Genetic Structure and Selection in Subdivided Populations

FRANÇOIS ROUSSET

MONOGRAPHS IN POPULATION BIOLOGY • 40

Genetic Structure and Selection in Subdivided Populations

Genetic Structure and Selection in Subdivided Populations

FRANÇOIS ROUSSET

PRINCETON UNIVERSITY PRESS
Princeton and Oxford

Contents

List of Figures and Tables

Acknowledgments

The writing of this book was initiated at the suggestion of Sam Elworthy, science editor at Princeton University Press. Collaborations and discussions with Sylvain Gandon, Henri Leturque, Virginie Ravigné, Ophélie Ronce, Denis Roze, and Renaud Vitalis have helped shape my ideas about many of the topics addressed in this book. Thomas Bataillon, Sylvain Gandon, Olivia Judson, Michel Raymond, Renaud Vitalis, Mike Whitlock, and the reviewers for Princeton University Press have provided helpful comments on various drafts. I also thank Nick Barton, Troy Day, Michael Doebeli, Stefan Geritz, Sylvain Glémin, Robert Griffiths, Raphaël Leblois, Sido Mylius, Hidenori Tachida, Jinliang Wang, and John Wakeley for answering various queries or providing unpublished information.

Until recently, French research institutions have offered only shamefully deficient bibliographic tools to their scientists. I thank Nicole Pasteur for her heroic efforts to solve this problem, and Isabelle Olivieri for a temporary but very useful fix. I also thank Valérie Durand for help in getting references.

I would like to thank those who found sensible words in difficult times.

Preface

WHAT IS AND IS NOT THERE

This book reviews the theory of spatial genetic structure, with some focus on the application of its concepts and results to the analysis of phenotypic evolution. It presents basic concepts, as well as some more advanced results that are relevant to the evaluation of current research, and emphasizes a few efficient reasoning tools. For each topic reviewed, I briefly discuss alternative formulations, bringing to light potential sources of confusion. I have not covered a number of important topics, such as conditions for local adaptation in variable environments and in host-parasite systems, and I have not attempted to review the huge literature on all aspects of genetic population structure. References are not comprehensive, but the reader should have access to the most important literature through those given.

ASSUMED BACKGROUND

I assume the reader has some basic interest in population biology, and some knowledge of the population genetics of panmictic populations. The contents of this book are of variable technical level, but within each chapter, the main arguments, results, and implications are summarized separately from the technical sections.

The book aims to be self-contained, particularly on topics that are not easily accessible in the primary literature or are common sources of confusion; hence, it contains detailed derivations of most major results. However, except in the technical appendices and footnotes, I have used only a minimum set of mathematical techniques that are already well established among population biologists. The exposition is based on discrete-time models. The mathematical prerequisites include the usual tools of calculus, particularly partial derivatives and Taylor expansions,

and some matrix algebra. A vague recollection of eigenvectors and eigenvalues will be helpful to the reader, but all important results based on these quantities are summarized in the main text or in the Mathematical Appendix (Appendix A). For a helpful textbook, see Horn and Johnson (1985), and for a biologically motivated summary of matrix algebra, see Caswell (2001). An elementary but firm understanding of the basic concepts of probability is essential to the field, and I have also assumed knowledge of the concept of the Markov chain: see, for example, Feller (1968) or Grimmett and Stirzaker (1992) for introductions to these topics. Diffusion approximations are considered, but no knowledge of the underlying mathematical theory is required.

Notation used consistently throughout the book is summarized in Table A.1 in the Mathematical Appendix.

OF *GENE* AND *FITNESS*

The words "gene," "allele," and "locus" will be used as follows. A gene is any copy of some inherited information, and an allele is a class of genes. At each autosomal locus, you have received one gene from your mother and one homologous gene from your father. Etymologically, *allele* means "different," and hence you have not received two alleles unless the genes received from each parent bear different information.

In some current usage, the word "fitness" is used to refer, for example, to the number of juveniles produced. Here *fitness* will be the expected number of adult offspring of an adult. This definition may need qualification when considering age-structured populations, for example. However, what matters, particularly in the context of spatially subdivided populations, is to evaluate fitness as the number of descendants of an individual after one full iteration of the life cycle.

A *fitness measure* is any measure of the effects of differences in fitness on the fate of alleles inducing these differences.

CHAPTER ONE

Introduction

GENETIC STRUCTURE IN RELATION TO SELECTION

Initial interest in spatial genetic structure came out of the idea that dispersal prevents genetic differentiation of populations. To some extent, it may prevent adaptation to the local environment. More efficiently, it limits the random genetic divergence of populations and thus the first step of allopatric speciation. In most species, however, dispersal is spatially restricted, and individuals interact and compete preferentially with their neighbors.

According to Wright, limited dispersal was important for adaptation. He considered that adaptation occurred mainly through the evolution of coadapted gene complexes, that is, of alleles at different loci, which would be deleterious when combined with alleles of another coadapted gene complex. These deleterious effects would prevent adaptive evolution from one coadapted complex to another in a large, randomly mixing population. The selection against deleterious intermediates would be more easily overcome by random effects in small populations.

Thus, to explain adaptation, Wright proposed what he called the "shifting-balance" theory (e.g., Wright 1931b; Provine 1986). According to this theory, limited exchanges between local populations allowed local allele frequencies to randomly fluctuate (the phenomenon of genetic drift), and thus different coadapted gene complexes could be fixed in different populations. The better adapted gene complexes made populations more fecund, and a small amount of dispersal allowed them to spread in the total population (a step that Wright called interdeme selection). These multilocus processes were and still are difficult to quantify, so Wright mainly developed one-locus models quantifying

the extent of local differentiation due to drift. Later works have quantified the probability of fixation of various kinds of mutants in a single deme, and various approximations have been suggested for subdivided populations. It is clear that limited dispersal may slow, but not prevent, the spread of advantageous mutants, so the main output of such models concerns the extent to which spatial structure affects the selective advantage experienced by a mutant, particularly when this advantage depends on its local frequency.

Today, there is little evidence supporting Wright's shifting-balance theory and the idea that population structure is essential for adaptation (Coyne et al. 1997). However, the question remains of how demography, including the extent of dispersal or the regulation of population sizes, can affect selection, and this question has been addressed by population geneticists using the concepts developed by Wright. For example, there have been renewed attempts to compute fixation probabilities of mutants, their fixation times, and allele frequency distributions by way of diffusion approximations, an approach also largely initiated by Wright.

However, the demographic factors, such as dispersal rates, which may affect adaptation in natural populations, are difficult to estimate by demographic methods. To overcome this difficulty, the genetic structure itself has been used to infer dispersal and other demographic events in natural populations. Such studies were initiated by Dobzhansky and Wright (1941), and their number exploded after electrophoresis was developed to measure enzyme polymorphisms in natural populations, and more recently with microsatellite genotyping.

Out of such studies has grown a belief (e.g., Lewontin 1974; Slatkin 1987; Whitlock and McCauley 1999) that observed genetic structure is often not consistent with the expectations of the simple models developed by Wright, Malécot, and their followers. Various explanations have been proposed for these discrepancies, including selection on the genetic markers used and past demographic processes such as range expansions. At lower spatial and temporal scales, increased attention has also been given to "metapopulation" processes of extinction and recolonization and to their consequences on genetic differentiation.

These factors undoubtedly have an effect on the genetic structure of natural populations. With respect to selection, although statistical analyses of molecular polymorphisms have often been inconclusive (Kre-

itman and Akashi 1995), enzyme polymorphisms often appear subject to selection when they are studied using a combination of biochemical studies and field experiments (e.g. Feder et al. 1997; Eanes 1999). Comparable studies have not yet been conducted for microsatellites. Likewise, range expansions and contractions are parts of a species' history as a result of, for example, climate changes or changing interactions with competitors and parasites. Together, these factors put into question our ability to estimate demographic parameters from studies of population structure.

Another concern that has generated much speculation about population structure is the occurrence of supposedly altruistic behaviors. By altruistic we mean behaviors that increase some group's production of offspring to the detriment of the fitness of the individuals expressing them. The occurrence of altruism in natural populations was little questioned and attracted little interest until some authors, including Hamilton (1963, 1964), Maynard Smith (1964), and Williams (1966) felt that altruism needed a special explanation. What needed a special explanation was how genes increasing the expression of a behavior that supposedly decreases the fitness of altruists could spread into the population.

To define conditions under which cooperative or altruistic behaviors would be selected for or against, Hamilton (1964 and later works) developed the theory of inclusive fitness. This theory accounts for the fact that individuals who make altruistic fitness gifts may also be the main recipients of such gifts. This will occur when interactions occur between individuals more genetically similar than those who do not interact, so that an individual who bears genes increasing the expression of a fitness gift behavior interacts with individuals who tend to make fitness gifts to it. Then the individual fitness of altruists may on average be higher than that of nonaltruists.

Although inclusive fitness theory did not initially focus on spatially localized dispersal, it was recognized that such limited dispersal results in a genetic structure by which spatially close individuals are more genetically similar than spatially distant individuals. Thus, population structure has often been suggested as an explanation for the evolution of altruistic behaviors.

Whether many behaviors once viewed as altruistic are actually so has been repeatedly questioned. But inclusive fitness is a conceptual

framework that applies identically, in principle, to behaviors that are altruistic or not, and to a wide diversity of life history traits, such as sex ratios, and parental investment. Any behavior that affects the fitness of neighbors, that is, essentially all behaviors in spatially subdivided populations, will be affected by spatial genetic structure. Limited dispersal adds a new dimension, as dispersal is itself a selected trait, and it interacts with many other traits of ecological importance. Selfing or outcrossing, the evolution of sex ratios when dispersal is sex biased, and the investment in reproduction versus survival when only juveniles disperse are only a few examples. Further analysis is required to assess when the consequences of these traits on dispersal will have important implications on their evolution.

To say that neighbors are genetically more similar than distant individuals is essentially the same as saying that there are fluctuations of allele frequencies among different populations. Hence, Hamilton saw a connection between inclusive fitness and Wright's measures of population structure, and the idea has been put to work in some models. However, application of the inclusive fitness framework to spatially subdivided populations has had a checkered history and no consistent development until recently. While inclusive fitness theory has recently benefited from important methodological insights, allowing a wide range of selection processes to be analyzed using a few well-defined tools, its basic concepts remain commonly misunderstood.

Inclusive fitness is used mostly to determine the direction of phenotypic evolution in a game theoretical perspective that seeks what will be the "unbeatable strategies" to which a population will evolve. Evolutionary game theory grew out of the works of, for example, Hamilton (1967), Maynard Smith (e.g., 1982), and Eshel (1983), and recent developments have been publicized as "adaptive dynamics." However, the relationship of inclusive fitness to adaptive dynamics has also been questioned (e.g., Doebeli 1999). Indeed, it has been hard to find both general and consistent arguments for the use of inclusive fitness measures in game theory.

These difficulties with inclusive fitness are illustrative of a more general problem, in which different schools of thought address the same question with different languages and may each ignore results easily obtained by another approach. In contrast, reliable science is possible only when scientists base their work on widely shared con-

cepts. A major objective of this book is to show how the different approaches mentioned above relate to the theory of genetic structure and complement each other.

Therefore, this book presents a synthesis of the different approaches that have been developed to understand evolution in subdivided populations. This book:

- reviews models of genetic differentiation in subdivided populations. These form the basis for appreciating the potential for random differentiation under different circumstances. What can or cannot be reliably concluded from statistical analyses of population structure will be highlighted based on these models.
- shows how the selection process can be analyzed in terms of measures of spatial structure. I will show how this can be done with minimal effort under a variety of population structures, what it implies in terms of social evolution, and how it relates to game theoretical approaches. Unambiguous and widely applicable concepts of genetic similarity will come out of these analyses.
- shows, in a nontechnical way, how the same methods and the concept of effective size can be used for a finer description of the selection process, focusing for illustration on the question of fixation probabilities in subdivided populations.

This book also contains some new results. In particular, it discusses simple methods to construct better diffusion approximations for subdivided populations. Most of the results for effective size in metapopulations, and their derivations, also depart from earlier work.

PLAN OF THE BOOK

Chapter 2 illustrates basic concepts used by Wright and Hamilton to quantify evolutionary processes. It provides a more formal exposition of the ideas presented in this introduction and sets a statistical framework in which the concepts will be developed.

Chapter 3 develops some basic models of population structure: the finite island model and models of isolation by distance. It provides both a self-contained presentation of unusual and often poorly accessible algebra and a less technical discussion of the broad implications of the models for the interpretation of patterns observed in natural populations.

Chapter 4 reviews the biological significance and properties of the different measures of identity and of population structure in the context of coalescence theory. It emphasizes a distinction between models that have a hierarchical structure with two time scales and those that do not. This chapter departs from some tradition in emphasizing definitions and interpretations that apply equally under different dispersal models and to finite and infinite populations, that is, to the full range of scenarios considered in applications of the measures of population structure, rather than simply the infinite island model.

Chapter 5 introduces the different measures of evolutionary convergence and evolutionary stability of a phenotype (or "behavior" or "strategy") used in game theory and discusses how they can be computed in subdivided populations.

Chapter 6 demonstrates a mathematically and conceptually minimal way of determining convergence to evolutionarily stable strategies in a subdivided population, and illustrates its application in the example of dispersal rates.

Using simple models of cooperative or altruistic interactions in a subdivided population, chapter 7 shows how these results relate to the concepts of inclusive fitness theory, such as costs, benefits, and altruism.

The following chapters detail the application and generalization of these ideas to progressively more complex models.

Chapter 8 deals with the specifics of diploid populations. It analyzes patterns of genetic structure in plant populations with selfing, and shows how to measure fitness effects, whether the trait selected is under parent or offspring control in sexual diploid populations.

Chapter 9 develops concepts of effective size. It compares the different definitions of effective size used in the literature. It shows how effective size can be used, together with the techniques developed in previous chapters to quantify selection, to obtain better approximations of allele frequency dynamics in subdivided populations. It emphasizes the interpretation of effective size in terms of elementary events affecting gene lineages, an interpretation that easily generates a number of formulas for effective size, as illustrated in particular for populations with different classes of individuals.

Chapter 10 extends the analysis of effective size to populations with demographic fluctuations, in particular to metapopulations with local

extinctions and recolonization, for which some new results and new interpretations are presented.

Chapter 11 shows how selection can be measured in class-structured populations or metapopulations by the same concepts developed in previous chapters.

Much of the book is based on highly simplified models. The concluding chapter considers possible developments and discusses what conclusions can and cannot safely be drawn from the models. In particular, it highlights a number of common misconceptions of analyses of genetic diversity in natural populations. There is currently much interest in "coalescent" methods, and this chapter explains the logic of these methods. I have refrained from an extensive review of these methods, because their practical reliability is unknown, so little can be safely concluded at present.

In summary, this book aims to show how a range of questions can be efficiently addressed using a limited number of concepts and technical tools, and to provide a self-contained account of the basic models of the genetic structure of populations. There are two alternative paths through this book. One path temporarily ignores much of the neutral theory of population structure, jumping from chapter 2 to chapter 5 and later chapters on selection. On the other hand, the readers who want to make themselves more familiar with the computation and interpretation of measures of population structure may read chapters 3 and 4 before chapter 5.

