

2008-09-24

# Connectivity and Genetic Structure in Coral Reef Ecosystems: Modeling and Analysis

Johnathan Kool

*University of Miami*, [jkool@rsmas.miami.edu](mailto:jkool@rsmas.miami.edu)

Follow this and additional works at: [https://scholarlyrepository.miami.edu/oa\\_dissertations](https://scholarlyrepository.miami.edu/oa_dissertations)

---

## Recommended Citation

Kool, Johnathan, "Connectivity and Genetic Structure in Coral Reef Ecosystems: Modeling and Analysis" (2008). *Open Access Dissertations*. 157.

[https://scholarlyrepository.miami.edu/oa\\_dissertations/157](https://scholarlyrepository.miami.edu/oa_dissertations/157)

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact [repository.library@miami.edu](mailto:repository.library@miami.edu).

UNIVERSITY OF MIAMI

CONNECTIVITY AND GENETIC STRUCTURE IN CORAL REEF ECOSYSTEMS:  
MODELING AND ANALYSIS

By

Johnathan T. Kool

A DISSERTATION

Submitted to the Faculty  
of the University of Miami  
in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2008

©2008  
Johnathan T. Kool  
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of  
the requirements for the degree of  
Doctor of Philosophy

CONNECTIVITY AND GENETIC STRUCTURE IN CORAL REEF ECOSYSTEMS:  
MODELING AND ANALYSIS

Johnathan T. Kool

Approved:

---

Robert K. Cowen, Ph.D.  
Professor of Marine Biology  
and Fisheries

---

Terri A. Scandura, Ph.D.  
Dean of the Graduate School

---

Claire B. Paris, Ph.D.  
Assistant Professor of  
Applied Marine Physics

---

Douglas L. Crawford, Ph.D.  
Professor of Marine Biology  
and Fisheries

---

John W. McManus, Ph.D.  
Professor of Marine Biology  
and Fisheries

---

Geoff Sutcliffe, Ph.D.  
Associate Professor of  
Computer Science

---

Paul H. Barber, Ph.D.  
Associate Professor of Ecology and  
Evolutionary Biology  
University of California Los Angeles



KOOL, JOHNATHAN T.  
Connectivity and Genetic Structure in  
Coral Reef Ecosystems: Modeling and Analysis

(Ph.D., Marine Biology and Fisheries)  
(December 2008)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Robert K. Cowen.  
No. of pages in text. (137)

This dissertation examines aspects of the relationship between connectivity and the development of genetic structure in subdivided coral reef populations using both simulation and algebraic methods. The first chapter develops an object-oriented, individual based method of simulating the dynamics of genes in subdivided populations. The model is then used to investigate how changes to different components of population structure (e.g., connectivity, birth rate, population size) influence genetic structure through the use of autocorrelation analysis. The autocorrelograms also demonstrate how relationships between populations change at different spatial and temporal scales. The second chapter uses discrete multivariate distributions to model the relationship between connectivity, selection and resource use in subdivided populations. The equations provide a stochastic basis for multiple-niche polymorphism through differential resource use, and the role of scale in changing selective weightings is also considered. The third chapter uses matrix equations to study the expected development of genetic structure among Caribbean coral reefs. The results show an expected break between eastern and western portions of the Caribbean, as well as additional nested structure within the Bahamas, the central Caribbean (Jamaica and the reefs of the Nicaraguan Rise) and the Mesoamerican Barrier Reef. The matrix equations provide an efficient means of

modeling the development of genetic structure in subdivided populations through time. The fourth chapter uses matrix equations to examine the expected development of genetic structure among Southeast Asian coral reefs. Projecting genetic structure reveals an expected unidirectional connection from the South China Sea into the Coral Triangle region via the Sulu Sea. Larvae appear to be restricted from moving back into the South China Sea by a cyclonic gyre in the Sulu Sea. Additional structure is also evident, including distinct clusters within the Philippines, in the vicinity of the Makassar Strait, in the Flores Sea, and near Halmahera and the Banda Sea. The ability to evaluate the expected development of genetic structure over time in subdivided populations offers a number of potential benefits, including the ability to ascertain the expected direction of gene flow, to delineate natural regions of exchange through clustering, or to identify critical areas for conservation or for managing the spread of invasive material via elasticity analysis.

*For my family.*

## ACKNOWLEDGEMENTS

First, I must thank Bob Cowen, for taking me into his lab, giving me the freedom to pursue my own independent research, and for providing valuable advice whenever it was asked for. I also thank Claire Paris for helping me to grow both as a scholar and as a person. Her experience and ideas are the foundation of this dissertation. Thank you to Doug Crawford, who spurred me to think critically and creatively, and was always willing to listen to strange musings. I am also grateful to John McManus, who first brought me to Miami, and introduced me to individual-based models. They have had a profound influence on me since, and my research would not have been possible without his expertise. Thank you to Geoff Sutcliffe, who bravely crossed over from the sensible world of computer science into the strangeness of marine biology – his perspectives were both refreshing and insightful. I am also grateful to Paul Barber, who provided me with excellent advice as well as a window on Southeast Asia from my seat behind a computer in Miami.

Throughout my time here, I was supported by a number of students, many of whom have become very close friends. The members of the Cowen and Sponaugle labs have been a constant source of encouragement, and my especial thanks go to the original NCORE lab members, Marilyn Brandt, Alette Yñiguez and Wade Cooper who were there at the beginning, end and all points in-between.

I must also acknowledge several individuals, whose professional contributions and advice were invaluable – Subramanian Ramakrishnan for patiently explaining

elements of probability and matrix algebra, and helping to develop many of the mathematical proofs contained in the dissertation; Serge Andréfouët and Christine Kranenburg for providing Millennium Coral Reefs data in advance of its public release; and Klaus Meffert for his advice and development of JGAP.

But most of all, I would like to say thank you to my family, who were unwavering in their support and love, and provided me with the strength needed to see the dissertation through. Without them, this would only have remained a dream.

## TABLE OF CONTENTS

	Page
LIST OF FIGURES .....	vii
LIST OF TABLES .....	ix
 Chapter	
1 BACKGROUND AND LITERATURE REVIEW .....	1
2 AN OBJECT-ORIENTED, INDIVIDUAL-BASED APPROACH FOR SIMULATING THE DYNAMICS OF GENES IN SUBDIVIDED POPULATIONS .....	12
3 CONNECTIVITY, SELECTION AND RESOURCE USE IN SUBDIVIDED POPULATIONS – THEORY AND IMPLICATIONS.....	41
4 SIMULATING THE GENETIC STRUCTURE OF CARIBBEAN CORAL REEF ECOSYSTEMS.....	63
5 SIMULATING THE GENETIC STRUCTURE OF SOUTHEAST ASIAN CORAL REEF ECOSYSTEMS .....	82
6 SYNTHESIS AND CONCLUSIONS .....	102
 Appendices	
A STANDARD BAR-DOT NOTATION .....	117
B PROOF OF THE GENETIC PROJECTION FORMULA .....	118
C PROOF OF THE EQUIVALENCE OF HAPLOID AND DIPLOID ALLELE FREQUENCIES ARISING FROM RANDOM SAMPLING UNDER THE ASSUMPTION OF RANDOM MATING.....	120
D DISCRETE MULTIVARIATE DISTRIBUTIONS .....	122
E DEVELOPMENT OF THE RESOURCE-WEIGHTED VERSION OF WALLENIUS’ NON-CENTRAL HYPERGEOMETRIC DISTRIBUTION .....	123
References .....	124

## LIST OF FIGURES

### Chapter 2

2.1	Class diagram of the object-oriented, individual-based modeling framework .....	32
2.2	Construction of space-time autocorrelograms .....	33
2.3	Distance input matrices .....	35
2.4	Transition matrices .....	36
2.5	Space-time genetic structure plots (autocorellograms) for the simulations	38

### Chapter 3

3.1	Twelve populations and their respective allele frequencies connected by a circulant migration pattern .....	57
3.2	Illustration of time-frequency plot construction .....	58
3.3	Allele frequency probability plots for different levels of stepping stone migration and mortality models .....	59
3.4	Twelve populations and their respective allele frequencies connected by a circulant migration pattern, with the addition of an exterior circle showing relative proportion of available resources .....	60
3.5	Allele frequency probability plots for 13 populations with a circulant migration structure, altering proportional resource availability .....	61
3.6	A representation of how selective weightings may change according to scale .....	62

### Chapter 4

4.1	Map of derived polygon boundaries for the Caribbean divided into subregions .....	78
4.2	Transition matrix and projections .....	79
4.3	Similarity matrices resulting from random sampling .....	81

## Chapter 5

5.1	Map of derived polygon boundaries for Southeast Asia divided into subregions .....	96
5.2	Transition matrix, projections and derived results.....	97
5.3	Kernel density plots for selected regions in Southeast Asia.....	99
5.4	UPGMA clustering of the Similarity matrix at $t=100$ .....	100
5.5	Quantiled expected diversity values for Southeast Asian coral reef communities.....	101

## Chapter 6

6.1	Schematic representation of life cycle stages, associated processes and mathematical transformations.....	115
6.2	Cattell's Data Box and associated modes .....	116



## LIST OF TABLES

### Chapter 2

2.1	Input values for the simulation runs.....	31
-----	---	----

## **CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW**

### **Background**

Developing an understanding of population genetic structure and the forces that shape it is of considerable importance for both conservation and research. Genetic diversity provides the raw material for species diversification, and is an area of concern for managers when designing marine reserves and networks. From a research perspective, studying population genetic structure provides insight into evolutionary and biogeographic processes and the forces that shape them. There have been a growing number of studies of population genetic structure in marine environments, and there is clear evidence of genetic patterns at the regional scale (Benzie 1992; Benzie and Williams 1995; Doherty, Planes, and Mather 1995; Palumbi 1997; Planes and Fauvelot 2002; Taylor and Hellberg 2003) Questions remain, however, regarding the specific processes responsible for generating the observed structure.

Connectivity (the degree of exchange between subpopulations) is believed to play a critical role in generating genetic structure at the regional level, and has been identified as one of the most critical areas of research for understanding and managing populations of marine organisms (Sale et al. 2005). This has been reflected by a rapidly growing interest in probing the “black box” of the process of larval transport, and the factors that affect it (Underwood and Keough 2001; Armsworth 2002; Cowen 2002; Leis and Carson-Ewart 2002; Thorrold et al. 2002; Warner and Cowen 2002; Kinlan and Gaines 2003; Sale and Kritzer 2003). Other forces could also play a significant role in generating regional-scale genetic structure. For example, metapopulations with

high levels of productivity have the potential to contribute unequally to the gene pool (Hedgecock 1994), or evolutionary time-scale processes may have lasting influence on the system (McManus 1985; Benzie 1999). Moreover, multiple factors may be acting in concert with one another in a complex manner. The degree of influence of these various forces on population genetic structure and how they interact with one another is currently not well understood.

One of the greatest obstacles in identifying the roots of genetic structuring in marine environments is the scale of the system. Marine organisms produce large quantities of larval young which have the potential to disperse significant distances, and are typically subject to extremely high levels of mortality (Boehlert 1996). The logistical difficulties associated with physically tracking marine larvae are not trivial. Although new techniques continue to be developed for monitoring the fate of marine larvae (otolith elemental analysis - Campana 1999, dye experiments - Kingsford unpublished; e.g. chemical labeling Jones et al. 1999; drifters - Domeier 2004), presently, empirical large-scale larval tracking remains an elusive prospect.

As an alternative, researchers developed computer-based simulations that couple oceanographic models to models of larval dispersal in order to investigate the influence of physical processes on complex patterns of larval recruitment (Wolanski 2001; Armsworth 2002; James et al. 2002; Paris et al. 2002). These kinds of models have considerable value because of their ability to evaluate the outcome of different ecological scenarios at large scales, a task which would be logistically impossible to replicate experimentally.

With the development of the new dispersal models there is an unprecedented opportunity to quantitatively test various theories regarding the processes governing regional-scale genetic structuring. The dispersal simulations can be used in conjunction with quantitative population models to show the creation and flow of genes at ecosystem scales. The key is to develop a framework in which it is possible to integrate aspects of migration, demographics, behavior and genetic processes.

## **History**

Pioneered by the work of Fisher, Haldane and Wright (Fisher 1922, 1930; Wright 1931; Haldane 1932; Wright 1937; Haldane 1939; Wright 1943) population genetics is the key field for studies of the dynamics of genes within populations. Population geneticists have developed a number of equations and computer simulations to quantify the flow of genetic material through multiple generations. Some important issues still require attention however. The first is the influence of complex spatial connectivity (i.e. migration) on population genetic structure. Traditionally, the method of assessing the effects of migration on population genetic structure is to apply Wright's "Island Model" (1931). Wright's model makes the assumption that all populations are connected to one another, and that all have an equal degree of connectivity. In most natural systems however, this assumption is grossly violated (Epperson 2003; Rice 2004). Although there have been attempts to incorporate weighted migration effects (Malécot 1969; Wright 1969; Endler 1977; Maruyama 1977; Epperson 2003), algebraic analysis of spatially explicit population genetic processes can be complex. A second issue stems

from the fact that population geneticists place emphasis on investigating genetically-driven processes as opposed to demographically-driven processes, and in many cases, omit the role of fluctuating population sizes. For instance, in some models it is assumed that processes of interest are taking place over a long period of time (i.e. evolutionary scale), and therefore the state of a population can be regarded as being a relatively transitory phenomenon (Nei 1987). Nei also wrote however that genetic polymorphism and long-term evolution are merely two different phases of the same evolutionary process (Nei 1987). If patterns of genetic polymorphism can be influenced by short-term population dynamics, one might reasonably expect that the effects of some short-term processes have the potential to cascade out and have impact on longer time scales (Slatkin 1977; Wade and McCauley 1988; Whitlock and McCauley 1990; Whitlock 1992).

The role of connectivity has been addressed in population dynamics studies, particularly those relating to population viability analysis (PVA - Beissinger 2002) and metapopulation analysis (Levins 1969; Hanski and Gilpin 1997; Hanski and Gaggiotti 2004). PVA chiefly concerns the assessment of whether or not a population or set of populations will persist through time, and there are several existing PVA software packages that incorporate aspects of spatial structure (Lacy 1993; Kingston 1995; Possingham and Davies 1995). However, any genetic characteristics of populations are only incorporated in the form of inbreeding depression, if at all (Groom and Pascual 1998). There is a growing body of research regarding the genetic structure of metapopulations (Whitlock 1992; Gaggiotti 1996; Gaggiotti and Smouse 1996; Whitlock and Barton 1997; Rousset 1999, 1999). As with population genetics studies however,

research tends to focus on a single measure of the system (e.g. effective population size - Whitlock and Barton 1997), or else changes to a single population (e.g. Gaggiotti and Smouse 1996).

Considerable work has gone into the development of equations based on population genetic theory, but assembling a treatment of the system where it is possible to examine the interplay of different factors and the scales at which they exert their influence poses some challenges. Analytic treatment of the problem leads to equations that tend to be tightly focused in terms of their scope, or alternatively require a sophisticated degree of mathematical understanding for both use and interpretation. Gaps exist between the complexity of real-world systems and what is mathematically tractable. Part of the difficulty lies in the methods that are used to describe the system. Many of the analytical techniques commonly used are designed for deterministic systems. As stated by the population biologist Richard Lewontin however – “Biology, especially population biology is neither particle physics nor solar system astronomy. Organisms and their populations are a nexus of a large number of individually weakly determining interacting forces. One consequence is that different cases have different dynamics and that simple general functional forms may miss the important action.” He goes on to state that “Population biology can only be built by breaking down the distinction between the study of population processes and the study of individual properties” (Lewontin 2004).

Recently, there has been growing interest in the development of individual-based models (IBMs), also known as agent-based models (ABMs – although more correctly, IBMs are a specific type of ABM) or object-based models (OBMs). Individual-based modeling makes use of object-oriented programming techniques to define objects or

“agents” with particular attributes and behavior, allowing for variation among individual objects. The individual agents can then interact with one another within a given system, giving rise to emergent large-scale processes. IBMs have garnered attention because of their ability to simulate complex behavior of systems through the application of simple underlying rules (first principles). In addition, they do not require as many assumptions as more classical modeling frameworks. For example, agent based models do not require homogeneity across the domain of study, and are capable of incorporating demographic stochasticity as well as rare and conditional events within the system (DeAngelis and Gross 1992). In addition, IBMs can be coupled with Geographic Information Systems (GIS) in order to define a virtual environment that closely resembles real-world conditions. Individual-based models provide an ideal environment for investigating complex, spatially explicit population genetic systems.

Individual-based models are limited in an important respect however, in that they can be computationally intensive. If the system is spatially and temporally extensive, it may be necessary to simulate billions or trillions of individual objects which, even given large increases in computing power, may be difficult to implement. If simplifying assumptions can be made however, multivariate stochastic theory can be applied to greatly improve the efficiency of the model. Structured systems of stochastic branching processes in turn lead to the development of matrix-based models (Caswell 2001). The two approaches are complementary, the choice between them depending on the level of realism desired from the models, and the degree to which simplifying assumptions hold. The power of individual-based models lies in their flexibility of design, whereas the

power of matrix-based models lies in their efficiency and ability to be manipulated algebraically.

## **Objectives**

The purpose of the dissertation is to determine the relationship between aspects of connectivity and genetic structure, towards the goal of identifying the processes that create and maintain genetic structure in subdivided populations.

To investigate the dynamics of systems in which population genetics, connectivity and population demographics interact in a dynamic manner, a theoretical framework for studying the dynamics of gene flow in marine metapopulations is developed, including both agent-based and analytical components. Connectivity between populations is obtained by applying a dispersal-recruitment algorithm to the agents driven by a combination of physical oceanographic information and individual behavior.

Using the results of the models, it is possible to:

- 1) Investigate the expected spatial and temporal scale of genetic transport within coral reef ecosystems, as well as the relative importance of demographic parameters (birth rates, mortality rates, carrying capacities) and metapopulation structure (configuration, strength of migration).



- 2) Incorporate the effects of selection, and study how it interacts with migration to create structure.
- 3) Evaluate the form, extent and spread of genetic material through a metapopulation framework by conducting simulated genetic tracer experiments.
- 4) Examine differences between model results and empirical observations. This will address the relative importance of neutral processes within the system.

## **Outline**

The individual chapters are organized as follows:

Chapter 2: An individual-based simulation approach is developed for modeling the dynamics of genes in subdivided populations. The model is then used to examine how changes to basic aspects of connectivity structure lead to changes in genetic structure.

Chapter 3: Discrete multivariate distributions are used to study how selection, migration and resource availability operate in concert to maintain genetic polymorphism in subdivided populations within a stochastic context.

Chapter 4: Matrix-based projections of genetic structure are developed for Caribbean coral reef ecosystems, and elasticity analysis is used to identify populations that would have the greatest degree of influence when subjected to perturbations.

Chapter 5: Matrix-based projections of genetic structure are developed for Southeast Asian coral reef ecosystems. The directionality of the matrices and derived genetic distances are used to identify likely sources of genetic structure and diversity in the region

### **Significance**

Knowing the functional genetic consequences of different connectivity patterns is essential for a number of reasons, particularly with regard to designing marine protected areas (MPAs). While it may be possible to set up a system of MPAs that maintains a sufficient population size, the genetic effects of implementing such a system may not be as clearly defined. Protected area locations should be selected that in the long run avoid preserving extensive tracts of homogenous genetic material. Also, if populations are highly segregated, they will be more vulnerable to extinction and over time may become genetically isolated and inbred. The models will also be able to identify “keystone” metapopulations – those that play a critical role in maintaining genetic structure, acting as an important connector, or as a major source of genetic diversity.

Another important practical application of the research will come from evaluating the potential spread of invasive genes. Frequently, as part of restoration efforts, populations are supplemented using individuals foreign to the environment. This can generate significant changes in genetic structure, as was recently discovered by a recent study that showed that terrapin turtles from northern Atlantic populations (South Carolina to New York) were more closely related Texas populations than those from Florida, presumably due to transplants that took place in the early 20<sup>th</sup> century (Hauswaldt and Glenn 2005). The models that were developed are able to generate expected trajectories of this kind of genetic mixing.

There has been a longstanding debate regarding the relative importance of neutral processes in evolution versus selection. The Neutral Theory of Genetic Evolution was put forward by Kimura (Kimura 1983) who proposed that the majority of mutations are expected to have a relatively insignificant effect on organisms and therefore gene frequencies in populations are not determined by natural selection, but rather by a balance between the effects of mutation and random genetic drift. Assuming that connectivity of coral reef fish populations is primarily physically driven (in conjunction with behavior), genetic structure should be resolved to some degree using realistic demographic parameters in conjunction with an oceanographically-derived connectivity matrix. If the model results do not match the expected results, this would be a positive indication that processes other than neutral and migratory ones are of significance, such as selection.

Under the proposed framework, three major ecological components will be integrated: demographics, connectivity and genetics. However, other areas remain that have the potential to significantly influence the system, such as trophodynamic processes,

habitat, larval energetics and interspecific interactions (including predation). Using either agent-based or analytical approaches, it is possible to incorporate these kinds of processes as well. The level of simplicity or complexity of the model can be adjusted to a level suitable for studying the problem at hand.

## **Summary**

Understanding the relationship between connectivity and genetic structure is important for conservation management and scientific understanding of evolutionary processes. Combining bio-oceanographic larval dispersal models with population genetic theory presents a new opportunity for studying the dispersal of genetic material through time and the development of genetic structure over large spatial and temporal scales. Object-oriented individual-based simulation as well as algebraic methods will be developed and used for this purpose, and applied to identify the manner in which various aspects of connectivity affect genetic structure, and how genes are expected to spread in real-world environments.

## **CHAPTER 2: AN OBJECT-ORIENTED, INDIVIDUAL-BASED APPROACH FOR SIMULATING THE DYNAMICS OF GENES IN SUBDIVIDED POPULATIONS**

### **Background**

Preservation of genetic diversity is an important consideration when managing populations (Chesser, Olin E. Rhodes, and Smith 1996; Avise 1998; Frankham, Ballou, and Briscoe 2002). High levels of genetic diversity are expected to confer a greater degree of resilience to environmental perturbations (Amos and Balmford 2001; Reed and Frankham 2003; Hughes and Stachowicz 2004), and a diverse gene pool provides fuel for evolutionary processes (Amos and Harwood 1998). Understanding how genetic structure is created and maintained takes on even greater significance in light of increasing concerns regarding habitat fragmentation (Andren 1994; Young, Boyle, and Brown 1996; Fahrig 2002), management of invasive organisms (Huxel 1999; Clavero and Garcia-Berthou 2005) and potential habitat shifts due to climate change (McCarty 2001; McLaughlin et al. 2002; Hughes et al. 2003).

Identifying the processes responsible for creating and maintaining genetic structure can be challenging. Studies of a variety of organisms have demonstrated genetic structure at extensive scales (Benzie et al. 1994; Barber et al. 2002; Baums, Miller, and Hellberg 2005; Hauswaldt and Glenn 2005; Purcell et al. 2006), and empirical studies of large spatial extents with a fine degree of resolution may not be feasible, particularly when dealing with multiple species. Temporal scaling is another important consideration. Changes in genetic structure may result from shifting

environmental conditions or may simply be a consequence of the amount of time required for material to diffuse through the system. Demographic processes may also play an important role; levels of fecundity, mortality, immigration and emigration alter the degree of genetic exchange between populations and are capable of feeding back into one another (e.g. through density-dependent interactions) (Milligan, Leebensmack, and Strand 1994; Aars and Ims 2000). The dynamics of genes in population networks (metapopulations - Levins 1969; Hanski and Gaggiotti 2004) can be extremely complex, especially in areas with extensive migration. Although many genetic models incorporating migration have been developed (e.g. - Bodmer and Cavalli-Sforza 1968; Maruyama 1977; Nagylaki 1996; Fu, Gelfand, and Holsinger 2006), simplifying assumptions are often used to ensure analytic tractability, only simple connectivity structures are considered, or else the focus is on a very specific aspect of the system, leaving out potentially important interacting factors. A means of studying complex interactions in population genetic networks at various spatial scales in a simple, yet flexible manner is needed.

Individual-based models (IBMs) provide an effective means of studying complex systems (Doligez, Baril, and Joly 1998; Bousquet and Le Page 2004; Grimm et al. 2005; Breckling, Middelhoff, and Reuter 2006). IBMs operate by programmatically defining objects and assigning particular properties or behavior to them (Grimm and Railsback 2005). Interactions between individual agents then give rise to emergent large-scale processes. While the use of simulation for studying genetic systems is not new, object-oriented individual based modeling does provide some key advantages. The first is that it does not require homogeneity of the system; members can exhibit different properties and

behavior according to their unique environment. IBMs also tend to be easier to parameterize, because input values correspond to the characteristics of individuals, the level at which data collection typically takes place. Finally, abstraction and implementation (described below) allow for different components of the model to be altered without having to redesign the entire system.

The purpose of this chapter is twofold. First – to develop an object-oriented framework for modeling spatially explicit population genetic processes, and second – to demonstrate how the approach can be applied by evaluating the relationship between fundamental aspects of connectivity structure (e.g. number of connections, strength of connections, number of populations, configuration etc.) and genetic structure. The latter goal is accomplished by generating and comparing Moran's *I*-based spatial and temporal correlation surfaces. In order to connect the approach with existing theory, algebraic formulae describing the expected trajectory of the system are also developed and discussed.

## **Model Development**

The modeling framework (Fig. 2.1) was constructed using the Java programming language (v. 1.5) (Sun Microsystems 2006) and is built upon the development of the *Organism* interface (italicized words refer to program objects). An interface is an abstract class; a class is a program object containing variables (properties) and methods (behavior). Abstraction provides the ability to generically specify object behavior, while leaving the detail as to how they are performed to a separate implementing class. This is

an important point as it means that different implementations may be developed and swapped into the model independently of other components. This allows for flexible customization of the model system. *Organism* encapsulates the properties and behavior of an individual member of a species (e.g. a single plant or animal). In the sample implementation, each *Organism* has a unique identifier, sex, method of mating and *Genotype*. A *Genotype* is composed of *Chromosome* pairs containing *Genes* (JGAP 2007) and each gene acts as a container for particular allele values. Note that the appellation “*Gene*” is related to the class’ origin as part of a genetic algorithms package. Ideally, *Gene* should contain one or more instances of a *Locus* class which would in turn be used to contain the allele values, however this was not adopted here since it would require extensive modifications to the existing JGAP classes. Instead, different loci can be implemented as different *Genes*. Allele values can take different forms, for example, boolean (true/false) values in the case simple dominant-recessive allele types or integers for microsatellite sequences. Collections of interbreeding *Organisms* form discrete *Populations*, and a set of *Populations* make up a *Metapopulation*.

Members of individual *Populations* (the *Organisms*) mate with one another using the *Reproduction* interface. Reproduction was implemented using a *MonogamousSexualReproduction* class, which pairs organisms with a single member of the opposite sex at each time step (i.e. no overlapping generations) on a random basis. Each pairing generates progeny according to an exact number (or a distribution if desired). Each parent produces a gamete, which is generated through crossover and independent segregation of *Chromosome* pairs. Although mutation is an important part of population genetic processes, it is not included in the simulations performed here.



Although it is possible for many mutations to take place during the course of a given simulation, the probability of their survival is relatively low ( $1/2N$  assuming a diploid population, where  $N$  is the number of organisms in the population). Furthermore, the results are used to identify relative differences between populations. If there is no spatial or temporal bias in the mutation rate, then the probability of a new mutation arising will be uniform across all populations, leaving the relative patterns unchanged. Selection was also not addressed here, since most of the molecular markers commonly used in population genetics studies are assumed to be neutral.

Once progeny have been produced for each population, they recruit back to their original population or migrate to a different one according to a given type of *Movement*. *Movement* was implemented using a transition probability matrix (connectivity matrix) where columns of the matrix represent source populations, rows represent destination populations, and cell contents describe the probability with which a given *Organism* migrates from a source population to a destination. Elements along the diagonal of the matrix correspond to levels of self recruitment within a population. Mathematically, this can be represented by  $\mathbf{y}_{t+1} = \mathbf{M}\mathbf{y}_t$  where  $\mathbf{M}$  is a matrix of transition probabilities between populations and  $\mathbf{y}$  is a column vector containing the collection of *Organisms* eligible for migration at each location at a given time  $t$ . The structure of the connectivity matrix can be manipulated in a variety of ways, including expanding the neighborhood of distribution or altering the strength of connections. This approach is similar to the one used in other existing transition matrix-based models (Bodmer and Cavalli-Sforza 1968; Fu, Gelfand, and Holsinger 2006), though here there is a difference in that *Organisms* are transported, not simply gene frequencies. This is a key distinction - it means that in

addition to the genetic information, the characteristics and behavior of the genetic material's host are also retained (e.g. age, condition, parentage). *Mortality* was implemented at each time interval by randomly re-sampling a set number of individuals from each population following reproduction. Time was considered to be discrete, and at each step results generated by the model were saved to a database. The Java-based Colt package (CERN 2004) was used for generating random numbers.

## **Simulations**

Simulations were carried out using the model implementation in order to evaluate how changes in connectivity structure or demographic parameters in the model generate changes in spatial and temporal genetic structure. As a reference simulation, 10 populations were constructed in a ring formation (1-dimensional stepping-stone), with 50 male members and 50 female members within each population. Exactly ten progeny were produced by each pairing, and populations were randomly culled to maintain exactly 100 individuals in each population while iterating through the model for 144 generations. All transitioning members were assumed to arrive at their intended destination (i.e. no transitional mortality). For all simulations, initial members of populations were assigned genetic values corresponding to their population of origin (i.e. *Organisms* in population 1 at the start of the simulation were assigned *Genes* with a value of 1). Although only the dynamics of a single *Gene* are considered here, it is possible to investigate multiple *Genes* and their interactions by adding multiple versions to a given *Chromosome*. The populations were not structured according to age; all members had an

equal probability of mortality and each simulation was run 100 times. The input parameters for the different simulations are summarized in Table 2.1.

