),
likely resulting in less gene flow between core and edge populations which could result in less
variation in evolutionary trajectories at the expansion edge.

Previous work has suggested that gene surfing might increase variation in expansion
speed in addition to dispersal evolution (Weiss-Lehman et al. 2017; Chapter 4). However, results
from the simulation experiment and recent research suggest gene surfing is unlikely to be as
important in heterogeneous landscapes. Gilbert et al. (2017), for example, found that landscape
heterogeneity reduced the buildup of deleterious alleles due to gene surfing. While the
quantitative genetic framework of our model does not allow for the explicit tracking of
individual alleles as done by Gilbert et al. (2017), other signatures of gene surfing can be
detected. For example, in homogeneous landscapes, edge populations appear to lag slightly
behind core populations in adapting to the optimum trait value over intermediate values of fitness
heritability (Fig. 5.7). This could be due to the lack of genetic variation in edge populations,
which is characteristic of gene surfing (Hallatschek & Nelson 2008), hindering adaptation at the
edge except for high levels of heritability which aid adaptation or low levels of heritability which
inhibit adaptation across the landscape. Similarly, the deviation of edge populations from their
phenotypic optimum in heterogeneous landscapes could be at least partially due to gene surfing,
but it is likely not the dominant force behind this pattern. The deviation of edge populations in
heterogeneous landscapes is similar in magnitude to the lag observed in homogeneous
landscapes, but gene surfing is likely to play a larger role in homogeneous landscapes (Gilbert et
al. 2017). Furthermore, since edge populations are small, they are more susceptible to low
amounts of gene flow, for example from individuals with phenotypes more adapted to the poor
quality habitat (Lenormand 2002). This could result in different equilibrium frequencies of
breeding values, with a greater proportion of individuals adapted to poor quality habitat than in
the range core. This hypothesis would also explain why edge populations in heterogeneous landscapes maintain more genetic variation for their fitness trait than edge populations in homogeneous landscapes (Fig. 5.7), a result that agrees with previous experimental work in stationary populations (Huang et al. 2015).

It is critical to understand the sources of variation in range expansions as they can be very important in conservation settings (e.g. invasive species or species responding to climate change). Previous work has suggested that evolution during range expansions will lead to more variable expansion dynamics (Phillips 2015; Ochocki & Miller 2017; Weiss-Lehman et al. 2017). The results presented here support this conclusion, but illustrate an important caveat: spatial evolution during range expansions can interact with other factors to produce counterintuitive effects on the variability of expansion speed. Based on the experimental results alone, we might expect more variability in expansion speed of an invader through pristine, rather than fragmented habitat. However, the model results suggest this prediction should only hold when the role of dispersal evolution is greater than the role of fitness evolution. It is critical to understand the interaction of landscape structure and evolutionary forces in altering the intrinsic variability of range expansions. Whether making predictions about the spread of invasive species through pristine habitats or climate induced range shifts of populations through patchy landscapes, it is important to understand the inherent uncertainty associated with such predictions. The model results demonstrate that, regardless of landscape type, dispersal evolution in particular can still increase the fundamental limit to predictability in range expansions. While we focused on the importance of landscape structure here, other ecological factors will likely interact with spatial and adaptive evolution during range expansion as well. It is crucial for future
conservation work to understand the contributions of interacting evolutionary and ecological factors to the intrinsic uncertainty of range expansions.
CHAPTER 6

Conclusion

In my dissertation, I have demonstrated the critical role that population spatial structure can play in extinctions and range expansions. In extinctions, the spatial occupancy structure of a population can provide important information for estimating extinction risk in a robust and accurate manner. Importantly, this type of information is substantially easier to collect, both in terms of time and monetary investment, than other population metrics of extinction risk such as abundance (Fieberg & Ellner 2000) or genetic diversity (Hedrick 1994). Additionally, spatial occupancy data is particularly amenable to collection via citizen scientist programs (van Strien et al. 2013), meaning the model I derive in Chapter 2 could easily be used in conjunction with large scale citizen science conservation programs to assess extinction risk for species without sufficient available funding for a formal scientific survey.

In range expansions, the spatial population structure formed via the expansion process can lead to unique and rapid evolutionary dynamics with the potential to drastically increase variance in expansion speed (Ochocki & Miller 2017; Weiss-Lehman et al. 2017; Chapter 3). Importantly, I demonstrate in Chapter 4 that both deterministic, adaptive processes and stochastic, neutral processes are important in driving evolution of populations at the expansion edge, thus providing mechanistic insight to the processes driving the increased variance in expansion speeds observed in Chapter 3. However, in Chapter 5, I also showed that these rapid evolutionary mechanisms brought on by the spatial population structure of range expansions can interact with the physical spatial structure of the landscape. This interaction can lead to counterintuitive results, such as a reduction of variance in expansion speed through heterogeneous landscapes. By combining a variety of empirical and theoretical approaches in
Chapters 3-5, I am able to identify the important role of spatial population structure in driving evolution of range expansions, provide mechanistic insight into the underlying processes, and extrapolate to explore the interaction of those processes with other types of spatial structure, such as fragmented spatial habitat structure.