CHAPTER TWO

Selection and Drift

The measures of population structure of interest in quantifying selection are measures of the extent of genetic drift, that is, of random fluctuations of allele frequencies, either at a local geographical scale or at the level of a total population. This chapter describes how drift and its consequences have been measured in some simple models.

One measure of drift is the variance of allele frequency change over one generation at the level of the total population. In the simplest case, it is directly related to the population size, and more complex scenarios lead to the concept of effective size.

At a local scale, drift expresses itself as variation in allele frequencies among different subpopulations or groups of individuals. It may be quantified by Wright's F_{ST} or by Hamilton's concepts of relatedness. Hamilton's theory implies that this local drift affects the expected change in allele frequency of a nonneutral allele, rather than only its variance. In this chapter, we present these different approaches and concepts, which will be generalized in later chapters. We also provide definitions of measures of population subdivision that will be useful for such generalizations.

SELECTION IN PANMICTIC POPULATIONS

A simple way to quantify selection is to describe the expected change in allele frequency over one generation. However, it is quickly realized that this is insufficient. In particular, the random sampling of descendant genes causes genetic drift, by which the favored allele may be lost.

The simplest results illustrating the importance of drift concern the probability of fixation of mutants in the Wright-Fisher model. This

model (Wright 1931a; Fisher 1958) considers a population in which each individual produces a very large (ideally infinite) number of juveniles, and the next generation is derived by randomly sampling N adults among these juveniles. This amounts to sampling genes with replacement from the N parents.

We consider here a haploid population of N adults. In the absence of mutation and selection, if two alleles A and a segregate among the adults, in frequency p and $1-p$, the frequency p' in the next generation[1] will follow a binomial distribution, with mean p and variance $p(1-p)/N$. This variance quantifies genetic drift, that is, the fluctuations of allele frequencies due to random sampling of gametes.

If the allele A has a fecundity advantage, such that a bearer of A produces $1+s$ juveniles for any juvenile produced by an a bearer, the frequency of the A allele among juveniles is $\tilde{p} = (1+s)p/(1+sp)$, and the frequency p' in the next generation will follow a binomial distribution, with mean $\tilde{p}$ and variance $\tilde{p}(1-\tilde{p})/N$. The expected change in allele frequency over one generation is

$$\mathrm{E}(p'|p) - p = \frac{sp(1-p)}{1+sp} = sp(1-p) + O(s^2), \qquad (2.1)$$

where $\mathrm{E}(.)$ denotes expectation and the term $O(s^2)$ is small relative to s (see p. 215 for this notation). Then the fixation probability Π of the A mutant is approximately $2s$ for "large" s, in which case it is independent of the population size N. This result may be obtained through methods of branching processes (e.g., Fisher 1922; Haldane 1927; Ewens 1979; Gale 1990).

A more general result for the fixation probability is

$$\Pi \approx \frac{1-e^{-2s}}{1-e^{-2Ns}} \qquad (2.2)$$

(e.g., Malécot 1952; Ewens 1979; see also Wright 1942). This is approximately $2s$ for $s > 1/N$ and decreases approximately exponentially with s for $s < 0$. This result can be obtained by diffusion approximations, which will be considered in chapter 9.

These results assume that the number of juveniles of each parent is Poisson distributed, and that its number of adult offspring is binomially distributed. If the distribution of the number of juveniles has higher

[1]Primes ($'$) are used throughout the book to denote any variable considered at the later time in a recursion.

variance V than the Poisson distribution, the probability of fixation of advantageous mutants is roughly $2s/V$ (Ewens 1979, p. 23), and the more general diffusion result is likewise affected.

Thus, stochastic fluctuations, such as those due to random sampling of N offspring, as well as higher than Poisson variance in number of juveniles, affect the probability of fixation. The extent of such fluctuations is commonly measured in terms of *effective size*. For example, in the Wright-Fisher model, if A is neutral ($s = 0$), the variance of its change in frequency over one generation is

$$p(1-p)/N, \tag{2.3}$$

the variance due to the binomial sampling. With a non-Poisson distributed number of juveniles, it is $Vp(1-p)/N$: the effective size is N/V. The concept of effective size will be developed in chapters 9 and 10.

We have considered only the probability of fixation, but other features, such as the average fixation time or the distribution of allele frequency under recurrent mutation, can also be expressed in terms of the expected change and variance in allele frequency.

EVOLUTION IN SPATIALLY STRUCTURED POPULATIONS

As usual, the question is whether we can generalize the above results while retaining their simplicity. In other words, do structured populations behave like the Wright-Fisher model, with appropriately defined selection and drift parameters in place of s and N? A positive answer has been given for a range of models, in particular for life cycles that create some form of genetic structure, such as that created by sib mating before random dispersal or selfing before random dispersal of seeds (e.g., Ethier and Nagylaki 1980, 1988; Pollak and Sabran 1992; Caballero and Hill 1992b). We may refer to such life cycles as *random mixing* populations, by which we mean populations with one step of spatially random dispersal in the life cycle, in contrast to *panmixia*, which is more restrictive (e.g., excluding selfing).

There have also been some attempts to extend diffusion approximations to spatially subdivided populations (e.g., Wright 1931a; Slatkin 1981; Maruyama 1983; Barton 1993; Cherry and Wakeley 2003; Whitlock 2003), but none of these attempts provides a systematic framework

for a variety of ecological scenarios. Here we first consider how one might generalize the expressions (2.1) and (2.3) for expected change and variance in allele frequency.

Selection and Local Drift

KIN SELECTION AND RELATEDNESS

Inclusive fitness deduces the direction of selection on a behavior from its effect (conventionally written $-c$) on the number of offspring of a "focal" individual that expresses the behavior, from its effect (written b) on the number of offspring of some other "recipient" individuals, and from "relatedness" R, a measure of genetic similarity between focal and recipient individuals (Hamilton 1964). Hamilton's argument was that changes in allele frequency over one generation could be written

$$\mathrm{E}(p'|p) - p \approx p(1-p)\,(-c + Rb) \tag{2.4}$$

for some c, b, and R independent of allele frequency. Thus, the change in allele frequency is of the same form as in equation (2.1), with the selection coefficient s replaced by $(-c + Rb)$.

Hamilton's initial model considered interactions between family members and therefore relatedness between family members (e.g., full sisters, half sisters). Methods for computing such relatedness may be historically traced back at least to Fisher's (1918) models of correlations between relatives and Wright's (1922) coefficients of inbreeding, and can be found in many population genetics textbooks (e.g., Hartl and Clark 1997; Lynch and Walsh 1998). However, Hamilton also suggested that inclusive fitness theory could be extended to subdivided populations and considered that the appropriate relatedness parameters would be analogous to Wright's F_{ST} parameter of population structure (e.g., Hamilton 1971, pp. 89–90; Michod and Hamilton 1980).

WRIGHT'S F_{ST}

To quantify the random differentiation between subpopulations, Wright considered the quantity $\mathrm{Var}(p)/[\bar{p}(1-\bar{p})]$ (e.g., Dobzhansky and Wright 1941; Wright 1943), later known as F_{ST} (e.g., Wright 1951, 1969).

Here $\mathrm{Var}(p)$ is the variance of allele frequency among different subpopulations subject to the same conditions, and $\bar{p}$ is the average allele frequency among subpopulations. The more genes are alike within subpopulations and different between subpopulations, the larger F_{ST} becomes. Thus, in an inclusive fitness perspective, F_{ST} may measure to what extent genes within subpopulations are more alike than genes in different subpopulations: it measures the relatedness of genes within subpopulations.

Wright obtained a simple expression for F_{ST} in the *infinite island model* (Wright 1931a). In this model we consider an infinity of subpopulations, called *demes*, of N adults. We assume that gametes may disperse between these demes. The probability that a gamete is an immigrant is denoted m. We also assume that immigrants come with equal probability from any other deme and that all gamete dispersal events are independent from each other. Under these assumptions,

$$F_{\mathrm{ST}} = \frac{(1-m)^2}{2N-(2N-1)(1-m)^2} \tag{2.5}$$

for N diploid adults (Wright 1943). For large N and small m, this can be approximated by $F_{\mathrm{ST}} \approx 1/(1+4Nm)$. Thus, it is often said that the number of migrants Nm, determines differentiation. Wright also obtained the latter result from a diffusion approximation for the distribution of allele frequency within a deme (Wright 1931a, 1937). The higher Nm is, the lower the variance of the distribution. This is expected: higher N means fewer random fluctuations due to genetic drift, and higher m means a higher input of migrants, bringing the local allele frequency to its total population average $\bar{p}$.

Effective Size in Subdivided Populations

If we could write the variance $\mathrm{Var}(\bar{p}' - \bar{p}|\bar{p})$ of change in allele frequency by analogy with equation (9.11), as

$$V(\bar{p}) = \bar{p}(1-\bar{p})/N_{\mathrm{e}}, \tag{2.6}$$

where N_{e} is some effective size, then we could combine it (at least tentatively) with an expression for expected change $M(\bar{p})$ in allele frequency and obtain the probability of fixation of a mutant and various other quantities that can be computed using diffusion approximations.

The concept of effective size and its first applications are also due to Wright. He derived the formula

$$N_{\mathrm{e}} = \frac{N_{\mathrm{T}}}{1 - F_{\mathrm{ST}}}, \tag{2.7}$$

where N_{T} is the total population size (Wright 1943). Thus, the variance $V(\bar{p})$ may be quantified by F_{ST} much as the expected change $M(\bar{p})$ may be, supporting the idea that F_{ST} may play an important role in a theory of selection in subdivided populations.

MEASURING POPULATION STRUCTURE

However, it is easy to get into trouble with Wright's definition of F_{ST}. Wright could express the quantity $\mathrm{Var}(p)/[\bar{p}(1-\bar{p})]$ as a function of subpopulation size and migration rate between subpopulations in a specific case where $\bar{p}$ and $\mathrm{Var}(p)$ were fixed quantities, not random variables. Namely, this was the infinite island model, with enough mutation to keep different alleles at stable frequencies in the total population, although frequencies fluctuate in each subpopulation. However, we may consider more general models where $\bar{p}$ and $\mathrm{Var}(p)$ are not fixed quantities. This is so, for example, for finite populations with drift or when the allele frequency considered is the frequency of an allele going to fixation. In such cases, there have been problems with different, conflicting definitions of relatedness (e.g., Seger 1981). These difficulties are easily resolved by using alternative definitions of relatedness. In the next section, we will present a generic way of defining measures of differentiation and relatedness that are free from such problems and applicable to a wide range of models, and we will specify the statistical concepts involved in such definitions.

Genetic Identity

IDENTITY BY DESCENT

We now consider a pair of homologous genes. A coalescence event occurs when their two ancestral lineages merge in a common ancestor. Mutations may have occurred since the coalescence of ancestral lineages (fig. 2.1), and we are interested in the probability that the genes

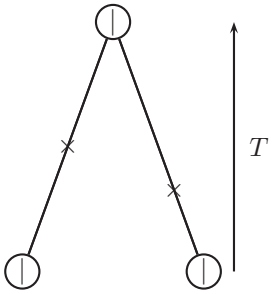

FIGURE 2.1. Coalescence of a pair of ancestral lineages. Two individuals have a common ancestor T generations ago, and mutations ($\times$) occurred since the common ancestor.

descend without mutation from their most recent common ancestor. This probability was first considered by Malécot (see Nagylaki 1989) and is one of the parameters known as *identity by descent* (IBD), an expression due to Crow (1954) (Kempthorne 1954 also introduced the expressions *alike by descent* and *alike in state* in reference to Malécot's coefficients).

We suppose that the individuals produce a very large number of gametes, which undergo mutations independently, each with probability ("rate") μ. Identity by descent can be expressed in terms of the probability C_t that coalescence occurred t generations ago, and of the probability of identity given t, which is $(1-\mu)^{2t}$, as

$$\dot{Q} = \sum_{t=0}^{\infty} C_t (1-\mu)^{2t} \tag{2.8}$$

(e.g., Malécot 1975; Slatkin 1991). It can be computed by simple algebra. For example, we consider a population of N haploid adults, following the Wright-Fisher model. Let $\dot{Q}$ be the probability of identity of two different genes in some parental generation. The probability of identity in the next generation, $\dot{Q}'$, is $(1-\mu)^2$ if the two genes have the same parent, and it is $(1-\mu)^2\dot{Q}$ if they have different parents. If each parent produces the same number of gametes, that is, a fraction $1/N$ of the total number of gametes, the two cases occur, respectively, with probability $1/N$ and $1-1/N$. Then $\dot{Q}' = (1-\mu)^2[1/N + (1-1/N)\dot{Q}]$ (Malécot 1948). The value of $\dot{Q}$ at equilibrium ($\dot{Q}' = \dot{Q}$) is $(1-\mu)^2/[(1-\mu)^2 + N(1-(1-\mu)^2)]$. An approximation for this formula is $1/(1+2N\mu)$ for small μ.

For N diploid individuals, $\dot{Q} \approx 1/(1 + 4N\mu)$ (Malécot 1948), because there are $2N$, rather than N, homologous genes at the locus. This result actually depends on additional assumptions, such as gametic dispersal. So we will rather consider the simpler example of haploid populations throughout.

In the infinite island model, we argue in the same way (Malécot 1948). As before, let m be the probability that a gamete is an immigrant. The probability that two genes are issued from nondispersed gametes is therefore $(1 - m)^2$. Now we will suppose (rather than formally show, see next chapter) that two randomly chosen genes in different demes are always different. Then the recursion for the probability of identity of pairs of genes from adults within demes is $\dot{Q} = (1-\mu)^2(1-m)^2[1/N + (1-1/N)\dot{Q}]$. Let $\gamma \equiv (1-\mu)^2$.[2] The value of $\dot{Q}$ at equilibrium is

$$\dot{Q} = \frac{\gamma(1-m)^2}{\gamma(1-m)^2 + N[1-\gamma(1-m)^2]}, \tag{2.9}$$

which is approximated by $1/[1+2N(\mu+m)]$ for small μ and small m. For N diploid adults, this is rather $1/[1 + 4N(\mu + m)]$, which is the same as Wright's approximation for F_{ST} (Wright 1931a), so that identity by descent and F_{ST} have often been equated.

IDENTITY IN STATE

Identity by descent may not be directly observable. Alternatively, we may classify genes among different allelic types according to the technology used for typing, and ask whether the genes are of the same allelic type. When we say that a pair of genes are *identical in state* (IIS), we simply mean that the genes have, for example, the same sequence if alleles are defined from DNA sequence data, the same length if alleles are defined by the number of repeats of a microsatellite motif, or the same electrophoretic mobility if alleles are typed by electrophoresis. Although IIS is directly measurable, IBD is not. The probability of IIS between genes in two sets of individuals X and Y can be expressed in terms of the frequencies p_{iX} and p_{iY} of allelic types i in the two sets,

[2]The equivalence symbol $\equiv$ will be used throughout the book for definitions.

as the expectation

$$\mathrm{E}\left(\sum p_{iX}p_{iY}\right) \tag{2.10}$$

where the sum is over all allelic types.