### Space-time autocorrelation surfaces

One way of identifying population genetic structure is through the use of autocorrelation statistics (Sokal and Wartenberg 1983; Sokal, Jacquez, and Wooten 1989; Slatkin and Arter 1991; Neigel 1997). Autocorrelation examines the relationships between different objects and their descriptors with respect to a given distance measure. One of the more commonly used metrics is Moran's  $I$  (Cliff and Ord 1981; Barbujani 1987; Epperson and Li 1996). Moran's  $I$  is a product-moment correlation defined by the equation

$$I(d) = \frac{\frac{1}{W} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{n} \sum_{j=1}^n (y_j - \bar{y})^2} \quad \text{where } i \neq j \quad (\text{Eq. 1}).$$

$W$  is the total number of connections between populations,  $n$  is the total number of populations,  $w_{ij}$  is a matrix of connection strengths between populations  $i$  and  $j$ , and  $y$  is a vector of observed values. In this case, the values are the frequencies of occurrence of a particular allele type over all model runs. If measuring spatial autocorrelation of multiple loci, the  $R_G$  statistic developed by Smouse and Peakall (1999) might also be used. Distance classes correspond to the number of joins traversed between populations, although alternative distance measurements may also be used. It is important to make a distinction between the connectivity matrix and the distance matrix between populations. The connectivity matrix ( $M$  in Table 2.1) describes the degree of migration taking place

between populations, whereas the distance matrix ( $D$  in Table 2.1) establishes the relative position of the populations according to a particular measure, such as geographic distance. Although the two are expected to be correlated, this is not a given. Corrections to the Moran's  $I$  values were made to account for small numbers of populations and confidence intervals were also determined (Cliff and Ord 1981). Distance-based correlograms were constructed for each allele at each time step using the Moran's  $I$  metric over non-accumulative distance classes (e.g. 1 to 2 and 2 to 3, as opposed to  $\leq 2$  and  $\leq 3$ ), and then the correlograms at each time step were then combined to create a space-time correlogram surface (Fig. 2.2). The locations at which the correlogram intercepts the upper and lower confidence limit are shown, resulting in outlines of peaks or troughs of significant space-time autocorrelation. All calculations were carried out using Matlab (The Mathworks 2007).

A unique advantage of the autocorrelation surface is that it provides a concise means of showing the relationship between the various populations at different spatial and temporal scales. The  $I$  values indicate the degree of relatedness between populations at the specified time and distance interval, where values near to 1 represent populations with similar values, and values close to -1 represent populations dissimilar in values (n.b. if extreme  $y$  values are heavily weighted, it is possible for  $I(d)$  values to exceed -1 and 1). Although the autocorrelation values for the different alleles were averaged in order to provide an overall measure of the genetic structure of the system, it is also possible to generate separate autocorrelation surfaces for each allele form.

### Algebraic formulation of the model implementation

If the simulation model is implemented as described above, it is also possible to provide equations describing the expected trajectory of the system. Although Moran's  $I$  typically takes the form shown in eq.1, it can also be expressed using vector-matrix notation as

$$I(d) = \frac{\mathbf{z}'\mathbf{W}\mathbf{z}}{\mathbf{z}'\mathbf{z}} \quad (\text{Eq. 2})$$

where  $\mathbf{z}$  is the centered vector of the observed variables (i.e.  $\mathbf{y} - \bar{\mathbf{y}}$ ) and  $\mathbf{W}$  is the connection matrix between populations ( $w_{ij}$  in eq. 1). The  $n$  and  $W$  terms are folded into  $\mathbf{W}$  by multiplying and dividing  $w_{ij}$  by these values respectively.

The expected population structure  $\mathbf{Q}$  at the after each time step is given by

$$\mathbf{Q}_{t+1} = \mathbf{K}(\overline{\mathbf{MBQ}_t + \mathbf{Q}_t})_i \quad (\text{Eq. 3 - See Appendix A for bar-dot notation}),$$

where  $\mathbf{K}$  is a diagonal matrix of population carrying capacities, and  $\mathbf{B}$  is a diagonal matrix of birth rates for each population. The equation formalizes that the state of the system at  $t+1$  is the result of migrating progeny ( $\mathbf{MBQ}_t$ ) combined with the original parent generation ( $+\mathbf{Q}_t$ ), converted to probabilities via row standardization (bar-dot), and multiplied by carrying capacity ( $\mathbf{K}$ ). If  $\mathbf{M}, \mathbf{B}$  and  $\mathbf{K}$  can be assumed to be stationary, then

$$\mathbf{Q}_t = \mathbf{K}(\overline{\mathbf{AK}})_i^{t-1} \overline{\mathbf{A}\mathbf{Q}_{0,i}} \quad \text{where } \mathbf{A} = \overline{\mathbf{MB} + \mathbf{I}}. \quad (\text{Eq. 4 - See Appendix B for proof})$$

where  $\mathbf{I}$  is an identity matrix having the same dimensions as  $\mathbf{M}$  and  $\mathbf{B}$ .

To use this relationship in conjunction with Moran's  $I$ , the vector form of the equation needs to be modified to use matrices as opposed to vectors. A partitioned matrix  $\mathbf{Q}$  can be defined as the set of  $\mathbf{z}$  vectors such that

$$\mathbf{Q} = \{z_1, z_2 \dots z_\alpha\} \quad (\text{Eq. 6})$$

where  $\alpha$  is the number of different allele types in the population, indexed by  $h$ , and therefore

$$I(d)_h = \frac{\mathbf{Q}_h' \mathbf{W} \mathbf{Q}_h}{\mathbf{Q}_h' \mathbf{Q}_h} \quad (\text{Eq. 7}).$$

The formulas integrate several basic aspects of population genetics: genetic structure ( $I(d)$ ), demographic structure ( $\mathbf{Q}$ ,  $\mathbf{K}$  and  $\mathbf{B}$ ), migration ( $\mathbf{M}$ ), and spatial structure ( $\mathbf{W}$ ).  $\mathbf{Q}$  describes the state of the system in terms of observations on a given variable (e.g. allele counts),  $\mathbf{K}$  contains information on carrying capacity and  $\mathbf{B}$  captures fecundity.  $\mathbf{M}$  describes the degree of migration taking place between populations.  $\mathbf{W}$  provides information on the relative positions of the different populations according to a specified distance. This distance measure may be a linear measurement (e.g. geographic distance) or non-linear (e.g. distance according to flow). Folding the number of connections ( $W$  in eq. 1) into  $w_{ij}$  and dividing by the total number of populations ( $n$ ) standardizes the  $\mathbf{W}$  matrix.

In the form of a diagram (i.e. Fig. 2.2),  $\mathbf{Q}$  represents the pie charts and their fractional content,  $\mathbf{M}$  represents the arrows indicating movement between populations, and  $\mathbf{W}$  describes how the different populations are spatially arranged. The terms of the algebraic equation also map to individual classes of the simulation model.  $\mathbf{M}$  corresponds to the *Movement* interface,  $\mathbf{W}$  is accounted for by the *Habitat* interface, and the matrix  $\mathbf{Q}$  corresponds to *Metapopulation*, which contains *Populations* (row vectors) which in turn consist of *Organisms* and their genetic complement (entry counts). The  $\mathbf{B}$  matrix and the  $\mathbf{K}$  matrix map to the *Reproduction* and *Mortality* classes respectively.

The formulas provide the expected outcome, but with any stochastic process there is also a variance component. Finding an analytical solution for the variance structure of the system can be difficult. Pollard's method (1969, Caswell 2002) can be used to project the variance structure of the system under conditions where the transition matrix is stationary through time, but in cases where the conditions change in a complex, nonlinear manner, numerical methods (i.e. simulation) provide the only practical means of evaluating the covariance structure of the system.

## Results

If populations are completely isolated from one another, the connectivity matrix takes the form of an identity matrix and all progeny produced return to their sources of origin. If individual populations begin with a mixture of different allele types, each will independently drift towards fixation of a single allele type with a probability equal to its initial frequency, assuming selective neutrality (Nei 1987). If all populations are initially in a state of fixation and no allele forms are shared between them, no relationship will exist between the populations, and the autocorrelation structure will be undefined (because relationships between populations and themselves are not considered, all values of  $y$  will effectively be 0). Populations that are completely and evenly interconnected will have a connectivity matrix where all values are equal (the transition probability will be  $1/n$ , where  $n$  is the number of populations). Although it is technically possible for the autocorrelation value to be undefined if all values are exactly the same (the expected values of  $y_i$  and  $y_j$  would equal  $\bar{y}$ ), it is more likely that there would be some variance in

the  $y$  terms and consequently  $I(d)$  would tend towards zero. Here too, populations will eventually drift towards fixation of a single allele type, although since the populations are equally and evenly connected, they will function as a single population. As a single unit, the populations will have a higher effective population size, and consequently the time to fixation will be much longer.

More meaningful patterns are produced when migration between populations is restricted. In the reference simulation (Fig. 2.5.1a), negative autocorrelation is initially more prevalent, but positive autocorrelation in the system quickly increases until a state of quasi-equilibrium is reached, with positive autocorrelation values at distances less than 2, negative values at distances greater than 3 and 0 at a distance of 2.5. This is to be expected of a 10-population ring; opposite sides (i.e. distance of 5) are expected to show the greatest differences in composition, and the relationships would be balanced at mid-distance. Eventually the patterns become chaotic as the populations in the system begin to resemble one another and spatial structure dissolves into stochastic noise. The jaggedness in the lines of significance is the result of variability in the degree of variance of the autocorrelation values, and would be reduced by a greater number of simulation replicates.

Decreasing the relative strength of connections between populations extended the space-time autocorrelation pattern along the time axis, indicating a lengthening the amount of time required for genetic material to diffuse through the system (Fig. 2.5.1a, b, c). In contrast, increasing the range at which populations were connected (Fig. 2.5.2a, b, c relative to 2.5.1a,b,c) caused the autocorrelation pattern to compress along the time



axis, meaning that although the spatial relationships between populations at each time interval remained the same, the overall duration of the process was shortened.

With a greater number of populations, the autocorrelation pattern stretched along the time axis (Fig 2.5.3a,b,c), and the confidence interval also narrowed. This is the result of effectively increasing the sample size. By sampling a greater number of objects (in this case, populations) the overall variance in the system decreases. Increasing the fecundity of members equally for all populations did not appear to affect the characteristics of the surface (Fig. 2.5.4a, b, c). The populations reached a stable pattern at approximately the tenth generation, and the autocorrelation structure dissolved near generation fifty.

In all of the simulations up to this point, the spatial arrangement of the populations (i.e. ring form) remained unchanged, and when adjusting model parameters, all populations were altered equivalently. As a result, long-term changes to the autocorrelation structure only occurred in the temporal domain. By completely restricting gene flow between different portions of the metapopulation, it was possible to generate change in the overall structure spatial domain (Fig. 2.5.6a,b,c). Differences developed not only in the location of the zero contour line, which decreased along the spatial axis, but also in the strength of the autocorrelation values, which also decreased. Altering the arrangement of the populations generated changes in both the spatial and temporal domain, resulting in more complicated spatial-temporal genetic structures (Figs. 2.5.7 and 2.5.8 a,b,c). In the simulations using networked structures, autocorrelation structure was maintained for a longer period of time, although the zero line continued to remain near the center of the plot. In contrast, using unbalanced population structures led to shifts in

the position of the zero contour line along the spatial axis, changes in the strength of autocorrelation values, and particularly in the last case, a considerably more complicated structural pattern.

## **Discussion**

The autocorrelation surfaces demonstrate how the space-time structure of simple systems is affected by changes to connectivity and demographic structure. The simulation results are meaningful for two reasons. First, they provide evidence that the object-oriented model is sound. Increasing the number and strength of connections to neighboring populations leads to a decrease in mixing time; increasing the number of populations increases mixing time (prolonging structure) and reduces the band of non-significance around the zero line; partitioning populations leads to a reduction in the distance over which populations are similar, as well as a decrease in the magnitude of the autocorrelation values. These findings are consistent with the behavior expected of the systems studied. The results also have a more important second function however, in that they demonstrate that scalar changes to the system only affect the time scale of the autocorrelation surface. Changes in the spatial domain occur when inter-population differences exist as a result of unequal migration, distance or demographic composition. Intuitively, this makes sense, because with structure, it is relative differences that are responsible for generating structure, as opposed to absolute ones. This is also evident in the autocorrelation equations as well (e.g. 7). Scalar multipliers factor out, and are eliminated by dividing out the terms in the numerator and denominator. This is the

reason that changing the number of births and population size did not have a significant effect on the overall structure (although). An important consequence of this is that if time scale is not of importance, but rather the manner in which the system changes through time, then the system can be scaled. In other words, it is possible to reduce population sizes and the birth rate, as long as all the different elements are scaled in proportion to one another. Note that this does not mean that birth and population size are unimportant. The time interval over which a process occurs can be crucial (whether genetic spread occurs over the course of tens, hundreds or millions of generations is a significant distinction), and higher order aspects of the system, such as variance of the autocorrelation values will be affected as well, as was evident in the case where the number of populations was increased.

The dissolving of autocorrelation structure that is evident towards the end of some of the simulations occurs due to the discrete nature of the populations. With discrete values, the minimum detectable scaled difference between populations will be  $1/2N$  (again assuming a diploid population). As  $N$  increases, this difference decreases. In discrete populations, as populations become well-mixed, the differences between them fall to zero. In the algebraic formulation, elements of  $\mathbf{Q}_t$  can take on real (i.e. decimal) values, meaning that the minimum detectable difference can be infinitely small, and therefore structure can be maintained over a longer time interval, as was seen when the number of individuals per population was increased.

The results also highlight the importance of scaling considerations when dealing with subdivided populations. The space-time autocorrelation surfaces explicitly show that relationships between populations change depending on distance and time. As

systems transition from being relatively closed to open (i.e. isolated to connected), statistical properties such as variance, covariance and correlation are significantly altered, especially in patchy environments. Levin (1992) has noted that “The problem of pattern and scale is the central problem in ecology, unifying population biology and ecosystems science, and marrying basic and applied ecology.” By examining autocorrelation patterns, not only is it possible to determine whether structure exists, but also the characteristic scales at which structure is evident in both space and time. A critical aspect of this is that the scale of genetic processes within a system may be very different from the scale of demographic processes. Host organisms are transient, but the information contained in genetic material is persistent through time. Although spatial and temporal scales of genetic processes are affected by the life-history of the host, the spatial and temporal scales at which a biological organism functions will likely be quite different from those of the genetic material that it carries. Typically, management activities are carried out with the goal of preserving population numbers, but if genetic diversity is to be preserved as well, then consideration should be given to the processes that determine genetic structure, and the scales at which those processes operate.

The algebraic formulas provide an efficient way of representing the system, which raises the question as to whether there is a need for the simulation model. Indeed, for some of the simulations provided, analytic solutions have already been found (e.g. stepping-stone/ring populations - Kimura and Weiss 1964; e.g. stepping-stone/ring populations - Malécot 1969; Malécot 1973, 1975; Maruyama 1977). The structures examined were selected for illustrative purposes however, and are simple by design; real-world systems are almost certainly much more complex, involving mixtures of source

and sink populations with differing degrees of migration taking place between them according to local conditions. Individual organisms can also be expected to engage in context-specific behaviors based on their environment, including interactions with other organisms. The equations define the system in a concise manner, but as the model continues to increase in terms of complexity, interpreting the system becomes more and more difficult, particularly as more probabilistic sources of variation are added, and higher order aspects such as variance exert greater influence on the system. An advantage of the individual-based approach lies in its ability to encapsulate the characteristics and behavior of the host organism.

In addition to its use in exploring theoretical concepts, this type of modeling approach can be applied practically as well. Recently, models have been developed that are able to predict the dispersal of organisms from one location to another based on actual environmental conditions (e.g. Cowen, Paris, and Srinivasan 2006). The approach described here extends these models further by translating the movement of those individuals into movement of genetic material between populations. These results can then be used to assess the population genetic consequences of habitat loss in real-world systems (removal of matrix elements), to determine characteristic spatial and temporal scales at which they operate (using the analytical methods shown in this manuscript) or to possibly to determine reserve designs that would maintain the greatest amount of biological diversity (set an objective function and simulate across a range of conditions).

Cavalli-Sforza and Bodmer (1971) noted that the migration matrix may change with time, and in dynamic environments, this will almost certainly occur. While this would lead to considerable analytical complexity, the simulation model is capable of

addressing this situation through the use of scheduling. A lookup table can be created associating different transition matrices with a date-time code, and as the simulation progresses through time, transition matrices change accordingly. In such a manner, it becomes possible to evaluate the expected trajectory of even highly variable systems.

Object-oriented programming also provides the advantage of partitioning the problem into smaller modular units. Because elements of the model such as *Organism*, *Reproduction* and *Movement* are interfaces, it is possible to create different implementations without needing to re-define the model. For example, a simple implementation of *Organism* could be developed requiring only basic information such as its mode of reproduction and manner of movement. A more realistic implementation might include methods for resolving interspecific interactions, reactions to given predator densities, energetics, and age structure. The “realistic” class would presumably deliver more accurate results, but would require more parameterization. Conversely, the generic implementation might require less input, but would only be capable of generating coarse approximations. By having interfaces that can accept different implementations, the model can be configured to require the minimum amount of detail in order to sufficiently address the question at hand. In many respects, biological systems are ideally suited for object-oriented modeling, since well-developed hierarchical classifications of organisms already exist (e.g. functional taxonomic classifications). A further advantage of a modular-type approach is that it becomes easier for teams to work independently on small pieces of a much larger problem. The pieces can then be assembled to form a greater whole. Open-source code provides transparency, and individual classes can be reviewed and tested rigorously through unit testing.

## Conclusions

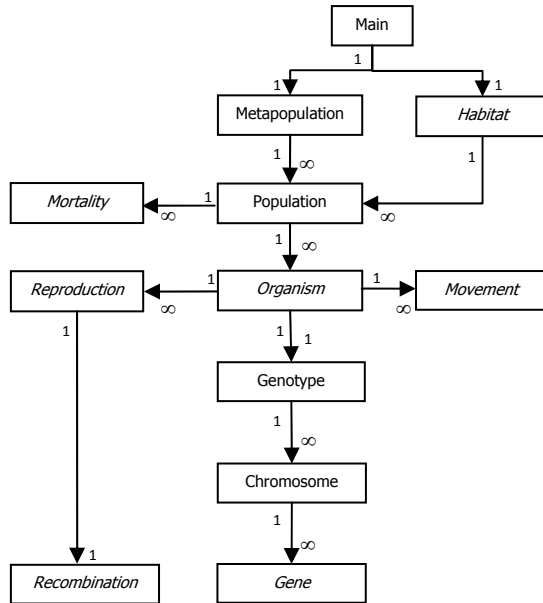
Object-oriented individual-based models provide a powerful and versatile means of exploring the interactions between migration, spatial structure and demographic parameters. The space-time autocorrelation surfaces capture changes in genetic structure over space and time, and show that uniform changes to population structure only lead to changes in the temporal domain. Changes in the spatial domain require relative differences between populations. The matrix formulas relate between migration, demographic structure, geographic structure and genetic structure in an analytical manner, however an advantage of the object-oriented, individual-based approach is that it is able to integrate multiple sources of complex, non-linear variation without altering model structure through the use of abstraction and implementation. The model has practical application for evaluating the development of genetic structure in real-world ecosystems by using real oceanographic data and realistic life-history parameters as input.

**Table 2.1.** Input values for the simulation runs. *S* – Set number, *Sub* – Subset, *D* – form of the distance matrix/spatial configuration (Fig. 2.3), *M* – form of the connectivity/transition matrix (Refer to Fig. 2.4 for matrix structure), *b* – number of offspring produced per pairing, *n* – total number of populations, *K* - carrying capacity of all individual populations. \*These transition matrices have the same circulant form as *M(a)*, but have a greater number of populations.

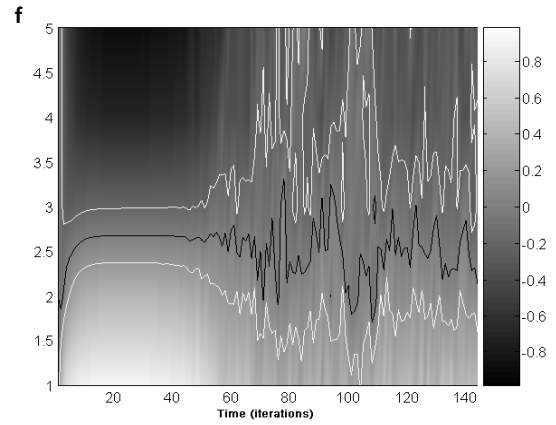
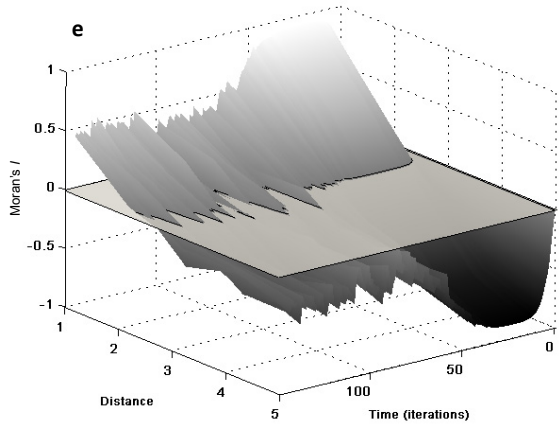
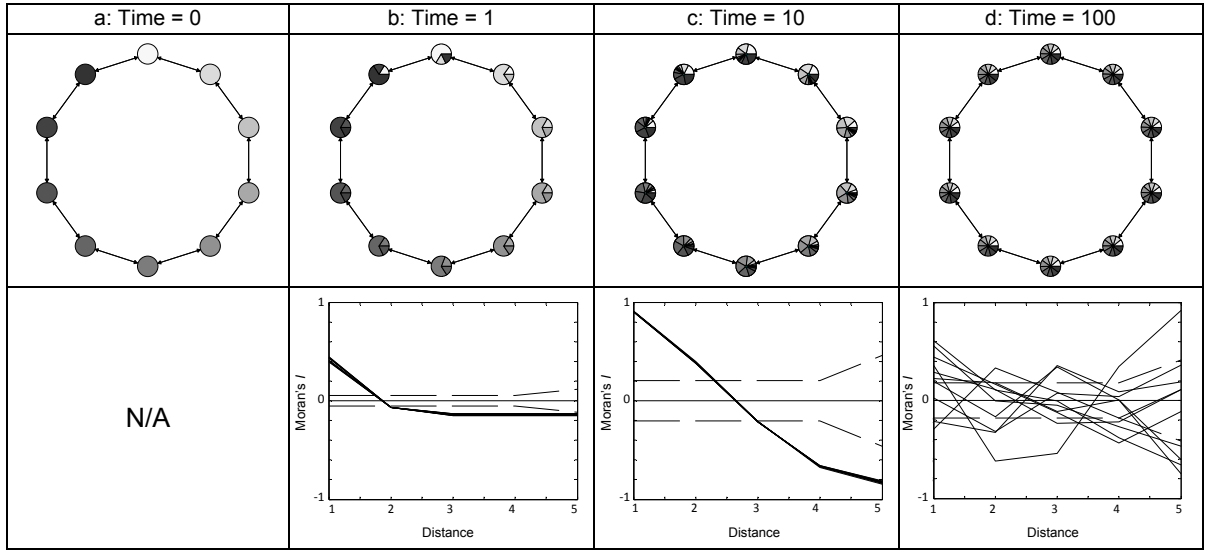
<b>S</b>	<b>Set Description</b>	<b>Sub</b>	<b>D</b>	<b>M</b>	<b>n</b>	<b>b</b>	<b>K</b>
1	1 <sup>st</sup> order neighbors	<i>a</i>	<i>a</i>	<b>a</b>	10	10	100
		<i>b</i>	<i>a</i>	<b>b</b>	10	10	100
		<i>c</i>	<i>a</i>	<b>c</b>	10	10	100
2	2 <sup>nd</sup> order neighbors	<i>a</i>	<i>a</i>	<b>d</b>	10	10	100
		<i>b</i>	<i>a</i>	<b>e</b>	10	10	100
		<i>c</i>	<i>a</i>	<b>f</b>	10	10	100
3	# Populations	<i>a</i>	<i>a</i>	<i>a</i> *	<b>15</b>	10	100
		<i>b</i>	<i>a</i>	<i>a</i> *	<b>20</b>	10	100
		<i>c</i>	<i>a</i>	<i>a</i> *	<b>40</b>	10	100
4	Fecundity	<i>a</i>	<i>a</i>	<i>a</i>	10	<b>5</b>	100
		<i>b</i>	<i>a</i>	<i>a</i>	10	<b>20</b>	100
		<i>c</i>	<i>a</i>	<i>a</i>	10	<b>50</b>	100
5	# Individuals per population	<i>a</i>	<i>a</i>	<i>a</i>	10	10	<b>200</b>
		<i>b</i>	<i>a</i>	<i>a</i>	10	10	<b>400</b>
		<i>c</i>	<i>a</i>	<i>a</i>	10	10	<b>1000</b>
6	Partition	<i>a</i>	<i>a</i>	<b>g</b>	20	10	100
		<i>b</i>	<i>a</i>	<b>h</b>	20	10	100
		<i>c</i>	<i>a</i>	<b>i</b>	20	10	100
7	Network	<i>a</i>	<b>b</b>	<b>j</b>	20	10	100
		<i>b</i>	<b>c</b>	<b>k</b>	20	10	100
		<i>c</i>	<b>d</b>	<b>l</b>	20	10	100
8	Irregular	<i>a</i>	<b>e</b>	<b>m</b>	20	10	100
		<i>b</i>	<b>f</b>	<b>n</b>	20	10	100
		<i>c</i>	<b>g</b>	<b>o</b>	20	10	100



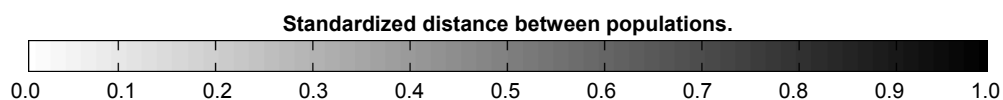
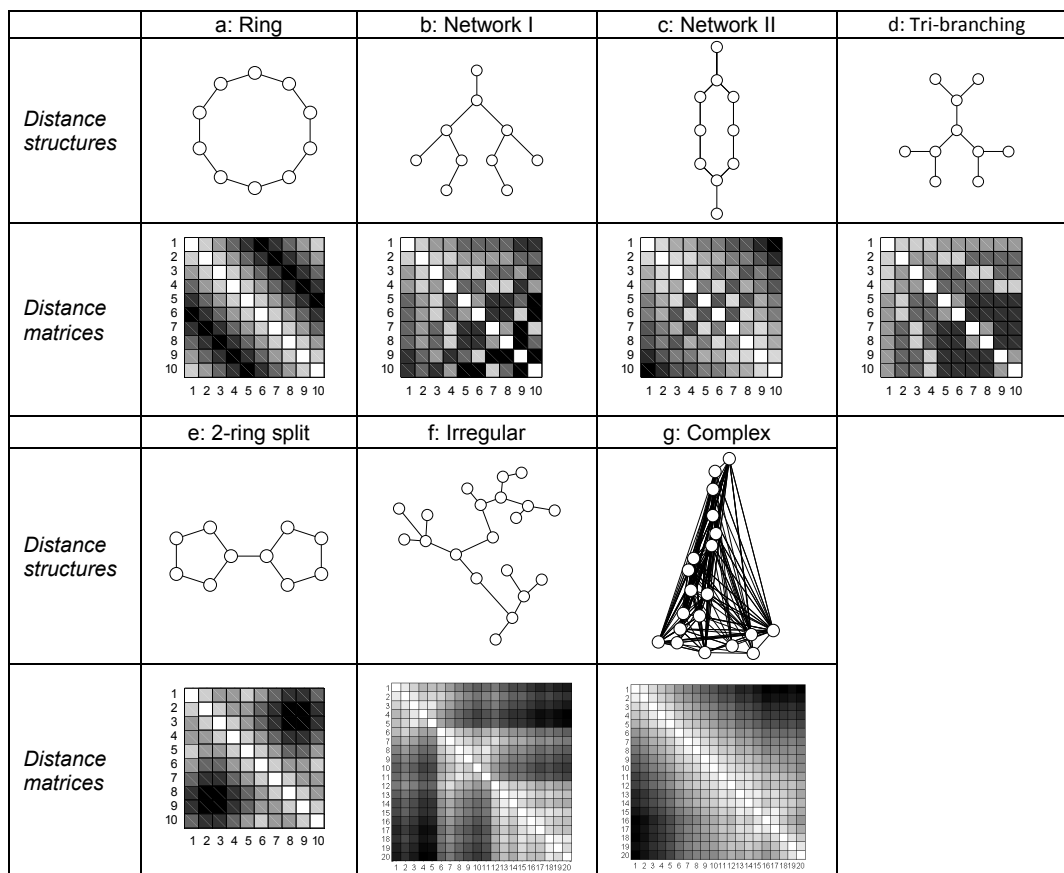
**Fig. 2.1.** Class diagram of the object-oriented, individual-based modeling framework. Relationships between classes are indicated by the arrows (e.g. a *Population* may contain several *Organisms*). Numbers indicate the nature of the relationship: one to many (1 -  $\infty$ ) or one to one (1 - 1) e.g. an *Organism* may have several manners by which it is able to reproduce (sexual reproduction vs. asexual reproduction through fragmentation or vegetative growth) but only one unique *Genotype*. Italicized classes indicate interfaces.