One overarching theme from my dissertation is the critical importance of considering both the ecological and evolutionary consequences of spatial population structure. Ecological and rapid evolutionary processes have the potential to feedback on each other, with ecological processes affecting selection pressures and evolutionary processes changing ecological conditions (Turcotte et al. 2013). The interplay of ecological and evolutionary processes resulting from spatial population structure due to range expansion is a particularly important example of this (Fronhofer & Altermatt 2015). The simple ecological processes of population growth and dispersal lead to a particular spatial population structure characteristic of range expansions (i.e. an expansion wave) (Hastings et al. 2005). This spatial population structure, the outcome of ecological processes, can directly cause rapid evolution of both dispersal and individual fecundity, which in turn feedback to affect the speed of expansion and hence the expansion wave (Ochocki & Miller 2017; Weiss-Lehman et al. 2017; Chapter 3). This feedback between ecological and evolutionary processes during range expansion can, over time, lead to accelerating rates of spread (Phillips et al. 2006; Shine et al. 2011) and dramatically increased variability around an expected outcome (Phillips 2015; Weiss-Lehman et al; 2017, Ochocki & Miller 2017). The importance of these feedbacks between ecological and evolutionary processes in range expansions suggest that other situations defined by distinct patterns of spatial population structure will have similarly important eco-evolutionary feedbacks.
For example, both ecological and evolutionary processes can be important in setting stable range limits for many species (Holt 2003). Theoretical work has shown that purely ecological mechanisms, such as Allee effects and source-sink metapopulation dynamics, can form stable range boundaries (Holt & Keitt 2000; Holt et al. 2005). However, evolutionary mechanisms are also known to cause the formation of stable range boundaries by preventing adaptation to conditions outside the range via genetic linkage of key traits (Moeller et al. 2011) and gene flow of maladapted genotypes from the range core (Kirkpatrick & Barton 1997). Given the potential importance of both ecological and evolutionary mechanisms in forming the spatial structure of stable range limits, it seems likely that feedbacks and interactions between them, as seen for the expansion of a range (Ochocki & Miller 2017; Weiss-Lehman et al. 2017; Chapters 3-5) are important as well.

This suggests that one critical area for future work is understanding the interacting roles of the spatial population structure of a species’ stable range and the spatial structure formed during a range shift in response to climate change. I have demonstrated in my dissertation that if species expand their range in response to a changing climate, evolution due to the resultant expansion wave will likely play an important role. However, the spatial population structure of a species’ original stable range may very well interact with the formation of an expansion wave, depending on the specific ecological and/or evolutionary mechanisms causing the formation of the previously stable range limits. For example, populations with range limits formed by Allee effects may be hindered in the formation of an expansion wave due to the negative impacts of Allee effects on speed in range expansions (Lewis & Kareiva 1993). Alternatively, if a population’s original range limits were formed via source-sink metapopulation dynamics, populations at the range edge could be characterized by low genetic diversity (Eckert et al.
This could lead to more detrimental outcomes due to gene surfing, an inability to evolve heightened dispersal, or both in a subsequent range shift driven by the low diversity edge populations. It will be important to consider both the ecological and evolutionary mechanisms involved in climate induced range shifts to understand and make predictions about the outcomes (Valladares et al. 2014).

Integrating complex processes such as ecological and evolutionary causes and consequences of spatial population structure requires the coupling of experimental and theoretical approaches, as I demonstrate in my dissertation. Well replicated experiments can provide a check on theoretical predictions, as well as provide new directions for research via the discovery of surprising or unintuitive results (e.g. Melbourne & Hastings 2009). However, it is necessary to couple such experiments with appropriate theoretical and statistical models to provide mechanistic insight as well as generality for the results. Thus, it will be important to continue using both empirical and theoretical approaches to test the role of spatial population structure in the extinction and expansion dynamics of populations in a continually changing world.
Bibliography


Elton, C.S. 1958. The ecology of invasions by animals and plants. The University of Chicago Press, Chicago, IL.


Peischl, S., Kirkpatrick, M. & Excoffier, L. 2015. Expansion load and the evolutionary dynamics...


APPENDIX A

Derivation of extinction risk model from Chapter 2

Derivation of the standard likelihood equation

Let \( y_{i,t} \) be the recorded observation (either one or zero) for location \( i \) at time \( t \), \( N \) be the number of locations, \( T \) be the number of time points, and \( k_i \) be the time point of the last observed presence of the species in location \( i \) (i.e. the data point for location \( i \) at time \( k_i \) is one and all subsequent data for patch \( i \) are zeros).