IBD is a specific case of IIS, in which each allele produced by mutation differs from preexisting alleles in the population (the *infinite allele model*, IAM). When the distinction is useful, we will generically write Q for probabilities of identity in state and $\dot{Q}$ for probabilities of identity by descent. Likewise, we will use a dot ($\dot{}$) to distinguish any function of probabilities of IBD from equivalent functions of probabilities of IIS. In most of the later chapters, however, we consider only identity by descent, and denote it simply Q.

Genetic Correlations and Inbreeding Coefficients

Consider a population structured in some way, for example, by age-structure or by spatially restricted dispersal between different demes. We may define a probability of identity Q_{w} within a class of genes, for example, among individuals of the same age or among individuals within the same deme. Likewise, we may define a probability of identity Q_{b} between classes, here the identity of genes from individuals in different age classes or in different demes. Generically, we may define a parameter F as

$$F \equiv \frac{Q_{\mathrm{w}} - Q_{\mathrm{b}}}{1 - Q_{\mathrm{b}}}. \tag{2.11}$$

This definition can also be written in the form

$$F \equiv \frac{H_{\mathrm{b}} - H_{\mathrm{w}}}{H_{\mathrm{b}}}, \tag{2.12}$$

where H_i, here defined as $1 - Q_i$, is known as gene diversity or as "heterozygosity."

The above definitions may be considered as a general definition of so-called "inbreeding" coefficients. For example, we consider Wright's F-statistics: F_{ST}, F_{IS}, and F_{IT}. Let Q_1 be the probability of identity of genes from adults sampled within a deme, and let Q_2 be the probability of identity of genes from different demes. The parameter F_{ST} originally considered by Wright (e.g., 1951, 1969) is best defined as

$$F_{\mathrm{ST}} \equiv \frac{Q_1 - Q_2}{1 - Q_2}. \tag{2.13}$$

Such a parameter has been described as an "intraclass" correlation of genes within demes with respect to genes between demes (e.g., Cockerham and Weir 1987).

Likewise, let Q_0 be the probability of identity of two homologous genes in a diploid individual. Then the parameter F_{IS} also considered by Wright is best defined as

$$F_{\mathrm{IS}} \equiv \frac{Q_0 - Q_1}{1 - Q_1}. \tag{2.14}$$

F_{IS} is the correlation of two gene copies within an individual relative to genes within demes, and is classically used to quantify deficits in heterozygote frequencies. In a population with two alleles in frequencies p and $q = 1 - p$, the frequency of heterozygotes can be written $2pq(1 - F_{\mathrm{IS}})$ and the frequencies of the different homozygotes can be written $p^2 + pqF_{\mathrm{IS}}$ and $q^2 + pqF_{\mathrm{IS}}$. In an infinite randomly selfing hermaphrodite population, genotype frequencies are in the well-known Hardy-Weinberg frequencies, that is, $F_{\mathrm{IS}} = 0$: the probability of identity of pairs of genes is the same whether genes are sampled in the same or in different individuals, $Q_0 = Q_1 = 1 - 2pq$.

For completeness, we mention Wright's F_{IT}, which may be defined as $(Q_0 - Q_2)/(1 - Q_2)$. F_{IT} is the correlation of two gene copies within an individual relative to genes between demes. The relationship $(1 - F_{\mathrm{IT}}) = (1 - F_{\mathrm{IS}})(1 - F_{\mathrm{ST}})$ immediately follows from the definitions.

Multilocus Associations

Although this book is largely concerned with one-locus models, associations between different loci cannot be ignored. They can be described by parameters similar to those outlined above. In particular, gametic disequilibrium (or "linkage" disequilibrium) describes the nonindependent association between alleles at different loci. Given the frequency p_{AB} of "gametes" or "haplotypes" with allele A at a first locus and allele B at a second locus, nonindependence can be measured as $D \equiv p_{AB} - p_A p_B$, where p_A and p_B are allelic frequencies at each locus. In a stochastic model with drift and mutation, disequilibrium parameters are, for example, $\mathrm{E}(D)$ and $\mathrm{Var}(D)$. If the direction of allele frequency changes at each locus is independent (in particular if the alleles are neutral), there is no average association between alleles at the

different loci: $\mathrm{E}(D) = 0$. However, there may be a random association between alleles at any time, so $\mathrm{Var}(D) > 0$. Parameters comparing disequilibrium within and among subpopulations in a way more or less analogous to F parameters, were defined and discussed by, for example, Ohta (1982) and Tachida and Cockerham (1986). However, despite their work, there is no simple expression for expected values of, for example, $\mathrm{Var}(D)$ for linked loci under simple assumptions such as the island model, which makes it difficult to account for the consequences of gametic associations in subdivided populations.

Statistical Concepts of Equilibrium and Population

The first definitions of F in the form of equation (2.12), $(H_\mathrm{b} - H_\mathrm{w})/H_\mathrm{b}$ (Nei's G_ST, e.g., Nei 1973; Takahata and Nei 1984) raised a debate over the estimators of F (e.g., Nei 1986; Cockerham and Weir 1986, 1987, 1993). One issue was the suitability of the formalism of analysis of variance used by Cockerham and Weir. For example, this formalism seems inappropriate to describe negative F_IS values in terms of positive expected mean squares. Such problems are resolved by recognizing that the model of analysis of variance usually considered makes restrictive assumptions (Rousset 2003b). Another issue was the concept of "population" inherent in different definitions of gene diversities. Independently of this past controversy, the concept of "population" used here and the related concept of "equilibrium" deserve some explanation.

We have considered models in which the probability of genetic identity was a function of the parameters N, μ, and m. This means that we evaluate Q under a probability distribution determined by these parameters. On the other hand, by completely sampling a biological population, we can observe the frequency p of an allele and the frequency q of identical pairs of genes. As a result of drift, these frequencies will not simply be functions of demographic and genetic parameters: they will be random variables. The realized values (sometimes called "actual" values) of p and q need not match their expectations, the expected allele frequency $\mathrm{E}(p)$ and the probability of identity Q, which are functions of the parameters N, μ, and m.

The mathematical analysis of natural processes and the statistical inferences about them are based on probability distributions under dif-

ferent hypotheses. In this example, these hypotheses are characterized by N, μ, and m, which are not variable. The word *parameter* best applies to such quantities. If the deme size N is a random variable, N is not a parameter; but the parameters of the probability distribution of N will play a role similar to N above.

The distinction between parameters and random variables parallels a distinction between two different concepts of "population." The first takes "population" to refer simply to a collection of observable things, such as animals in a particular locality. The second takes "population" to refer to the different possible samples and to their probability under some hypothesis. To avoid confusion, we will use the word "population" for a collection of individuals rather than in the latter probabilistic sense. On the other hand, a parameter is here defined by reference to the probability distribution of samples under a hypothesis. Hence, contrary to a common idea, the value in the (biological) population is not the parameter value. Rather, the biological process being investigated determines what should be considered a parameter. Likewise, what characterizes random effect models in the analysis of variance is that population values (where "population" is a collection of observable things) are random variables, not parameters. The true parameter considered in such models is the variance of the population values.

It is sometimes assumed that a population "at equilibrium" is a biological population in which the value of some variable is equal to the expectation of this variable. But in fact, in considering the variable in a population at statistical equilibrium, we actually consider the expected value of the variable under a given set of parameters. For example, we have computed equilibrium values of probabilities of identity. They were also computed at stationary equilibrium, as the expectation Q was assumed to be the same in two successive generations.

In the theory of spatially subdivided populations, the parameters F are generally known as *F-statistics*, thereby conflicting with the above distinction between parameters and statistics. The probabilities of identity Q_{w} and Q_{b} are defined as parameters of a stochastic process, and F as defined above is of the same nature. The sufficient justification for this choice is that parametric definitions are useful for the analysis of natural processes, based on the models developed in the following chapters.

SUMMARY

In this chapter, we have briefly reviewed some approaches to modeling drift and selection. F_{ST} can be used to quantify local differentiation due to drift, but Wright's definition of it was restrictive and more general definitions have been given. Descriptions of mean and variance in allele frequency changes offer a framework for developing approximation of the dynamics of mutants in finite populations, but it remains to show how and to what extent they can be developed for spatially structured populations.

Although these concepts may provide a basis for developments we will consider later, the questions addressed remain limited. In particular, only a fixed set of alleles is considered. Understanding the long-term evolution of populations requires consideration of the range of possible phenotypic effects of alleles that may be created by mutations. If there are stable states in the dynamics of a population evolving under recurrent mutations, the population is likely to be close to them. The current framework for finding such stable states is game theory (or adaptive dynamics), as will be described in more detail in chapter 5.

CHAPTER THREE

Spatially Homogeneous Dispersal: The Island Model and Isolation by Distance

Patterns of genetic structure may provide information about the importance of population structure in evolutionary processes. They may also bring information about the parameters of population structure, such as deme sizes and dispersal rates. But the spatial patterns under different conditions are not easily predicted and must be derived from models.

Wright's island model is the easiest to analyze, and today this model remains useful for understanding the logical consequences of limited dispersal. However, from the start Wright noticed the weakness of the defining assumption of the island model. It assumes that immigrant individuals come with equal probability from any of the other demes. Hence, it is unable to describe localized dispersal between demes. But, in nature, dispersal is generally localized in space, and one expects that genetic similarity will be larger between individuals from closer subpopulations. This expected genetic consequence of localized dispersal is known as "isolation by distance." The mathematical results and practical interpretation of models developed to quantify isolation by distance will be reviewed.

Our previous account of the island model was heuristic. Here we reconsider it in the same framework used to analyze more general models of population structure. All of the models considered in this chapter assume that demes have identical demographic characteristics (e.g., the same number of adults and the same distribution of dispersal distance from any of them), so they will be referred to as spatially homogeneous. They also assume that genetic variation is neutral; that is, it has no effect on the number of offspring of individuals.

ISLAND MODELS

Consider n_{d} demes, each occupied by N haploid adults. As in the previous chapter, we assume that each adult produces a very large number of gametes, each of which is subject to independent mutation events at rate μ. This assumption is not quite realistic, as different gametes usually have common ancestral cells for some time during the development of the germ line of an individual, so they will share mutations occurring during this time. But this does not significantly affect the conclusions. Hence, we assume that mutations occur after gamete production, in which case they are independent.

Let m be the dispersal rate between the demes, that is, the proportion of juveniles that move to the other demes before competition. Thus, the number of adults that are of immigrant origin is binomially distributed, with mean Nm: m is not the realized dispersal rate.[1] An immigrant comes from any one of the $n_{\mathrm{d}} - 1$ other demes with probability $m/(n_{\mathrm{d}} - 1)$. Competition among juveniles reduces the number of individuals to N adults per deme.

Figure 3.1 shows the life cycle assumed and the notation used for probabilities of identity by descent of genes between adults (Q^{A}), between juveniles before dispersal (Q^{J}), and between juveniles after dispersal (Q^{D}). Q^{A} and Q^{D} are identical, provided adults can be compared. This raises an exception for Q^{A} within demes when $N = 1$. To handle both cases, we will give expressions for Q^{D} rather than Q^{A}; by default Q is Q^{D}. A prime ($'$) is used to denote the same variable after one iteration of the life cycle (one time unit). We will also use the notation Q^{R} for identity between adults, sampled with replacement. Thus, if we draw among N adults, $Q^{\mathrm{R}} = 1/N + (1 - 1/N)Q^{\mathrm{D}}$. A recursion for probabilities of identity by descent (IBD) is constructed by considering events backward in time, that is, dispersal preceded by mutation preceded by reproduction, as follows.

Dispersal. To describe the evolution of probabilities of identities of

[1] Stochasticity of the dispersal rate is easily incorporated in the infinite island model. If the immigration rate m fluctuates independently between generations, the recursion for identity within demes may be written $\dot{Q}_0' = \mathrm{E}[(1-m)^2][\dot{Q}_0 + (1-\dot{Q}_0)/N]$; hence, the differentiation is as in a population with dispersal rate m_{e} such that $(1-m_{\mathrm{e}})^2 = \mathrm{E}[(1-m)^2] = [1-\mathrm{E}(m)]^2 + \mathrm{Var}(m)$: differentiation is increased in comparison with a population with fixed rate $\mathrm{E}(m)$ (Whitlock 1992b).

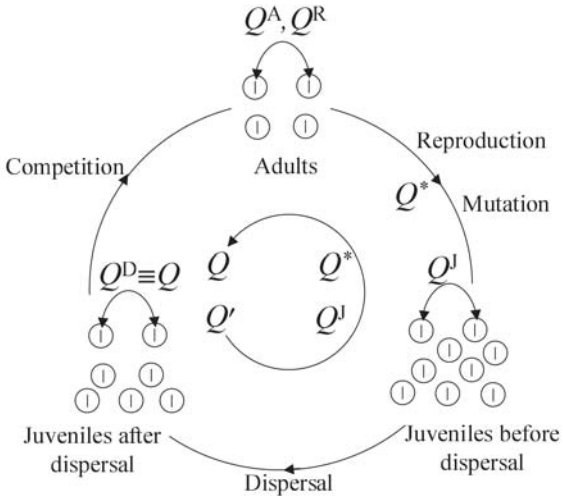

FIGURE 3.1. Assumed life cycle. The outer circle shows the order of events and the notation used for probabilities of identity at different stages of the cycle. The inner cycle shows the identities considered to establish the recursion equations when events are considered backwards: identity in the offspring generation after dispersal Q', preceded by identity before dispersal Q^J, preceded by identity among gametes before mutation Q^*, preceded by identity among parents after dispersal Q.

a pair of genes, we need to track the movements of pairs of gene lineages. For example, when there are only two demes, we need the probability that the parents of two genes presently in deme 1 were in the same deme in the previous generation (both in deme 1 or both in deme 2). This probability is $(1-m)^2+m^2 = 1-2m(1-m)$. More generally, we consider the transition probabilities between the states "1" (lineages in the same deme) and "2" (lineages in different demes). These probabilities may be written in matrix form. Element a_{11} is $(1-m)^2$[for no dispersal]$+m$[when first gene is immigrant]$\times m/(n_\mathrm{d}-1)$[when second gene is immigrant too, and from the same deme of origin]. So $a_{11} = (1-m)^2 + m^2/(n_\mathrm{d}-1)$. Likewise, the probability that two genes come from the same deme, given that they are in different demes, is $a_{21} = 2m(1-m)/(n_\mathrm{d}-1)$[when one gene did not move, and the other moved from the first one's deme] $+ m^2(n_\mathrm{d}-2)/(n_\mathrm{d}-1)^2$[when genes in demes k and $l \neq k$ moved from deme $i \neq k, l$]. Hence,

$a_{21} = [2 - mn_{\rm d}/(n_{\rm d}-1)]m/(n_{\rm d}-1) = (1-a_{11})/(n_{\rm d}-1)$ (Nagylaki 1983). Other elements of the matrix are $a_{12} = 1 - a_{11}$ and $a_{22} = 1 - a_{21}$. For example, if $n_{\rm d} = 2$,

$$\mathbf{A} \equiv (a_{ij}) \equiv \begin{pmatrix} 1-2m(1-m) & 2m(1-m) \\ 2m(1-m) & 1-2m(1-m) \end{pmatrix}. \tag{3.1}$$

Let $\dot{\mathbf{Q}}' \equiv (\dot{Q}_0', \dot{Q}_1')^\top$ be the vector of probabilities of identity either within ($\dot{Q}_0$) or among demes ($\dot{Q}_1$) among different adults or juveniles after dispersal but before competition at generation t.[2] We see that $\dot{\mathbf{Q}}' = \mathbf{A}\dot{\mathbf{Q}}^{\rm J}$, where $\dot{\mathbf{Q}}^{\rm J}$ is the column vector of IBD probabilities immediately before dispersal.