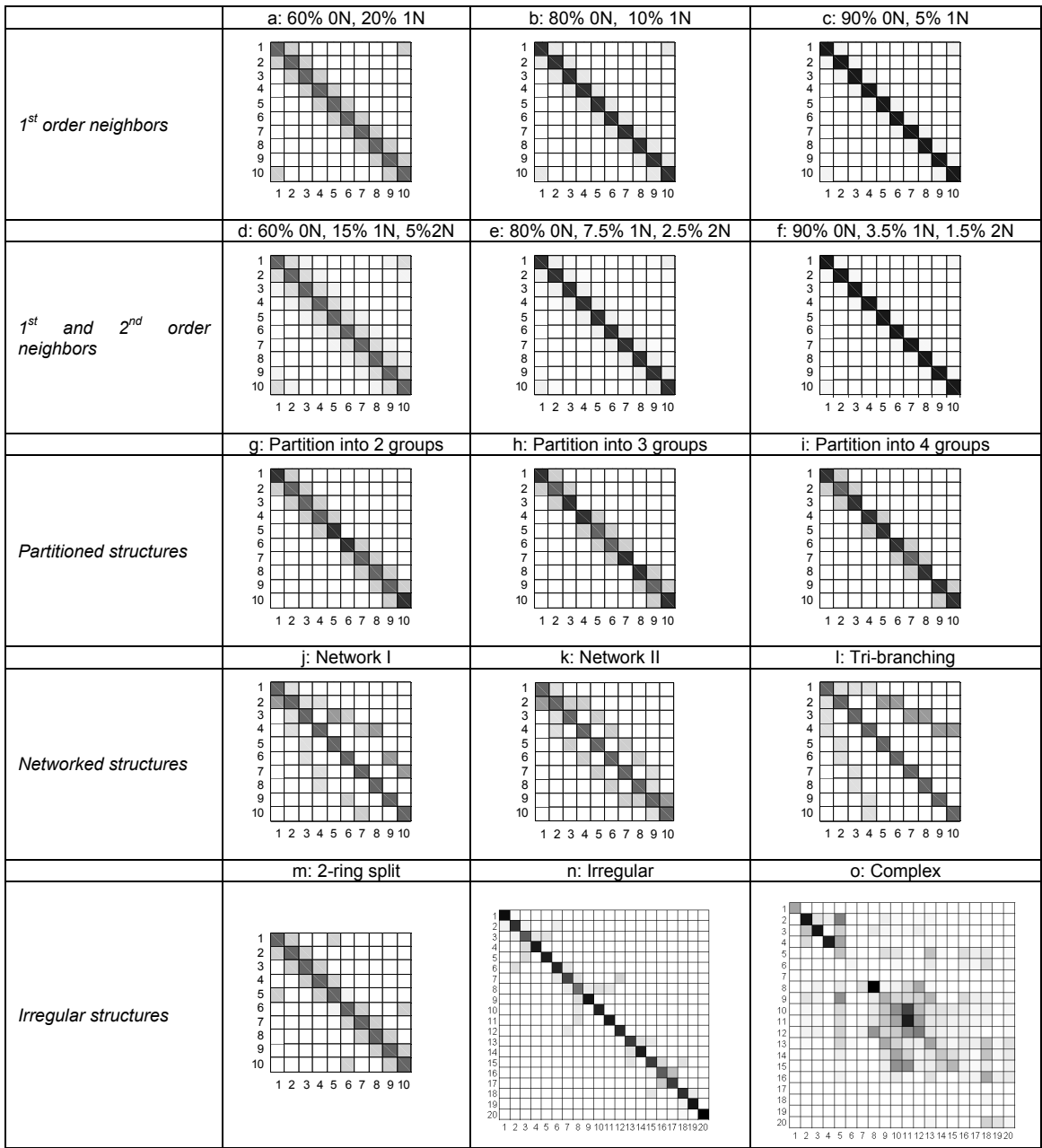
**Fig. 2.2.** A sample metapopulation is shown at different time steps along with corresponding autocorrelograms. Dashed lines in the autocorrelograms indicate the 95% confidence interval. All populations are initialized such that members are genetically labeled according to their population of origin. Therefore at time 0, all populations are 100% composed of the allele type corresponding to the population of origin (**a**). Following mating and dispersal of progeny, the genetic composition of the populations changes according to the transition matrix (**b – time 1**). **c** shows the population composition at time 10 and **d** the composition at time 100. Correlogram lines for each allele type are initially convergent, appearing as one line (**b and c**), but then separate over time (**d**). The autocorrelograms for each time step were then stacked to provide a surface (**e**). Axes on the 3-d surface were reversed to improve visibility. Average autocorrelation values are shown, but surfaces for individual allele forms can also be produced. **f** presents a 2-dimensional representation of **e**, corresponding to a view from above (but note re-orientation of axes). The black line indicates where the autocorrelation surface intersects the zero plane, and the white lines show the boundaries of the 95% confidence interval.



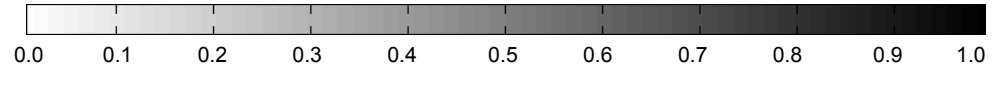
**Fig 2.3.** Distance input matrices. Distances for **a-e** are the number of links traversed between populations. The distance measurement for **f** is the path length, and for **g** it is Euclidean (straight line) distance between populations. For purposes of comparison, all matrices were standardized according to their respective maximum distance.



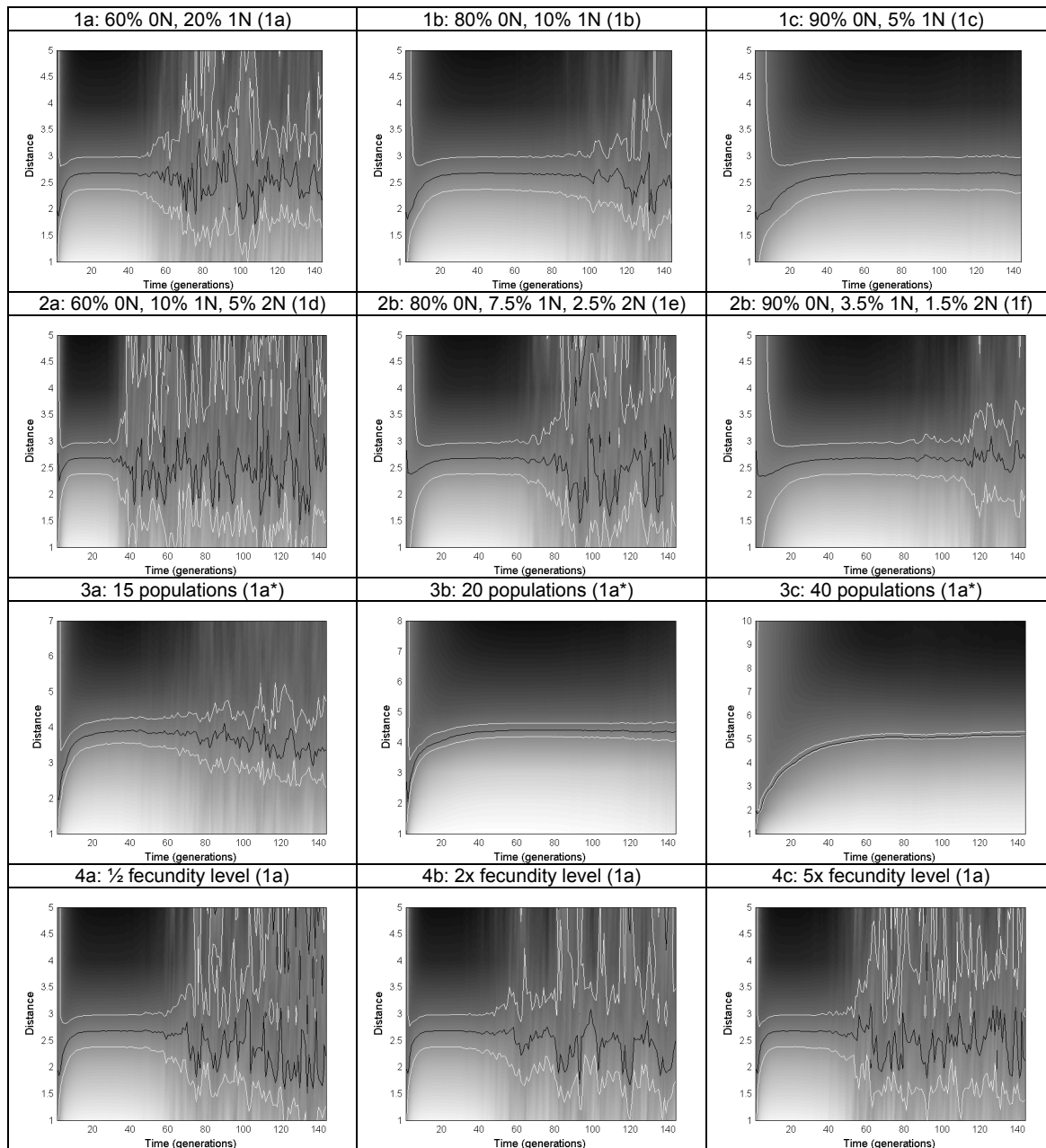
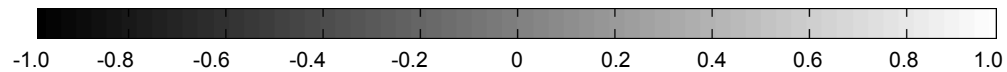
**Fig. 2.4.** Transition matrices. Each matrix indicates the degree of migration taking place between populations as a fraction of the source. Columns represent source populations, rows represent destination populations. Stronger levels of migration are indicated by darker colors. For the first two rows, 0N refers to the percentage of individuals migrating to the same population, 1N refers to the percentage of individuals traveling to either nearest neighbor, and 2N refers to the percentage of individuals traveling to either second neighbor.



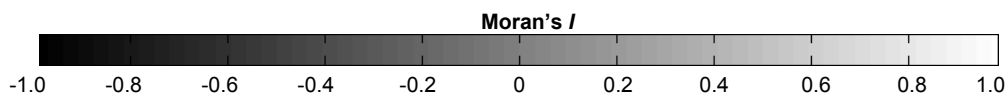
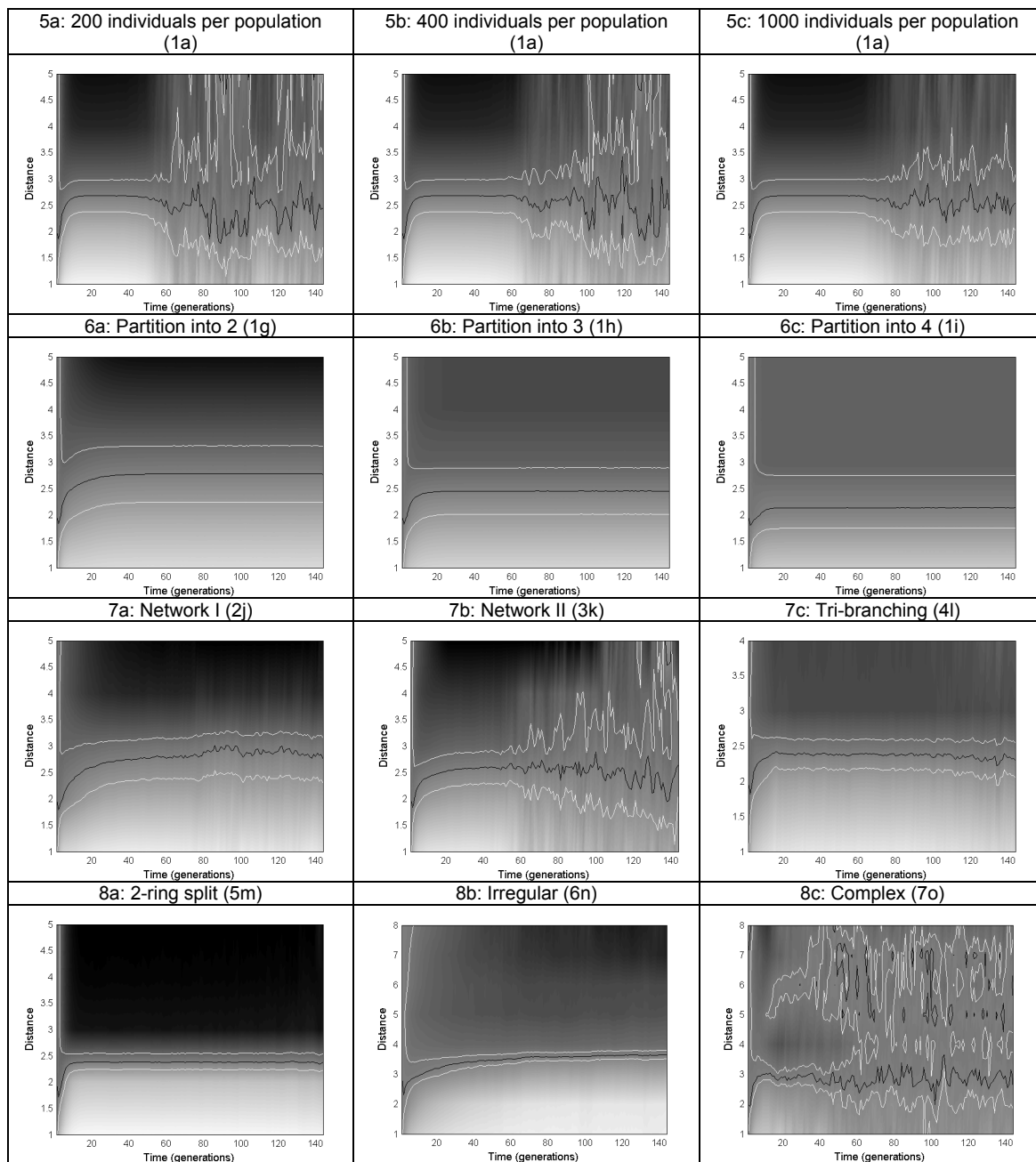
Degree of migration from one source population to a destination population as a proportion of source population size at the preceding time step.



**Fig. 2.5 (2 pages).** Space-time genetic structure plots (autocorrellograms) for the simulations. Number-letter combinations refer to the simulation set-subset code identified in Table I. Surfaces indicate the average amount of spatial autocorrelation present over space and time. Light shading indicates similarity between populations at the given distance and time, dark shading indicates dissimilarity. The black line designates an autocorrelation value of 0. Areas outside of the white lines indicate significant autocorrelation at a level of  $\alpha = 0.05$ . 0N designates the percentage of individuals returning to the same population following migration. 1N and 2N denote connectivity levels for 1<sup>st</sup> and 2<sup>nd</sup> order neighbors respectively.

Moran's  $I$ 





## **CHAPTER 3: CONNECTIVITY, SELECTION AND RESOURCE USE IN SUBDIVIDED POPULATIONS – THEORY AND IMPLICATIONS**

### **Background**

Understanding how migration and selection interact is an important goal for population geneticists and evolutionary biologists (Lacy 1987; Hastings and Harrison 1994; Chesser, Olin E. Rhodes, and Smith 1996; Barton and Whitlock 1997). This effort also has practical importance - as connectivity structure (the relative exchange of individuals among geographically separated subpopulations - Cowen et al. 2007) is increasingly altered by habitat fragmentation and loss (Andren 1994; Young, Boyle, and Brown 1996; Bender, Contreras, and Fahrig 1998; Goodwin and Fahrig 2002; Ryall and Fahrig 2006), environmental managers must be aware of potential problems in maintaining adequate levels of genetic diversity over the long-term (Templeton et al. 1990; Young, Boyle, and Brown 1996; Young and Clarke 2000). Connectivity research typically involves large spatial and temporal scales however, making direct experimentation difficult, if not impossible. Instead, models provide a more effective means of investigating these types of systems.

Genetic models for interconnected populations have been developed by a number of different authors, with much of the research focusing on the effects of migration. Some of the earliest research was performed by Wright (1931; 1937; 1943), and other approaches have subsequently been developed – e.g. stepping stone models, (Kimura and Weiss 1964; Malécot 1969; Maruyama 1977), diffusion models (Nagylaki 1978; Nagylaki and Moody 1980), and matrix models (Bodmer and Cavalli-Sforza 1968; Fu,

Gelfand, and Holsinger 2006). In addition to migration, selection is equally capable of playing a major role. Selection affects genetic structure by biasing the sampling process. Integrating both migration and selection into a metapopulation (network of populations) model presents some challenges due to the multivariate nature of the system; activities of multiple populations need to be processed in parallel. Although work has been done to address the joint effects of selection and migration (Maruyama 1977; Slatkin 1981; Barton 1993; Nagylaki 1996; Rousset 2004), their interaction in a metapopulation context has not been as well-studied (but see Barton and Whitlock 1997; Whitlock 2002), particularly when resource availability is also taken into consideration.

Discrete multivariate distributions provide a means of efficiently simulating the transport of genetic material between interconnected populations in a probabilistic manner. Using combinations of discrete multivariate distributions, it is possible to create a neutral model of gene flow between populations. Furthermore, by extending the model using weighted distributions, selection and the influence of resource availability can be incorporated as well. Taking this approach leads naturally to stochastic formulation of multiple-niche polymorphism (Levene 1953; Dempster 1955) within a metapopulation context. Relative differences in fitness values also affect the outcome of the evolutionary process, and this effect is shown to be scale-dependent.

## **Definitions**

A population state matrix  $\mathbf{Q}$  is defined, consisting of  $r$  populations (objects/rows) and  $s$  allele forms (descriptors/columns). Individual elements are the number of alleles of

the given form in the designated population. The  $\mathbf{Q}$  matrix represents the proportional allelic content of the various populations (Fig. 3.1 – circles). A stochastic transition matrix  $\mathbf{M}$  is also defined, containing the probability of individuals migrating from a population  $j$  to population  $i$ , where  $i$  and  $j$  are row and column indices respectively in the interval 1 through  $r$ . Columns of the matrix represent source populations and rows represent destination populations. The sum of each individual column should equal 1. This is commonly referred to as a “forward” transition matrix, and is distinct from the row-standardized “backwards” transition matrix more commonly encountered in population genetics (e.g. Nagylaki 1982). If loss of genetic material occurs during the transition (e.g. emigration, transitional mortality), this can be addressed by including an additional row containing the remaining probability (i.e.  $1 - m_{\cdot j}$ ; refer to Appendix A for bar-dot notation conventions). In Fig. 3.1,  $\mathbf{M}$  is represented by the arrows between populations. Lower case, bolded letters with subscripts (e.g.  $\mathbf{q}_i$ ) refer to the vector of the corresponding matrix at the designated row or column subscript.

### **Demographic processes**

Two important demographic processes are birth and mortality. Birth is linked with recombination and segregation; mortality provides a second sampling process, and is linked with selection. To account for the birth process, a random vector  $\mathbf{b}$  may be defined, where  $b_i$  is the number of births in the  $i$ th population. Values of  $\mathbf{b}$  were assigned using a degenerate distribution to control for variance resulting from the birth process, however alternative distributions (e.g. binomial, Poisson) could also be used. For all

models, mating is assumed to occur randomly, and organisms are assumed to be diploid and hermaphroditic (random sexual pairing produces identical expected values - see Appendix B). Determining the number and type of alleles produced as a result of reproduction involves more than just multiplying  $b_i$  and  $q_i$ , since resampling takes place due to segregation. The number of new alleles produced of each type can be determined using the multinomial distribution  $mn(k, \mathbf{p})$ , where  $k$  is the number of items (alleles) being sampled and  $\mathbf{p}$  is a vector of probabilities corresponding to the likelihood of a particular outcome (Johnson, Kotz, and Balakrishnan 1997). In the following examples,  $k_i = u_i = \frac{1}{2} b_i q_i$ . The 1/2 enters into the equation under the assumption of a diploid population, since alleles migrate in pairs. Values for  $\mathbf{p}$  are obtained by standardizing row vectors by the sum of their elements (i.e.  $\mathbf{p}_i = \bar{q}_i$ ).

With mortality, the number of individuals sampled corresponds to the number of individuals surviving the mortality event multiplied by the ploidy level. If the numbers of individuals within each population are extremely large, then the probability distribution of the survivors will be approximately multinomial (sampling with replacement). If populations are of limited size however, then sampling should occur without replacement, which is instead given by the multivariate hypergeometric distribution  $mh(k, \mathbf{n})$ , where  $k$  is the total number of items being selected, and  $\mathbf{n}$  is a vector corresponding to the number of items in each category type (Johnson, Kotz, and Balakrishnan 1997).

### A neutral model with migration

An algorithm describing the transition of the state of the system at one time step to the next (with non-replacement of populations) is given by the following steps:

1. Determine the total number of progeny produced by population  $i$  at a

given time step:  $u_i = \frac{1}{2} b_i q_{i,t}$ .

2. Determine the number of individuals transitioning from all source populations to the set of destination populations:  $\mathbf{v}_j = \text{mn}(u_i, \mathbf{m}_j)$ .

3. Determine the genetic composition of the incoming cohort based on the probabilities of the source population:  $\mathbf{g}_i = \sum_{j=1}^s \text{mn}(2v_{ij}, \bar{q}_i)$ .

4. Reduce each population to its respective carrying capacity  $z_i$ :

$$\mathbf{q}_{i,t+1} = \text{mh}(z_i, \mathbf{g}_i).$$

The second step captures migration, the third – segregation, and the fourth - demographic mortality. The first step parameterizes the second. Both migration and segregation can be represented using the multinomial function since the probabilities do not change with each subsequent draw. Determining the genetic composition of the

cohort is performed after migration for computational efficiency. This is possible because individuals can be assigned to destinations independently of their genotype.

To observe how the system changes in response to different levels of connectivity, numerical simulations were carried out in Java (v. 1.6) (Sun Microsystems 2006) using algorithms translated from the C++ based **stocc** package developed by Fog (2006) (Java source available from the author upon request). The form of migration used in the simulations was a nearest-neighbor stepping-stone model (e.g. Fig. 3.1); however the approach can be used in conjunction with a connectivity structure of arbitrary form. Each population was assigned a carrying capacity of 100 individuals (200 alleles), and the number of births was set at 10. Allele counts for individual populations through time were written to an output file, and then imported into Matlab (The Mathworks 2007) for analysis. Allele counts were converted into frequencies by cross-tabulating runs, time steps, populations and allele counts. The resulting time-frequency surfaces were then averaged across all model runs (runs = 100) (Fig. 3.2). The results of the simulation are shown in Fig. 3.3 (first row). Note that because only two allele forms were used in the simulations, plots for only one form are shown, since plots for the second form would simply be a reflection of the first across  $y = 0.5$ .

In the event that there is no migration between populations, each individual population is independent of the rest ( $\mathbf{M}$  is an identity matrix and the associated sampling process can be eliminated from the equation). Each population drifts towards fixation of a particular allele form with a probability equal to the initial frequency of that allele (Nei 1987). On the other extreme, assuming all populations are equally connected, then the entire group functions as a single, larger population. The populations still drift towards

fixation or loss, however they do so jointly and at a slower rate. The pattern is stretched along the time axis; allele frequencies remain concentrated near middle values for a longer period of time (Fig. 3.3 – Neutral). The appearance of the cloud is consistent with what would be expected from a diffusion-type process.

### **Incorporating selection into the model**

Fitness and selective advantage have been well described, and form the basis of an extensive amount of population genetic theory (Haldane 1927; Fisher 1930; Wright 1931; Kimura and Ohta 1971; Whitlock 2002). In this situation however, there are additional complications in that multiple populations are being considered, it is possible to have multiple allele forms, and the fitness of a genotype at one location may be different at another. For example, organisms that take up water rapidly might have an advantage in a humid environment, but might be at a disadvantage in a sere one. These differences in relative fitness can be addressed by defining an  $n \times m$  matrix  $\mathbf{\Omega}$ , where the entries correspond to the fitness of allele  $j$  at population  $i$ . The multivariate hypergeometric function assumes that all allele forms have identical weight, and therefore to incorporate selective advantage, a different model is required.

Fisher (1934; 1935) developed a noncentral hypergeometric model (NCH) that incorporates weighting, and the distribution was extended into its multivariate form by McCullagh and Nelder (1989; also see Fog 2008). This function will be designated as  $\text{mf}(k, \mathbf{n}, \mathbf{\omega})$ , where  $k$  is the number of items being sampled,  $\mathbf{n}$  is a vector corresponding



to the number of items in each category type, and  $\omega$  corresponds to the fitness values for each category type.

Fisher's NCH assumes that the relative weights of organisms do not change as members are selected. A second function which does adjust the weighting scheme as sampling proceeds is Wallenius' NCH distribution (Wallenius 1963, Manly 1985). The multivariate form of Wallenius' NCH distribution was developed by Chesson (1976; also see Fog 2008). This function will be designated by  $mw(k, \mathbf{n}, \omega)$ . As the difference between  $k$  and  $N$  becomes large, Wallenius' NCH is approximated by Fisher's NCH.

The relative fitness values within each population ( $\omega$ ) can be defined arbitrarily, but to proceed further, the concept of resource use is now introduced. To provide a simple, illustrative scenario, an additive, one-to-one correspondence between genotype and phenotype is assumed (i.e. heterozygotes have intermediate fitness); an individual uses a unit of resource in proportion to its genotype (i.e. an individual with a genotype of  $a_1a_1$  would require two units of Type I resource (white),  $a_2a_2$  two units of Type II resource (black), and  $a_1a_2$  or  $a_2a_1$  one unit of Type I and one unit of Type II resource) (Fig. 3.4). Therefore in a population with 150- units of Type I resource and 50 units of Type II resource, the fitnesses for  $a_1$  and  $a_2$  would be would be 0.75 and 0.25 respectively, or standardizing the maximum fitness level to 1, 1 and 1/3 respectively. In effect, this defines the niche space (albeit a simple one) available to the population (Hutchinson 1957, 1965; Leibold 1995). Additional complexity may be introduced by increasing the number of different resource types (e.g. Type III, IV etc.). The setup used in the simulations is shown in Fig. 3.4, with 250 units of Type I resource and 150 units of

Type II resource in the first population (i.e. standardized weightings of 0.625 and 0.375), and alternating values for each subsequent population.

In model runs involving selection and no migration, populations are quickly driven to fixation. As the system shifts from closed (no migration) to open (equal migration), the system converges on the relative proportions of resource type (Fig. 3.3, rows 2 and 3). Wallenius' NCH follows the same basic pattern as Fisher's NCH, but concentrates the allele frequencies to a greater degree and is slightly offset from the proportional resource values.

### **Demand-side selection**

Fisher's NCH, assumes that the relative fitnesses of different allele forms and genotypes do not change. With the implementation of Wallenius' NCH used above, weightings change with each draw, but the change is based on the characteristics of the population from which the sample is being drawn. One must also consider the possibility however that the weightings might be affected by the characteristics of the habitat patch the organisms will be occupying. Consider the analogy of a board containing five square holes and five round holes. If the ten slots are to be filled from a source population of square and round pegs with the condition that the board can only be filled by pegs of the correct type, the resulting sample will always be five square pegs and five round pegs, regardless of the source population size and composition (as long as the number of source pegs of each type is greater than the number of corresponding destination spaces). To address this, Wallenius' NCH may be applied in a slightly different manner, such that

weightings are dependent on the characteristics of available resource rather than the composition of the population the sample is being drawn from (hence the appellation demand-side selection). In this case  $n_i = \rho_i / \nu_i$ , where  $\rho_i$  is the amount of resources of type  $i$  in the population, and  $\nu_i$  is a usage coefficient of that resource type (i.e. one instance of  $n$  will consume  $\nu$  units of resource type  $i$ ). Note that  $n_i$  is no longer the number of individuals in the source population, but rather the number of individuals the environment is capable of supporting, given the available resources. The formula becomes

$$\Pr[\mathbf{K} = \mathbf{k}] = \left( \prod_{i=1}^c \binom{n_i}{k_i} \right) \int_0^1 \prod_{i=1}^c (1 - t^{\nu_i \omega_i / d})^{k_i} dt \quad (\text{Eq. 1}),$$

$$d = \omega_i (\rho_i - \nu_i k_i),$$

$$\sum k_i = n, \sum n_i = N$$

which will be designated by:  $\text{mr}(k, \mathbf{n}, \boldsymbol{\rho})$ , where  $\boldsymbol{\rho}$  is a vector containing the amounts of available resources, and  $c$  is the number of different allele forms (the proof is given in Appendix D).

In contrast to the invariant weightings of Fisher's NCH or the source-driven frequency-dependence of the previous application of Wallenius' NCH, here the outcome is influenced by the opportunities provided by the available niche space. In the pegboard example, there was a stringent condition in that the number selected is equal to the total amount of resource. Again, as the difference between the total amount of available resources and the total number of individuals increases, the constraint is relaxed, until eventually there is convergence back to Fisher's NCH. Note that resources are assumed

to be pooled; individuals do not have exclusive access to their respective resources, generating the co-dependency of the resource weightings.

The simulations using the resource-based weighting scheme show a pattern similar to those obtained from the multivariate Fisher's NCH and the previous Wallenius' NCH runs (Fig. 3.3, last row), though allele frequencies tend to be closer to the proportional availability of resources. There is a key difference however in the case with no migration. When using Fisher's and the previous Wallenius' NCH as the mortality function, fixation is quickly reached. Using resource-weighted sampling, even if an allele is at a considerable disadvantage in the general sense, it is able to persist for an extended period of time.

### **Fitness values and scale dependency**

The results to this point have demonstrated the effects of altering the form of the mortality function and the strength of migration, however differences in fitness values within and between populations also have an effect. If  $\Omega$  is uniform (i.e. all selective weights are equal), then  $\Omega$  has no effect on the outcome. The result is a neutral process. This relationship should not be surprising, since a lack of differences in fitness values is the very definition of neutrality. In contrast, as the elements of the weighting matrix become increasingly dissimilar, they exert a proportionally greater degree of influence on the system, which translates into an increasing degree of fixation (Fig 3.5). The result is a smooth transition between neutrality and natural selection conditioned on relative fitness weightings, which in turn are dependent on resource availability.

However, if one accepts that the outcome of the evolutionary process can be mediated by relative differences in fitness values, and that fitness values may be affected by resource availability, one must also consider that the potential exists for the results to be scale dependent, since resource structure can change according to scale. As an illustration, consider a situation in which there are four habitat patches, two of which have only Type I resource and two only have Type II resource, and there is equal migration between all populations (Fig. 3.6). Under these conditions, using any of the weighted mortality functions, the allele frequencies will quickly be driven to fixation. If the scale of the system changes however (e.g. combining patches or assuming that the organism has a greater range – Fig. 3.6 dashed boxes), this will generate a neutral outcome instead.

## **Discussion**

Discrete multivariate models give concrete form to two important aspects of the relationship between connectivity structure, selection and resource availability. The first is the ability for resource structure to influence the persistence of allele forms, allowing for survival even in the presence of what might be considered to be considerable fitness disadvantage. In the example provided, the initial relative difference between the two allele forms in fitness was 0.4, orders of magnitude larger than what one would expect to find in real environments, but in spite of this, the sub-optimal allele form was able to maintain itself over an extended period of time because of the frequency-dependence of the fitness values. The findings are consistent with formulas developed by Levene (1953)

and Dempster (1955), who generated deterministic models resulting in multiple-niche polymorphism. Discrete multivariate distributions provide a stochastic foundation for those equations. Stochastic formulations are relevant in that it is not only expectations that are important for projections, but higher order aspects of the process as well, such as variance. The equations given could be used to generate weighted multi-type branching processes in the same way the multinomial distribution is used to generate standard multi-type branching processes (Caswell 2001). Support for such a linkage is provided by Bulmer (1972), who noted that Levene's formulas could be expressed in terms of matrices corresponding to a weighted version of Bodmer and Cavalli-Sforza's model (1968). Bulmer's equations are deterministic however, and do not address the potential stochastic nature of the component matrices.

Although multiple-niche polymorphism resulting from demand-side selection will be obscured by factors such as greater population numbers, less strict resource requirements and more complex genotype-fitness mapping, the fact that unoccupied niche space is capable of harboring allele forms that otherwise might be considered to be at an overall selective disadvantage warrants attention, particularly in environments where resources are extremely limited. It should be emphasized that if selection is frequency dependent on resource availability, it is not the total niche space in a destination that is important, but rather the *available* niche space. Consequently, the effect of demand-side selection may be significant, even in systems with a large resource base.