Consider first the likelihood for one location, \( i \), where the species was observed to be present at least one time. In that case, we know that the species had not gone permanently extinct before the time \( k_i \). The probability, \( P(A) \), of the time series of observations up to and including \( k_i \) is

\[
P(A) = \prod_{t=1}^{k_i} (1 - \omega) \phi^{y_{i,t}} (1 - \phi)^{1-y_{i,t}}, \quad \text{(Eqn. S1)}
\]

where \( 1 - \omega \) is the probability that the species did not go permanently extinct from one time to the next, and \( \phi \) and \( 1 - \phi \) are respectively the probabilities of observing or not observing the species if it has not gone permanently extinct at time \( t \). The rest of the time series is a series of zeros. However, it is not possible to know at which time the species may have gone permanently extinct in this series of zeros. It could have happened immediately, the species could have gone unobserved for a period of time before becoming permanently extinct at location \( i \), or it could have not gone permanently extinct but remained unobserved. We must therefore consider each possibility. Since the possibilities are mutually exclusive, their probabilities are additive. For each possible time of permanent extinction at location \( i \), we consider the probability of the species not going extinct but being unobserved until that time. The probability, \( P(B) \), of the time series of zeros after \( k_i \) is therefore:
\[ P(B) = \sum_{e=1}^{T-k_i} (1 - \omega)^{e-1} \omega (1 - \phi)^{e-1} + (1 - \omega)^{T-k_i} (1 - \phi)^{T-k_i}. \]

(Eqn. S2)

In Eqn. S2, \( e \) represents the amount of time after the last successful observation for which the species may have remained unobserved at location \( i \), before going permanently extinct at time \( k_i + e \). The probability for the complete time series, both before and after \( k_i \), includes the probability that the site was occupied at least once, thus

\[ \nu P(A) P(B). \]  

(Eqn. S3)

In this simpler case in which the species was observed to be present at least once, we scale \( P(A) \) and \( P(B) \) by the probability of the focal species occupying location \( i \) at least once (\( \nu \)) since we have evidence the species occurred in this location.

Consider next the likelihood for one location, \( i \), where the species was never observed to be present (i.e. the sum of \( y_{i,t} \) is zero). Then there are two possibilities. Either the species does not occur there, or it occurred there but was never observed. Since these events are mutually exclusive, the probability is

\[ (1 - \nu) + \nu P(B), \]  

(Eqn. S4)

where \( P(B) \) accounts for all the possible ways to observe a series of zeros, as before (Eq. S2). We combine Eq. S2 and Eq. S3 by including the indicator function,

\[ \left\lfloor \sum_{t=1}^{T} y_{i,t} = 0 \right\rfloor, \]  

(Eqn. S5)

which evaluates to one if the time series contains all zeros, or zero if the time series contains at least one observed presence. Multiplying the individual likelihoods for all the sites gives Eq. 1 in the main text.
Derivation of the likelihood equation to account for missing data

To account for missing data, we consider both possible realizations of each missing data point (i.e. zero or one). There are two different situations. Either a missing data point occurs before $k_i$ or it occurs after $k_i$. If a missing data point occurs before $k_i$, then the likelihood is changed to reflect the fact that there is no information on $\phi$ for that time. However, the probability of not going permanently extinct, $1 - \omega$, is still included for that time, since it is known that the species has not gone permanently extinct. If a missing data point occurs after $k_i$, we consider the possibility that the missing data point could have been a one (i.e. an observed presence) and therefore could have changed the time of the last observed presence, $k_i$. This necessitates an extra term that sums the probabilities for each of the different possibilities for $k_i$. These considerations modify Eq. 1 to yield Eq. 2 in the main text.
Figure B1: Average depth profile of the aligned sequencing reads. The shaded grey region corresponds to the depth cutoffs used in our analyses (4 and 22).
<table>
<thead>
<tr>
<th>landscape</th>
<th>generation</th>
<th>treatment</th>
<th>location</th>
<th>Number of bases (Mbp)</th>
<th>Mean Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>After</td>
<td>Before</td>
<td>After</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>landscape</td>
<td>location</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B1: Number of bases and Phred quality score for each population before and after trimming and alignment.
APPENDIX C

Results for the X chromosome from Chapter 4

Figure C1: Mean nucleotide diversity ($\pi$) across the X chromosome calculated in the same way as for autosomes (Fig. 1). The raw datum for each replicate is plotted as a point and the distribution among replicates is shown with a standard Tukey box plot. Structured landscapes are shown in red and shuffled landscapes in blue.
Figure C2: Differences in mean \( \pi \) values for the X chromosome in individual landscapes over 8 generations. Differences were calculated in the same manner as for autosomes (Fig. 2). Structured landscapes are shown in red and shuffled landscapes in blue. Individual data points are shown as well as a mean and 95% confidence interval estimated from a linear model including the effect of location on changes in mean \( \pi \) values.
Figure D1: Results from the sensitivity analysis examining the effect of parameter values for $k$, $\eta$, and $\gamma$ and choice of initial conditions for $V_{p,R}$ and $V_{p,D}$. 500 simulations were run for each
parameter combination and the default values used for the analyses presented in the main text are indicated with vertical dotted lines. Each row corresponds to a different parameter as indicated by axis labels and the mean distance spread after seven generations is shown in the left column and the coefficient of variation in distance spread is shown in the right column. Results for homogeneous landscapes are shown in green and heterogeneous landscapes are shown in purple.