Mutation. We consider the vector $\dot{\mathbf{Q}}^* \equiv (\dot{Q}_0^*, \dot{Q}_1^*)^\top$ of IBD probabilities immediately after gamete production but before mutation, within ($\dot{Q}_0^*$) or among demes ($\dot{Q}_1^*$). In the infinite allele model, the identity between juveniles is reduced by mutation, on average by a factor of $\gamma \equiv (1-\mu)^2$. Hence, $\dot{\mathbf{Q}}^{\rm J} = \gamma\dot{\mathbf{Q}}^*$.

Reproduction. Two juveniles in a deme before dispersal may be the offspring of the same haploid adult (i.e., coalescence occurs) with probability $1/N$, assuming that each adult produces equal numbers of juveniles. Then $\dot{Q}_0^* = \dot{Q}_0^{\rm R} \equiv \dot{Q}_0(1-1/N)+1/N = \dot{Q}_0+(1-\dot{Q}_0)/N$. On the other hand, two juveniles in different demes before dispersal cannot be the offspring of the haploid adult, so $\dot{Q}_1^* = \dot{Q}_1$.

Combining the three steps, the recursion over one generation is

$$\dot{\mathbf{Q}}' = \gamma\mathbf{A}\left(\dot{\mathbf{Q}} + \mathbf{c}\right), \tag{3.2}$$

where

$$\mathbf{c} \equiv \begin{pmatrix} \frac{1-\dot{Q}_0}{N} \\ 0 \end{pmatrix}. \tag{3.3}$$

The solution of equation (3.2) at equilibrium,

$$\dot{\mathbf{Q}} = \gamma\mathbf{A}\left(\dot{\mathbf{Q}} + \mathbf{c}\right), \tag{3.4}$$

is given by

$$\dot{\mathbf{Q}} = \gamma\left(\mathbf{I} - \gamma\mathbf{A}\right)^{-1}\mathbf{A}\mathbf{c}, \tag{3.5}$$

[2]The $^\top$ notation will be used throughout for transpose: see p. 215.

where $\mathbf{I}$ is the identity matrix.[3] We will express it in terms of the eigenvalues of $\mathbf{A}$, which are $l_0 = 1$ and $l_1 = a_{11} - a_{21} = [1 - mn_\mathrm{d}/(n_\mathrm{d} - 1)]^2$, and of its eigenvectors

$$\mathbf{e}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \mathbf{e}_1 = \begin{pmatrix} n_\mathrm{d} - 1 \\ -1 \end{pmatrix}. \tag{3.6}$$

Since

$$\mathbf{c} = \frac{1 - \dot{Q}_0}{N} \frac{1}{n_\mathrm{d}} (\mathbf{e}_0 + \mathbf{e}_1), \tag{3.7}$$

then

$$\mathbf{Ac} = \frac{1 - \dot{Q}_0}{N} \frac{1}{n_\mathrm{d}} (\mathbf{e}_0 + l_1 \mathbf{e}_1) \tag{3.8}$$

and, using equation (A.10),

$$\dot{\mathbf{Q}} = (\mathbf{I} - \gamma \mathbf{A})^{-1} \gamma \mathbf{Ac} = \frac{1 - \dot{Q}_0}{N} \frac{1}{n_\mathrm{d}} \left(\frac{\gamma}{1 - \gamma} \mathbf{e}_0 + \frac{\gamma l_1}{1 - \gamma l_1} \mathbf{e}_1 \right), \tag{3.9}$$

which yields $\dot{Q}_0$ and $\dot{Q}_1$. In particular, we find that

$$\frac{\dot{Q}_0}{1 - \dot{Q}_0} = \frac{1}{N n_\mathrm{d}} \left(\frac{\gamma}{1 - \gamma} + (n_\mathrm{d} - 1) \frac{\gamma l_1}{1 - \gamma l_1} \right) \tag{3.10}$$

and

$$\frac{\dot{Q}_1}{1 - \dot{Q}_0} = \frac{1}{N n_\mathrm{d}} \left(\frac{\gamma}{1 - \gamma} - \frac{\gamma l_1}{1 - \gamma l_1} \right). \tag{3.11}$$

When there is no mutation, all genes become identical. Thus, when the mutation rate $\mu \to 0$, the denominator $1 - \dot{Q}_0$ vanishes and the above ratios diverge to infinity. Yet if we compute the difference between the two above expressions,

$$\frac{\dot{F}_\mathrm{ST}}{1 - \dot{F}_\mathrm{ST}} = \frac{\dot{Q}_0 - \dot{Q}_1}{1 - \dot{Q}_0} = \frac{1}{N} \frac{\gamma l_1}{1 - \gamma l_1}, \tag{3.12}$$

we see that it does not diverge when $\mu \to 0$, because $l_1 < 1$ and because the $1 - \gamma$ denominator no longer appears in the right-hand side of the equation. The above result has a limit

$$\frac{1}{N} \frac{l_1}{1 - l_1} \tag{3.13}$$

[3] This notation will be used throughout the book: see p. 215.

as $\mu \to 0$. This is a function of the demographic parameters, here N, m, and n_{d}. The variation of F_{ST} as a function of n_{d} is relatively weak because n_{d} appears only in the ratio $n_{\mathrm{d}}/(n_{\mathrm{d}}-1)$. In the limit case where both $n_{\mathrm{d}} \to \infty$ and $\mu \to 0$, we recover Wright's famous result, in the form

$$\dot{F}_{\mathrm{ST}} = \frac{\dot{Q}_0 - \dot{Q}_1}{1-\dot{Q}_1} = \frac{(1-m)^2}{(1-m)^2 + N[1-(1-m)^2]} \tag{3.14}$$

$$\approx \frac{1}{1+2Nm}. \tag{3.15}$$

Remember that we previously obtained this result assuming that the identity of genes from different demes is zero (p. 16), that is, that $\dot{Q}_1 \to 0$ when $n_{\mathrm{d}} \to \infty$. Does this still hold in the low mutation limit? This may be shown to depend on the product of the total population size $N_{\mathrm{T}} \equiv N n_{\mathrm{d}}$ and of the mutation rate μ. In contrast to $\dot{Q}_0$ and $\dot{Q}_1$, $\dot{F}_{\mathrm{ST}}$ has the same value whether $N_{\mathrm{T}}\mu \to \infty$ or not in the limit. We previously (p. 16) computed $\dot{F}_{\mathrm{ST}}$ as the probability of IBD within demes, because we implicitly assumed $N_{\mathrm{T}}\mu \to \infty$. In this case, $\dot{Q}_1 \to 0$. Equation (3.14) can likewise be obtained by assuming $N_{\mathrm{T}}\mu \to \infty$ and $\mu \to 0$, which allows us to write

$$\dot{Q}_0' = (1-m)^2 \dot{Q}_0^{\mathrm{J}} \text{ and } \dot{Q}_0^{\mathrm{J}} = \dot{Q}_0^{\mathrm{R}} = \dot{Q}_0 + (1-\dot{Q}_0)/N. \tag{3.16}$$

ISOLATION BY DISTANCE

In natural populations, dispersal preferentially occurs between geographically close subpopulations. We first illustrate this fact with a brief overview of observed dispersal distributions. Next, we focus on the models of "isolation by distance," which take it in account. The analysis of such models was first attempted by Wright (1943, 1946) and by Malécot (1948), but all rigorous analyses are based on the lattice model formulated by Malécot (1950). This model can consider an arbitrary distribution of dispersal distances, with the island model as a special case. This generality is useful, as dispersal distributions are very diverse in natural populations. Further, it allows us to distinguish results that are valid under more or less general forms of dispersal distributions.

Dispersal in Natural Populations

WHAT IS "DISTANCE"?

In two dimensions, dispersal can be described by two coordinates (x, y) (projections on orthogonal axes), which define the vectorial distance. x or y are known as axial dispersal distances and may be negative. Distance, in everyday language, refers to the Euclidian distance $r \equiv \sqrt{x^2 + y^2}$.

MOMENTS OF DISPERSAL DISTRIBUTIONS

Dispersal distributions may also be characterized by their moments, defined as $\mathrm{E}[X^k] = \sum_x x^k \Pr(X = x)$ (noncentral moments) or as $M_k \equiv \mathrm{E}\{[X - \mathrm{E}(X)]^k\}$ (central moment, i.e., around the mean). The variance is the second central moment, $\mathrm{Var}(X) = M_2 - M_1^2$. The kurtosis may be defined as $M_4/M_2^2 - 3$ (the "-3" is a convention that makes the kurtosis of the normal distribution zero). Among different distributions with the same variance, those with a higher kurtosis tend to have more dispersal at long distances ("longer tail"), more at short distances, and less at intermediate distances. A distribution with a high kurtosis is often said to have a long tail.

Dispersal is symmetrical if the distribution of $-x$ is the same as that of x. In this case, the mean of x is zero and its variance is measured around this zero mean, rather than around the mean of unsigned distance $|x|$. There is no simple relationship between axial or Euclidian distances, except for mean squared values. For example, assume that parent-offspring distance is described by (x, y), where x and y are independent and normally distributed with mean zero and variance σ^2. Denote $g(x)$ the density of this normal distribution, $g(x) = e^{-x^2/2\sigma^2}/(\sqrt{2\pi}\sigma)$. The mean squared Euclidian distance is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)g(y)(x^2 + y^2)\, dy\, dx = 2\sigma^2, \qquad (3.17)$$

which is twice the mean squared axial distance. This is not the square of the mean Euclidian distance, this mean being

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)g(y)\sqrt{x^2 + y^2}\, dy\, dx = \sqrt{\pi/2}\sigma \approx 1.25\sigma. \qquad (3.18)$$

Thus, assuming normal dispersal, a σ^2 of $1\,\mathrm{km}^2$ implies a mean Euclidian distance of $1.25\,\mathrm{km}$. These relationships are not conserved for other dispersal distributions.

Note that the variance of Euclidian distance is, by definition of the variance, $\mathrm{Var}(r) \equiv \mathrm{E}\{[r - \mathrm{E}(r)]^2\} = \mathrm{E}(r^2) - \mathrm{E}(r)^2$. For normally distributed dispersal, this is $\approx 0.43\sigma^2$. There has been some confusion between this "variance of dispersal" and the mean squared dispersal distance σ^2, which is the only "variance" considered below.

Observed Dispersal

There are still remarkably few complete data on dispersal analyzed by appropriate quantitative methods. Most demographic studies have focused on specific components of dispersal, such as natal dispersal (dispersal before reproduction), dispersal of seeds, or dispersal of pollen. Endler (1977) provides an extensive review of the earlier literature, showing that dispersal is often quite localized. Analysis of clines of selected alleles have provided estimates of σ^2: Endler (1977, p. 156–162) reviews rough estimates from the older literature. Most of them are within a kilometer.

The physical ability to disperse appears to be a poor predictor of dispersal distributions. For example, birds, even migratory ones, do not always disperse far from their natal site (e.g., Friesen et al. 1996; Doncaster et al. 1997). In lions, males disperse within a few home ranges from their natal territory (Spong and Creel 2001), while females disperse even less (Packer and Pusey 1993). In some New Guinea forest people, mean squared parent-offspring distance was $3.1\,\mathrm{km}^2$ for females and $0.76\,\mathrm{km}^2$ for males (Wood et al. 1985).

Dispersal of insects is notoriously difficult to measure. Occasional long-distance dispersal is easily documented through colonization of habitat distant from nearest populations, but the bulk of the dispersal distribution is generally poorly known. Mark-recapture studies generally do not estimate parent-offspring dispersal distances. Some recent studies of clines have provided relatively high estimates of σ in insects (*Pontia* butterflies, Porter et al. 1997; and *Culex pipiens* mosquitoes, Lenormand et al. 1999). In *Culex*, mean squared parent-offspring distance may be $86\,\mathrm{km}^2$ during the reproduction season.

In plants, dispersal of both pollen and seeds may be quite limited.

There are few satisfactory data sets documenting both in the same species in the literature, and higher pollen dispersal is the most common pattern (Crawford 1984; Beattie and Culver 1979; Fenster 1991; Eguiarte et al. 1993; Dutech et al. 2002; Boshier et al. 1995), with exceptions (Oddou-Muratorio et al. 2001; Ruckelshaus 1996).

Fitting to Theoretical Distributions

There have been many attempts to fit observed dispersal distributions to various theoretical distributions. The normal distribution has often been considered. Its main quality for fitting purposes is its convexity at short distances (i.e., it is curved downward at the origin); it may thus be used to fit to observed convex dispersal distributions. But it does not capture other important features. For example, observed dispersal distributions are also often leptokurtic (e.g., Bateman 1950; reviews and data in Endler 1977; Portnoy and Willson 1993; Clark et al. 1999), while the normal distribution is not leptokurtic.

To describe dispersal distances, various other dispersal distributions have been proposed, either derived from mechanistic models of dispersal or not (see Turchin 1998 for a review of much of this work). Of course, such functions can be used to describe the distribution within the observed range, but do not tell the shape of its tail at larger distances unless the same mechanism of dispersal underlies the observed and unobserved parts of the distribution and generates a distribution of the form assumed. A family of distributions allowing both a high kurtosis and a convex shape of the distribution at short distances is the bivariate Student distribution (Clark et al. 1999), whose probability density at a point at distance r is $f(r) = p/[\pi u(1 + r^2/u)^{p+1}]$. Here p is a parameter controlling the shape of the dispersal curve and u is a scale parameter. Families of distributions may be classified according to the shape of their tail, which determines whether the higher moments are finite or not. The finiteness of the second and third moments is important for deducing some patterns of population structure. A distribution whose tail decreases as r^{-n} has moments of order $k < n - 1$ finite; the higher moments are infinite. A distribution whose tail decreases faster than r^{-n} for all n has all moments finite. Examples are the normal, the lognormal, or any distribution with an exponential tail, of form e^{-r}. In

the case of the bivariate Student, the tail is of order r^{-2p-1},[4] so moments of order $\geq 2p$ are infinite. For $p > 10$, the bivariate Student is practically indistinguishable from the Gaussian.