The second point is that in heterogeneous environments, fitness weightings may change according to scale, and as a consequence the balance between whether an

evolutionary process is effectively neutral or driven by natural selection may also be scale-dependent. At what scale then should fitness values be measured? The answer depends on the scale at which the host organism interacts with its environment, as it is the range of an organism that determines the extent to which resources are considered to be pooled. The minimum degree of resolution with which the environment is resolved is referred to as the *grain*, and in conjunction with the *extent* (the range of the system), the two form the components of scale (Jenerett and Wu 2000). The importance of scaling considerations has been recognized by ecologists (Levin 1992), however the potential effect of scale on selective weightings does not appear to have been given the same degree of attention. An implicit assumption of many discrete-patch systems is that each individual patch is homogenous in composition, and with homogeneity comes scale-independence, and as a result scaling effects can be ignored. Real environments are rarely completely homogeneous however, and both the grain and extent of the environment from an organism's perspective will likely change through time. The discrete multivariate approach provides a direct means of incorporating scaling effects into the evolutionary process by providing a mechanism by which selective weights can be changed relative to one another by lumping and splitting the resource content of different patches.

The analysis was facilitated by the direct, simple mapping between genotype and fitness values. The reality is undoubtedly much more complex, involving multiple environmental factors, energetic costs as well as interactions at the level of the allele, locus and gene. Although this genotype-fitness mapping is an important part of the system, it presents some challenges in terms of development. Lewontin (1974) stated

that “To the present moment no one has succeeded in measuring with any accuracy the net fitness of genotype for any locus in any environment in nature.” Although research into fitness values of artificial and natural populations has taken place since (Bijlsma-Meeles and Bijlsma 1988; Fowler et al. 1997; Ochando and Ayala 1999; Shaver et al. 2002), the prospects for meeting Lewontin’s challenge for individual loci on a widespread basis remain distant. One possibility however is that the mapping may be flattened, meaning that it is not necessary to have information on the fitness values of individual allele forms and all of their potential interactions, only a more general relationship between genotypes and fitness. Advances in high throughput genotyping (Kwok 2000; Pastinen et al. 2000; Richardson et al. 2007) could be used in conjunction with various measures of fitness (e.g. growth, lipid content or DNA to RNA ratios - Bergeron 1997; Buckley, Caldarone, and Ong 1999) in order to generate these kinds of mappings.

The use of discrete multivariate sampling provides a means of simulating the genetic consequences of different connectivity regimes. One limitation however is that the use of transition matrices in studying evolutionary processes is limited by their specificity; results cannot be generalized to other systems (Epperson 2003). Nevertheless, as physical environmental models become more available and accurate, generating these types of matrices as needed becomes a very real possibility (e.g. Cowen, Paris, and Srinivasan 2006). One must be careful however to make the distinction between projection and forecasting (Caswell 2001). The model only provides the expected trajectory of populations under a set of given expectations. There may be a considerable amount of variation in an individual run. Still, the approach is able to

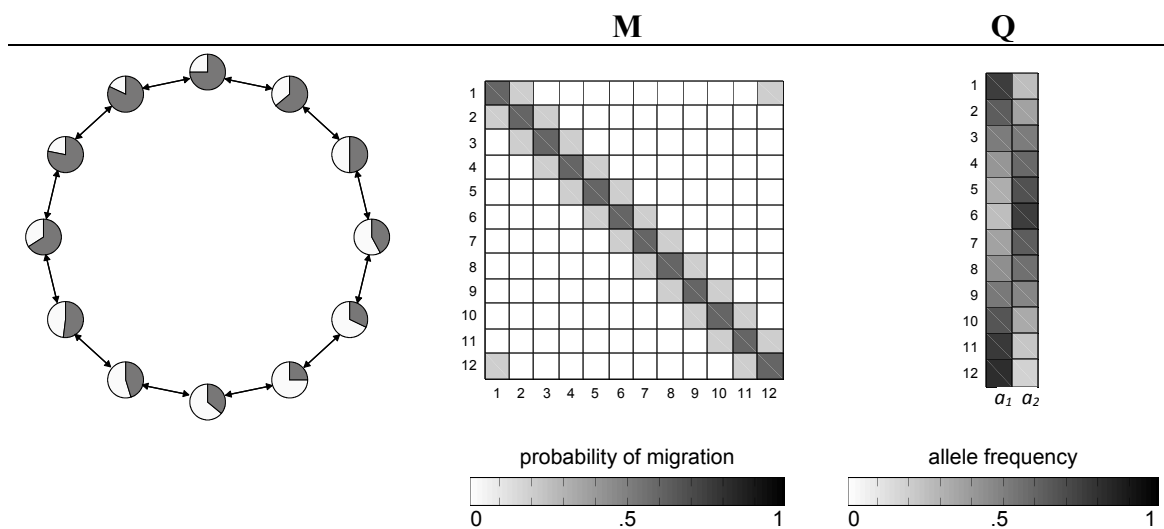


provide a picture of what one might expect of a given system, and just as importantly, what is unexpected.

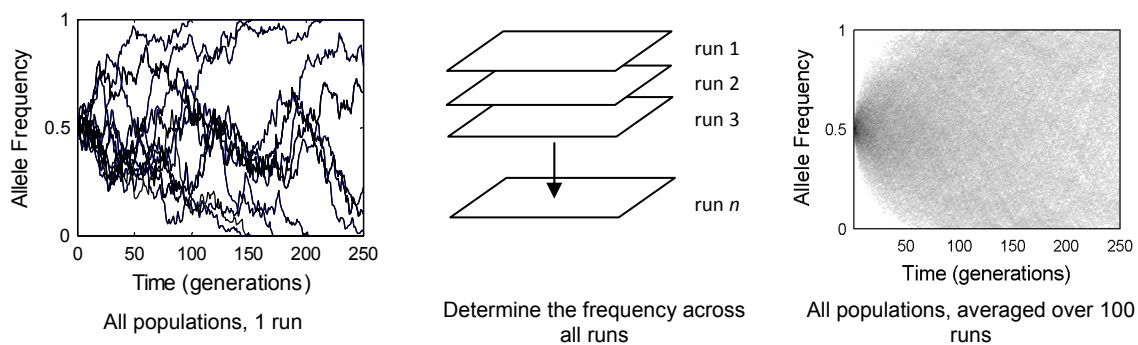
## **Conclusions**

Discrete multivariate distributions provide a mechanism of unifying aspects of migration, selection and resource availability in metapopulations. Different forms of mortality lead to different evolutionary outcomes, particularly in the case where there is differential resource use and resources are limiting. The degree of similarity in fitness values affects the outcome of the evolutionary process (i.e. neutrality, natural selection), and since fitness values in heterogeneous environments are scale dependent, the outcome will also be scale-dependent.

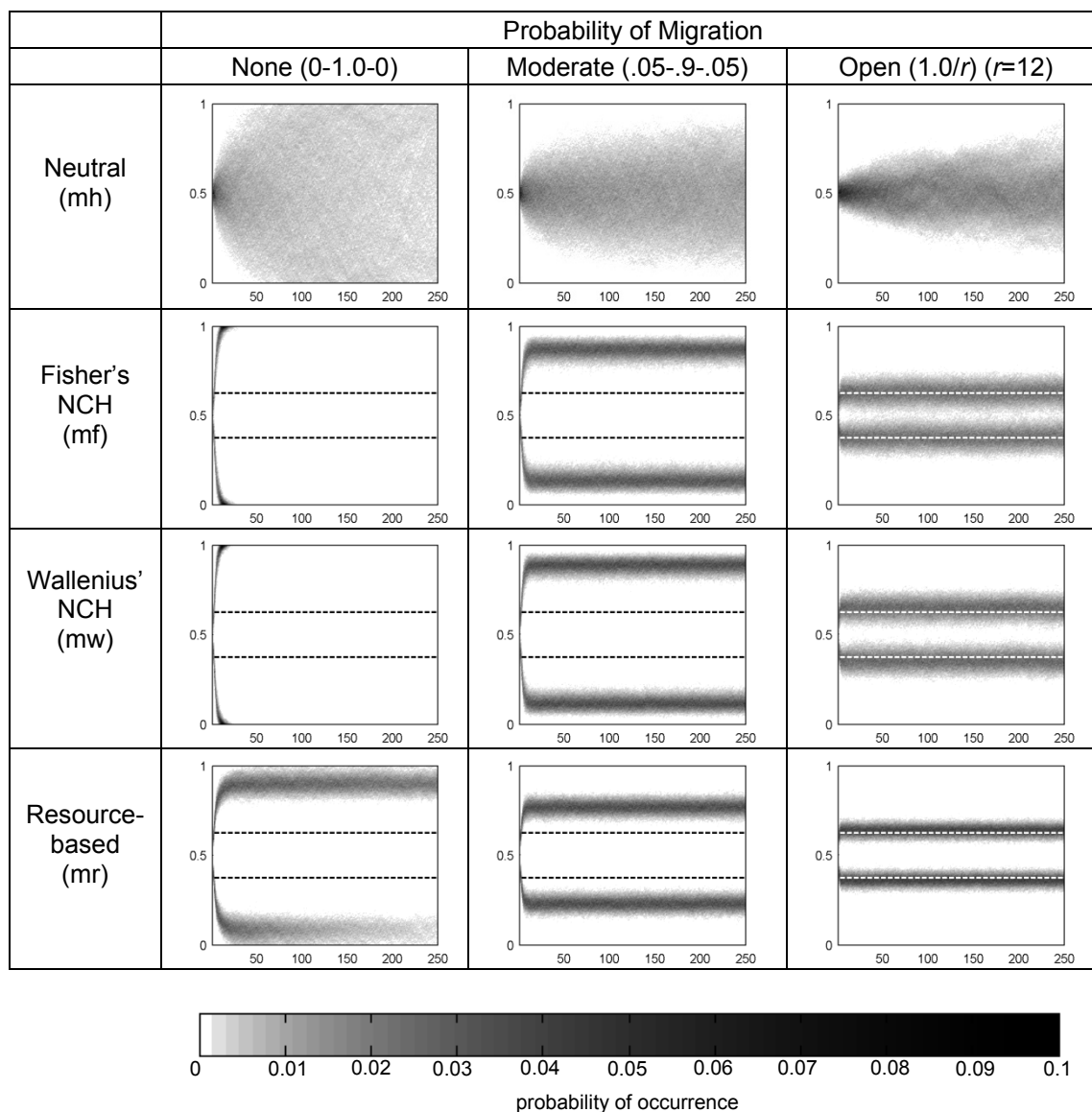
**Fig. 3.1.** Schematic representation of twelve populations and their respective allele frequencies connected by a circulant migration pattern. The **M** matrix contains probabilities of migrating between a source population (columns) and a destination population (rows), and is represented in the population diagram by the arrows between populations (self-recruitment arrows are not shown). The **Q** matrix contains information regarding allele frequencies (columns) in each population (rows). In the population diagram, frequencies of allele type 1 ( $a_1$ ) are represented by shaded pie portions, and frequencies of allele type 2 ( $a_2$ ) are represented by the light pie portions.



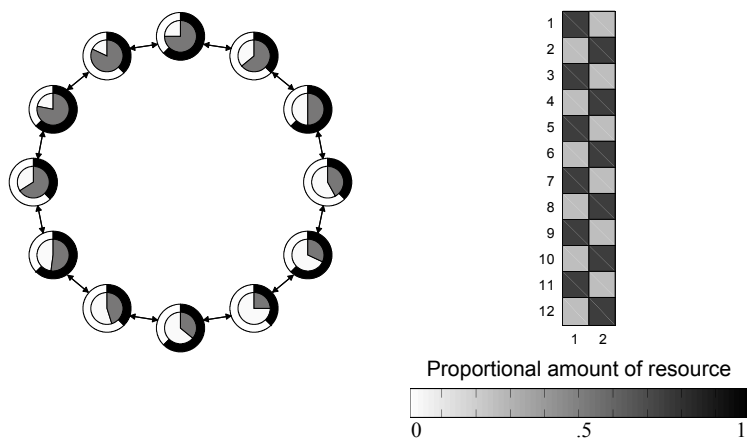
**Fig. 3.2.** Illustration of time-frequency plot construction.



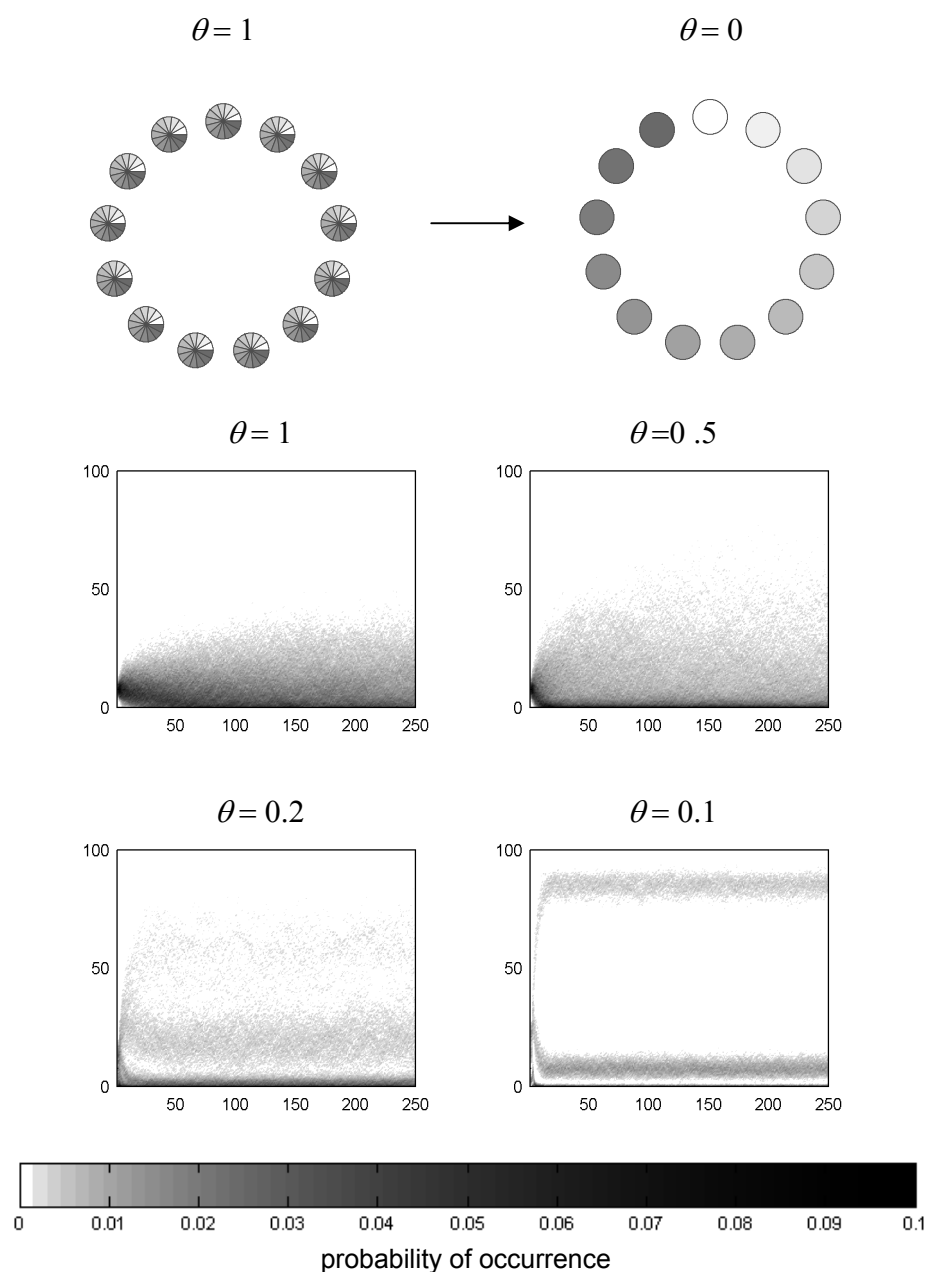
**Fig. 3.3.** Allele frequency probability plots for different levels of stepping stone migration and mortality models. Horizontal axis is time in generations, vertical axis is the frequency of the allele in the population, and the shaded value is the probability of occurrence (at each time step all probabilities sum to 1). Dashed lines indicate the proportions of available resource (as shown in Fig. 3.4), 0.625 in odd-numbered populations, 0.375 in even-numbered populations.



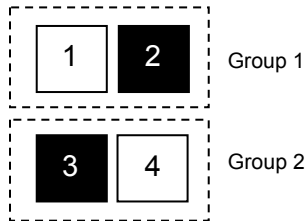
**Fig. 3.4.** Twelve populations and their respective allele frequencies connected by a circulant migration pattern, with the addition of an exterior circle showing relative proportion of available resources. The proportion of resource type I ( $\rho_1$ ) is represented by the dark portion, whereas the proportion of resource type II ( $\rho_2$ ) is given by the light portion. Interior circles represent allele frequencies, identical to Fig. 3.1.



**Fig. 3.5.** Allele frequency probability plots for 13 populations with a circulant migration structure, with 90% self-recruitment and 5% recruitment to either nearest neighbor (mf mortality). Resource quantities were distributed among populations using a discretized normal distribution. The variance parameter ( $\sigma$ ) was determined as  $\sigma = \tan(\pi\theta/2)$ , with  $\theta$  ranging between 0 and 1. Changing the value of  $\theta$  alters the distribution of resources between an equal and even distribution of resources, and one where resources are unique to each population. 13 populations were used instead of 12 since an even number of populations would split the peak into two portions. The number of different allele forms was increased to be equal to the number of populations (13) so that it would be possible to generate completely orthogonal resource vectors.  $\theta=0$  is not shown as all values fall on either the top or bottom boundary line.



**Fig. 3.6.** A representation of how selective weightings may change according to scale. T-I and T-II refer to the proportional amount of Type I (white) and Type II (black) resources respectively. By combining groups (e.g. due to increased range or by pooling populations), the proportion of resources changes accordingly.



Group	Population	T-I	T-II
1	1	1	0
1	2	0	1
2	3	0	1
2	4	1	0

Group	Populations	T-I	T-II
1	1&2	0.5	0.5
2	3&4	0.5	0.5

## CHAPTER 4: SIMULATING THE GENETIC STRUCTURE OF CARIBBEAN CORAL REEF ECOSYSTEMS

### Background

Understanding the development and form of large-scale population genetic structure is an area of growing interest for both scientists and managers (National Research Council (NRC) 2001; Roberts et al. 2003; Hedgecock, Barber, and Edmands 2007). The interest can be attributed in part to improvements in the ability to process large amounts of genetic data using multiple genetic markers (e.g. Richardson et al. 2007), and also to increasing recognition of the importance of spatial aspects of conservation management, such as the degree of connectivity between subpopulations (Cowen et al. 2007).

Evaluating genetic structure at large scales can be challenging due to the effort required to cover large spatial extents with a meaningful degree of resolution. Nevertheless, there have been several studies that have examined genetic patterns of coral reef organisms within the Caribbean. Among the more consistent findings is evidence of a distinct genetic break between eastern and western populations in the Caribbean. This has been observed for populations of *Acropora palmata* (Baums, Miller, and Hellberg 2005) as well as *Montastrea annularis* (Foster et al. unpublished). Additional structure is also apparent within the Bahamas as well. Taylor and Hellberg (2003; 2006) have proposed additional breaks along the length of the Bahamas, and an examination of the gorgonian *Pseudopterogorgia elisabethae* by Gutierrez-Rodriguez and Lasker (2004) identified differences among populations located at Exuma Sound, in outlying areas, and near San Salvador Island. Purcell *et al.* (2006) analyzed microsatellite data for *Haemulon*



*flavolineatum* and found significant, but weak structure. The authors proposed that the weakness in signal strength might be caused by overlapping populations.

In addition to evaluating the form of regional-scale genetic pattern however, it is important to consider the processes that create and maintain them as well. Reconciling the factors believed to be responsible for generating genetic structure with real-world genetic patterns will lead to a better understanding of not only what the important processes actually are, but also the expected consequences if they were to change.

Recently, models have been developed that couple oceanographic data with biological behavior of marine larvae (Cowen, Paris, and Srinivasan 2006). Using these models in conjunction with matrix analysis, it is possible to determine how genetic structure is expected to develop as a result of natural processes. This is accomplished by simulating the dispersal of marine larvae according to their own unique life-history traits, generating transition probability matrices, and projecting them through time to evaluate the expected development of genetic structure through time. Coral reefs and their associated populations provide an ideal environment for applying this approach. Coral reef communities are discrete habitats, and can be identified with a high degree of resolution over large spatial extents using remotely sensed imagery (Andréfouët et al. 2006). Corals and many of the organisms associated with them have a sessile adult stage and a pelagic larval stage, minimizing the potential for confounding effects from active behavioral migration as adults. Furthermore, taken on a regional level (e.g. the Caribbean) they form a closed system, with no immigration or emigration. Coral reefs also have considerable commercial importance as ecotourism destinations, and are the target of many marine protected area efforts (ISRS 2004).

I develop here an analytical means of using matrix analysis to evaluate the expected development of genetic structure in discretely subdivided populations. Although the approach is built on existing methods (Bodmer and Cavalli-Sforza 1968; Fu, Gelfand, and Holsinger 2006), because new recruits merge into existing populations rather than replacing them, changes to the form of the model were required. To provide biological context, matrices derived from a bio-oceanographic larval dispersal model were used to project the expected development of genetic structure in Caribbean coral reef ecosystems. The resulting patterns conform to field-based observations of genetic structure, suggesting that for some species, contemporary migration patterns may play a significant role in generating some of the genetic structure evident in Caribbean coral reef populations. Elasticity analysis is also used to identify critical areas for connectivity in the region.

## **Methods**

A GIS data layer of coral reefs and associated habitats were developed as part of the Millennium Reefs Assessment (Andréfouët et al. 2004). The data was restricted to Caribbean environments that would be typically considered as coral reef (i.e. patch reef, spur and groove, not seagrass or shelf). The subset was buffered to a distance of 5 kilometers to account for the ability for larvae to engage in directed settlement behavior, and the resulting coverage was then split into discrete polygon units following Paris *et al.* (2005) using a tolerance level of 10 km (Fig. 4.1). Tolerance in this case is defined as the minimum allowable distance between any two vertices along an arc, and was used to

provide a balance between preserving perimeter length and polygon area. At the time at which the polygon layer was developed, Millennium Coral Reef data was not available for Venezuela. Coral reef location information developed by UNEP-WCMC (Spalding, Ravilious, and Green 2001) was used for this area instead.

Transition matrices describing the probability of migrants moving from one population to another (or self-recruiting) under different life-histories were constructed using a Java-based port of the bio-oceanographic dispersal model developed by Cowen, Paris and Srinivasan (Cowen, Paris, and Srinivasan 2006). Current velocity information was obtained from the Hybrid Ocean Coordinate Model (HYCOM), using daily offline output for the time interval January 2, 2003 through December 25<sup>th</sup>, 2005. 1000 individuals were released from each population at 30 day intervals. Individuals became competent to settle after a period of 15 days (S1) or 5 days (S2), and were monitored for up to 30 days. A mortality rate ( $z$ ) of 0.2 per day was also applied to each transiting individual (Houde 1989). The results were combined into a single transition matrix describing the number of survivors from a given source population (columns) arriving at the designated destination population (rows). Another way to interpret the matrix is that the row index indicates the position of the base of an arrow and the column index indicates the position of the tip of an arrow connecting the two populations. The value of the matrix at the intersection indicates the strength of the connection.

To project expected genetic structure forward in time, a modified version of the matrix-based approach developed by Bodmer and Cavalli-Sforza (Bodmer and Cavalli-Sforza 1968) was used. The transition matrix  $\mathbf{M}$  is applied to a state matrix  $\mathbf{Q}_t$  to yield the expected state of the population at time  $t+1$ , where  $\Delta t$  encompasses a discrete

reproductive cycle. Through recursive substitution, it can also be shown that  $\mathbf{Q}_t = \mathbf{M}'\mathbf{Q}_0$ . This assumes that all elements of  $\mathbf{Q}$  undergo transition however, which is not the case here. Instead, it is only newly produced individuals which are transported, which then merge with existing populations. To account for this, the formula may be re-expressed such that

$$\mathbf{Q}_{t+1} = \mathbf{K}(\overline{\mathbf{MBQ}_t + \mathbf{Q}_t})_{i\bullet}. \quad (\text{Eq. 1})$$

where  $\mathbf{K}$  is a diagonal matrix of carrying capacity values for each population. This simply states that the state of the system at  $t+1$  is the result of migrating progeny ( $\mathbf{MBQ}_t$ ) combined with the original parent generation ( $+\mathbf{Q}_t$ ), converted to probabilities via row standardization (bar-dot), and multiplied by carrying capacity ( $\mathbf{K}$ ). The conversion of  $\mathbf{MBQ}_t + \mathbf{Q}_t$  to probability values and multiplication by  $\mathbf{K}$  produces expected values for the matrix the same way expected values of the vector-based multinomial distribution are derived by multiplying  $n\mathbf{p}$ , where  $n$  is the number of objects sampled, and  $\mathbf{p}$  is a vector of their probability of being sampled. The ordering of the multiplicative terms is significant. From right to left, they represent the order in which the processes occur: birth first, then migration, followed by scaling to carrying capacity. Using iterative substitution, it can be shown that

$$\mathbf{Q}_t = \mathbf{K}(\overline{\mathbf{AK}})_{i\bullet}{}^{t-1} \overline{\mathbf{AQ}_0}_{i\bullet}, \quad \text{where } \mathbf{A} = \overline{\mathbf{MB} + \mathbf{I}}_{i\bullet}. \quad (\text{Eq. 2 - See Appendix B for proof})$$

If  $\mathbf{Q}_0$  is equal to  $\mathbf{K}$ , as would be the case if individuals were uniquely labeled according to their population of origin, the equation simplifies to  $\mathbf{Q}_t = \mathbf{K}(\overline{\mathbf{AK}})_t$ . If gene frequencies are of interest rather than expected allele counts,  $\mathbf{Q}_t$  may be divided by the carrying capacity  $\mathbf{K}$  in which case the equation further reduces to  $\mathbf{Q}_t = (\overline{\mathbf{AK}})_t$ . The projections would therefore represent the overall probability that genetic material would transition from a source population to a destination population. Note that although convergence over time is determined by the eigenvalues of  $\mathbf{MB}$ , the ratio between  $\mathbf{MB}$  and  $\mathbf{I}$  affects the rate at which convergence occurs. To provide a representative time scale, a reproductive output value of 500 per individual was used based on values for *Stegastes partitus* (~1000 successful hatches per female per spawning event (Cole and Sadovy 1995) - the probability of an individual within a population being male or female was assumed to be equal). Relative carrying capacity was calculated based on reef area available within a given polygon. Relative carrying capacity can be used instead of absolute carrying capacity due to the progressive row-normalization of the equation. The most commonly used genetic markers in population genetic studies are assumed to be neutral, therefore the potential effects of selection were not considered. The role of mutation was also not addressed, however the effects could be incorporated by multiplying the  $\mathbf{MBQ}$  term by an additional matrix (e.g.  $\mathbf{MBQV}$ ) describing the probability of switching from one allele form to another (Fu, Gelfand, and Holsinger 2006). The degree of variance in the projected allele frequencies may be determined using Pollard's model for projecting the moments of a multitype branching process (Pollard 1969; Caswell 2001), though this was not carried out due to the large number of populations. Elasticity analysis (Caswell 2001) was also performed to identify the relative

importance of linkages within the system. Elasticity analysis measures the proportional response of matrix eigenvalues to proportional perturbations using partial derivatives.

Elasticity was calculated as

$$\mathbf{E} = \begin{pmatrix} \phi_{ij} & \delta\lambda \\ \lambda & \delta\phi_{ij} \end{pmatrix} \text{ (Eq. 3)}$$

where  $\Phi = \overline{(\mathbf{AK})}_i$ , and  $\lambda$  is the dominant eigenvalue of the  $\Phi$  matrix. The dominant eigenvalue is the primary factor determining how the matrix changes upon exponentiation, and consequently the changes that exert the greatest effect on it will also have the greatest effect on the system as a whole.

## Results

Transition matrices showing the exponentiation of  $\overline{(\mathbf{AK})}_i$  were generated using the input parameters provided. Individuals tended to settle near to their population of origin, although there were exceptions for the reefs of Cay Sal (< LBB), eastern Cuba, the northern portion of Hispaniola (HPN), and for the chain of reefs along the Nicaraguan Rise south of Cuba, including the Cayman Islands and Swan Island (CCB-JAM) (Fig 4.2a). Decreasing the time to competency of the larvae resulted in a greater proportion of individuals settling back to or near their natal habitat (there was a 54% increase in the probability of settlement to reefs less than or equal to 50 km radial distance from the source). The decrease in the values of the off-diagonal elements and the corresponding

increase in values near the diagonal demonstrate increased retention, although the general form of the matrix remained the same (Fig 4.2b).

Projecting expected genetic connectivity forward through time, links between disparate populations become evident, represented as blocks in the resulting matrices (Fig 4.2c,d). Over time, populations become increasingly genetically connected with one another, and small-scale differences in allele frequencies erode (Fig 4.2d). Even though many of the probabilities of the projected matrices were extremely low, relative differences between populations do persist (Fig 4.2e), providing the basis for the development of genetic structure. The most obvious and persistent feature in the matrices is a break that begins near Puerto Rico, and encompasses the islands of the Lesser Antilles through to the Gulf of Venezuela. The similarity matrix indicates that nested structure is also expected to develop within the Bahamas and in portions of the Western Caribbean. A relationship between southern Cuba, Jamaica and the reefs of the Nicaraguan Rise is anticipated, as is a strong link between the Mesoamerican Barrier Reef and the Florida Keys (Fig 4.3).

The elasticity matrix (Fig 4.2f) indicates that in general, proportional changes to any single population do not have a large effect on the dominant eigenvalue ( $\lambda$ ) of the row-normalized **AK** matrix (average degree of proportional change = 2.85E-7), however the relative pattern indicates that changes to values in the Lesser Antilles (LAN), near Panama (PAN) and central Nicaragua (CCB), and in the Mesoamerican Barrier Reef (MBR) have the greatest influence on  $\lambda$ , and through it, the projected matrix values.