Despite many fitting attempts, the shape of observed distributions remains poorly characterized. For example, using the bivariate Student model, Clark et al. (1999) analyzed several seed dispersal distributions and found a number of observed distributions with $p < 1/2$, which implies an infinite variance and even an infinite mean Euclidian distance. They also found many distributions with $p > 10$ and relatively few with $1/2 < p < 10$. However, fitting with the bivariate Student distribution has some important limitations. First, the parameter p, which determines the shape of the tail, also describes the shape of the distribution near the origin. Thus, inferences about the tail will depend on the shape at short distances. Hence, it remains unclear how often dispersal distributions have tails best fitted by different values of p. Further, in some cases the inferred median dispersal distance was close to the maximum distance over which seed dispersal was measured. In these cases the observed "tails" may ignore a large fraction of seeds. Finally, neither this model nor the other families of distributions combine the properties of maximum dispersal at an intermediate distance and of variable tail behavior. In contrast, axial distributions with a minimum at the origin may be generated by wind (e.g., the lognormal model, Greene and Johnson 1989; Stoyan and Wagner 2001) or by ballistic dispersal (Neubert et al. 1995).

In practice, dispersal distributions are very diverse, and we must address this diversity in modeling the consequences of localized dispersal on genetic differentiation. Therefore, we will not focus on results for the normal distribution or any other single distribution, but rather consider results common to different dispersal distributions.

The Lattice Models

The major analytical tool in the analysis of the lattice models, Fourier analysis, can be understood as a special case of the same matrix algebra as that used for the island model. Here we use matrix analysis,

[4]The tail of $f(r)$ is of order r^{-2p-2}, but the distribution of Euclidian distance is obtained by integrating $f(r)$ over all points at distance r, so its tail is that of $rf(r)$.

FIGURE 3.2. The spatial settings assumed in the lattice models. Demes are regularly spaced on a circle or on a torus.

following Maruyama (1970b). Fourier analysis is developed in the first appendix to this chapter.

Spatial homogeneity implies no edge effects. Thus, in one dimension the demes can be represented on a circle, and in two dimensions they can be represented on a torus (fig. 3.2). On a circle, demes may be labeled clockwise, and dispersal between them may be described by the probabilities f_{ij} that a gene sampled in deme i had its parent in deme j one unit time before. For four demes, these probabilities form the matrix

$$\mathbf{F} = (f_{ij}) = \begin{pmatrix} 1-m & m/2 & 0 & m/2 \\ m/2 & 1-m & m/2 & 0 \\ 0 & m/2 & 1-m & m/2 \\ m/2 & 0 & m/2 & 1-m \end{pmatrix}. \qquad (3.19)$$

Here the dispersal rate is m, with dispersal between adjacent demes on the circle only. From this matrix, we can compute the probabilities $a_{\mathbf{ij}}$ that two gene lineages, at $\mathbf{i}$ steps from each other after dispersal, were at $\mathbf{j}$ steps from each other before dispersal. Here we use bold notation ($\mathbf{i}$ and $\mathbf{j}$) for vectorial distances in one or two dimensions, and it is convenient to use the same notation for the position on a circle, with $\mathbf{i} = i, \mathbf{j} = j$.

For example, a_{00} is the probability that two lineages in the same deme were in the same lineage one generation before. This is the probability that neither gene moved $[(1-m)^2]$ plus the probability that both moved clockwise $[(m/2)^2]$ plus the probability that both moved counterclockwise $[(m/2)^2]$. More generally, $a_{\mathbf{ij}} = \sum_{\mathbf{r}=0}^{n_\mathrm{d}-1} f_{\mathbf{ir}} f_{\mathbf{i}-\mathbf{j}+\mathbf{r}}$ (with subscripts read on a circle: e.g., $f_{1,-1} = f_{1,n_\mathrm{d}-1}$). For $n_\mathrm{d} = 4$,

these probabilities form the matrix

$$\mathbf{A} \equiv (a_{\mathbf{ij}}) = \mathbf{F}^\top \mathbf{F} \tag{3.20}$$

$$= \begin{pmatrix} (1-m)^2+m^2/2 & (1-m)m & m^2/2 & (1-m)m \\ (1-m)m & (1-m)^2+m^2/2 & (1-m)m & m^2/2 \\ m^2/2 & (1-m)m & (1-m)^2+m^2/2 & (1-m)m \\ (1-m)m & m^2/2 & (1-m)m & (1-m)^2+m^2/2 \end{pmatrix}. \tag{3.21}$$

We label probabilities of identity according to the distance between genes, $\dot{Q}_0, \ldots, \dot{Q}_3$. Two juveniles in a deme before dispersal are the offspring of the same haploid adult with probability $1/N$, so the recursion for these probabilities takes the form $\dot{\mathbf{Q}}' = \gamma \mathbf{A} \left(\dot{\mathbf{Q}} + \mathbf{c} \right)$ for $\mathbf{c} = [(1 - \dot{Q}_0)/N, 0, \ldots, 0]^\top$. Hence, the recursion at equilibrium is of the same form as equation (3.4) and can be analyzed by the same algebraic techniques. We need the eigenvectors and eigenvalues of $\mathbf{A}$. $\mathbf{A}$ is a circulant matrix (each row is a cyclic permutation of the previous one, cycled forward one step). For such matrices, eigenvectors and eigenvalues take a relatively simple form, using complex numbers. The eigenvalues are, for $j = 0$ to $j = n_\mathrm{d} - 1$,

$$l_j = \sum_{k=0}^{n_\mathrm{d}-1} a_{0k} e^{\imath 2\pi kj/n_\mathrm{d}} = \sum_{k=0}^{n_\mathrm{d}-1} \sum_{r=0}^{n_\mathrm{d}-1} f_{0r} f_{0,r-k} e^{\imath 2\pi kj/n_\mathrm{d}} \tag{3.22}$$

$$= \left(\sum_{r=0}^{n_\mathrm{d}-1} f_{0r} e^{\imath 2\pi rj/n_\mathrm{d}} \right) \left(\sum_{l=0}^{n_\mathrm{d}-1} f_{0l} e^{-\imath 2\pi lj/n_\mathrm{d}} \right). \tag{3.23}$$

In the stepping stone model, each of the latter two sums equals

$$1-m+\frac{m}{2} e^{-\imath 2\pi j/n_\mathrm{d}} + \frac{m}{2} e^{\imath 2\pi j/n_\mathrm{d}} = 1-m[1-\cos(2\pi j/n_\mathrm{d})]. \tag{3.24}$$

Hence,

$$l_j = \{1 - m[1 - \cos(2\pi j/n_\mathrm{d})]\}^2. \tag{3.25}$$

Likewise, the eigenvectors are

$$\mathbf{e}_j \equiv (e_{jk}) = \left(1, \ldots, e^{\imath 2\pi jk/n_\mathrm{d}}, \ldots, e^{\imath 2\pi j(n_\mathrm{d}-1)/n_\mathrm{d}}\right)^\top \tag{3.26}$$

for $j, k = 0, \ldots, n_\mathrm{d} - 1$. These vectors are orthogonal. That is, the scalar product $\mathbf{e}_j \cdot \bar{\mathbf{e}}_k \equiv \sum_{l=0}^{n_\mathrm{d}-1} e_{jl} \bar{e}_{kl}$ is zero for any pair of different

vectors ($j \neq k$), where $\bar{e}_{kl}$ is the complex conjugate of e_{kl} (here, it is $e^{-\imath 2\pi kl/n_\mathrm{d}}$).[5] On the other hand, for any k,

$$\mathbf{e}_k \cdot \bar{\mathbf{e}}_k = n_\mathrm{d} \tag{3.27}$$

[see equation (A.4)]. Then, by equation (A.2),

$$\mathbf{c} = \sum_{k=0}^{n_\mathrm{d}-1} \frac{\mathbf{e}_k \cdot \mathbf{c}}{\mathbf{e}_k \cdot \bar{\mathbf{e}}_k} \mathbf{e}_k = \frac{1-\dot{Q}_0}{N n_\mathrm{d}} \sum_{k=0}^{n_\mathrm{d}-1} \mathbf{e}_k. \tag{3.28}$$

Proceeding as in equation (3.9),

$$\dot{\mathbf{Q}} = (\mathbf{I} - \gamma \mathbf{A})^{-1} \gamma \mathbf{A} \mathbf{c} = \frac{1-\dot{Q}_0}{N n_\mathrm{d}} \sum_{k=0}^{n_\mathrm{d}-1} \frac{\gamma l_k}{1-\gamma l_k} \mathbf{e}_k \tag{3.29}$$

$$= \frac{1-\dot{Q}_0}{N n_\mathrm{d}} \sum_{k=0}^{n_\mathrm{d}-1} \frac{\gamma\{1-m[1-\cos(2\pi k/n_\mathrm{d})]\}^2}{1-\gamma\{1-m[1-\cos(2\pi k/n_\mathrm{d})]\}^2} \mathbf{e}_k. \tag{3.30}$$

Equation (3.30) holds for nearest-neighbor dispersal, while equation (3.29) holds for any distribution of dispersal distance, with l_k different from equation (3.25). Equation (3.29) directly extends into equations (3.61) and (3.62) for two dimensions, which give the identity $\dot{Q}_\mathbf{r}$ as a function of the vectorial distance $\mathbf{r} = (x, y)$ between the genes sampled.

Differentiation under Isolation by Distance

From these formulas, one can deduce the value of various population structure parameters. Kimura and Weiss (1964) and Weiss and Kimura (1965) considered a "correlation function" that they noted $r(\mathbf{k})$, where $\mathbf{k}$ is the vectorial distance between genes. In terms of probabilities of identity between adults, $r(\mathbf{k}) = Q^\mathrm{R}_\mathbf{k}/Q^\mathrm{R}_\mathbf{0}$ for the haploid model. With $N = 1$ haploid adult per deme, this is simply $Q^\mathrm{R}_\mathbf{k}$. In addition, they focused on the asymptotic decrease of $r(\mathbf{k})$ at large distances. We will focus on differentiation at shorter distances and consider instead

$$\mathsf{a}(\mathbf{r}) \equiv \frac{Q_\mathbf{0} - Q_\mathbf{r}}{1 - Q_\mathbf{0}}. \tag{3.31}$$

We consider $\mathsf{a}(\mathbf{r})$ for convenience, as expressions for it are relatively simple. The pairwise F_ST, comparing identity of genes within demes

[5] The dot notation will be used throughout for the scalar product: see p. 215.

to identity of genes at distance $\mathbf{r}$, is

$$\frac{Q_{\mathbf{0}} - Q_{\mathbf{r}}}{1 - Q_{\mathbf{r}}} = \frac{\mathsf{a}(\mathbf{r})}{1 + \mathsf{a}(\mathbf{r})}. \tag{3.32}$$

Other statistics have been considered, in particular autocorrelation statistics, such as "Moran's I" which may be viewed as an estimator of $(Q_{\mathbf{r}} - \bar{Q})/(1 - \bar{Q})$, where $\bar{Q}$ is an average identity across individuals in a sample (Hardy and Vekemans 1999). This average is an expected value that depends on the sampling design, introducing unnecessary complications.

Expressions for $\mathsf{a}(\mathbf{r})$ follow from those of the previous section. Equation (3.29) implies that, for demes at distance r in a one-dimensional habitat,

$$\mathsf{a}(r) = \frac{\dot{Q}_0 - \dot{Q}_r}{1 - \dot{Q}_0} = \frac{1}{Nn_{\mathrm{d}}} \sum_{k=0}^{n_{\mathrm{d}}-1} \frac{\gamma l_k}{1 - \gamma l_k} \left(\mathbf{e}_{k0} - \mathbf{e}_{kr}\right). \tag{3.33}$$

In the one-dimensional stepping stone model, for low mutation and an infinite number of demes, this reduces to

$$\mathsf{a}(r) = \frac{1}{4N} \left(-4 + \frac{1}{\sqrt{1-m}} - \frac{\left(\frac{-2+2\sqrt{1-m}+m}{m}\right)^r}{\sqrt{1-m}} + \frac{2\,r}{m} \right) \tag{3.34}$$

$$\rightarrow \frac{1}{N} \left(-1 + \frac{1}{4\sqrt{1-m}} + \frac{r}{2m} \right) \text{ for large } r. \tag{3.35}$$

For a two-dimensional habitat, figure 3.3 compares values of $\mathsf{a}(\mathbf{r})$ at different distances and for different dispersal distributions of dispersal distance, with the same total dispersal probability but with more or less distant dispersal. This figure illustrates two of the simplest results that emerge from the analysis of the models. First, when total dispersal probability is low, local differentiation is close to that predicted from the island model. Second, in a two-dimensional habitat, differentiation increases linearly with the logarithm of spatial distance. In the one-dimensional habitat, differentiation increases linearly with distance, as equation (3.35) shows. These results are now developed.

AVERAGE DIFFERENTIATION UNDER ISOLATION BY DISTANCE

Whatever the dispersal distribution, F_{ST} will be approximately the same as that in the island model, at least when the total dispersal rate

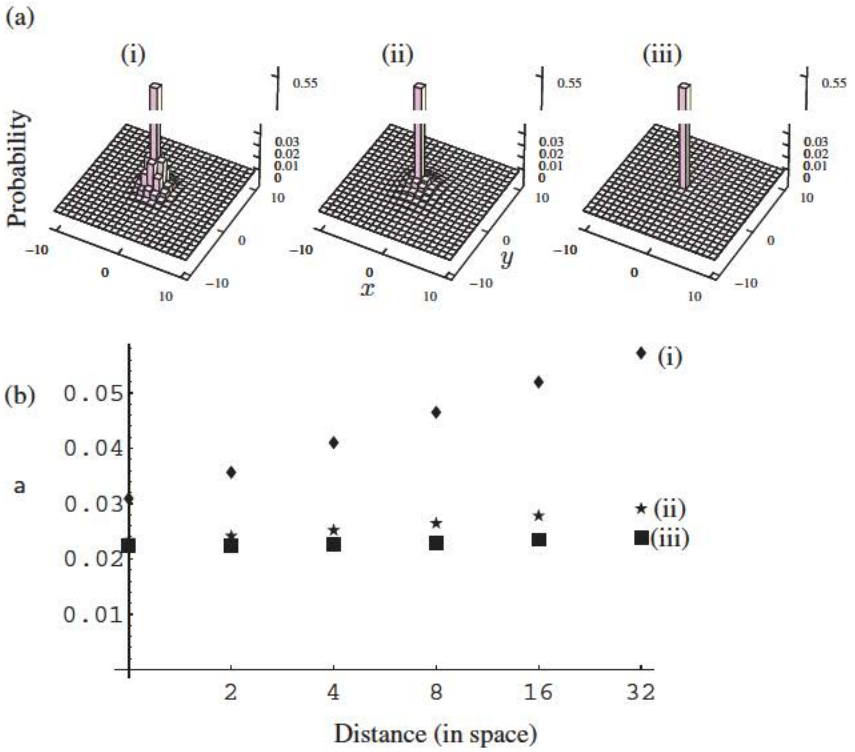

FIGURE 3.3. Differentiation as a function of distance. (a) Three two-dimensional dispersal distributions are shown as probabilities of dispersal to distances $\mathbf{r} = (x, y)$; all have the same total dispersal rate $m = 4/9$. (b) Differentiation is shown for these three distributions [with symbols ♦, ★, and ■ for distributions (i), (ii), and (iii), respectively] and for $N = 10$ diploid adults per deme. Notice the logarithmic scale for spatial distance. With the same values $m = 4/9$ and $N = 10$, $F_{ST} = 0.0218$ in an island model; this is close to the value obtained at short distances in the different cases.

is low. This result had been noticed in simulations by Kimura and Maruyama (1971). This can be shown mathematically for a weighted average of differentiations at different distance, where weights are proportional to the dispersal rates:

$$\bar{\mathsf{a}} \equiv \frac{\sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}}\mathsf{a}(\mathbf{r})}{\sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}}} = \frac{\sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}}\mathsf{a}(\mathbf{r})}{m}. \tag{3.36}$$

for the total dispersal rate $m \equiv \sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}}$. Thus, this is a local average of differentiation when dispersal is localized. Equation (3.75) shows that $\sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}}\mathsf{a}(\mathbf{r}) \approx 1/(2N)$, and hence $\bar{\mathsf{a}} \approx 1/(2Nm)$. Then $\bar{\mathsf{a}}/(1+\bar{\mathsf{a}}) \approx 1/(1+2Nm)$, or $1/(1+4Nm)$ as usual in the corresponding diploid model. Note that

$$\frac{\bar{\mathsf{a}}}{1+\bar{\mathsf{a}}} = \frac{Q_{\mathbf{0}} - \sum_{\mathbf{r}\neq\mathbf{0}} Q_{\mathbf{r}} m_{\mathbf{r}}/m}{1 - \sum_{\mathbf{r}\neq\mathbf{0}} Q_{\mathbf{r}} m_{\mathbf{r}}/m}. \tag{3.37}$$

Hence, it can be viewed as an F_{ST}, measuring identity within demes relative to a weighted identity among demes.

Spatial Genetic Patterns

The evaluation of spatial patterns from the expressions developed above is complicated. A first "simplification" is to consider populations with an infinite number of demes (infinite lattices). When the lattice size increases indefinitely ($n_x \to \infty, n_y \to \infty$), the sum in equation (3.29) and its two-dimensional equivalent converge to integrals detailed in the appendix (p. 49). Many works have proposed more or less general expressions for such integrals, or for ensuing expressions for probabilities of identity (e.g., Malécot 1950, 1951, 1967, 1975; Maruyama 1972; Nagylaki 1974, 1976, 1989; Sawyer 1977). We review here the most useful approximations that have emerged from these analyses.

The following results are based on a few assumptions in terms of the absolute moments of order k of the dispersal distribution, $\mathrm{E}(|X|^k) \equiv \sum_x |x|^k \Pr(X = x)$. In particular, it is assumed that the absolute moments of order k are finite for all $k \leq \zeta$ for some $\zeta > 3$. This is a way of saying that most dispersal is local, although the fourth moment, hence the kurtosis, can be infinite. It is definitely not assumed that dispersal follows a normal distribution. We also assume that dispersal is independent and identically distributed in two dimensions, but results similar to those presented below may be obtained under more general assumptions (Sawyer 1976, section 4.2.2; Sawyer 1977).

Of particular importance is the second moment of axial distance, σ^2, best described as the mean squared parent-offspring axial dispersal distance, and more confusingly described as the "variance of dispersal" (see p. 30). For symmetric dispersal distributions, the second moment is also the variance of signed axial distance, but it is not the variance of

unsigned distance. A useful interpretation of σ^2 is that σ^2 is a measure of the speed at which two gene lineages issued from an ancestor move away from each other, since it is the rate at which the mean squared axial distance between these two lineages increases per time unit.

The most detailed results are given by Sawyer (1977), which should be consulted for derivation of the formulas below. *In one dimension*,

$$\frac{\dot{Q}_r}{1-\dot{Q}_0} \approx \frac{e^{-\sqrt{2\mu}r/\sigma}}{2N\sigma\sqrt{2\mu}}. \tag{3.38}$$

This is an approximation for large geographic distances.[6] For $r = 0$,

$$\frac{\dot{Q}_0}{1-\dot{Q}_0} \approx \frac{1}{2N\sigma\sqrt{2\mu}} + \frac{A_1}{2N\sigma}, \tag{3.39}$$

where A_1 is a constant determined by the dispersal distribution, but not by N or μ.[7] Its definition is given by Sawyer (1977, equation 2.4):

$$A_1 = 2\sigma\left(\frac{1}{\pi}\int_0^\pi \frac{\psi^2(x)}{1-\psi^2(x)} - \frac{1}{\sigma^2x^2}\,dx - \frac{1}{\pi^2\sigma^2}\right). \tag{3.40}$$

In two dimensions, for two genes at distance $r \equiv \sqrt{r_x^2 + r_y^2}$, one has

$$\frac{\dot{Q}_{\mathbf{r}}}{1-\dot{Q}_{\mathbf{0}}} \approx \frac{K_0(\sqrt{2\mu}r/\sigma)}{2N\pi\sigma^2}, \tag{3.41}$$

where $K_0(x) \equiv \int_0^\infty \cos(xt)/\sqrt{t^2+1}\,dt$ is the modified Bessel function of second kind and zero order (e.g., Abramovitz and Stegun 1972). As in one dimension, this is an approximation for large r; a different formula must be considered when $\mathbf{r} = \mathbf{0}$:

$$\frac{\dot{Q}_{\mathbf{0}}}{1-\dot{Q}_{\mathbf{0}}} \approx \frac{-\ln(\sqrt{2\mu}) + 2\pi A_2}{2N\pi\sigma^2}. \tag{3.42}$$

A_2 is of the same nature as A_1 above. Its definition is given by Sawyer (1977, equation 3.4). For independent and identically distributed dispersal in each dimension, it reduces to

$$A_2 = \frac{\sigma^2}{\pi^2}\int_0^\pi\int_0^\pi \frac{\psi^2(x)\psi^2(y)}{1-\psi^2(x)\psi^2(y)} - \frac{1}{\sigma^2(x^2+y^2)}\,dx\,dy + \frac{\ln(2\pi\sigma)}{2\pi} - \frac{\lambda}{\pi^2}, \tag{3.43}$$

[6]The error on $\dot{Q}_r/(1-\dot{Q}_0)$ is $o(1)$ uniformly in μ as $r \to \infty$.
[7]The error on $\dot{Q}_0/(1-\dot{Q}_0)$ is $o(1)$ as $\mu \to 0$.

where $\lambda = 0.9159...$ is Catalan's constant.

Approximations for $\dot{a}(\mathbf{r})$ follow from the previous expressions. In one dimension,

$$\dot{a}(r) \approx \frac{A_1}{2N\sigma} + \frac{1 - e^{\frac{-\sqrt{2\mu}r}{\sigma}}}{2N\sigma\sqrt{2\mu}} \tag{3.44a}$$

$$\overset{(r\sqrt{\mu}) \text{ small}}{\approx} \frac{A_1}{2N\sigma} + \frac{r}{2N\sigma^2} \approx \frac{A_1}{2D\sigma} + \frac{r}{2D\sigma^2}, \tag{3.44b}$$

and in two dimensions,

$$\dot{a}(\mathbf{r}) \approx \frac{-\ln(\sqrt{2\mu}) - K_0(\sqrt{2\mu}r/\sigma) + 2\pi A_2}{2N\pi\sigma^2} \tag{3.45a}$$

$$\overset{(r\sqrt{\mu}) \text{ small}}{\approx} \frac{\ln(r/\sigma) - 0.116 + 2\pi A_2}{2N\pi\sigma^2} \tag{3.45b}$$

$$\approx \frac{\ln(r/\sigma) + 2\pi A_2'}{2D\pi\sigma^2}, \tag{3.45c}$$

where the A_2' constant depends on the spatial unit used to measure density and dispersal distance.

In both equations (3.44) and (3.45), the first expression is given for σ measured in the length unit of the model (i.e., one interdeme distance on the lattice), the second is the small distance/low mutation limit of the first, and the third is the second for any length unit. The third is in terms of population density D per length or surface unit. Note that this represents density over the whole area considered, not the density in the favorable environments represented as demes.

SIGNIFICANCE OF THE RESULTS

The approximations demonstrate a linear relationship between $\dot{a}(r)$ and geographical distance in one dimension, and between $\dot{a}(r)$ and the logarithm of geographical distance in two dimensions (as further illustrated by Rousset 1997, 1999b, and fig. 3.3). The slope of this linear relationship is given by $D\sigma^2$ [being either $1/(2D\sigma^2)$ or $1/(2D\pi\sigma^2)$ in the above haploid models]. Note the nature of the approximations: at one point we assumed r is "large" and at another that $r\sqrt{\mu}$ is "small." The linearity will occur at intermediate distances: at too low distances, the first assumption is not valid.

In practice, it seems that the approximations are reasonably accurate within the range of distances from σ to $0.2\sigma/\sqrt{2\mu}$ in one dimension, or

to $0.56\sigma/\sqrt{2\mu}$ in two dimensions (Rousset 1997). At larger distances, mutation must be taken into account. Thus, it has often been noticed, both from theory and observation, that $F_{\rm ST}$ correlates with diversity, and the latter will correlate with mutation rate and total population size (e.g., Nagylaki 1998b; Hedrick 1999). Note that, in more complex models, $F_{\rm ST}$ values computed at a local scale may also be negatively correlated with variation in diversity across different demes. This is expected when such local $F_{\rm ST}$ depends on local deme sizes and on local dispersal rates (Rousset 1999b). Here mutation is not the important factor.

These results suggest a way to estimate $D\sigma^2$ from data. But they also show that differentiation is a function of the A constants, which are not only functions of σ^2 but also of other features of the dispersal distribution. It is tempting to neglect the terms A_1 and A_2 or to assume that they are functions of $N\sigma^2$. However, for $F_{\rm ST}$ the resulting approximations are not good. In fact, when the total migration rate is low, the differentiation between adjacent subpopulations is close to that expected under an island model with the same total number of migrants, as already shown on p. 36. This confirms that σ^2 is not the only relevant parameter of the dispersal distribution. Thus, it is often mistaken to assume that the "neighborhood size" $4D\pi\sigma^2$ determines differentiation [see Rousset 1997, for further discussion of Wright's (1946) definitions of neighborhood size]. Obviously, if one considers only a family of dispersal distributions characterized by σ^2, any property of the models can be related to σ^2. Statements about the role of σ^2 have meaning when different distributions, not characterized solely by σ^2, are considered.

This theoretical caveat is relevant in practice. Dispersal distributions with high kurtosis often have high A values. Stepping stone dispersal may also yield large A_2 values despite having low kurtosis (Rousset 1997). High A values might also result from continuous dispersal distributions with local maxima at nonzero distances.

LOCAL DISPERSAL DETERMINES LOCAL DIFFERENTIATION

The above formulas involve asymptotic approximations at "large" distances, so they may hold only at very large distances. In practice, both genetic differentiation and dispersal will be observed at limited distances, which may not be large enough for the approximations to be

accurate. Further, it is not possible to assess the assumption that a moment of a distribution is infinite, since realized moments are necessarily finite (Clark et al. 2001). Then how to interpret the above results? In particular, can we relate differentiation over some maximal distance d_{g} to dispersal observed over some maximal distance d_{d}?

The following argument suggests we can do so by using an estimate $\hat{\sigma}$ of σ obtained from observations of dispersal only within some maximal distance d_{d}, provided the fraction of migrants m beyond d_{d} is limited. If these migrants go to an infinite distance (so that the "true" σ^2 is infinite and does not predict any local pattern of differentiation), such migrants will be unrelated to their neighbors, and these migration events are analogous to mutation events (as in the infinite island model). Hence, we can apply, for example, equation (3.45a) with $\hat{\sigma}$ for σ and m for μ, so that

$$\dot{\mathsf{a}}(\mathbf{r}) \approx \frac{-\ln(\sqrt{2m}) - K_0(\sqrt{2m}r/\hat{\sigma})}{2D\pi\hat{\sigma}^2} + \text{constant}. \tag{3.46}$$

From this expression we can deduce as in equation (3.45c) that there is an approximately linear increase of differentiation, determined by $D\hat{\sigma}^2$, roughly up to distance $0.56\hat{\sigma}/\sqrt{2m}$ [deduced analogously to the bound $0.56\sigma/\sqrt{2\mu}$ for (3.45c) as an approximation to (3.45a)]. For example, if we miss 1% of migrants in a demographic study that estimates $\hat{\sigma}^2 = 10$, we can predict spatial patterns up to a distance $d_{\mathrm{g}} = 39.6 = 12.5\hat{\sigma}$ reasonably well.

These results show that local dispersal determines local differentiation. It should be remembered, however, that the above equation relies on assumptions about the moments of the dispersal distribution. The third moment should be finite; hence, the fraction of migrants m beyond d_{d} must fall faster than d_{d}^{-2}. This is consistent with numerical results that showed that local patterns are not well predicted otherwise (Rousset 2001).

CONTINUOUS MODELS

In more realistic models, one would like to incorporate several features. First, there would not necessarily be any "deme," that is, any patch where individuals are equivalent to each other when considered from outside the patch. In comparison, in the lattice model, the N

individuals in a deme are equivalent, for example, in terms of probability of identity with members of another deme. The absence of a demic structure is achieved in a lattice model only when $N = 1$, which has therefore been viewed as an approximation for populations without a deme structure (Malécot 1975; Slatkin and Barton 1989; Rousset 2000). Second, individuals would be allowed to settle in any position in continuous space. Then the position of individuals would be variable between generations and local density would fluctuate.

Models in continuous space have been formulated (e.g., Wright 1943, 1946; Malécot 1967, section 2; Sawyer 1977), but they substitute a continuous density of "individuals" for the actual distribution of individuals at any time. For this and additional reasons, it has long been known that these models do not follow from a well-defined set of biological assumptions (Maruyama 1972; Felsenstein 1975, 1976). Moreover, formulas derived from some of these models are incoherent. An approximation of Malécot's "continuous" model by a differential equation (Malécot 1967, equation 3.7) leads to a result similar to (3.41), but for any distance r, while (3.41) gives a limit value for $r \to \infty$, $\mu \to 0$, and $r\sqrt{\mu}$ fixed on a discrete lattice. The differential equation (rather than the continuous model itself) would therefore imply that the probability of identity $\dot{Q}_r$ diverges to infinity at short distances (Nagylaki 1974, p. 2934). Neither the discrete model nor Malécot's original continuous model show such aberrant behavior.

In a realistic continuous model, each individual is placed in different conditions characterized by its distance from its neighbors. This can be viewed as a class-structured model, where spatial patterns are approximated in terms of "effective" rates of dispersal and of coalescence (chapter 9) substituted to σ^2 and N, but it is not known here how to compute these effective rates in terms of the demographic parameters of the model. Another approach is to estimate "effective" σ^2, "effective" density D, or their product by fitting simulation results to heuristic expressions for genetic identity (Barton et al. 2002). However, because the actual relationship between identities and the effective parameters is not known, the effective parameters may not be a function of the demographic parameters only. For example, Barton et al.'s "effective density" is a function of a time scale of observation, and when applied to the lattice model, its definition does not reduce to the actual density, even for large observation times.

SUMMARY

Qualitatively, the main result of the models of isolation by distance is that moderate amounts of localized dispersal are sufficient to limit neutral differentiation to low levels. More quantitatively, as simplistic as the assumptions of the models may be, the results are not easily summarized. There is no simple relationship describing differentiation in terms of σ^2. Rather, for a low total dispersal probability, differentiation is a function of the number of migrants Nm, as in the island model. What σ^2 helps us to determine is the increase of differentiation with distance.