## Discussion

The matrices show the development of expected genetic connectivity patterns in the Caribbean through time, using the input parameters provided. Although populations may not be connected in an obvious manner demographically, they may be linked genetically. In particular, the projections show the formation of regional clusters, including the Lesser Antilles, the Bahamas and northern Cuba, and among the reefs of Panama and the Nicaraguan Rise. The results are dependent on the input parameters used, however there is consistency with runs using other input parameters (e.g. 2b). Individuals persistently recruit back to or near their population of origin, and the structure of the matrix retains a similar overall form, indicating that for Caribbean coral reef ecosystems the overall characteristics of the projections are robust. The projections demonstrate that a genetic break is expected from Puerto Rico down to the Gulf of Venezuela, which is consistent with observations from field studies (Starck and Colin 1978; Taylor and Hellberg 2003; Baums, Miller, and Hellberg 2005) and previous independent simulations (Cowen, Paris, and Srinivasan 2006). Note that although the break appears in the projected matrices at Puerto Rico and the Mona Passage, the distance matrix groups Puerto Rico more closely with Hispaniola and the Western Caribbean. The Mona Passage is an area with highly variable current structure (Metcalf, Stalcup, and Atwood 1977), and the work by Baums *et al.* (2005) identified extensive mixing in this area as well. It is possible that the variability in current structure results in a “leaky” northern boundary. Simulations using specific life-history information as well as longer-term and higher resolution oceanographic data in this area will be required to



resolve the precise nature of the genetic break. There also appears to be support for the existence of a weaker break in the Central Bahamas consistent with observations made by Taylor and Hellberg (2003; 2006). The projections suggest that part of the weakness of this break stems from the fact that allele frequencies change along the length of the Bahamas, forming a gradient, although the sharpest transition does occur at the boundary between the Northern Bahamas (GBW, GBE) and the Southern Bahamas (MBH). For at least some species, it appears that oceanographic currents play a significant role in shaping the genetic structure of Caribbean coral reef communities.

The high degree of spread of genetic material across populations after only a few generations may seem questionable, however the powered matrix values of  $\mathbf{AK}$  indicate the probability that *any* form of genetic material from a given source population reaches the corresponding destination. The probability with which a new mutant would arise in a population would be  $1/2N$  (assuming a diploid population), where  $N$  is the number of individuals in the population (which in this case equals the corresponding diagonal element of  $\mathbf{K}$ ). This results in a decreased probability of transition, and consequently the amount of time it would take for a similar allele to spread to other populations. The model also does not take into account mortality due to intra-population demographic processes. It is likely that juveniles would experience greater mortality than established adults, which would further increase the time required for novel genes to spread. Assigning  $\mathbf{Q}_0$  to be equal to  $\mathbf{K}$  may also seem unusual, since populations are more likely to begin in a mixed state rather than in a unique condition. However, if all populations are completely identical in terms of their composition, for structure to appear, new material would have to be injected into the system (e.g. via mutation). Assuming that the

probability of the mutation being generated is uniform across time and space, then the probability of arising in any particular population will be equal, and if the new material is only generated in a single population at a time, then the result is equivalent to using an identity matrix, i.e. row-normalized  $\mathbf{K}$ .

Given the significant degree of structure expected to be present in the system from the simulations, why are strong breaks between populations not more evident in observations of natural populations (e.g. Shulman and Bermingham 1995; Purcell et al. 2006)? One possible answer to this lies in the relationship between the number of individuals sampled and the probability of sampling the available allele forms. The matrix probabilities are the allele frequencies that would be expected from sampling the populations with a high (theoretically infinite) degree of accuracy. Sampling a limited number of individuals increases perceived distances between populations, and decreases the amount of visible structure (Fig. 4.2f). Structure may be present, but relative differences between populations might not be discernable given the available sampling power. Typically, this problem is addressed by sampling more individuals or loci, and the matrix projections can be used to help identify the sample sizes required to bring genetic structure into focus. It must be noted that the “markers” used by the matrix are idealized in the sense that each different form arises independently in a different population, and all populations have a unique marker associated with them. In reality, some populations may not produce a unique mutation, or there may be homoplasy, reducing the amount of observable structure in the system. It is also likely that additional processes are also contributing variability to natural populations, such as density

dependent effects, trophodynamics or disease, which could dampen structure even further.

Elasticity analysis identifies the Lesser Antilles, Panama, the Mesoamerican Barrier Reef, and the chain of reefs spanning the Nicaraguan Rise as areas that would experience the greatest proportional change due to perturbations in connectivity structure. The importance of the eastern Antilles stems from their position upstream of other populations in the Caribbean; they act as a source to other areas of the Caribbean without themselves being a sink. Panama and the Mesoamerican Barrier Reef appear to be important due to a combination of extensive reef area capable of acting as a reservoir for genetic material, in conjunction with their proximity to gyres capable of dispersing and recirculating that material.

The matrix model provides a powerful and concise means of evaluating the expected genetic structure of subdivided populations through time. Although the form of the equation is similar to other approaches (Bodmer and Cavalli-Sforza 1968; Fu, Gelfand, and Holsinger 2006) there are also some critical differences. Classic models assume that all elements of the  $\mathbf{Q}$  matrix undergo transition, whereas here, migrants merge into previously existing populations. There is an important consequence of this. Although the eventual trajectory of the population is ultimately determined by the combined effects of migration, birth and carrying capacity ( $\mathbf{MB}$  and  $\mathbf{K}$ ), the magnitude of the difference between  $\mathbf{MB}$  and  $\mathbf{I}$  affects the rate at which convergence occurs. Therefore, if the time scale over which genetic structure develops is of importance, demographic factors (i.e. levels of birth and mortality relative to population size) must be taken into account.

An analytical solution for projecting expected genetic structure has several advantages over simulation-based approaches. The results are more accurate and computationally efficient, and can be easily executed using a standard matrix package (e.g. Matlab, Scilab). Although simulation-based frameworks can be robust and easy to conceptualize, the aforementioned importance of demographics with regards to time scale means that to capture the full dynamics of the populations using simulation, it would be necessary to simulate actual numbers of individuals present in the system. For regional-scale studies of marine systems, this could require simulating billions or trillions of individuals through time repeatedly, which would be challenging, even given exponential increases in computing power. An analytical solution simplifies the calculations in the same way that the binomial function simplifies repeated sampling of a binary event. The analytical model does have its own limitations in that the various matrices (**M**, **B** and **K**) are assumed to remain constant through time. Although this is unlikely to be strictly true, if it can be assumed that events are independent of one another, have finite variance, and are generated through an identical process, then the central limit theorem will apply, meaning that the values will be approximately normally distributed about the mean of the set. Evaluating the trajectory of a system in which connectivity changes over time in a complex, autocorrelated manner would likely require simulation (e.g. using individual-based methods).

It is important to keep in mind that the results are projections, not predictions. The model is not designed *a priori* to fit with any particular set of observations. Instead, the model provides an expected set of patterns, and these expected results may be compared with observed results to test whether the assumptions regarding the system are

correct. Here, the assumption is that oceanographic-driven migration is the primary force shaping genetic structure of populations associated with coral reef ecosystems, but the comparisons with existing studies do appear to provide support for this. There are other factors that also have the potential to influence population genetic structure however, including age structure, biased patterns of mutation and demographic bottlenecks that could be explored as well.

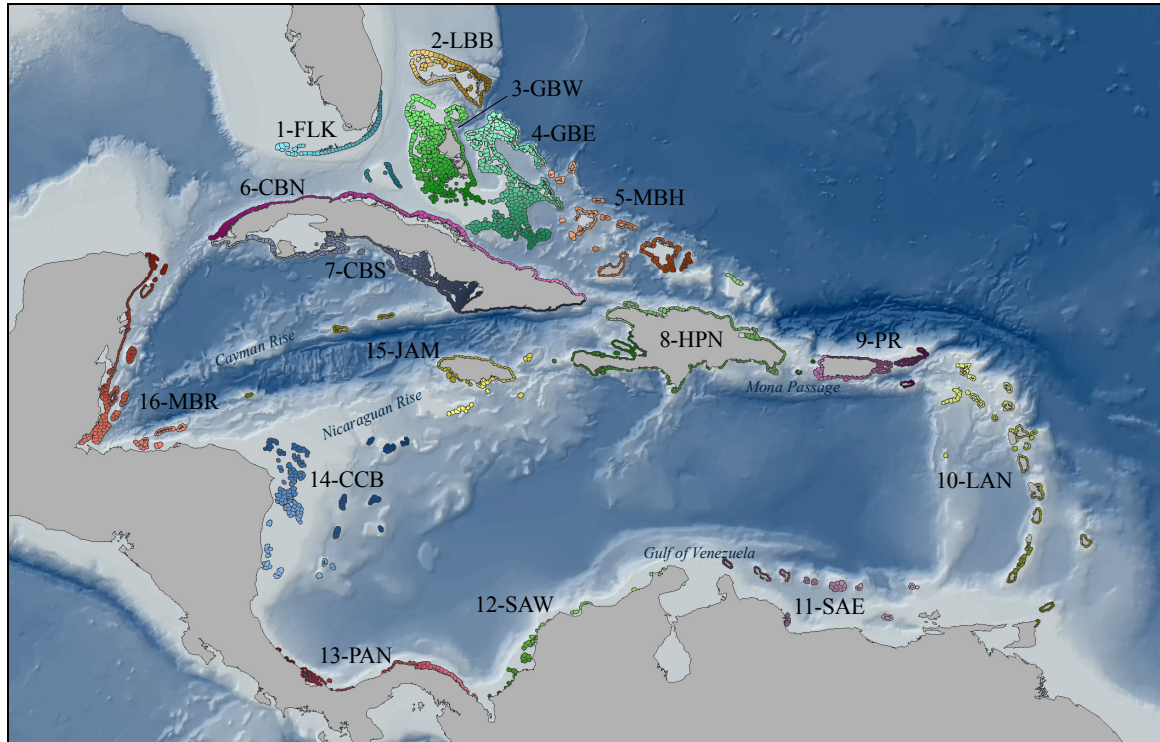
The use of matrix methods in conjunction with the output of bio-oceanographic larval dispersal models provides an effective means of evaluating the expected development of genetic structure over time. With ongoing efforts to evaluate regional-scale genetic structure of marine populations, the ability to cross-validate expectations from physical models with field-based observations will be a valuable tool for studying how genetic diversity is created and maintained in marine populations.

## **Conclusions**

Matrix analysis was coupled with a bio-oceanographic larval dispersal model to evaluate the expected development of genetic structure in coral reef ecosystems. Applying this approach to Caribbean coral reef ecosystems demonstrates a clear expected genetic break between eastern and western portions of the Caribbean, which is concordant with previous simulations and field studies. Other potentially important relationships include linkages among the reefs of the Nicaraguan Rise, and a strong connection between the Mesoamerican Barrier Reef and the Florida Keys. Elasticity analysis suggests that the reefs that would generate the greatest proportional changes in

the matrix in response to perturbations are located in the Eastern Antilles, Panama, the Mesoamerican Barrier Reef and along the Nicaraguan Rise. A matrix-based approach to modeling genetic connectivity provides a concrete means of evaluating the relationship between genetic theory and field-based evidence in regional-scale environments, and provides a powerful way to study and visualize the development of genetic structure in large-scale ecosystems through time.

**Fig. 4.1.** Map of derived polygon boundaries for the Caribbean divided into subregions. FLK – Florida Keys; LBB – Little Bahama Bank; GBW – Grand Bahama West; GBE – Grand Bahama East; MBH – Mid-Bahamas; CBN – Cuba North; CBS – Cuba South; HPN Hispaniola; PR – Puerto Rico and the Virgin Islands; LAN – Lesser Antilles; SAE – South America East; SAW – South America West; PAN – Panama; CCB – Central Caribbean, including the Nicaraguan Rise; JAM – Jamaica, the Cayman Islands and Swan Island; MBR – Roatan and the Mesoamerican Barrier Reef. Shading from light to dark indicates the ordering of the polygons within each region.

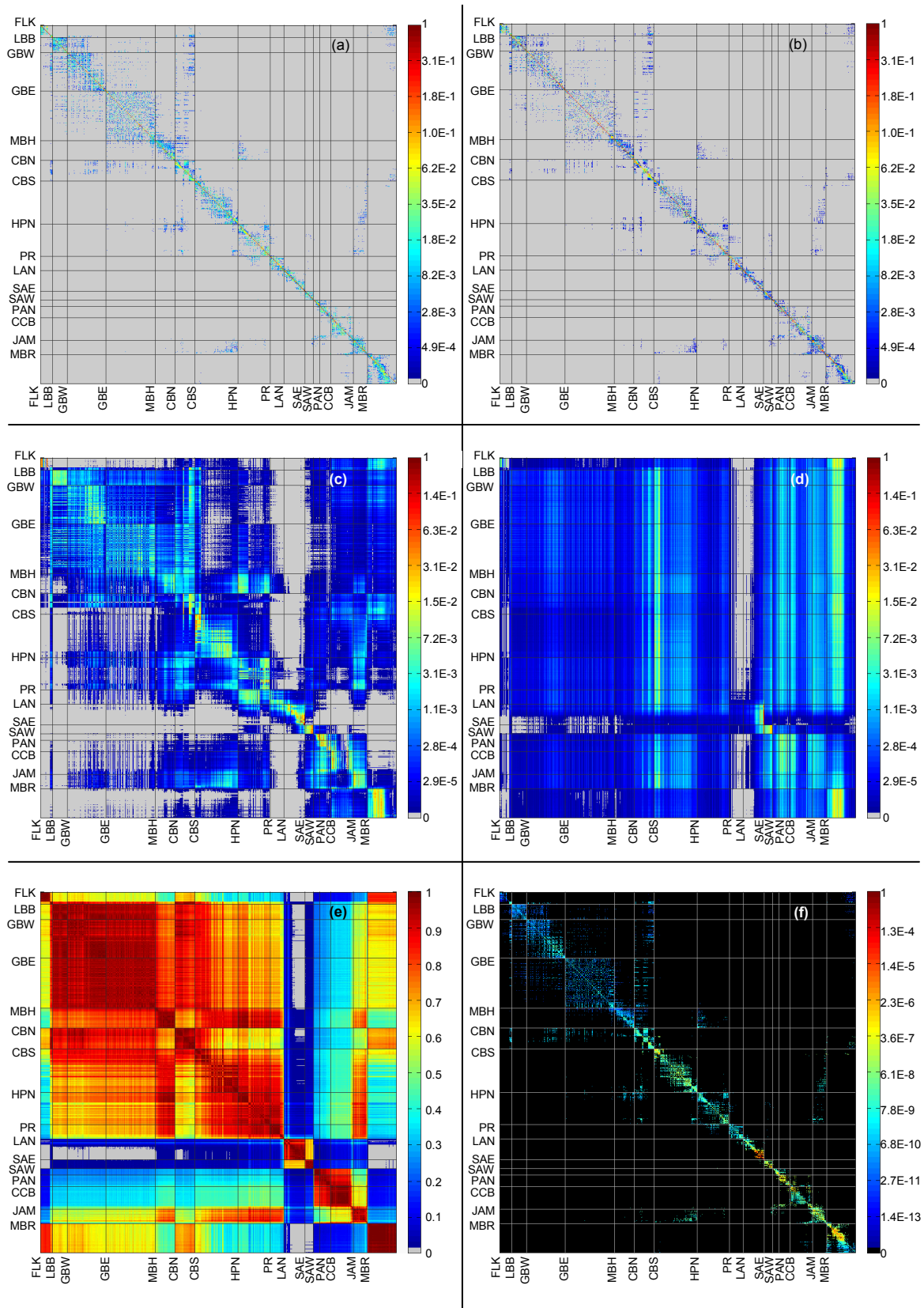


**Fig. 4.2.** Transition matrix, projections and derived results. The entries of the matrix represent the probability of finding genetic material that originated in the source population (columns) in the destination population (rows) at the given time. **(a)** Transition matrix **A** for Simulation 1 (S1): competent to settle at day 15, 30 day pelagic larval duration. **(b)** Transition matrix **A** for Simulation 2: competent to settle at day 5, 30 day pelagic larval duration. **(c)** Probability of allele occurrence at  $t=10$  for S1. **(d)** Probability of allele occurrence at  $t=100$  for S1. **(e)** Similarity matrix for S1 at  $t=100$ .

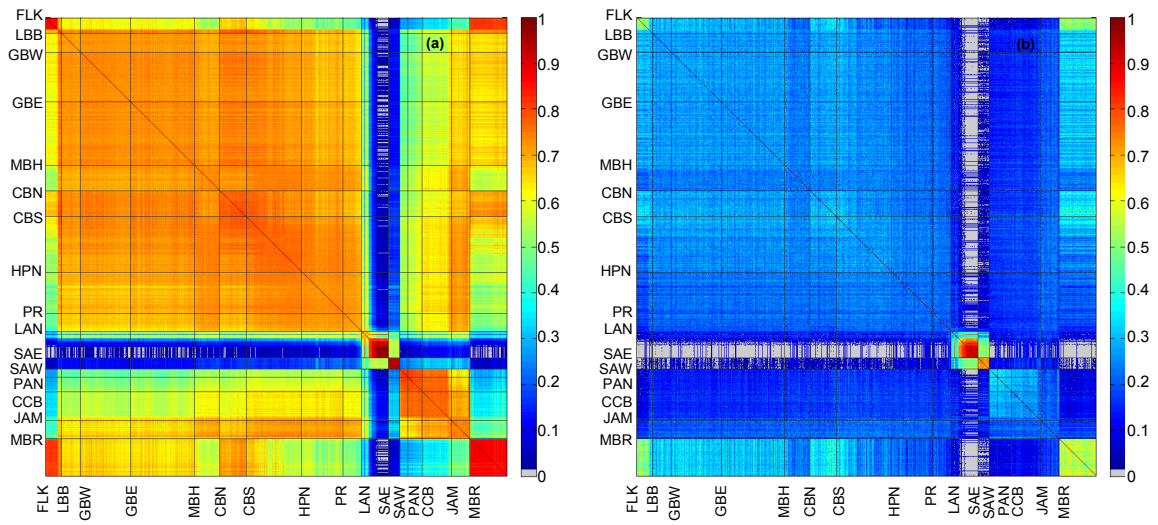
Similarity was calculated as  $\sum_{k=1}^r \sqrt{x_{ik}y_{ik}}$ , where  $x$  and  $y$  are allele frequency vectors of two

different populations (Nei 1987) **(f)** Elasticity values derived from transition matrix **A** for S1. Axis values refer to the subregion beginning immediately after the line break. See Fig. 4.1 for abbreviations.





**Fig. 4.3.** (a) Similarity matrices resulting from randomly sampling S1 at  $t=100$  using 1000 individuals per population (b) using 100 individuals per population.



## **CHAPTER 5: SIMULATING THE GENETIC STRUCTURE OF SOUTHEAST ASIAN CORAL REEF ECOSYSTEMS**

### **Background**

The coral reefs of Southeast Asia are among the most diverse marine communities on earth (Roberts et al. 2002), and are key areas of interest for conservation (Burke, Selig, and Spalding 2002). The processes leading to this high level of diversity remain unclear however. Genetic data from the field has been used to infer connectivity, but the patterns appear to be complicated, possibly involving bi-directional interchanges between reefs of the Indo-West Pacific (IWP) and the open Pacific (Barber and Bellwood 2005). Nevertheless, given the limited resources available for conservation, it is imperative to find ways of identifying coral reef patches that play key roles in creating and maintaining long-term genetic diversity in the region. Otherwise, they may be lost, adversely impacting levels of biological diversity and possibly ecosystem resilience as well (McClanahan, Polunin, and Done 2002; Hughes and Stachowicz 2004). Although obtaining data from the field is essential, it is unreasonable to expect this to be possible for the entire region with a high degree of resolution over a long period of time. Modeling the system offers a more productive approach for understanding how the various populations relate to one another, and how genetic structure develops throughout the region.

With recent developments in coupling biological larval dispersal models to hydrodynamic models, the possibility exists for determining the probability of migration between populations through simulation. Matrices of these probability values can then

be used in conjunction with life-history information to project the expected genetic structure of the system through time, and the expected structure can be compared with field-derived data to test whether assumptions regarding how the system functions are correct. One significant advantage of this approach is that it provides a way of identifying the expected direction of gene flow over time. An additional benefit comes from being able to assess the relative importance of contemporary migration patterns versus historical ones. Benzie (1999) noted that genetic structures of *Tridacna maxima*, *T. gigas*, *T. derasa*, *Acanthaster plancki* and *Linckia laevigata* appeared to be the result of historic dispersal events rather than contemporary gene flow. If this is the case, then one might expect that genetic structure derived from field-based observations will have a significantly different form from one based on the oceanographic-based simulations.

Here, matrix methods coupled with bio-oceanographic larval dispersal models are used to study the development of genetic structure in the Southeast Asian region resulting from oceanographic transport of marine larvae. The matrices are used to determine the expected degree of relatedness between populations, as well as the direction of gene flow. The values can also be converted into diversity measurements to provide an estimate of what the pattern of diversity in the region would be expected to look like on the basis of contemporary migration patterns and unbiased mutation. Elasticity analysis is also used to identify areas that would have the greatest effect on the system if perturbed.

## Methods

A GIS data layer of coral reef locations developed by Spalding *et al.* (2001) was used as the base coral reef layer, and was restricted to an area bounded by 15°S-30°N and 95°E-140°E. The subset was buffered to a distance of 5 kilometers to account for the ability for larvae to engage in directed settlement behavior, and the resulting coverage was then split into discrete polygon units at a tolerance level of 10 km (Fig. 5.1). Tolerance in this case is defined as the minimum allowable distance between any two vertices along an arc, and was used to provide a balance between preserving perimeter length and polygon area.

Transition matrices describing the probability of migrants moving from one population to another (and self-recruiting) under different life-histories were constructed using a Java-based port of the bio-oceanographic dispersal model developed by Cowen, Paris and Srinivasan (Cowen, Paris, and Srinivasan 2006). Current velocity information was obtained from the Hybrid Ocean Coordinate Model (HYCOM), using daily offline output for the time interval January 2, 2003 through December 25<sup>th</sup>, 2005. 1000 individuals were released from each population at 28 day intervals. Individuals became competent to settle after a period of 15 days, and were monitored for up to 30 days, values that are consistent with values for *Halichoeres* sp. of the Pacific region (minimum - 20.8-4.9, average - 24+6.3 (Victor 1986)). A mortality rate ( $z$ ) of .2 per day was also applied to each transiting individual (Houde 1989). The results were combined into a single transition matrix describing the number of survivors from a given source population (columns) arriving at the designated destination population (rows). Another

way to interpret the matrix is that the row index indicates the position of the base of an arrow and the column index indicates the position of the tip of an arrow connecting the two populations. The value of the matrix at the intersection indicates the strength of the connection.

To project expected genetic structure forward in time, a modified version of the matrix-based approach developed by Bodmer and Cavalli-Sforza (1968) was used. A sub-stochastic transition matrix  $\mathbf{M}$  is applied to a state matrix  $\mathbf{Q}_t$  to yield the expected state of the population at time  $t+1$ , where  $\Delta t$  encompasses a discrete reproductive cycle. The individual elements of  $\mathbf{M}$  contain the probability of transitioning (migrating) from population  $j$  to population  $i$ , where  $i$  and  $j$  are row and column indices respectively. Through recursive substitution, it can also be shown that  $\mathbf{Q}_t = \mathbf{M}' \mathbf{Q}_0$ . This assumes that all elements of  $\mathbf{Q}$  undergo transition however, which is not the case here. Instead, it is only newly produced individuals which are transported, which then merge with existing populations. To account for this, the formula may be re-expressed such that

$$\mathbf{Q}_{t+1} = \mathbf{K}(\overline{\mathbf{Q}_t + \mathbf{M}\mathbf{B}\mathbf{Q}_t})_{i\bullet}. \text{ (Eq. 1)}$$

From right to left, the terms represent the order in which the processes occur: birth first, then migration, integration into existing populations, and scaling to carrying capacity.

Using iterative substitution, the recursion formula is determined as

$$\mathbf{Q}_t = \mathbf{K}(\overline{\mathbf{A}\mathbf{K}})_{i\bullet}^{t-1} \overline{\mathbf{A}\mathbf{Q}_{0i\bullet}}$$

$$\mathbf{A} = \overline{\mathbf{M}\mathbf{B} + \mathbf{I}}_{i\bullet}$$

If  $\mathbf{Q}_0$  is equal to  $\mathbf{K}$ , the equation simplifies to  $\mathbf{Q}_t = \mathbf{K}(\overline{\mathbf{AK}})_i^t$ . If gene frequencies are of interest rather than expected allele counts,  $\mathbf{Q}_t$  may be divided by the carrying capacity  $\mathbf{K}$  in which case the equation reduces to  $\mathbf{Q}_t = (\overline{\mathbf{AK}})_i^t$ . Although the eventual outcome is determined by the eigenvalues of  $\mathbf{MB}$ , the ratio between  $\mathbf{MB}$  and  $\mathbf{I}$  affects the rate to convergence. To provide a representative time scale, a reproductive output value of 500 per individual was used, the probability of an individual within a population being male or female was assumed to be equal). The most commonly used genetic markers in population genetic studies are neutral, therefore the potential effects of selection were not considered. The role of mutation was also not addressed, although the effects could be incorporated by multiplying the  $\mathbf{MBQ}$  term by an additional matrix (e.g.  $\mathbf{MBQV}$ ) describing the probability of switching from one allele form to another (Fu, Gelfand, and Holsinger 2006). If it can be assumed that there is no spatial or temporal bias to the mutations however, then the mutation matrix would contain identical values, and would be effectively eliminated upon row-normalization. Similarity matrices were calculated as  $1 - D_A$ , Nei's genetic distance (Nei 1972),  $S_{i,j} = \sum_{k=1}^s \sqrt{q_{ik}q_{jk}}$ , where  $q$  is an individual element of  $\mathbf{Q}$ ,  $i$  and  $j$  are different row indices,  $k$  is a column index, and  $s$  is the number of columns. The similarity/distance values were used in conjunction with UPGMA (unweighted pair group method with arithmetic mean) clustering to generate a spatial map of areas expected to have similar genetic composition. The projected genetic matrix was also converted into diversity values by calculating one minus the concentration value for a given population (Legendre and Legendre 1998),  $1 - \sum_{j=1}^s q_{ij}^2$ . Elasticity analysis (Caswell 2001) was carried out to identify the relative importance of linkages within the

system. Elasticity measures the proportional degree of change in matrix eigenvalues generated by proportional changes to individual matrix elements. The results can be used to identify which populations are critical connectivity junctions, and should be preserved if current conditions are to be maintained.

## **Results**

The transition matrix shows that a high overall level of self-recruitment is expected within the region (Fig. 5.2a). Off-diagonal elements are evident in areas of exchange between the Spratly Islands (SIS), Palawan (PLW) and the Sulu Archipelago (SAR). Projecting the genetic exchange matrix through time (Figs 5.2b and c), two extensive blocks become evident, the first consisting of reefs of the South China Sea (SCS; SCW-SAR), the second consisting of reefs of the Coral Triangle (CT), loosely defined as the area bounded by the Philippines, Borneo (Kalimantan) and New Guinea (SAR-NGN). Reefs in the Gulf of Thailand, the Gulf of Tonkin and the Southwestern edge of Hainan, as well as those in the far northeast of the study region near Taiwan and the Ryukyu Islands do not appear to contribute significantly to genetic structure in the South China Sea. The reefs of Vietnam do appear to play an active role, interacting with the Spratly Islands, the Philippines and the Sulu Archipelago. There is an obvious asymmetry in the projected matrices, indicating that genetic material from the SCS is expected to flow into the CT, but not the reverse, with the exception of some contributions from reefs in the southern Philippines. The key connection between the two regions appears to be the reefs of the Sulu Archipelago. This can be confirmed using



kernel plots for particles released from the Spratly Islands, Sulu Archipelago and Makassar Strait (Fig 5.3). The majority of the larvae originating in the Sulu Archipelago appear to be prevented from reaching Palawan by a cyclonic gyre in the Sulu Sea, whereas larvae originating in the SCS are able to pass through the Balabac Strait into the Sulu Sea, where they merge into flows leading through the Archipelago and continuing on southwards or eastwards.

In addition to the macro-scale break between the SCS and the CT, nested structure is expected within these regions as well (Fig. 5.2b,c,d). In the South China Sea, the Philippines form a distinct block, and substructure is associated with several of the islands. Within the Coral Triangle, the reefs of the Makassar Strait (MKS) show divergence from other reefs of the CT region. Halmahera and Ceram also appear to form a cohesive unit, although connectivity drops off towards the southern islands in the eastern portion of the Banda Sea (BSE). The associations with northern New Guinea reefs (NGN) are strong however, suggesting that connectivity is stronger along the northern coast. This is reflected in the kernel plots (Fig 5.3), which show Halmahera and Northern New Guinea connected by an anticyclonic gyre. The reefs of the Andaman Sea and southern Java are expected to be highly isolated from other areas within the region.

Plotting the major clusters (Fig. 5.4) reveals a number of distinguishable regions, including: the Spratly Islands, Halmahera, Southern Sumatra, Southern Java, Flores Sea, Makassar Strait, Eastern Sulawesi and Banda Sea, as well as reefs of the northwest Arafura Sea and various subdivisions within the Philippines. The results show some correspondence with field observations made by Ablan *et al.* (2002), which indicated four major groups, the first two composed of reefs in the north and south of the SCS

respectively, the third made up of reefs the southern Philippines, Borneo and the eastern Indonesian Islands, and the fourth comprised of the eastern edge of the Philippines and outlying Pacific Islands based on 16 populations of *Dascyllus trimaculatus*.

Converting the expected genetic structure into diversity values (Fig. 5.5) indicates that the areas expected to accumulate the greatest amount of diversity are the reefs near Halmahera and northern New Guinea. This result is primarily a function of these areas existing in convergence zones. The CT region also appears to develop a greater amount of diversity over time than the SCS, due in part to the unidirectional flow of genetic material from the SCS into the CT.

Elasticity analysis of the projection matrix (Fig. 5.2e) shows that the reefs of the Sulu Archipelago are expected to have a significant effect on the dominant eigenvalue of the transition matrix and consequently the genetic diffusion process as whole, however Philippine reefs appear to have an even greater degree of influence. Populations associated with the Strait of Malacca have the least effect on the system.

## **Discussion**

The projected matrices provide insight into how genetic material is expected to spread throughout the region under current conditions. Of considerable significance is the possible one-way link observed between the South China Sea and Coral Triangle regions. The expected unidirectional flow of genetic material from the SCS into the CT indicates that the SCS may be acting as an upstream source of genetic diversity, in the same way that the Lesser Antilles are an upstream source of genetic material in the

Caribbean. It is possible that some of the biological diversity found within the CT may be supported in part by the Spratly Islands and other areas within the SCS. As a result, conservation of reefs in the SCS may extend beyond being a local concern to one affecting the entire IWP. A more detailed examination of this connection using high-resolution oceanographic models and sophisticated larval behavior is certainly warranted.