No species has populations homogeneously distributed as in the models. Thus, to apply the results to any species, one must consider which are the robust results. Overall, it appears that local dispersal (and local densities) determines local differentiation, so the results can be applied at a local spatial scale. Indeed, the comparison of estimates of density and of σ^2 obtained by demographic and genetic analyses based on the above models has shown that the models predict the local increase in differentiation rather well (Rousset 1997, 2000; Sumner et al. 2001; Fenster et al. 2003). A few studies support this conclusion for average levels of local differentiation as well (e.g., Whitlock 1992a; Ingvarsson et al. 1997; Lewis et al. 1997; Rousset 2003b).

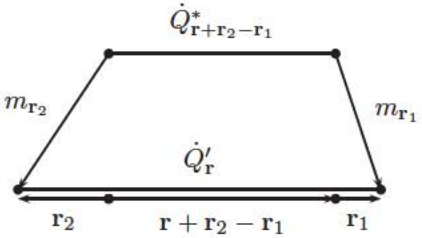

FIGURE 3.4. Graphical depiction of equation (3.47).

APPENDIX 1: GENERAL ANALYSIS OF THE LATTICE MODEL

We consider a population subdivided in different demes regularly spaced on a circle, or on a torus. There are N haploid adults ("genes") in each deme. The position of each deme on the lattice is described by a pair of coordinates, $\mathrm{r} \equiv (r_x, r_y)$. The length unit is the distance between adjacent nodes on the lattice. "Distance" here is vectorial distance. Let $\dot{Q}_{\mathrm{r}}$ be the probability of identity of a pair of genes at distance r from each other on the lattice at time $t-1$. Each individual produces an infinity of juveniles. Let $\dot{Q}^{\mathrm{J}}_{\mathrm{r}}$ be the probability of identity of two such juveniles. With probability μ, a mutation occurs between $t-1$ and t at the locus considered, independently for each juvenile. The juveniles disperse independently from each other. Competition between juveniles arriving to a deme leaves N adults alive at t. Their probability of identity is then $\dot{Q}'_{\mathrm{r}}$.

We assume that the distribution of dispersal distance is the same for any deme (homogeneous dispersal). Let m_{r} be the probability that the parent of an individual was at distance r from the present position of the individual. In other words, the probability that the parent of a gene in position r' was in position r^* is $m_{\mathrm{r}'-\mathrm{r}^*}$. After mutation and dispersal, the probability of identity of two genes is, for any r,

$$\dot{Q}'_{\mathrm{r}} = \sum_{\mathrm{r}_1}\sum_{\mathrm{r}_2} m_{\mathrm{r}_1} m_{\mathrm{r}_2} \dot{Q}^{\mathrm{J}}_{\mathrm{r}+\mathrm{r}_2-\mathrm{r}_1} = \gamma \sum_{\mathrm{r}_1}\sum_{\mathrm{r}_2} m_{\mathrm{r}_1} m_{\mathrm{r}_2} \dot{Q}^{\mathrm{R}}_{\mathrm{r}+\mathrm{r}_2-\mathrm{r}_1}. \tag{3.47}$$

Each term of this sum is obtained as the probability that two juveniles have dispersed by r_1 and r_2, respectively, in which case the parental lineages were distant by $\mathrm{r}+\mathrm{r}_2-\mathrm{r}_1$, as shown in figure 3.4. Their probability of identity was then $\dot{Q}^*_{\mathrm{r}+\mathrm{r}_2-\mathrm{r}_1}$.

As in the island model, two juveniles in a deme before dispersal

may be the offspring of the same haploid adult with probability $1/N$. Hence, $\dot{Q}^{\mathrm{R}}_{\mathbf{r}} = \dot{Q}_{\mathbf{r}}$ if $\mathbf{r} \neq \mathbf{0} \equiv (0,0)$, and $\dot{Q}^{\mathrm{R}}_{\mathbf{0}} = 1/N + (1-1/N)\dot{Q}_{\mathbf{0}}$. Let $[X = Y]$ take value 1 if $X = Y$ and take 0 otherwise. Then, for any $\mathbf{r}$, $\dot{Q}^{\mathrm{R}}_{\mathbf{r}} = \dot{Q}_{\mathbf{r}} + [\mathbf{r} = \mathbf{0}](1 - \dot{Q}_{\mathbf{0}})/N$. So

$$\begin{aligned}\dot{Q}'_{\mathbf{r}} =& \gamma \sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2} m_{\mathbf{r}_1} m_{\mathbf{r}_2} \left(\dot{Q}_{\mathbf{r}+\mathbf{r}_2-\mathbf{r}_1} + [\mathbf{r}+\mathbf{r}_2-\mathbf{r}_1 = \mathbf{0}] \frac{1-\dot{Q}_{\mathbf{0}}}{N} \right) \\ =& \gamma \sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2} m_{\mathbf{r}_1} m_{\mathbf{r}_2} \dot{Q}_{\mathbf{r}+\mathbf{r}_2-\mathbf{r}_1} + \gamma \sum_{\mathbf{r}_1} m_{\mathbf{r}_1} m_{\mathbf{r}_1-\mathbf{r}} \frac{1-\dot{Q}_{\mathbf{0}}}{N}.\end{aligned} \tag{3.48}$$

(Malécot, 1951, 1975, equation 2). There is one equation of this form for each distance $\mathbf{r}$. Using matrix notation, they can be gathered in a single expression of the form (3.2). In a one-dimensional habitat, at equilibrium

$$\dot{\mathbf{Q}} = \gamma \mathbf{A} \left(\dot{\mathbf{Q}} + \mathbf{c} \right), \tag{3.49}$$

where $\mathbf{A} \equiv \mathbf{F}^\top \mathbf{F}$ for $\mathbf{F}$ the matrix with elements $f_{i,i+r} = m_r$, and $\mathbf{c} \equiv [(1-\dot{Q}_0)/N, 0, \ldots, 0]^\top$. In two dimensions, $\mathbf{F}$ must be defined as a tensor product of migration matrices in each dimension; see Maruyama (1970b) for details.

Fourier analysis, as described below, is a classical tool for doing such algebra.

Fourier Analysis as Matrix Algebra

Let $\theta(k) \equiv 2\pi k/n$. For given variables $(f_0, \ldots, f_{n-1}) \equiv \mathbf{f}$, let $\mathcal{F}[\theta(k)] \equiv \mathbf{f} \cdot \mathbf{e}_k = \sum_{l=0}^{n-1} f_l e^{\imath l \theta(k)}$. The individual terms f_j may be recovered from $\mathcal{F}$ values by computing

$$\mathcal{L}_j(\mathcal{F}) = \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{F}[\theta(k)] e^{-\imath j \theta(k)} \tag{3.50}$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{f} \cdot \mathbf{e}_k \bar{e}_{jk} = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} f_l e_{lk} \bar{e}_{jk} \tag{3.51}$$

$$= \frac{1}{n} \sum_{l=0}^{n-1} f_l \mathbf{e}_l \mathbf{e}_j = f_j \tag{3.52}$$

from the orthogonality of eigenvectors and from equation (3.27). All the f_j's can be extracted from the function $\mathcal{F}$ by Fourier inversion; hence, manipulating a system of equations for the f_j's is equivalent to manipulating a single equation for $\mathcal{F}$. $\mathcal{F}(z)$ is known as the Fourier transform in z of the f_j's. Equation (3.50) defines the inverse Fourier transform of $\mathcal{F}$, which we denote $\mathcal{L}_j(\mathcal{F})$. In the following, $n = n_{\mathrm{d}}$ for a one-dimensional habitat.

In two dimensions, we extend these definitions as follows. We use vectorial indices $(j_x, j_y) \equiv \mathbf{j}$, $(k_x, k_y) \equiv \mathbf{k}$, $(l_x, l_y) \equiv \mathbf{l}$. For a torus of size $n_x \times n_y$, for $\mathbf{z} \equiv (x, y)$, and for given variables $f_{0,0}$, $\ldots$, f_{n_x-1,n_y-1}, the Fourier transform is defined as $\mathcal{F}(\mathbf{z}) \equiv \sum_{\mathbf{l}} f_{\mathbf{l}} e^{\imath \mathbf{l}\cdot\mathbf{z}}$ where $\mathbf{l}\cdot\mathbf{z}$ is the scalar product $l_x x + l_y y$. Let

$$\boldsymbol{\theta}(\mathbf{k}) \equiv 2\pi(k_x/n_x, k_y/n_y). \tag{3.53}$$

Any $f_{\mathbf{l}}$ may be recovered by the inverse transformation

$$\mathcal{L}_{\mathbf{l}}(\mathcal{F}) \equiv \frac{1}{n_x n_y} \sum_{\mathbf{k}=(0,0)}^{(n_x-1,n_y-1)} \mathcal{F}[\boldsymbol{\theta}(\mathbf{k})] e^{-\imath \mathbf{l}\cdot\boldsymbol{\theta}(\mathbf{k})}. \tag{3.54}$$

Since the variables of interest here are the probabilities of identity, we define $\mathcal{Q}(\mathbf{z})$, the Fourier transform of the probabilities of identity:

$$\mathcal{Q}(\mathbf{z}) \equiv \sum_{\mathbf{r}} \dot{Q}_{\mathbf{r}} e^{\imath \mathbf{r}\cdot\mathbf{z}} = \sum_{\mathbf{r}} \dot{Q}_{\mathbf{r}} \xi^{\mathbf{r}}, \tag{3.55}$$

for $\xi^{\mathbf{r}} \equiv e^{\imath \mathbf{z}\cdot\mathbf{r}}$.

Likewise, for given probabilities $m_{\mathbf{r}}$ of dispersal at distance $\mathbf{r}$, we consider the function

$$\psi(\mathbf{z}) \equiv \sum_{\mathbf{r}} m_{\mathbf{r}} e^{\imath \mathbf{r}\cdot\mathbf{z}} \equiv \sum_{\mathbf{r}} m_{\mathbf{r}} \xi^{\mathbf{r}}. \tag{3.56}$$

With this definition, the eigenvalues (3.23) of $\mathbf{A}$ are of the form $\psi[\boldsymbol{\theta}(\mathbf{k})]$ $\psi[-\boldsymbol{\theta}(\mathbf{k})]$. $\psi(\mathbf{z})$ is the Fourier transform of the distribution of dispersal distance. For example, with nearest-neighbor dispersal in one dimension, $\psi(x) = 1 - m[1 - \cos(x)]$. In a slightly more involved example, one may consider the distribution defined by the total dispersal rate m and by a geometric distribution of dispersal distance among genes that effectively disperse. This distribution has $m_0 \equiv 1 - m$ and $m_l = (1 - q)q^{l-1}m/2$ for $l \neq 0$, where q is the parameter of the geometric distribution. Then some algebra shows that $\psi(e^{ix}) = 1 - \{1 - (1 - q)[\cos(x) - q]/[1 - 2\cos(x)q + q^2]\}m$.

Note that

$$\sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2} m_{\mathbf{r}_1} m_{\mathbf{r}_2} \xi^{\mathbf{r}_1-\mathbf{r}_2} = \sum_{\mathbf{r}_1} m_{\mathbf{r}_1}\xi^{\mathbf{r}_1} \sum_{\mathbf{r}_2} m_{\mathbf{r}_2}\xi^{-\mathbf{r}_2} = \psi(\mathbf{z})\psi(-\mathbf{z}). \tag{3.57}$$

Further, using equation (3.48), $\mathcal{Q}'(\mathbf{z})$ can be written

$$\begin{aligned}
\sum_{\mathbf{r}} \dot{Q}'_{\mathbf{r}}\xi^{\mathbf{r}} =& \gamma\left(\sum_{\mathbf{r}}\sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2} m_{\mathbf{r}_1} m_{\mathbf{r}_2}\dot{Q}_{\mathbf{r}+\mathbf{r}_2-\mathbf{r}_1}\xi^{\mathbf{r}}\right. \\
&\left. \sum_{\mathbf{r}}\sum_{\mathbf{r}_1} m_{\mathbf{r}_1} m_{\mathbf{r}_1-\mathbf{r}}\xi^{\mathbf{r}}\frac{1-\dot{Q}_{\mathbf{0}}}{N}\right) \\
=& \gamma\left(\sum_{\mathbf{r}}\sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2} m_{\mathbf{r}_1}\xi^{\mathbf{r}_1} m_{\mathbf{r}_2}\xi^{-\mathbf{r}_2}\dot{Q}_{\mathbf{r}+\mathbf{r}_2-\mathbf{r}_1}\xi^{\mathbf{r}+\mathbf{r}_2-\mathbf{r}_1}\right. \\
&\left. + \sum_{\mathbf{r}_2}\sum_{\mathbf{r}_1} m_{\mathbf{r}_1} m_{\mathbf{r}_2}\xi^{\mathbf{r}_2-\mathbf{r}_1}\frac{1-\dot{Q}_{\mathbf{0}}}{N}\right) \\
=& \gamma\left(\psi(\mathbf{z})\psi(-\mathbf{z})\mathcal{Q}(\mathbf{z}) + \psi(\mathbf{z})\psi(-\mathbf{z})\frac{1-\dot{Q}_{\mathbf{0}}}{N}\right) \\
=& \gamma\psi(\mathbf{z})\psi(-\mathbf{z})\left(\mathcal{Q}(\mathbf{z}) + \frac{1-\dot{Q}_{\mathbf{0}}}{N}\right)
\end{aligned} \tag{3.58}$$

(Malécot 1975, equation 14). Hence, at equilibrium,

$$\frac{\mathcal{Q}(\mathbf{z})}{1-\dot{Q}_{\mathbf{0}}} = \frac{1}{N}\frac{\gamma\psi(\mathbf{z})\psi(-\mathbf{z})}{1-\gamma\psi(\mathbf{z})\psi(-\mathbf{z})}. \tag{3.59}$$

Write $l_{\mathbf{k}}$ the eigenvalue $\psi[\boldsymbol{\theta}(\mathbf{k})]\psi[-\boldsymbol{\theta}(\mathbf{k})]$ of the migration matrix for pairs of genes. The above expression implies

$$\frac{\mathcal{Q}[\boldsymbol{\theta}(\mathbf{k})]}{1-\dot{Q}_{\mathbf{0}}} = \frac{1}{N}\frac{\gamma l_{\mathbf{k}}}{1-\gamma l_{\mathbf{k}}}. \tag{3.60}$$

The equilibrium $\dot{Q}_{\mathbf{r}}$ is then obtained by the inverse transform [equation (3.50)]:

$$\dot{Q}_{\mathbf{r}} = \frac{1-\dot{Q}_{\mathbf{0}}}{N} L_{\mathbf{r}} \tag{3.61}$$

where

$$L_{\mathbf{r}} \equiv \mathcal{L}_{\mathbf{r}}\left(\frac{\gamma\psi(\mathbf{z})\psi(-\mathbf{z})}{1-\gamma\psi(\mathbf{z})\psi(-\mathbf{z})}\right) = \frac{1}{n_x n_y}\sum_{\mathbf{k}=\mathbf{0}}^{(n_x-1,n_y-1)}\frac{\gamma l_{\mathbf{k}}}{1-\gamma l_{\mathbf{k}}}\bar{\mathbf{e}}_{\mathbf{k}}. \tag{3.62}$$

This is the generalization, for any dispersal distribution in one or two dimensions, of equation (3.29) (where the $\mathbf{e}_k$'s were used as a basis rather than the $\bar{\mathbf{e}}_{\mathbf{k}}$'s).