The results also indicate that extensive genetic connectivity is expected among coral reefs throughout the Southeast Asian region, and the development of strong structural elements within the South China Sea region and the Coral Triangle is anticipated as well. The lack of connectivity between areas east and west of the Strait of Malacca suggests that contemporary oceanographic conditions work against the introduction of genetic material from outlying areas in the west (e.g. Andaman Islands, Strait of Malacca, western Sumatra). The IWP has been cited as a possible area of accumulation (Jokiel and Martinelli 1992; Pandolfi 1992), meaning that the region is expected to be an area of convergence for dispersal. In order for the IWP to be gaining genetic material from external areas, migrants would need to be entering into the region principally from the east, however the transition matrix and kernel plots show that material is generally expected to be transported eastwards out of the region. Given the degree of mortality expected over large distances, it seems unlikely that neutral genetic material originating in the open Pacific is swamping the genetic content of the IWP, at least for species with a relatively short pelagic larval duration. It is plausible that the islands of the open Pacific may be acting as independent reservoirs of genetic material, and that some bi-directional exchange with the IWP takes place, however it appears as though the majority of the genetic exchange is expected to occur within the boundaries

of the IWP. Increasing the extent of the simulations to incorporate coral reefs of the open Pacific would provide a concrete means of addressing this question.

It must be emphasized that the results only pertain to expectations based on contemporary conditions, not what has taken place in the evolutionary past. To do otherwise would violate the assumption that the migration matrix is stationary through time, since major changes in oceanographic current structure have occurred over the long-term from sea level change and tectonic shifts. Simulating long-term changes is possible, however it would require using a paleo-oceanographic model, which would be complicated by the need to account for shifting bathymetric conditions.

According to the model, the areas expected to develop the greatest amount of genetic diversity occur near Halmahera and the Banda Sea. This appears to primarily result of this area being a convergence point for organisms leaving the IWP, either from the Celebes Sea in the north, or the Banda Sea to the south. It must be emphasized that genetic diversity does not map directly to biological diversity. The former occurs within a single species, and is the result of a combination of mutation, migration, drift and selection. The latter is a function of speciation, migration and evolutionary processes, but is also shaped by other forces such as interspecific competition and niche availability, and is also is the cumulative result of a diverse array of life history patterns. Nevertheless, there are some common elements, in that mutation and speciation are similar processes, each resulting in the appearance of a new object type within a population, albeit over different time scales. Because genes are carried by individuals, the migration pattern of genes would also be expected to correspond with the migration pattern of individuals. Because of this, the predictions for genetic diversity can be regarded as a type of neutral

model for biological diversity. If overall genetic structure is robust with respect to changes in life history patterns, then the model can be used to determine where biological diversity is expected to arise on the basis of migration patterns.

Diversity maps have been developed for the IWP for a variety of species associated with coral reef ecosystems (Veron 2000; Spalding, Ravilious, and Green 2001; Allen 2002; Roberts et al. 2002; Mora et al. 2003), and a consistent feature is that they the CT region as the primary location of diversity. The map of expected genetic diversity reflects this pattern as well, suggesting a role for contemporary oceanographic patterns in shaping biological diversity. One must be very cautious with this interpretation however, since diffusion of biological diversity operates on a much longer time scale than the diffusion of genetic diversity, and the assumption of a stationary migration matrix based on short-term simulation becomes much less tenable. There are also discrepancies between the model pattern and some of the existing studies. Work by Carpenter and Springer (2005) identified maximum levels of diversity near Luzon and the central islands of the Philippines as opposed to near Halmahera and the Banda Sea. Furthermore, Sumatra was also shown to have high levels of diversity, which was not the case with the current model. It is possible that the differences might be due to underrepresentation of some areas in the field sampling scheme due to lower levels of accessibility to some reef areas, however another distinct possibility is the influence of historical effects, such as eustatic sea-level fluctuations (McManus 1985; Potts 1985). Without a doubt, patterns of biological diversity in Southeast Asia are the result of many complex and interacting processes (Hoeksema 2007) and must be studied accordingly,

however the matrix projections are able to provide some quantitative insight into the potential role of contemporary connectivity and demography in generating diversity.

The formation of multiple clusters throughout the area, such as in the Philippines, in the Makassar Strait, in the Flores Sea and in the vicinity of Halmahera and the Banda Sea demonstrates that even given the production of large numbers of larvae throughout the region, there is still a considerable degree of regionalization, considerably more so than is evident in the Caribbean (previous chapter). It is likely that adding age structure into the model would only strengthen the expected degree of structure even further. The development of populations with a greater degree of panmixia would likely require longer pelagic larval durations, specific pelagic larval behavior, weaker settlement behavior, or possibly a combination of all three.

The elasticity matrix provides information on the relative influence of connectivity elements on the dominant eigenvalue of the projection matrix. Although the importance of Palawan and the Sulu Archipelago were previously observed, the coral reefs of the Philippines also appeared to have significant influence on the system, possibly due to being part of a tight, discretely linked network. The Philippines are also situated near an area of strong oceanographic divergence, and larvae transported from the Philippines can enter into the Sulu Sea and transported into the Makassar Strait, westwards into the South China Sea, or south towards Halmahera and the Banda Sea.

It is important to also point out that the results are projections, not predictions. The model results indicate what is expected of the system given particular input parameters. Field data is vital for comparison with the model. Having better data regarding current genetic structure would allow research to move beyond simply looking

at general dispersal patterns, and more towards actual predictive scenarios. Pattern is often used to infer process, but this assumes that a model exists that includes the relevant factors, and has the relationships between them properly defined.

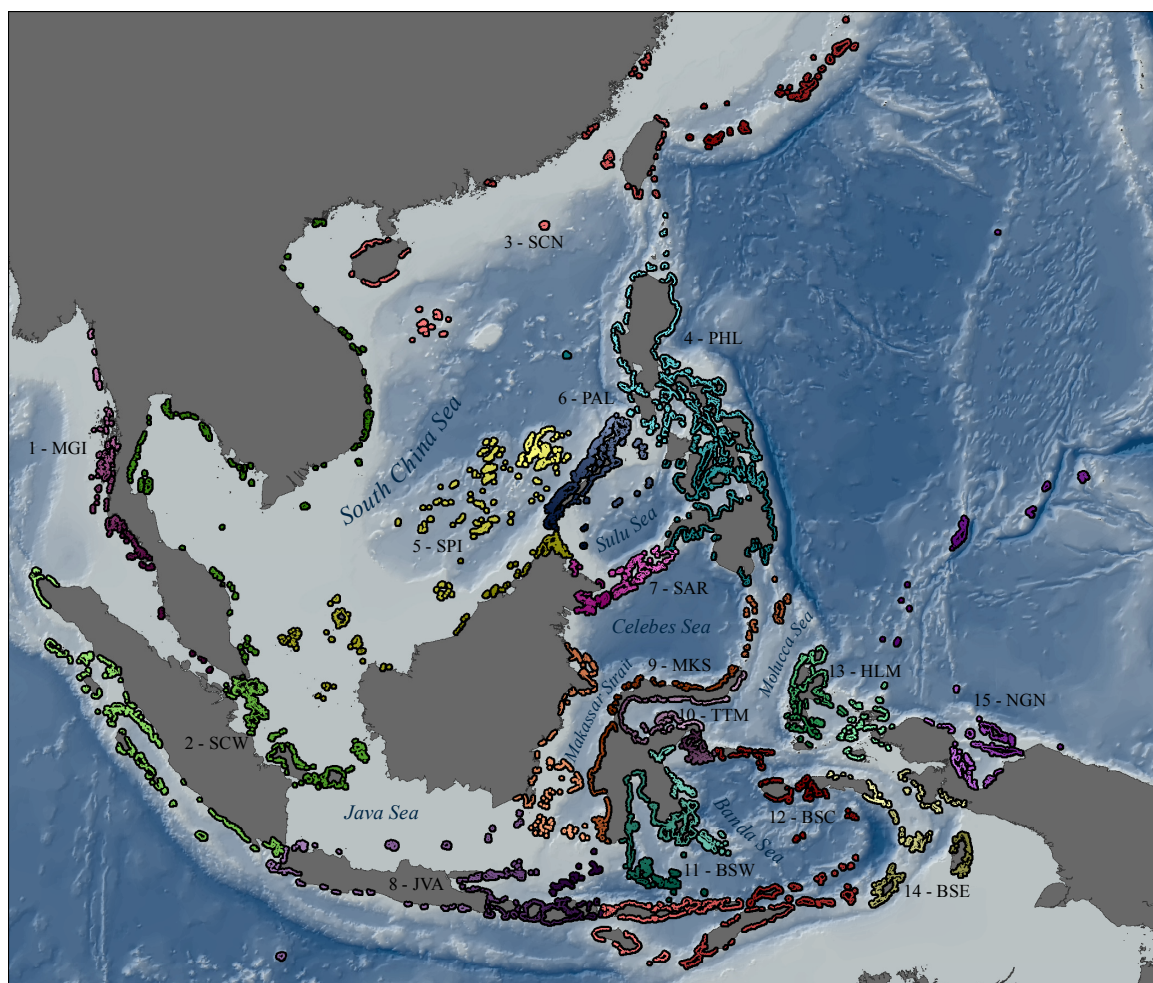
## **Conclusions**

The reefs out Southeast Asia have complex demographic and genetic connectivity structure. The genetic projections reveal that the flow of genetic material between the South China Sea and Coral Triangle is expected to be primarily unidirectional, with the Balabac Strait acting as a one way latch. Larvae are transported southwards through the Sulu Archipelago and into the Makassar Strait, and appear to be prevented from moving into the South China Sea by a cyclonic gyre in the Sulu Sea. The projected matrices also show the development of a number of regional clusters including groups of reefs near Halmahera, in the Flores Sea, in the Philippines, the Makassar Strait and in the Spratly Islands, among others. Converting the projection matrices into diversity values reveals that areas expected to develop the greatest levels of neutral diversity lie in convergence zones, such as in the Flores Sea, the Sulu Archipelago and near Kepulauan Sangihe. Elasticity analysis shows that critical populations in terms of maintaining matrix structure are found in the Philippines, including the central islands as well as the Sulu Archipelago. The use of bio-oceanographic larval dispersal models in conjunction with matrix methods for projecting gene flow through time presents a new and powerful means of studying the processes underlying the development of genetic structure, explicitly demonstrating the

expected direction of gene flow as well as the natural development of clusters under contemporary oceanographic conditions.

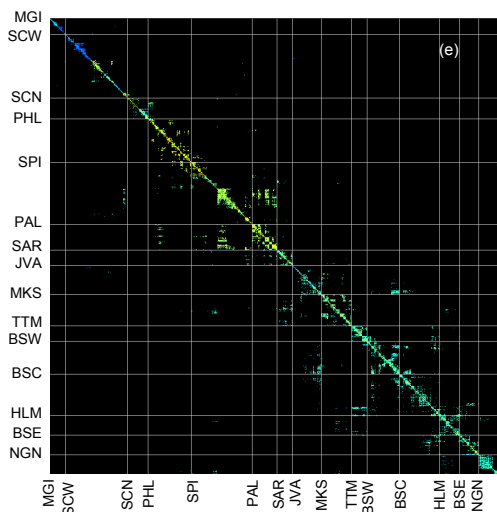
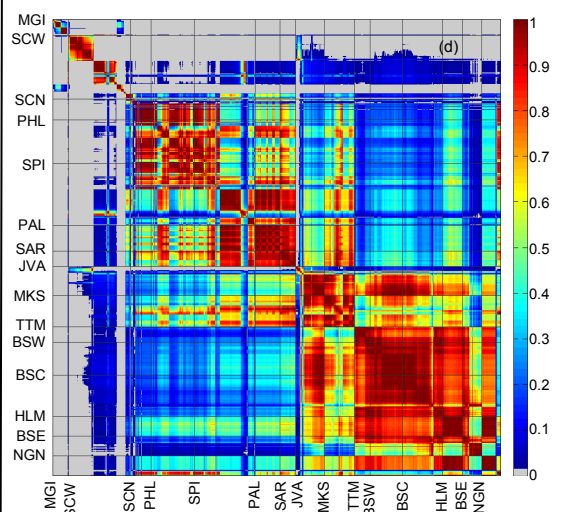
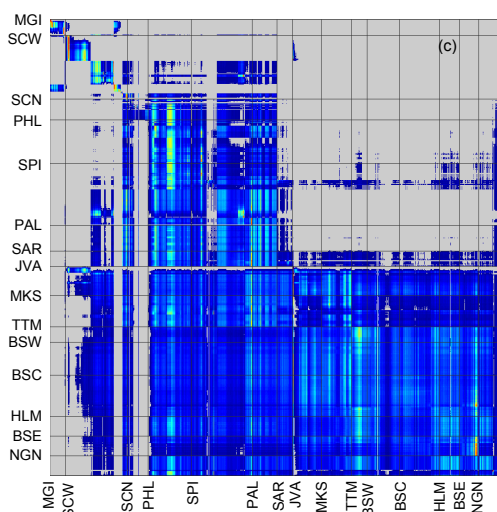
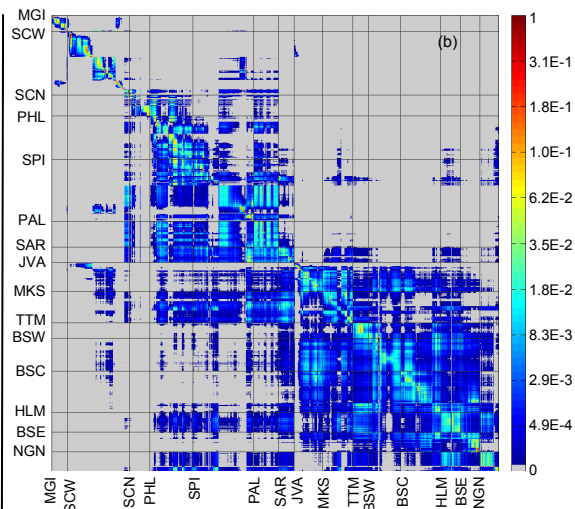
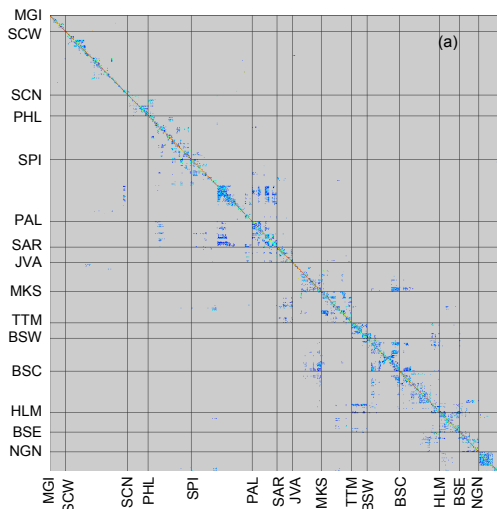


**Fig. 5.1** Map of derived polygons for Southeast Asia divided into subregions. Because of their high resolution, the dividing lines between polygons are not shown, however the polygons are shaded from light to dark according to their order. MGI – Mergui Archipelago, SCW – Sumatra and South China Sea West, SCN – South China Sea North, PHL – Philippines, SPI – Spratly Islands, PAL – Palawan, SAR – Sulu Archipelago, JVA – Java, MKS – Makassar Strait, TTM – Teluk Tomini, BSW – Banda Sea West, BSC – Banda Sea Central, HLM - Halmahera, BSE – Banda Sea East, NGN – New Guinea and the open Pacific.



**Fig. 5.2.** Transition matrix, projections and derived results. The entries of matrices a-c represent the probability of finding genetic material that originated in the source population (columns) in the destination population (rows) at the given time. **(a)** Transition matrix **A** for Simulation 1 (S1): competent to settle at day 15, 30 day pelagic larval duration. **(b)** Probability of allele occurrence at  $t=10$ . **(c)** Probability of allele occurrence at  $t=100$ . **(d)** Similarity matrix at  $t=100$ . The individual entries represent the degree of similarity between the row and column populations, and is symmetric.

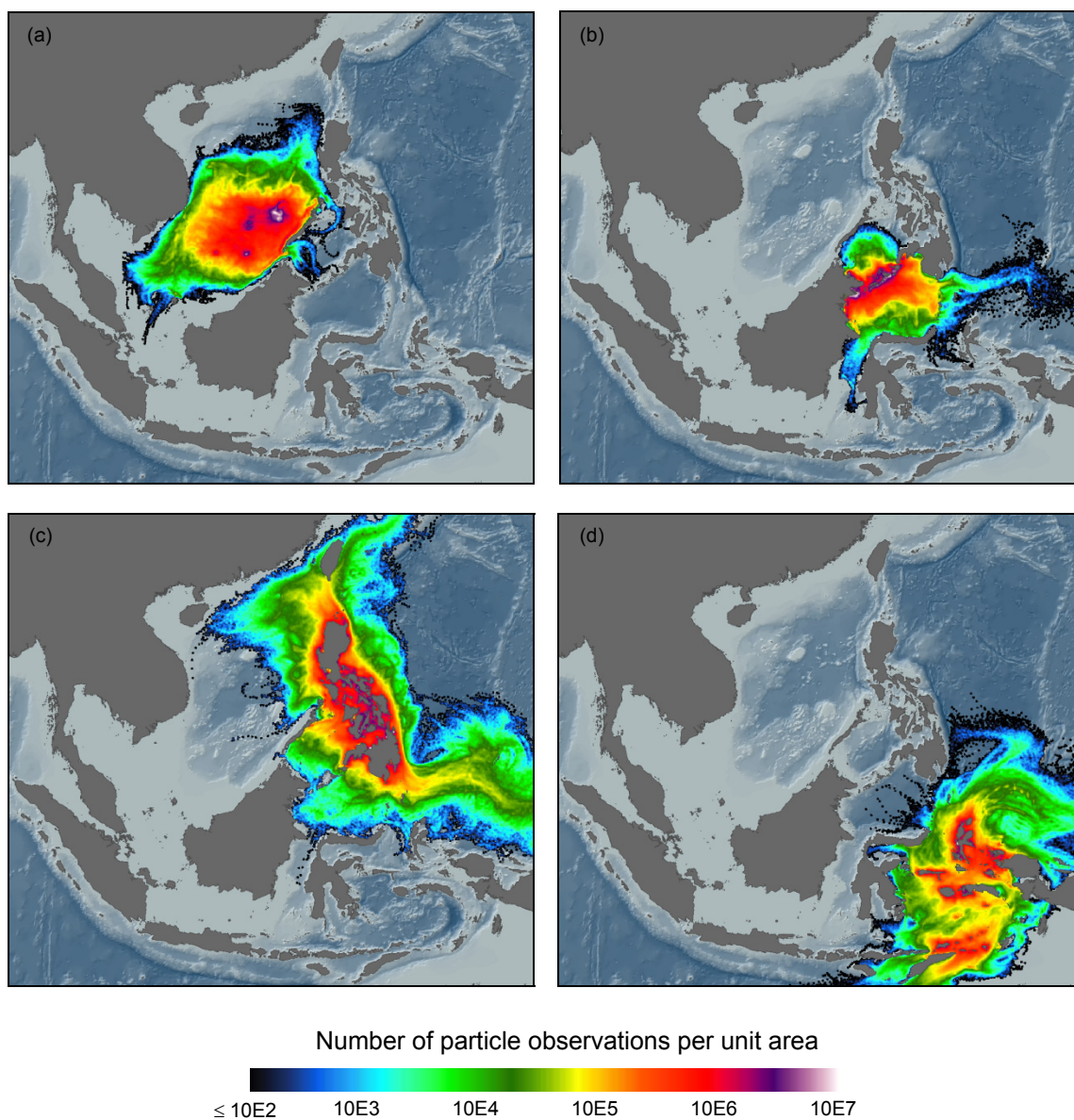
Similarity was calculated as  $\sum_{k=1}^r \sqrt{x_{ik}y_{ik}}$ , where  $x$  and  $y$  are allele frequency vectors of the two different populations (Nei 1987). **(e)** Elasticity magnitudes derived from transition matrix **A**. Axis values refer to the subregion beginning immediately after the line break. See Fig. 5.1 for abbreviations.



**Abbreviation List**

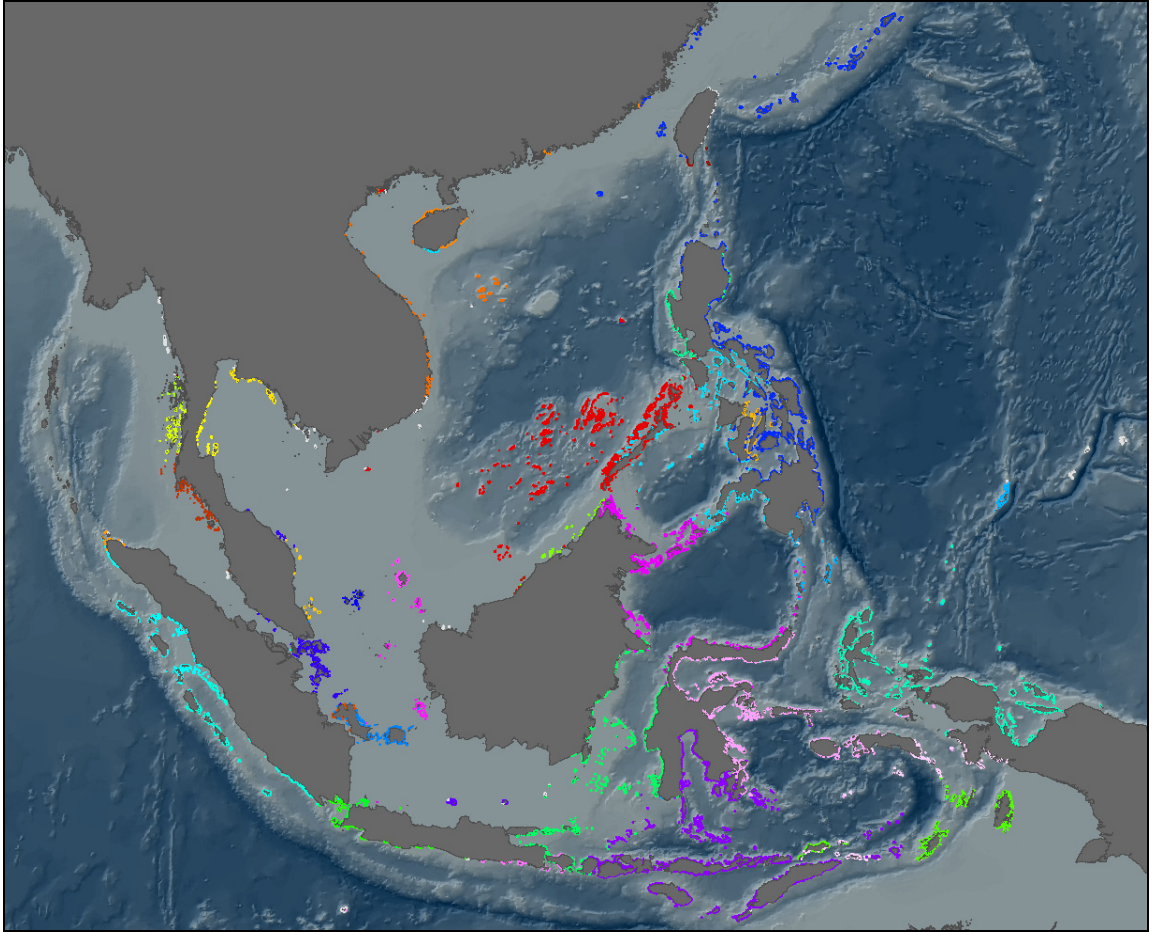
- MGI Mergui Archipelago
- SCW Sumatra and South China Sea West
- SCN South China Sea North
- PHL Philippines
- SPI Spratly Islands
- PAL Palawan
- SAR Sulu Archipelago
- JVA Java Sea
- MKS Makassar Strait
- TTM Teluk Tomini
- BSW Banda Sea West
- BSC Banda Sea Central
- HLM Halmahera
- BSE Banda Sea East
- NGN New Guinea North and Open Pacific

**Fig. 5.3.** Kernel density plots for selected regions in Southeast Asia. Kernel Density plots were constructed by calculating the density of particle tracking points eligible for settling (i.e. age  $\geq 15$  days) over a .1 decimal degree (dd) radius. Note that area measurements based on dd are not strictly equal, but since the region is near the equator, the distortional effects are minimal. (a) release locations in the Spratly Islands (b) release locations in the Sulu Archipelago (c) release locations near the main Philippine Islands (d) release locations near Halmahera, Ceram and the Banda Sea.

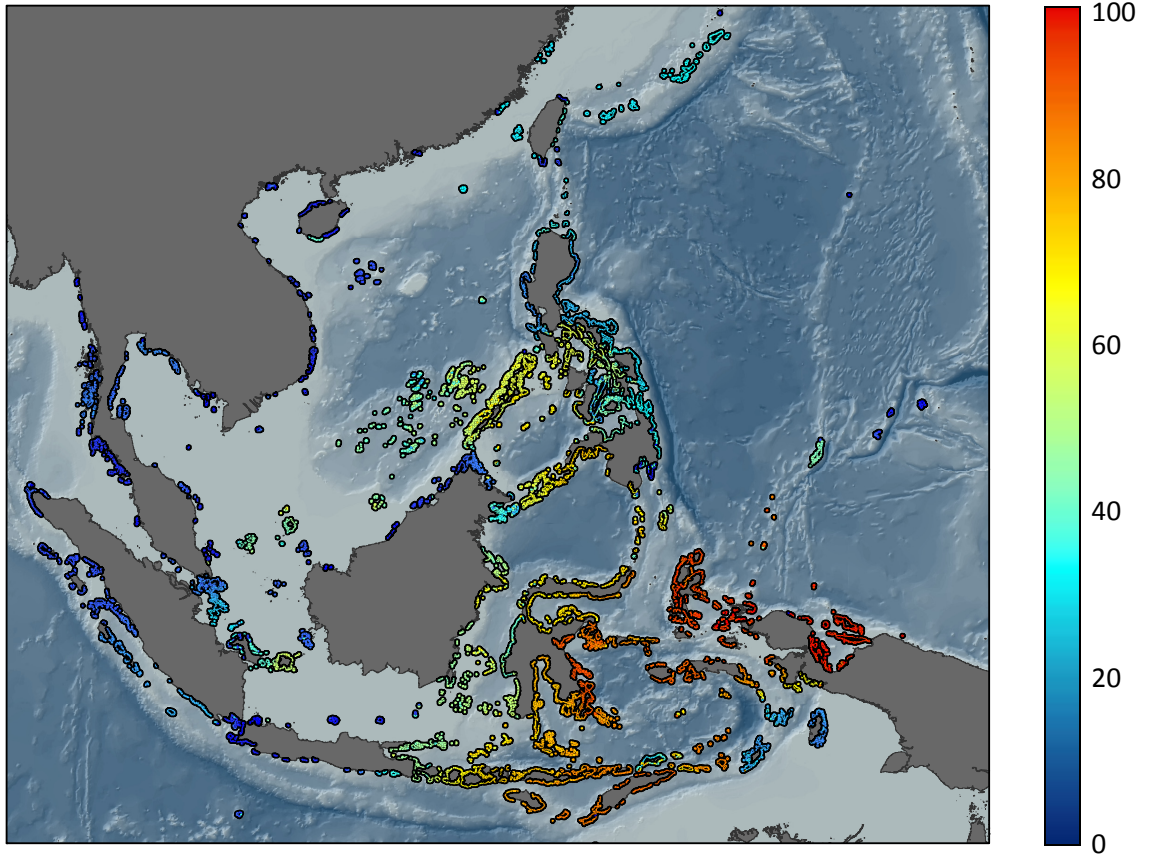




**Fig. 5.4.** UPGMA clustering of the similarity matrix (Fig. 5.4d) at  $t=100$ .



**Fig. 5.5.** Expected diversity values at  $t=100$  as percentiles for Southeast Asian coral reef communities. Diversity was measured as one minus the concentration value for a given population (Legendre and Legendre 1998),  $1 - \sum_{j=1}^s q_{ij}^2$ .



## CHAPTER 6: SYNTHESIS AND CONCLUSIONS

The common thread binding the elements of this dissertation together is the relationship between connectivity and genetic structure. The first chapter examined the manner in which genetic exchange occurs between populations on an individual basis, and the approach was linked to existing matrix-based theory in the context of spatial-temporal autocorrelation. The individual-based system was then used to evaluate fundamental relationships between connectivity structure and spatial-temporal genetic structure. Scalar changes were shown to only affect the temporal domain, whereas changes to matrix structure had both spatial and temporal effects. The second chapter addressed the potential role of selection in subdivided populations, emphasizing the effect of differential resource availability. The modeling approach was based on the use of discrete multivariate distributions, which led naturally to a stochastic formulation of multiple-niche polymorphism. The third and fourth chapters extended the stochastic approach developed in the second chapter into a full matrix-based treatment, which was used to examine expected genetic structure in Caribbean and Southeast Asian coral reef ecosystems respectively. In the Caribbean, the matrices gave an indication of structure, in particular the definitive presence of a break between eastern and western portions of the Caribbean, with the two endpoints of the dividing line at Puerto Rico and the Gulf of Venezuela. Elasticity analysis also revealed the populations that are most sensitive to proportional changes in values, the reefs of the Lesser Antilles, the Nicaraguan Rise and the Mesoamerican Barrier Reef. The fourth chapter applied the matrix model to the Southeast Asian region, showing the interconnections between various areas and discussing the consequences in the context of generating biological diversity.

All of the chapters address how the dynamics of genes in subdivided populations can be quantitatively modeled, as well as identifying some of the potential consequences that result from changes to connectivity structure. However, the findings can also be set within the context of a much larger picture. To proceed, it is important to begin by clarifying what is being discussed. Connectivity has been described within this dissertation as “the degree of exchange between subpopulations” (Introduction) or “the relative exchange of individuals among geographically separated subpopulations” (from Cowen et al. 2007), both of which are true, but the definitions also need to be set within the context of an organism’s life cycle. After all, what constitutes exchange? Is it any individual successfully transiting between populations? Is it only those that are successful colonizers, or is it only those that are successfully reproduce as adults? Connectivity might also be integrated through time. For example, populations could be considered to be connected even if they only interact sporadically. The nature of the connections must also be considered. Connections between populations may be binary (they exist, or do not), or quantitative (they have relative strengths). Formalizing the system mathematically gives substance and clear definitions to the concepts, and allows them to relate to one another in a common framework.

First, there must be some way of describing the system being studied. In this case, populations are characterized in terms of their allele frequencies. The frequencies can be arranged in a matrix where rows are populations, columns are alleles and entries are allele counts (although by row-normalizing, they can also be interpreted as allele frequencies). This is the  $\mathbf{Q}$  matrix referred to throughout the dissertation, and can be designated for a given time  $t$  by  $\mathbf{Q}_t$ . Several transformations are then subsequently



applied to this matrix (Fig. 6.1). First is birth (**B**), followed by migration (**M**), recruitment into destination populations (**+I**), and demographic (non-transitional) mortality (scaling to **K**). Note that these are simply linear transformations, and other functions can be included as well (e.g. selection, resources, mutation, age-based mortality), as long as their (mathematical) bases can be accommodated by the equation. Taken in this manner, different forms of connectivity can be seen as the successive products of these operations. The typical usage of connectivity is in the sense of demographic connectivity (i.e. the product of **MBQ**). The values are quantitative, but can be translated into binary values by imposing the logical condition **MBQ**>0. The result of **K(MBQ+Q)** corresponds to the number of individuals successfully reaching the age of reproduction (one turn of the life cycle has been completed). This has been referred to as reproductive connectivity (Pineda, Hare, and Sponaugle 2007), and also corresponds to the concept of effective population size ( $N_e$ ) from population genetics (i.e. effective connectivity). If connectivity is to be evaluated through time, it must include all of the components of the life cycle that affect the state matrix in a relevant way, therefore reproductive/effective connectivity is the most relevant matrix for discussing population genetic projections.

### **The roots of structure**

With the formalization of the life cycle, the next question to address is how the different processes act in concert (i.e. migration, demography, spatial structure and selection) to generate population genetic structure. Fundamentally, structure is the result

of relative differences within and between populations. This leads to the question of how to measure structure. Many different methods have been developed for this purpose (e.g.  $F_{ST}$ , distance metrics, autocorrelation). All are related by the fact that they define some sort of relationship between points on a hypervolume (vectors can be represented as points on a hypervolume displaced from the origin; normalizing standardizes the radius to 1). For example, Euclidean distance gives the straight line distance between points, Bhattacharya's distance is the angular measurement between two population vectors, Nei's Distance ( $D_A$ ) is the angle between square-rooted population vectors, and  $F_{ST}$  is the ratio of averaged concentration values. Other similarity and distance measurements are discussed in Legendre and Legendre (1998).

If all points are identical to one another, are orthogonal or are randomly distributed within the hypersphere, then the relationships (distances) between all points would be equal (or the probability distribution of the distances would be uniform in the case of random points) and the result would be an absence of structure; there would be no relative differences. Structural differences within the point cloud arise as a result of localized bias, or heterogeneity within the point cloud. With heterogeneity also comes scale dependence. Scaling re-partitions the data set, potentially changing the nature of the relationships between populations. Autocorrelation enters into the picture here as well, relating differences between populations with respect to a given distance measure. This can be seen in the matrix-based representation of Moran's I (Chapter 1), which projects the centered state matrix  $\mathbf{Q}$  through the distance matrix  $\mathbf{W}$ . Structure characterizes the relationships between elements of pattern, but there is also a question of

how pattern develops over time. This occurs through a recursive process, providing the basis for evolution.

### **Modeling evolution: different approaches, different perspectives**

Several different approaches have been used to model the evolutionary process. The most classic approaches use deterministic equations, and frequently involve simplifying assumptions such as equal population sizes, equal migration, or a two-allele system. Much of the research of the 1960s and 70s involved progressing into the use of differential equations and diffusion-based systems such as those developed by Kimura (1983). Malécot (1969) developed the backwards-looking identity by descent measure which provides the basis of coalescence theory, and Maruyama (1969; 1977) explored various aspects of stepping-stone migration models. At the same time, the foundations of matrix-based projections of genetic structure were laid (Bodmer and Cavalli-Sforza 1968), and were recently extended by the work on exact moment calculations by Fu *et al.* (2006). Individual-based models provide another means of studying the dynamics of genes in subdivided populations, and have their greatest value in addressing complex adaptive systems where individual differences are important. But in situations where objects can be classified into groups with identical properties, it is far more efficient to use probability distributions. This has the effect of not only reducing the number of calculations required to obtain a desired answer, but also provides quantitative insight into the development of higher-order aspects of the system, such as variance and covariance. Indeed, it is these higher-order aspects that hobble simulations, since their

analysis requires simulating the actual number of individuals present in the system. This can prove challenging in tracking billions or trillions of individuals through time with multiple loci, with multiple runs to account for variance. Even given exponential increases in computing power, this would be a formidable challenge.

The link between individual-based simulation of populations and matrix-based methods was explored by Caswell (2002) in the context of projecting demographic processes through time. Caswell discusses how improving the efficiency of individual based simulations leads to multivariate distributions (specifically the multinomial distribution), and branching processes, and from there to matrix projections. The development of this dissertation's chapters parallels this course as well, beginning with simulation, extending into multivariate distributions and then using matrix-based methods. By incorporating selection and resource use, the multinomial distribution no longer provides an adequate representation of the sampling process, and was instead replaced by the various forms of the multivariate hypergeometric distribution (Chapter 2). Projecting allele frequencies of marine populations through time also introduces some additional elements to the calculations resulting from the re-integration of new progeny into existing populations (the +I elements of the equations in Chapters 3 and 4).

For the most part, the various modeling approaches are simply variations on a common theme, depending mainly on whether the system is handled discretely or continuously, stochastically or deterministically, and if it has linear or non-linear components. To illustrate this point, Cattell's Data Box (Cattell 1966) provides a useful visual aid (Fig 6.2). Cattell's data box places data in the context of 3 different axes: objects, descriptors and time. Objects are the items under study, descriptors are the

parameters used to characterize the objects, and time represents different temporal stages at which observations were made. There are also 3 different “modes” associated with the box: Q-mode, R-mode and T-mode. The first tracks matrices of descriptors by time across different objects, the second – objects by time across different descriptors, and the third – objects and descriptors over time. Evolutionary models are primarily concerned with the development of structure over time, and therefore T-mode is the natural mode of operation for them. Obviously, the box can be reshaped and permuted into different forms (e.g. time could instead be a descriptor), however the original form provides an intuitive means of visualizing the differences between different modeling approaches.

The state matrix  $\mathbf{Q}$  has been defined previously and referenced to a given time step  $t$ . The  $\mathbf{B}$ ,  $\mathbf{M}$  and  $\mathbf{K}$  matrices (and any additional matrices) transform the state matrices, and project the state matrix forward in time (T-mode). This is simply a re-statement of the process described earlier in the chapter. Specifically, the transformations are *affine* transformations, commonly used when manipulating computer graphics (e.g. scaling, rotation and shear). Note that the elements are discrete. A continuous time axis requires the use of continuous-time-based Markov transitions or delay differential equations. A continuous state would require using integrodifference equations in the case of discrete time intervals or partial differential equations for continuous time intervals (e.g. Kimura’s diffusion-based approach). It is possible that insights regarding the relationship between discrete and continuous models could be drawn from methods of converting between analog and digital signals (interpolation and re-sampling). Methods for coupling integrodifference equations to matrix models have been explored by Neubert and Caswell (2000). Fundamentally however, genes, alleles and DNA are discrete units

and are best represented by a discrete state. Coalescence follows the most likely paths of arrows backwards through the transformations until they converge at a single common element. Incorporating stochasticity into the transformations through the use of probability distributions in the transformation matrix introduces variance and covariance into the system, degrading the average signal and decreasing the predictability of the system. With nonlinear components (i.e. the state of the system at different time steps cannot be represented properly using a function) direct calculations (simulation) are the only option.

There are many different ways to approach modeling the evolutionary process in interconnected populations, each involving tradeoffs between ease of use and understanding, realism of assumptions, and accuracy of the model. One potential means of integrating the various approaches in a common framework is through the use of object-oriented modeling.

### **Object-oriented architecture**

The basis of object-oriented modeling is the development of classes with properties and methods. Programming is most commonly used in conjunction with simulation, due to the latter's requirement for repetition and recursion, although object-oriented modeling can be used to implement classic models as well. One of the most classic and pervasive problems when modeling ecological systems is the difficulty in dealing with scale-dependent processes. For example, changes in nutrient abundance may be insignificant over small scales, but dramatic over large ones. Note however that

it is possible for changes at small scales to cascade out and affect larger processes (e.g. the potential effect of small scale eddies on the distribution of organisms). Classic analytical methods do not provide a straightforward mechanism for filtering parameters on the basis of scale. In contrast, object-oriented programming does provide this facility, through its use of interfaces (discussed in Chapter 1).

Object-oriented programming provides a natural fit for biological systems. Biological organisms are already classified hierarchically in the Linnaean system, lending themselves to object-oriented design with inheritance. Derived characteristics can be represented by extending a parent class (e.g. Perciformes *extends* Pisces). Functional differences would translate into unique properties (variables) or behavior (methods). Object oriented modeling also has the additional advantage of being well established in the open-source community. Open-source programming requires open access of information, transparency and constant scrutiny and revision. These characteristics are the very premise of scientific inquiry. Facilities such as Sourceforge (<http://sourceforge.net>) already exist as repositories for developing, managing and distributing these types of projects, and some of these ideas have also begun to enter into the consciousness of biologists as well (e.g. Bioforge – <http://www.bioforge.net>).

One drawback to using object-oriented modeling is that sensitivity testing must be performed component-wise. Each individual component class would need to be tested separately, and in the case of interactions, in conjunction with one another. In contrast, an analytical approach allows for concise and accurate testing through derivative-based sensitivity analysis. Another drawback is the inability to perform symbolic restructuring of the system, as is possible with an analytical model. However, if a mechanism could be

developed for translating numeric programming (i.e. numbers, arrays, loops) into symbolic language (i.e. numbers, vectors and matrices;  $\Sigma$  and  $\Pi$ ), this problem could be directly addressed. This idea is not completely unreasonable, considering that symbolic math programs and associated simplifying routines already exist in existing software (e.g. Matlab, Maple). Furthermore binary operations (implemented through machine language) are a common denominator, and provide a fundamental base for translating between programming syntax and arithmetic operations.

### **Applications**

This dissertation has devoted considerable attention to the development of quantitative approaches for studying the behavior of genes in subdivided populations, but what are the potential applications? Of what practical use is such an approach? One of the primary goals of conservation is the preservation and maintenance of biological diversity. Biologically diverse systems are capable of providing a greater array of services, are generally considered to be more robust and resilient, and tend to be more appealing in general. But at the root of biological diversity is genetic diversity. Biological diversity is merely the phenotypic translation of genetic diversity. Therefore, if scientists and managers are interested in maintaining biological diversity over the long term, an accounting should also be made for the optimal means of maintaining genetic diversity. The approaches developed here provide a way of quantitatively evaluating the expected behavior of the system given a certain set of assumptions regarding how the system functions (Chapters 3 and 4).



Determining the sensitivity of the system to disturbances is also of considerable importance for management. The effects of different levels of disturbance can be evaluated through sensitivity analysis of the models. For simulations, this is accomplished through systematic testing of the model objects and their interactions, and for matrix-based analytical systems this would be achieved through elasticity analysis. The results of sensitivity analysis can be used to identify which populations are critical connectivity junctions, and therefore should be preserved if the status quo is to be maintained. Alternatively, in the case of invasive organisms, the results could be used to identify where management efforts would likely be most effective in containing outbreaks.

### **Future directions and links to other disciplines**

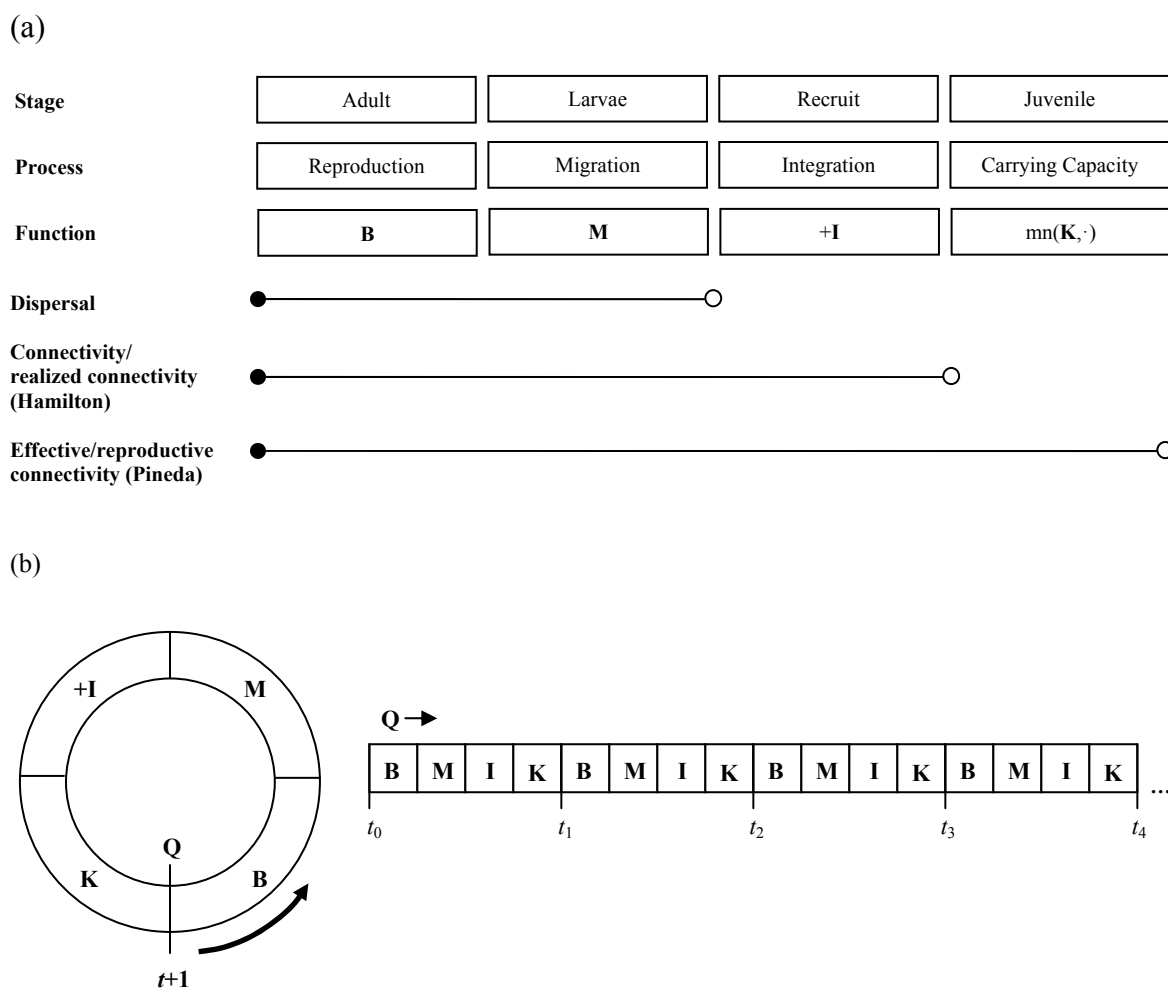
There are many potential directions in which to proceed with further research. One possibility would be to explore the role of demography (e.g. population fluctuations, age-structure, habitat quality), and more explicit individual interactions. The individual-based framework developed in the first chapter represents an ideal framework for exploring these types of questions. Greater availability of high-resolution data, as well as greater computing power for retrieving and processing data will also play a significant role in shaping the development of future research. More detailed information will provide insight into the importance of near-shore processes and fine-scale hydrodynamic features.

In the course of this dissertation, several intriguing links to other disciplines have emerged. In particular, there appear to be many links with quantum theory and relativity. For example, among the cornerstones of quantum theory are Hilbert Spaces. Hilbert spaces are related to the inner products of normed vector spaces. The normalization is performed explicitly in the equations, and the inner product is provided by examining the state of the genetic space as a genetic distance (Nei's distance, which is the inner product of the square root of the vector space). Another example is Minkowski distance (or p-norm) -  $\sum_{i=1}^n (|x_i|^p)^{1/p}$ . This formula provides a convenient means of scaling the connectivity matrices to highlight different levels of connectivity, but is also associated with special relativity. It is possible that the field of population genetics would benefit from development in this direction similar to the manner in which physics developed when it moved beyond deterministic Newtonian models into a probabilistic quantum framework. Many mathematical tools have been used in studying quantum theory, for instance tensors (Bowen and Wang 1976), quaternions (Hanson 2006) and geometric algebra (Hestenes 2004; Hestenes and Sobczyk 2004). Geometric algebra is of considerable interest as it appears to be a simpler, superseding framework with respect to the other approaches (i.e. tensors and quaternions) and is particularly well-suited for dealing with problems of space-time. Many of the common elements between population genetic modeling, quantum physics and computer graphics boil down to aspects of the dot product (angle) between normed vectors. The dot product can represent a wide array of operations including multiplication and summation, multiplication of a row vector by a transpose or column vector, correlation, conditional expectation or projection of a vector onto a different line.

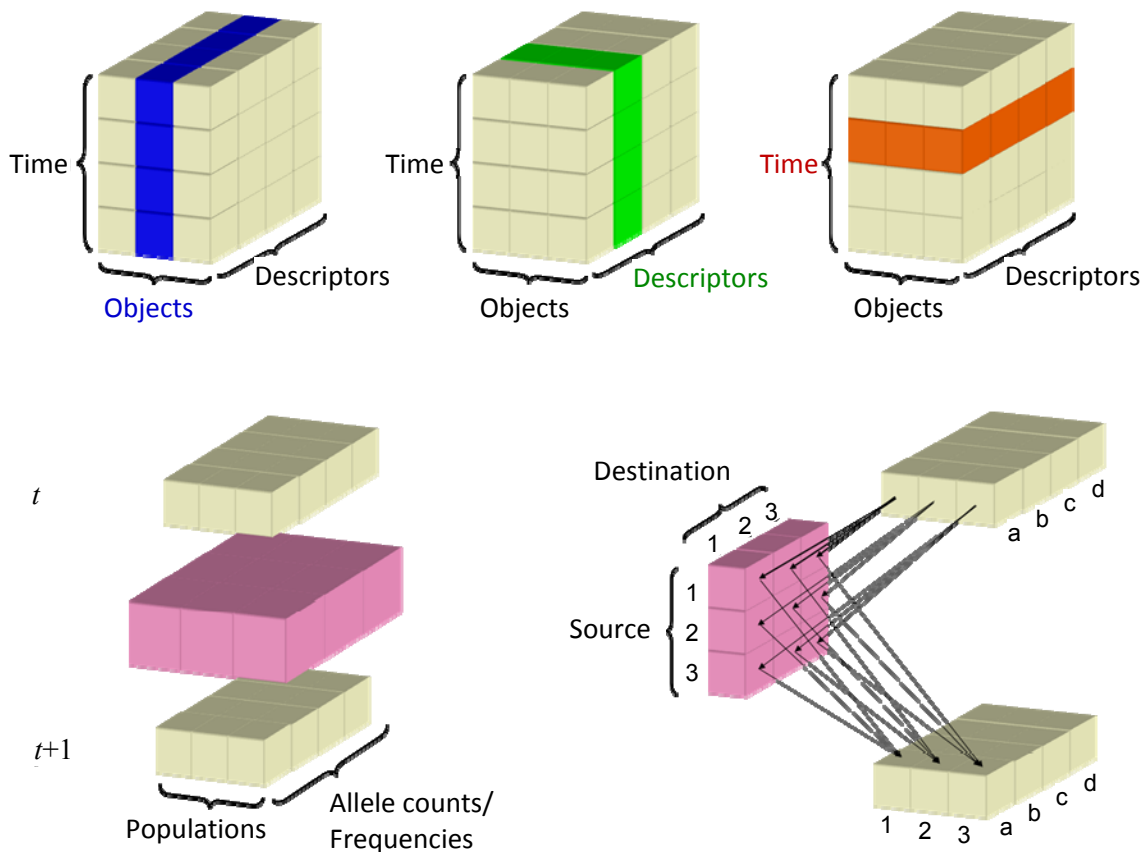
There are also links with active areas of research in computer science. Affine transformations, quaternions and tensors have all been extensively used in the manipulation of computer graphics (Pletinckx 1989; Foley et al. 1995; Nealen et al. 2006). Selection is an optimization-based problem, leading into a sequence of what are referred to as NP-hard problems in computer science. These include the Travelling Salesman Problem (TSP - Gutin and Punnen 2007), Quadratic Assignment Problem (QAP - Garey and Johnson 1979), and the Knapsack Problem (Kellerer, Pferschy, and Pisinger 2004). Calculating the shortest path between two points in the presence of obstacles (e.g. distance “as the fish swims”) provides a concrete link to a long-standing problem in computational geometry (Hershberger and Suri 1997). There also appears to be a link involving the Hadamard product and matrix product, serial versus parallel connections and quantum networks (Kauffman 1999). This could have potential relevance for optimization and selection in subdivided environments since selective weightings in different environments can be expressed in terms of a Hadamard product, and projection takes place using matrix products.

Clearly, there are many potential directions in which research may progress, and are likely beyond the scope of any single individual’s knowledge. Instead, a multidisciplinary, collaborative approach is warranted to explore these problems and the links between them. Identifying foundational links between such diverse fields will almost certainly have a significant impact on the development of science in general.

**Fig. 6.1.** (a) Schematic representation of life cycle stages, associated processes and mathematical transformations.  $\mathbf{B}$  is a diagonal matrix containing per capita birth coefficients,  $\mathbf{M}$  is the dispersal/migration matrix containing the probabilities of moving between populations,  $\mathbf{I}$  is an identity matrix,  $\mathbf{K}$  is a diagonal matrix containing carrying capacity values, and  $mn()$  is the multinomial function applied row-wise. The lines beneath the boxed items indicate that the definition is associated with the above stages, processes and functions. The closed dot at the beginning of the line indicates that the box is included in the definition. The hollow dot indicates that all boxes up to, but not including that point are included in the definition. (b) The life cycle continues iteratively through time, progressing through each function in sequence.



**Fig. 6.2.** Cattell's Data Box and associated modes. The data boxes have three axes: objects, descriptors and time. Q mode (blue) tracks descriptors versus time over different objects. R mode involves objects versus time over various descriptors. T mode tracks objects and descriptors through time. With discrete entities, transformation of the state of the system is achieved through matrix multiplication.



## APPENDIX A: STANDARD BAR-DOT NOTATION

Given a matrix  $\mathbf{X}$  with  $R$  rows and  $C$  columns,

$$X_{\cdot j} = \sum_{i=1}^R X_{ij}$$

$$X_{i\cdot} = \sum_{j=1}^C X_{ij}$$

$$\bar{X}_{\cdot j} = \frac{1}{R} \sum_{i=1}^R X_{ij}$$

## APPENDIX B: PROOF OF THE PROJECTION FORMULA

$$Q_1 = \overline{K(MBQ_0 + Q_0)}_{i.}$$

Let  $A = MB + I$

$$Q_1 = \overline{KAQ_0}_{i.}$$

$$Q_2 = \overline{KAQ_1}_{i.}$$

$$Q_2 = \overline{KAKAQ_0}_{i.i.}$$

$$Q_3 = \overline{KAKAKAQ_0}_{i.i.i.}$$

Re-normalization of rows yields the same result<sup>†</sup>, therefore

$$Q_3 = \overline{K(AKAK)}_{i.} \overline{AQ_0}_{i.}$$

and by extension

$$Q_t = \overline{K(AK)}_{i.}^{t-1} \overline{AQ_0}_{i.}$$

† Re-normalization Proof: For clarity the  $i.$  subscripts for the row-normalizations will be omitted.

For any combination  $\overline{\overline{XY}}$  where  $\sum_{c=1}^n y_{ic} = 1$ , since matrix multiplication is defined as

$$z_{ij} = \sum_{r=1}^n x_{ir} y_{rj}$$

Therefore

$$\overline{\overline{\mathbf{XY}}} = \frac{\sum_{r=1}^n x_{ir} \left( \frac{y_{rj}}{\sum_{c=1}^n y_{ic}} \right)}{\sum_{c=1}^n x_{ic} \sum_{r=1}^n x_{ir} \left( \frac{y_{rj}}{\sum_{c=1}^n y_{ic}} \right)}$$

By definition, every row of a row-normalized matrix will sum to 1, since

$$\frac{\sum_{j=1}^c y_{ij}}{\sum_{j=1}^c y_{ij}} = 1$$

Therefore if  $\mathbf{Y}$  is row-normalized

$$\begin{aligned} &= \frac{\sum_{r=1}^n x_{ir} \left( \frac{y_{rj}}{1} \right)}{\sum_{c=1}^n x_{ic} \sum_{r=1}^n x_{ir} \left( \frac{y_{rj}}{1} \right)} \\ &= \frac{\sum_{r=1}^n x_{ir} y_{rj}}{\sum_{c=1}^n x_{ic} \sum_{r=1}^n x_{ir} y_{rj}} \\ &= \frac{\mathbf{XY}}{\text{diag} \sum_{c=1}^n \mathbf{XY}} = \overline{\overline{\mathbf{XY}}} \end{aligned}$$

This completes the proof.



**APPENDIX C: PROOF OF THE EQUIVALENCE OF HAPLOID AND DIPLOID ALLELE FREQUENCIES ARISING FROM RANDOM SAMPLING UNDER THE ASSUMPTION OF RANDOM MATING**

Consider  $I$  individuals, each having two characters (repetition allowed) from a set of  $J$  possible characters. Let  $p_{ij}$  denote the proportion of occurrences of character  $j$  in individual  $i$ , for  $i = 1, \dots, I$  and  $j = 1, \dots, J$ . For example, if for the first individual, the associated characters are  $AB$ , then  $p_{11} = \frac{1}{2} = p_{12}$ , and  $p_{1j} = 0$  for  $j > 2$ , and if  $CC$  is associated with the second individual, then  $p_{23} = 1$  and  $p_{2j} = 0$  for  $j \neq 3$ . We can now verify that if  $p_{\cdot j}$  denotes the relative frequency of occurrence of character  $j$  in the whole population of  $I$  individuals, then we have for  $j = 1, \dots, J$ ,

$$p_{\cdot j} = \frac{1}{I} \sum_{i=1}^I p_{ij}.$$

Suppose that there are  $K$  progenies. Consider any sampling scheme in which, for any progeny  $k$ , the probability that an individual  $i$  is a parent, is the same for all  $i$ . In such a case, this probability will equal  $2/I$  since each progeny has two parents. Let  $n_{ijk}$  denote the count of character  $j$  inherited from individual  $i$  by progeny  $k$ . Thus each  $n_{ijk}$  has value 1 if individual  $i$  is a parent of progeny  $k$  and contributes character  $j$  to the progeny; otherwise  $n_{ijk}$  equals  $(2/I)p_{ij}$ . Also the proportion  $\hat{p}_{\cdot jk}$  of occurrences of character  $j$  in progeny  $k$  is given by

$$\hat{p}_{\cdot jk} = \frac{1}{2} \sum_{i=1}^I n_{ijk}.$$

Thus the expected value of  $\hat{p}_{\cdot jk}$  equals  $(1/2)(2/I)\left(\sum_{i=1}^I p_{ij}\right)$  which in turn equals  $p_{\cdot j}$ .

Since this answer is independent of  $k$ , and since the relative frequency of occurrence of character  $j$  in the collection of all progenies is a weighted average over  $k$  of  $\hat{p}_{\cdot jk}$ , the expected value of this relative frequency must equal  $p_{\cdot j}$ , the relative frequency of occurrence of character  $j$  in the population of  $I$  individuals.

## APPENDIX D: DISCRETE MULTIVARIATE DISTRIBUTIONS

The forms of the discrete multivariate distributions used in the manuscripts are given below, with  $k$  as the number of items being selected,  $i$  as the number of categories being selected from,  $p$  as a probability of being selected,  $n$  as the number of items being selected from  $k$ ,  $\omega$  as a fitness value,  $c$  as the number of different item types, and  $t$  representing the action of a single draw from the system.

Distribution name	Formula
Multinomial	$\Pr \left[ \bigcap_{i=1}^c (K_i = k_i) \right] = k! \prod_{i=1}^c (p_i^{k_i} / k_i!)$ $k_i \geq 0, \sum_{i=1}^c k_i = k$
Multivariate Hypergeometric	$\Pr[\mathbf{K} = \mathbf{k}] = \frac{\left( \prod_{i=1}^c \binom{n_i}{k_i} \right)}{\binom{n}{k}}$
Fisher's Noncentral Hypergeometric	$\Pr[\mathbf{K} = \mathbf{k}] = \frac{\left( \prod_{i=1}^c \binom{n_i}{k_i} \omega_i^{k_i} \right)}{\binom{n}{k}}$
Wallenius' Noncentral Hypergeometric	$\Pr[\mathbf{K} = \mathbf{k}] = \left( \prod_{i=1}^c \binom{n_i}{k_i} \right) \int_0^1 \prod_{i=1}^c (1 - t^{\omega_i/d})^{k_i} dt$ $d = \omega_i(n_i - k_i)$

**APPENDIX E: DEVELOPMENT OF THE RESOURCE-WEIGHTED VERSION OF WALLENIUS' NON-CENTRAL HYPERGEOMETRIC DISTRIBUTION**

Let  $\rho_i$  be the amount of resources of type  $i$  available within a given population. Let  $\nu_i$  be a usage coefficient representing the amount of resource consumed. Consequently,  $n_i$  will be the number of individuals capable of being supported by the population (assuming individuals cannot outstrip the available resources), and so  $\rho_i = \nu_i n_i$ .

Let  $d = \omega_i(\nu_i n_i - \nu_i k_i)$ , and let  $a = \omega_i(n_i - k_i)$ . Therefore  $a = d/\nu_i$ .

$$\begin{aligned} \Pr[\mathbf{K} = \mathbf{k}] &= \left( \prod_{i=1}^c \binom{n_i}{k_i} \right) \int_0^1 \prod_{i=1}^c (1 - t^{\omega_i/a})^{k_i} dt \quad (\text{Chesson, 1976}) \\ &= \left( \prod_{i=1}^c \binom{n_i}{k_i} \right) \int_0^1 \prod_{i=1}^c (1 - t^{\nu_i \omega_i/d})^{k_i} dt . \end{aligned}$$

Since  $d = \omega_i(\nu_i n_i - \nu_i k_i)$  and  $\rho_i = \nu_i n_i$ , therefore  $d = \omega_i(\rho_i - \nu_i k_i)$ .

## REFERENCES

- Aars, J., and R. A. Ims. 2000. Population dynamic and genetic consequences of spatial density-dependent dispersal in patchy populations. *American Naturalist* 155 (2):252-265.
- Ablan, M.C.A., J.W. McManus, C.A. Chen, K.T. Shao, J. Bell, A.S. Cabanban, V.S. Tuan, and I.W. Arthana. 2002. Meso-scale transboundary units for the management of coral reefs in the South China Sea area. *NAGA, WorldFish Center Quarterly* 25 (3 & 4):4-9.
- Allen, G. R. 2002. Indo-Pacific coral-reef fishes as indicators of conservation hotspots. *Proceedings of the Ninth International Coral Reef Symposium, Bali, 2000*, 2:921-926.
- Amos, W., and A. Balmford. 2001. When does conservation genetics matter? *Heredity* 87:257-265.
- Amos, W., and J. Harwood. 1998. Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society of London B* 353 (1366):177.
- Andréfouët, S., F. E. Muller-Karger, J.A. Robinson, C.J. Kranenburg, D. Torres-Pullizza, S.A. Spraggins, and B. Murch. 2006. Global assessment of modern coral reef extent and diversity for regional science and management applications: A view from space. *Proceedings of the Tenth International Coral Reef Symposium, Okinawa, 2004*, 1732:1745.
- Andren, H. 1994. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: A review. *Oikos* 71 (3):355-366.
- Armstrong, P. R. 2002. Recruitment limitation, population regulation, and larval connectivity in reef fish metapopulations. *Ecology* 83 (4):1092-1104.
- Avise, J. C. 1998. Conservation genetics in the marine realm. *Journal of Heredity* 89 (5):377-382.
- Barber, P. H., and D. R. Bellwood. 2005. Biodiversity hotspots: Evolutionary origins of biodiversity in wrasses (Halichoeres: Labridae) in the Indo-Pacific and new world tropics. *Molecular Phylogenetics and Evolution* 35 (1):235-253.
- Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2002. Sharp genetic breaks among populations of *Haptosquilla pulchella* (stomatopoda) indicate limits to larval transport: Patterns, causes, and consequences. *Molecular Ecology* 11 (4):659-674.

- Barbujani, Guido. 1987. Autocorrelation of gene frequencies under isolation by distance. *Genetics* 117 (4):777-782.
- Barton, N. H. 1993. The probability of fixation of a favored allele in a subdivided population. *Genetical Research* 62 (2):149-157.
- Barton, N.H., and M. C. Whitlock. 1997. The evolution of metapopulations. In *Metapopulation dynamics: Ecology, genetics and evolution*, edited by I. Hanski and M. Gilpin. London: Academic Press.
- Baums, I. B., M. W. Miller, and M. E. Hellberg. 2005. Regionally isolated populations of an imperiled Caribbean coral, *Acropora palmata*. *Molecular Ecology* 14 (5):1377-1390.
- Beissinger, Steven R. 2002. Pva: Past, present, future. In *Population viability analysis*, edited by S. R. Beissinger and D. R. McCullough. Chicago: University of Chicago Press.
- Bender, D. J., T. A. Contreras, and L. Fahrig. 1998. Habitat loss and population decline: A meta-analysis of the patch size effect. *Ecology* 79 (2):517-533.
- Benzie, J. A. H. 1992. Review of the genetics, dispersal and recruitment of crown-of-thorns starfish (*Acanthaster-planci*). *Australian Journal of Marine and Freshwater Research* 43 (3):597-610.
- . 1999. Genetic structure of coral reef organisms: Ghosts of dispersal past. *American Zoologist* 39 (1):131-145.
- Benzie, J. A. H., K. P. Black, P. J. Moran, and P. Dixon. 1994. Small-scale dispersion of eggs and sperm of the crown-of-thorns starfish (*Acanthaster planci*) in a shallow coral-reef habitat. *Biological Bulletin* 186 (2):153-167.
- Benzie, J. A. H., and S. T. Williams. 1995. Gene flow among giant clam (*Tridacna-gigas*) populations in Pacific does not parallel ocean circulation. *Marine Biology* 123 (4):781-787.
- Bergeron, J. P. 1997. Nucleic acids in ichthyoplankton ecology: A review, with emphasis on recent advances for new perspectives. *Journal of Fish Biology* 51 (sa):284-302.
- Bijlsma-Meeles, E., and R. Bijlsma. 1988. The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: Fitness measurements and predictions under conditions with no alcohol stress. *Genetics* 120 (3):743-753.
- Bodmer, Walter F., and Luigi L. Cavalli-Sforza. 1968. A matrix migration model for the study of random genetic drift. *Genetics* 59 (4):565-592.

- Boehlert, G. W. 1996. Larval dispersal and survival in tropical reef fishes. In *Reef fisheries*, edited by N. V. C. Polunin and C. M. Roberts. London: Chapman and Hall.
- Bousquet, F., and C. Le Page. 2004. Multi-agent simulations and ecosystem management: A review. *Ecological Modelling* 176:313-332.
- Bowen, Ray M., and C.C. Wang. 1976. *Introduction to vectors and tensors*. New York: Plenum Press.
- Breckling, Broder, Ulrike Middelhoff, and Hauke Reuter. 2006. Individual-based models as tools for ecological theory and application: Understanding the emergence of organisational properties in ecological systems. *Ecological Modelling* 194 (1-3):102.
- Buckley, L., E. Caldarone, and T.-L. Ong. 1999. RNA-DNA ratio and other nucleic acid-based indicators for growth and condition of marine fishes. *Hydrobiologia* 401:265-277.
- Bulmer, M.G. 1972. Multiple niche polymorphism. *The American Naturalist* 106 (948):254-257.
- Burke, L., E. Selig, and S. Spalding. 2002. *Reefs at risk in Southeast Asia*. Washington, DC: World Resources Institute.
- Campana, S. E. 1999. Chemistry and composition of fish otoliths: Pathways, mechanisms and applications. *Marine Ecology-Progress Series* 188:263-297.
- Carpenter, Kent E., and Victor G. Springer. 2005. The center of marine shore fish biodiversity: The Philippine islands. *Environmental Biology of Fishes* 72:467-780.
- Caswell, Hal. 2001. *Matrix population models: Construction, analysis, and interpretation*. 2nd ed. Sunderland, MA: Sinauer Associates.
- Cattell, R.B. 1966. The data box: Its ordering of total resources in terms of possible relational systems. In *Handbook of multivariate experimental psychology*, edited by R. B. Cattell. Chicago: Rand McNally & Co.
- Cavalli-Sforza, L. L., and W.F. Bodmer. 1971. *The genetics of human populations*. San Francisco: W. H. Freeman.
- CERN. 2004. The Colt project 1.2.0. <http://acs.lbl.gov/~hoschek/colt/>.
- Chesser, Ronald K., Jr. Olin E. Rhodes, and Michael H. Smith. 1996. Gene conservation. In *Population dynamics in ecological space and time*, edited by J. Olin E. Rhodes, R. K. Chesser and M. H. Smith. Chicago: University of Chicago Press.

- Chesson, Jean. 1976. A non-central multivariate hypergeometric distribution arising from biased sampling with application to selective predation. *Journal of Applied Probability* 13 (4):795-797.
- Clavero, M., and E. Garcia-Berthou. 2005. Invasive species are a leading cause of animal extinctions. *Trends in Ecology & Evolution* 20 (3):110-110.
- Cliff, A.D., and J.K. Ord. 1981. *Spatial processes - models and applications*. London: Pion.
- Cole, K. S., and Y. Sadovy. 1995. Evaluating the use of spawning success to estimate reproductive success in a Caribbean reef fish. *Journal of Fish Biology* 47 (2):181-191.
- Cowen, R. K. 2002. Larval dispersal and retention and consequences for population connectivity. In *Coral reef fishes: Dynamics and diversity in a complex ecosystem*, edited by P. F. Sale. San Diego, CA: Academic Press.
- Cowen, R. K., G. Gawarkiewicz, J. Pineda, S. R. Thorrold, and F.E. Werner. 2007. Population connectivity in marine systems: An overview. *Oceanography* 20 (3):14-21.
- Cowen, R. K., C. B. Paris, and A. Srinivasan. 2006. Scaling of connectivity in marine populations. *Science* 311 (5760):522-527.
- DeAngelis, Donald L., and Louis J. Gross. 1992. *Individual-based models and approaches in ecology: Populations, communities, and ecosystems*. New York: Chapman & Hall.
- Dempster, E.R. 1955. Maintenance of genetic heterogeneity. *Cold Spring Harbor Symposium Quantitative Biology* 20:25-32.
- Doherty, P. J., S. Planes, and P. Mather. 1995. Gene flow and larval duration in 7 species of fish from the Great-Barrier-Reef. *Ecology* 76 (8):2373-2391.
- Doligez, A., C. Baril, and H. I. Joly. 1998. Fine-scale spatial genetic structure with nonuniform distribution of individuals. *Genetics* 148 (2):905-919.
- Domeier, M. L. 2004. A potential larval recruitment pathway originating from a Florida marine protected area. *Fisheries Oceanography* 13 (5):287-294.
- Endler, J.A. 1977. *Geographic variation, speciation, and clines*. Princeton, N.J.: Princeton University Press.
- Epperson, Bryan K. 2003. *Geographical genetics*. Princeton, N.J.: Princeton University Press.



- Epperson, Bryan K., and Tianquan Li. 1996. Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *PNAS* 93 (19):10528-10532.
- Fahrig, L. 2002. Effect of habitat fragmentation on the extinction threshold: A synthesis. *Ecological Applications* 12 (2):346-353.
- Fisher, R.A. 1934. Two new properties of mathematical likelihood. *Proceedings of the Royal Society A - Mathematical Physical & Engineering Sciences* 144:285-307.
- . 1935. The logic of inductive difference (with discussion). *Journal of the Royal Statistical Society* 98:39-82.
- Fisher, Ronald A. 1922. On the dominance ratio. *Proceedings of the Royal Society of Edinburgh* 42:321-341.
- . 1930. *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Fog, Agner. 2008. Stocc. <http://www.agner.org/random/stocc.zip>.
- Fog, Agner. 2008. Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions. *Communications in Statistics, Simulation and Computation* 37 (2).
- Foley, J.D., Andries van Dam, Steven K. Feiner, and John F. Hughes. 1995. *Computer graphics: Principles and practice in c*. 2nd ed. Boston, MA: Addison-Wesley Professional.
- Fowler, K., C. Semple, N. H. Barton, and L. Partridge. 1997. Genetic variation for total fitness in *Drosophila melanogaster*. *Proceedings of the Royal Society B: Biological Sciences* 264 (1379):191.
- Frankham, Richard, Jonathan D. Ballou, and David A. Briscoe. 2002. *Introduction to conservation genetics*. Translated by D. A. Briscoe. New York: Cambridge University Press.
- Fu, Rongwei, Alan E. Gelfand, and K. E. Holsinger. 2006. Exact moment calculations for genetic models with migration, mutation, and drift. *Theoretical Population Biology* 63 (3):231-243.
- Gaggiotti, O. E. 1996. Population genetic models of source-sink metapopulations. *Theoretical Population Biology* 50 (2):178-208.
- Gaggiotti, O. E., and P. E. Smouse. 1996. Stochastic migration and maintenance of genetic variation in sink populations. *American Naturalist* 147 (6):919-945.
- Garey, Michael R., and David S. Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco, CA: W.H. Freeman.

- Goodwin, B. J., and L. Fahrig. 2002. How does landscape structure influence landscape connectivity? *Oikos* 99 (3):552-570.
- Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H. H. Thulke, J. Weiner, T. Wiegand, and D. L. DeAngelis. 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* 310 (5750):987-991.
- Grimm, Volker, and Steven F. Railsback. 2005. *Individual-based modeling and ecology*. Princeton: Princeton University Press.
- Groom, M.J., and M.A. Pascual. 1998. The analysis of population persistence: An outlook on the practice of viability analysis. In *Conservation biology for the coming decade*, edited by P. L. Fiedler and P. M. Kareiva. New York, NY: Chapman and Hall.
- Gutierrez-Rodriguez, C., and H. R. Lasker. 2004. Microsatellite variation reveals high levels of genetic variability and population structure in the gorgonian coral *Pseudopterogorgia elisabethae* across The Bahamas. *Molecular Ecology* 13 (8):2211-2221.
- Gutin, G., and A.P. Punnen. 2007. *The traveling salesman problem and its variations*. Berlin: Springer-Verlag.
- Haldane, J.B.S. 1927. The mathematical theory of natural and artificial selection. Part V. *Proceedings of the Cambridge Philosophical Society* 23:838-44.
- . 1932. *The causes of evolution*. London: Longmans and Green.
- . 1939. The equilibrium between mutation and random extinction. *Annals of Eugenics* 9:400-5.
- Hanski, Ilkka, and Oscar E. Gaggiotti, eds. 2004. *Ecology, genetics, and evolution of metapopulations*. Burlington, MA: Elsevier.
- Hanski, Ilkka, and Michael E. Gilpin. 1997. *Metapopulation biology: Ecology, genetics, and evolution*. San Diego, CA: Academic Press.
- Hanson, Andrew J. 2006. *Visualizing quaternions*. San Francisco, CA.
- Hastings, A., and S. Harrison. 1994. Metapopulation dynamics and genetics. *Annual Review of Ecology and Systematics* 25:167-188.
- Hauswaldt, J. S., and T. C. Glenn. 2005. Population genetics of the diamondback terrapin (*Malaclemys terrapin*). *Molecular Ecology* 14 (3):723-732.

- Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms. In *Genetics and evolution of aquatic organisms*, edited by A. R. Beaumont. London: Chapman and Hall.
- Hedgecock, D., P. H. Barber, and S. Edmands. 2007. Genetic approaches to measuring connectivity. *Oceanography* 20 (3):70-79.
- Hershberger, John, and Subhash Suri. 1997. An optimal algorithm for euclidean shortest paths in the plane. *SIAM Journal on Computing* 28 (6):2215-2256.
- Hestenes, David. 2004. *New foundations of classical mechanics*. 2nd ed. Dordrecht: Kluwer.
- Hestenes, David, and Garrett Sobczyk. 2004. *Clifford algebra to geometric calculus*. 1st ed. Dordrecht: Kluwer.
- Hoeksema, Bert W. 2007. Delineation of the Indo-Malayan centre of maximum marine biodiversity: The coral triangle. In *Biogeography, time and place - Distributions, barriers and islands*, edited by W. Renema. Dordrecht, The Netherlands: Springer.
- Houde, Ed. 1989. Comparative growth, mortality, and energetics of marine fish larvae: Temperature and implied latitudinal effects. *Fishery Bulletin* 87 (3):471-495.
- Hughes, A. R., and J. J. Stachowicz. 2004. Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 101 (24):8998-9002.
- Hughes, T. P., A. H. Baird, D. R. Bellwood, M. Card, S. R. Connolly, C. Folke, R. Grosberg, O. Hoegh-Guldberg, J. B. C. Jackson, J. Kleypas, J. M. Lough, P. Marshall, M. Nystrom, S. R. Palumbi, J. M. Pandolfi, B. Rosen, and J. Roughgarden. 2003. Climate change, human impacts, and the resilience of coral reefs. *Science* 301 (5635):929-933.
- Hutchinson, G.E. 1957. Concluding remarks. Paper read at Cold Spring Harbor Symposia on Quantitative Biology, at Cold Spring Harbor.
- . 1965. *The ecological theatre and the evolutionary play*. New Haven, CT: Yale University Press.
- Huxel, G. R. 1999. Rapid displacement of native species by invasive species: Effects of hybridization. *Biological Conservation* 89 (2):143-152.
- ISRS. 2004. *Marine protected areas (MPAs) in management of coral reefs. Briefing paper 1*: International Society for Reef Studies.
- James, M. K., P. R. Armsworth, L. B. Mason, and L. Bode. 2002. The structure of reef fish metapopulations: Modelling larval dispersal and retention patterns.

*Proceedings of the Royal Society of London Series B-Biological Sciences* 269 (1505):2079-2086.

- Jenerett, G.D., and J. Wu. 2000. On the definitions of scale. *Bulletin of the Ecological Society of America* 81 (1):104-105.
- JGAP. 2007. Java Genetics Algorithms Package 3.2. <http://jgap.sourceforge.net/>.
- Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1997. *Discrete multivariate distributions*. New York, NY: John Wiley & Sons, Inc.
- Jokiel, P., and F. J. Martinelli. 1992. The vortex model of coral-reef biogeography. *Journal of Biogeography* 19 (4):449-458.
- Jones, G. P., M. J. Milicich, M. J. Emslie, and C. Lunow. 1999. Self-recruitment in a coral reef fish population. *Nature* 402 (6763):802-804.
- Kauffman, Louis H. 1999. Combinatorics and topology-François Jaeger's work in knot theory. *Annales de l'institut Fourier* 49 (3):927-953.
- Kellerer, Hans, U. Pferschy, and D. Pisinger. 2004. *Knapsack problems*. Berlin: Springer Verlag.
- Kimura, M., and G.H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561-576.
- Kimura, Motoo. 1983. *The neutral theory of molecular evolution*. Cambridge [Cambridgeshire]; New York: Cambridge University Press.
- Kimura, Motoo, and Tomoko Ohta. 1971. *Theoretical aspects of population genetics*. Princeton, N.J.: Princeton University Press.
- Kingston, T. 1995. Valuable modeling tool. *Conservation Biology* 9 (4):966-968.
- Kinlan, B. P., and S. D. Gaines. 2003. Propagule dispersal in marine and terrestrial environments: A community perspective. *Ecology* 84 (8):2007-2020.
- Kwok, Pui-Yan. 2000. High-throughput genotyping assay approaches. *Pharmacogenomics* 1 (1):95-100.
- Lacy, R. C. 1987. Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology* 1 (2):143-158.
- Lacy, R.C. 1993. Vortex: A computer simulation model for population viability analysis. *Wildlife Research* 20 (1):45-65.
- Legendre, Pierre, and Louis Legendre. 1998. *Numerical ecology*. 2nd English ed. Amsterdam; New York: Elsevier.

- Leibold, Matthew A. 1995. The niche concept revisited: Mechanistic models and community context. *Ecology* 76 (5):1371.
- Leis, J. M., and B. M. Carson-Ewart. 2002. In situ settlement behaviour of damselfish (Pomacentridae) larvae. *Journal of Fish Biology* 61 (2):325-346.
- Levene, Howard. 1953. Genetic equilibrium when more than one ecological niche is available. *The American Naturalist* 87 (836):331.
- Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* 73 (6):1943-1967.
- Levins, Richard. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15:237-240.
- Lewontin, R. C. 1974. *The genetic basis of evolutionary change*. New York, NY: Columbia University Press.
- Lewontin, Richard C. 2004. Building a science of population biology. In *The evolution of population biology*, edited by R. S. Singh and M. K. Uyenoyama. Cambridge, UK: Cambridge University Press.
- Malécot, Gustav. 1969. *The mathematics of heredity*. Translated by D. M. Yermanos. San Francisco, CA: W.H. Freeman and Co.
- . 1973. Génétique des populations diploïdes naturelles dans le cas d'un seul locus, iii: Parenté, mutations et migration. *Annales Génétique. Sélection Animale* 5:333-361.
- . 1975. Heterozygosity and relationship in regularly subdivided populations. *Theoretical Population Biology* 8:212-241.
- Maruyama, T. 1969. Genetic correlations in the stepping stone model with non-symmetrical migration rates. *Journal of Applied Probability* 6:463-477.
- . 1977. *Stochastic problems in population genetics*. Berlin: Springer-Verlag.
- McCarty, J. P. 2001. Ecological consequences of recent climate change. *Conservation Biology* 15 (2):320-331.
- McClanahan, T., N. Polunin, and T. Done. 2002. Ecological states and the resilience of coral reefs. *Conservation Ecology* 6 (2).
- McCullagh, P., and J.A. Nelder. 1989. *Generalized linear models*. 2nd ed. London: Chapman and Hall.

- McLaughlin, J. F., J. J. Hellmann, C. L. Boggs, and P. R. Ehrlich. 2002. Climate change hastens population extinctions. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 99 (9):6070-6074.
- McManus, J. W. 1985. Marine speciation, tectonics and sea-level changes in Southeast Asia. *Proceedings of the Fifth International Coral Reef Symposium*, Tahiti, 1985, 4:133-138.
- Metcalf, William G., Marvel C. Stalcup, and Donald K. Atwood. 1977. Mona Passage drift bottle study. *Bulletin of Marine Science* 27 (3):586-591.
- Milligan, B. G., J. Leebensmack, and A. E. Strand. 1994. Conservation genetics - beyond the maintenance of marker diversity. *Molecular Ecology* 3 (4):423-435.
- Mora, C., P. M. Chittaro, P. F. Sale, J. P. Kritzer, and S. A. Ludsin. 2003. Patterns and processes in reef fish diversity. *Nature* 421 (6926):933-936.
- Nagylaki, T. 1978. A diffusion model for geographically structured populations. *Journal of Mathematical Biology* 6 (4):375-382.
- . 1996. The diffusion model for migration and selection in a dioecious population. *Journal Of Mathematical Biology* 34 (3):334-360.
- Nagylaki, Thomas. 1982. Geographical invariance in population genetics. *Journal of Theoretical Biology* 99 (1):159.
- Nagylaki, Thomas, and Michael Moody. 1980. Diffusion model for genotype-dependent migration. *Proceedings of the National Academy of Sciences of the United States of America* 77 (8):4842-4846.
- National Research Council (NRC). 2001. *Marine protected areas*. Washington, DC: National Academy Press.
- Nealen, Andrew, Matthias Müller, Richard Kaiser, Eddy Boxerman, and Mark Carlson. 2006. Physically based deformable models in computer graphics. *Computer Graphics Forum* 25 (4):809-836.
- Nei, Masatoshi. 1972. Genetic distance between populations. *American Naturalist* 106:283-292.
- . 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Neigel, J. E. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review Of Ecology And Systematics* 28:105-128.
- Neubert, M.G., and Hal Caswell. 2000. Density-dependent vital rates and their population dynamic consequences. *Journal of Mathematical Biology* 43:103-121.

- Ochando, Maria Dolores, and Francisco J. Ayala. 1999. Fitness of wild-caught *Drosophila melanogaster* females: Allozyme variants of GPDH, ADH, PGM, and EST. *Genetica* 105 (1):7-18.
- Palumbi, S. R. 1997. Molecular biogeography of the Pacific. *Coral Reefs* 16:S47-S52.
- Pandolfi, John M. 1992. Successive isolation rather than evolutionary centres for the origination of Indo-Pacific reef corals. *Journal of Biogeography* 19 (6):593-609.
- Paris, C. B., R. K. Cowen, R. Claro, and K. C. Lindeman. 2005. Larval transport pathways from Cuban snapper (Lutjanidae) spawning aggregations based on biophysical modeling. *Marine Ecology Progress Series* 296:93-106.
- Paris, C. B., R. K. Cowen, K. M. M. Lwiza, D. P. Wang, and D. B. Olson. 2002. Multivariate objective analysis of the coastal circulation of Barbados, West Indies: Implication for larval transport. *Deep-Sea Research Part I-Oceanographic Research Papers* 49 (8):1363-1386.
- Pastinen, Tomi, Mirja Raitio, Katarina Lindroos, Paivi Tainola, Leena Peltonen, and Ann-Christine Syvanen. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* 10 (7):1031-1042.
- Pineda, J., J. A. Hare, and S. Sponaugle. 2007. Larval transport and dispersal in the coastal ocean and consequences for population connectivity. *Oceanography* 20 (3):22-39.
- Planes, S., and C. Fauvelot. 2002. Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. *Evolution* 56 (2):378-399.
- Pletinckx, Daniel. 1989. Quaternion calculus as a basic tool in computer graphics. *The Visual Computer* 5 (1-2):2-13.
- Pollard, J.H. 1969. Continuous-time and discrete-time models of population growth. *Journal of the Royal Statistical Society* 132:80-88.
- Possingham, H. P., and I. Davies. 1995. Alex: A model for the viability analysis of spatially structured populations. *Biological Conservation* 73 (2):143-150.
- Potts, D. C. 1985. Sea-level fluctuations and speciation in the Scleractinia. Paper read at Fifth International Coral Reef Congress, at Tahiti.
- Purcell, J. F. H., R. K. Cowen, C. R. Hughes, and D. A. Williams. 2006. Weak genetic structure indicates strong dispersal limits: A tale of two coral reef fish. *Proceedings Of The Royal Society B-Biological Sciences* 273 (1593):1483-1490.
- Reed, D. H., and R. Frankham. 2003. Correlation between fitness and genetic diversity. *Conservation Biology* 17 (1):230-237.

- Rice, Sean H. 2004. *Evolutionary theory: Mathematical and conceptual foundations*. Sunderland, MA.: Sinauer Associates.
- Richardson, David E., Jeffrey D. Vanwye, Amy M. Exum, Robert K. Cowen, and Douglas L. Crawford. 2007. High-throughput species identification: From DNA isolation to bioinformatics. *Molecular Ecology Notes* 7 (2):199-207.
- Roberts, C. M., S. Andelman, G. Branch, R. H. Bustamante, J. C. Castilla, J. Dugan, B. S. Halpern, K. D. Lafferty, H. Leslie, J. Lubchenco, D. McArdle, H. P. Possingham, M. Ruckelshaus, and R. R. Warner. 2003. Ecological criteria for evaluating candidate sites for marine reserves. *Ecological Applications* 13 (1):S199-S214.
- Roberts, C. M., C. J. McClean, J. E. N. Veron, J. P. Hawkins, G. R. Allen, D. E. McAllister, C. G. Mittermeier, F. W. Schueler, M. Spalding, F. Wells, C. Vynne, and T. B. Werner. 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* 295 (5558):1280-1284.
- Rousset, F. 1999. Genetic differentiation in populations with different classes of individuals. *Theoretical Population Biology* 55 (3):297-308.
- . 1999. Genetic differentiation within and between two habitats. *Genetics* 151 (1):397-407.
- Rousset, François. 2004. *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.
- Ryall, K. L., and L. Fahrig. 2006. Response of predators to loss and fragmentation of prey habitat: A review of theory. *Ecology* 87 (5):1086-1093.
- Sale, P. F., R. K. Cowen, B. S. Danilowicz, G. P. Jones, J. P. Kritzer, K. C. Lindeman, S. Planes, N. V. C. Polunin, G. R. Ruse, Y. J. Sadovy, and R. S. Steneck. 2005. Critical science gaps impede use of no-take fishery reserves. *Trends in Ecology & Evolution* 20 (2):74-80.
- Sale, P. F., and J. P. Kritzer. 2003. Determining the extent and spatial scale of population connectivity: Decapods and coral reef fishes compared. *Fisheries Research* 65 (1-3):153-172.
- Shaver, Aaron C., Peter G. Dombrowski, Joseph Y. Sweeney, Tania Treis, Renata M. Zappala, and Paul D. Sniegowski. 2002. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162 (2):557-566.
- Shulman, M. J., and E. Bermingham. 1995. Early life histories, ocean currents, and the population genetics of caribbean reef fishes. *Evolution* 49 (5):897-910.
- Slatkin, M. 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. *Population Biology* 12:253-262.



- Slatkin, M., and H. E. Arter. 1991. Spatial autocorrelation methods in population genetics. *American Naturalist* 138 (2):499-517.
- Slatkin, Montgomery. 1981. Estimating levels of gene flow in natural populations. *Genetics* 99 (2):323-335.
- Smouse, P. E., and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82:561-573.
- Sokal, R. R., G. M. Jacquez, and M. C. Wooten. 1989. Spatial autocorrelation analysis of migration and selection. *Genetics* 121 (4):845-855.
- Sokal, Robert R., and Daniel E. Wartenberg. 1983. A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics* 105 (1):219-237.
- Spalding, Mark D., Corinna Ravilious, and Edmund P. Green. 2001. *World atlas of coral reefs*. Berkeley, CA: University of California Press.
- Starck, W.A., and P. L. Colin. 1978. *Gramma linki* - new species of grammid fish from tropical Western Atlantic. *Bulletin of Marine Science* 28:146-152.
- Sun Microsystems 2006. Java 1.5. <http://java.sun.com>.
- Taylor, M. S., and M. E. Hellberg. 2003. Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* 299 (5603):107-109.
- . 2006. Comparative phylogeography in a genus of coral reef fishes: Biogeographic and genetic concordance in the Caribbean. *Molecular Ecology* 15:695-707.
- Templeton, Alan R., Kerry Shaw, Eric Routman, and Scott K. Davis. 1990. The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden* 77 (1):13.
- The Mathworks. 2007. Matlab 7.4.0 (R2007a). <http://www.mathworks.com>.
- Thorrold, S. R., G. P. Jones, M. E. Hellberg, R. S. Burton, S. E. Swearer, J. E. Neigel, S. G. Morgan, and R. R. Warner. 2002. Quantifying larval retention and connectivity in marine populations with artificial and natural markers. *Bulletin of Marine Science* 70 (1):291-308.
- Underwood, A.J., and M.J. Keough. 2001. Supply-side ecology: The nature and consequences of variations in recruitment of intertidal organisms. In *Marine community ecology*, edited by M. D. Bertness, S. D. Gaines and M. E. Hay. Sunderland, MA: Sinauer Associates.
- Veron, J. E. N. 2000. *Corals of the world*. 3 vols. Townsville, MC: Australian Institute of Marine Science.

- Victor, B. C. 1986. Duration of the planktonic larval stage of one hundred species of Pacific and Atlantic wrasses (family Labridae). *Marine Biology* 90 (3):317-326.
- Wade, M. J., and D. E. McCauley. 1988. Extinction and recolonization - their effects on the genetic differentiation of local-populations. *Evolution* 42 (5):995-1005.
- Warner, R. R., and R. K. Cowen. 2002. Local retention of production in marine populations: Evidence, mechanisms, and consequences. *Bulletin of Marine Science* 70 (1):245-249.
- Whitlock, M. C. 1992. Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* 46 (3):608-615.
- . 2002. Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160 (3):1191-1202.
- Whitlock, M. C., and N. H. Barton. 1997. The effective size of a subdivided population. *Genetics* 146 (1):427-441.
- Whitlock, M. C., and D. E. McCauley. 1990. Some population genetic consequences of colony formation and extinction - genetic correlations within founding groups. *Evolution* 44 (7):1717-1724.
- Wolanski, Eric, ed. 2001. *Oceanographic processes of coral reefs: Physical and biological links in the great barrier reef*. Boca Raton, Fla.: CRC Press.
- Wright, Sewall J. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159.
- . 1937. The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America* 23:307-20.
- . 1943. Isolation by distance. *Genetics* 28:114-38.
- . 1969. *The theory of gene frequencies*. Vol. 2, *Evolution and the genetics of populations*. Chicago: Illinois: University of Chicago Press.
- Young, A., T. Boyle, and T. Brown. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends In Ecology & Evolution* 11 (10):413-418.
- Young, Andrew G., and Geoffrey M. Clarke, eds. 2000. *Genetics, demography and viability of fragmented populations*. Cambridge, U.K.; New York: Cambridge University Press.