INFINITE LATTICES

In one dimension, when $n_x \to \infty$, equation (3.50) converges to

$$\mathcal{L}_j(\mathcal{F}) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{F}(x)e^{-\imath jx}\,dx \tag{3.63}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{F}(x)e^{-\imath jx}\,dx. \tag{3.64}$$

If $\mathcal{F}(z)$ is an even function (i.e., for symmetric dispersal distributions), this is also

$$\mathcal{L}_j(\mathcal{F}) = \frac{1}{2\pi}\int_0^{\pi} \mathcal{F}(x)(e^{-\imath jx} + e^{\imath jx})\,dx = \frac{1}{\pi}\int_0^{\pi} \mathcal{F}(x)\cos(jx)\,dx. \tag{3.65}$$

In two dimensions, when $n_x \to \infty, n_y \to \infty$, equation (3.54) converges to

$$\mathcal{L}_{\mathbf{r}}(\mathcal{F}) = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} \mathcal{F}(\mathbf{z})e^{-\imath\mathbf{z}\cdot\mathbf{r}}\,dx\,dy, \tag{3.66}$$

which, for even $\mathcal{F}$, is

$$\mathcal{L}_{\mathbf{r}}(\mathcal{F}) = \frac{1}{\pi^2}\int_0^{\pi}\int_0^{\pi} \mathcal{F}(\mathbf{z})\cos(r_x x)\cos(r_y y)\,dx\,dy. \tag{3.67}$$

APPENDIX 2: MISCELLANEOUS RESULTS

We prove several results of technical interest.

Diversity in a Deme

The first result relates diversity in a deme to the total size of the population, as $1 - \dot{Q}_{\mathbf{0}} = 2N_{\mathrm{T}}\mu + O(\mu^2)$ (e.g., Nagylaki 1983, equation 46; Slatkin 1987; Strobeck 1987). We later use this result under the form

$$\lim_{\mu\to 0}\frac{1-\dot{Q}_{\mathbf{0}}}{1-\gamma} = N_{\mathrm{T}}. \tag{3.68}$$

It has been emphasized as an "invariant" property of structured populations, showing that gene diversity is a function of the product of total size and of mutation rate. However, this is only a low mutation approximation, which may be poor considering realistic mutation rates or considering that total population sizes of most species fluctuate widely over time.

A proof is as follows. Let $\mathbf{0}$ and $\mathbf{1}$ be the column zero and unit vectors.[8] Differentiate the recursion (3.4) for identities at equilibrium with respect to μ in $\mu = 0$. Using $\dot{\mathbf{Q}}|_{\mu=0} = \mathbf{1}$, $\mathbf{A1} = \mathbf{1}$ (as for any backward probability transition matrix), and $\mathbf{c}|_{\mu=0} = \mathbf{0}$, we obtain

$$\mathbf{0} = -2\mathbf{1} + (\mathbf{A} - \mathbf{I})\frac{d\dot{\mathbf{Q}}}{d\mu} + \mathbf{A}\begin{pmatrix} -\frac{d\dot{Q}_0}{d\mu}/N \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{3.69}$$

$\mathbf{1}^\top$ is a left eigenvector of $\mathbf{A}$, such that $\mathbf{1}^\top\mathbf{A} = \mathbf{1}^\top$. Premultiplying equation (3.69) by $\mathbf{1}^\top$, one then obtains $2n_\mathrm{d} + d\dot{Q}_0/Nd\mu = 0$, which yields the required result.

Average Diversity in a Population

The average identity between genes in the population, $\bar{\dot{Q}} \equiv \sum_{\mathbf{r}} \dot{Q}_{\mathbf{r}}/n_\mathrm{d}$, where n_d is the total number of demes, has a closely related expression. From equation (3.49), $\mathbf{1}^\top \cdot \dot{\mathbf{Q}} = \gamma\mathbf{1}^\top\mathbf{A}\left(\dot{\mathbf{Q}} + \mathbf{c}\right)$; that is, $n_\mathrm{d}\bar{\dot{Q}} = \gamma n_\mathrm{d}\bar{\dot{Q}} + \gamma(1 - \dot{Q}_\mathbf{0})/N$ or

$$\frac{\bar{\dot{Q}}}{1 - \dot{Q}_\mathbf{0}} = \frac{1}{N_\mathrm{T}}\frac{\gamma}{1-\gamma}. \tag{3.70}$$

This result has not found much use per se, but further dividing equation (3.49) by $1-\bar{\dot{Q}}$ yields a way to compute the mutation effective size, as defined in chapter 9. A slightly different method will be used there. Equation (3.70) was first noticed by Maruyama (1970a) and Crow and Maruyama (1971), although their proof was not very accurate (Nagylaki 1982). As Nagylaki noticed, for arbitrary migration matrices, more general results are obtained by premultiplying the recursion for identity

[8] These notations will be used throughout the book: see p. 215.

by the leading left eigenvector of $\mathbf{A}$, which in general will not be $\mathbf{1}^\top$. This technique will be used repeatedly in chapter 9.

Differentiation under Low Dispersal

The third result provides an approximation for differentiation independent of the dispersal distribution. The proof is relatively involved. We consider an average differentiation between demes, weighted according to the dispersal rates at different distances:

$$\sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}} \frac{\dot{Q}_{\mathbf{0}} - \dot{Q}_{\mathbf{r}}}{1-\dot{Q}_{\mathbf{0}}} = \sum_{\mathbf{r}} m_{\mathbf{r}} \frac{L_{\mathbf{0}} - L_{\mathbf{r}}}{N}. \tag{3.71}$$

We express this sum of inverse transforms as a sum on $\mathbf{z} = (q_x, q_y)$, as in equation (3.54). For each value of (q_x, q_y), there appears a factor

$$\sum_{\mathbf{r}\neq\mathbf{0}} m_{\mathbf{r}} \left[1 - \cos\left(2\pi\frac{q_x}{n_x} r_x\right) \cos\left(2\pi\frac{q_y}{n_y} r_y\right)\right] = 1 - \psi\left(2\pi\frac{q_x}{n_x}, 2\pi\frac{q_y}{n_y}\right) \tag{3.72}$$

and therefore,

$$\begin{aligned}
\lim_{\mu\to 0} \sum_{\mathbf{r}} m_{\mathbf{r}} \frac{\dot{Q}_{\mathbf{0}} - \dot{Q}_{\mathbf{r}}}{1-\dot{Q}_{\mathbf{0}}} &= \frac{1}{N_{\mathrm{T}}} \sum_{\mathbf{r}} m_{\mathbf{r}} \sum_{\mathbf{z}\neq\mathbf{0}} (1 - e^{-\imath\mathbf{z}\cdot\mathbf{r}}) \frac{\psi^2(\mathbf{z})}{1-\psi^2(\mathbf{z})} \\
&= \frac{1}{N_{\mathrm{T}}} \sum_{\mathbf{z}\neq\mathbf{0}} [1 - \psi(\mathbf{z})] \frac{\psi^2(\mathbf{z})}{1-\psi^2(\mathbf{z})} \\
&= -\frac{1}{2N_{\mathrm{T}}} + \frac{1}{N} \mathcal{L}_{\mathbf{0}}\left(\frac{\psi^2}{1+\psi}\right)
\end{aligned} \tag{3.73}$$

where $\mathcal{L}_{\mathbf{0}}$ is the inverse transform (3.54) at zero distance. We can deduce a low migration approximation to this expression. To that aim we write $m_{\mathbf{i}}$ ($\mathbf{i} \neq \mathbf{0}$) as $m b_{\mathbf{i}}$, and let the total backward dispersal rate $m \to 0$ for a fixed distribution of dispersal distance (i.e., $b_{\mathbf{j}}/b_{\mathbf{i}}$ constant for all $\mathbf{i}, \mathbf{j} \neq \mathbf{0}$). Then $\psi(\mathbf{z}) = 1 - m(1 - \sum_{\mathbf{i}} b_{\mathbf{i}} e^{\imath\mathbf{i}\cdot\mathbf{z}})$; hence,

$\mathcal{L}_{\mathbf{0}}(\psi) = 1 - m$, and

$$\begin{aligned}\frac{\psi^2}{1+\psi} =& \frac{1}{2} - (1-\psi)\frac{1-2\psi}{2(1+\psi)} = \frac{1}{2} - (1-\psi)\left[\frac{1}{4} + O(m)\right] \\ =& \frac{1}{2} - \frac{1-\psi}{4} + O(m^2).\end{aligned} \tag{3.74}$$

The $\mathcal{L}_{\mathbf{0}}$ transform of this expression is $1/2 - m/4 + O(m^2)$, since $\mathcal{L}_{\mathbf{0}}(1-\psi) = m$. Finally,

$$\sum_{\mathbf{r}} m_{\mathbf{r}} \frac{\dot{Q}_{\mathbf{0}} - \dot{Q}_{\mathbf{r}}}{1 - \dot{Q}_{\mathbf{0}}} \approx \frac{1}{N}\left(\frac{1}{2} - \frac{m}{4}\right). \tag{3.75}$$

This approximation holds no matter what the shape of the dispersal distribution is, provided the total dispersal rate m is small.

> I look at [my isolation papers] at the moment with a lack of enthusiasm closely approaching nausea.
> Wright, in Provine (1986, p. 379)

CHAPTER FOUR

Interpretations of Inbreeding and Relatedness Coefficients in Subdivided Populations

In the previous chapter, specific properties of parameters such as F_{ST} were noticed, particularly their weak dependence on mutation rate, at least at a local spatial scale. This property is important, as it will provide a connection with measures of genetic structure that do not depend on mutation and can therefore be involved in measures of the effect of selection on a trait.

To better understand such properties, it is useful to interpret the measures of population structure in terms of the coalescence and migration events that affect the genealogy of different individuals. In this chapter, we review such interpretations. A common interpretation of F_{ST}, or more generally of relatedness parameters, is as a probability of recent coalescence, where "recent" needs to be better defined. A more general interpretation, however, is that inbreeding and relatedness coefficients measure differences in the probability of recent coalescence of different pairs of genes. These ideas will be detailed based on "migration matrix" models of population structure. We describe these models first.

Two different classes of phenomena are described as "inbreeding" or "consanguinity" in common usage. One is the higher identity of some pairs of genes relative to some other ones in biological populations, as measured by, for example, F_{IS} or F_{ST}. The other is the idea of similarity of genes within populations relative to completely independent populations. These two classes of phenomena may be formally described using similar equations, but they have different biological

consequences, and lumping them together as "inbreeding" has been a source of continued confusion in the mind of scholars. The genealogical interpretation of F_{ST} discussed in this chapter is generally suitable for the first class of phenomena and also distinguishes F_{ST} from various measures of "genetic distance" that are not considered in this book. Similarity relative to independent populations is more conveniently understood in terms of the factors that affect the genetic diversity of a total population, which will be discussed at length in chapters 9 and 10.

PROBABILITIES OF COALESCENCE IN MIGRATION MATRIX MODELS

Migration Matrix Models: Formulation

In the previous chapter, we have seen that in the island and isolation by distance models, identity obeys a recursion of the form (3.4):

$$\dot{\mathbf{Q}}' = \gamma \mathbf{A}(\dot{\mathbf{Q}} + \mathbf{c}) \tag{4.1}$$

for suitable $\mathbf{A}$ and $\mathbf{c}$. The $\mathbf{A}$ matrix described the movements of pairs of genes and the vector $\mathbf{c}$ described the increase in identity by descent due to coalescence events. An equation of this form can also describe a population structure between any finite number of demes, provided deme sizes are constant (see chapter 9). We here assume that probabilities of identity obey a recursion of the more general form

$$\dot{\mathbf{Q}}' = \gamma(\mathbf{A}\dot{\mathbf{Q}} + \tilde{\mathbf{A}}\mathbf{c}) \tag{4.2}$$

where $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are two matrices[1] and $\mathbf{c}$ is a vector expressing the gain in genetic identity ensuing from coalescence events. Typically $\mathbf{c}$ contains elements c_i either null or of the form $(1 - \dot{Q}_i)/N_i$. It is necessary to distinguish between the two matrices $\mathbf{A}$ and $\tilde{\mathbf{A}}$ when there is some class of genes that cannot coalesce (e.g., in an age-structured population, genes in two old individuals cannot coalesce in a single young individual when they were younger; see p. 163 for more on this).

In equation (4.2), the $\dot{Q}$'s appear in the two terms $\mathbf{A}\dot{\mathbf{Q}}$ and $\tilde{\mathbf{A}}\mathbf{c}$. All terms involving the $\dot{Q}$'s can be gathered in a single expression $\mathbf{G}\dot{\mathbf{Q}}$,

[1]It is assumed that the Markov chain induced by $\mathbf{A}$ is aperiodic and has a single irreducible closed set (see Feller 1968, p. 391). Biologically, two distinct irreducible sets are two different species, so this assumption is generally inconsequential.

where $\mathbf{G}$ is a matrix whose elements g_{ij} are a_{ij} plus the factor of $\dot{Q}_j$ in the ith elements of $\tilde{\mathbf{A}}\mathbf{c}$. For example, in the finite island model,

$$\mathbf{G} = \begin{pmatrix} a_{11}(1-1/N) & a_{12} \\ a_{21}(1-1/N) & a_{22} \end{pmatrix}, \tag{4.3}$$

with the a_{ij}'s given on p. 26. Then equation (4.2) takes the form

$$\dot{\mathbf{Q}}' = \gamma(\mathbf{G}\dot{\mathbf{Q}} + \tilde{\mathbf{A}}\boldsymbol{\delta}) \tag{4.4}$$

where $\boldsymbol{\delta}$ is a vector with elements $\delta_j = 1/N_j$ if c_j is of the form $(1 - \dot{Q}_j)/N_j$, and $\delta_j = 0$ otherwise. In the island model, $\boldsymbol{\delta} = (1/N, 0)^\top$.

The term $\tilde{\mathbf{A}}\boldsymbol{\delta}$ can also be written $(\mathbf{I} - \mathbf{G})\mathbf{1}$, where $\mathbf{1}$ is the column vector $(1, \ldots, 1)^\top$. Then equation (4.4) can be written

$$\dot{\mathbf{Q}}' = \gamma\left[\mathbf{G}\dot{\mathbf{Q}} + (\mathbf{I} - \mathbf{G})\mathbf{1}\right] \tag{4.5}$$

or equivalently

$$\gamma\mathbf{1} - \dot{\mathbf{Q}}' = \gamma\mathbf{G}\left(\mathbf{1} - \dot{\mathbf{Q}}\right). \tag{4.6}$$

This expression shows that in the absence of mutation ($\gamma = 1$), the $\mathbf{G}$ matrix describes the decrease of gene diversities (or "heterozygosities") $1 - \dot{Q}_i$ (Hill 1972). In particular, cumulative probabilities of coalescence over time obey equation (4.6), with $\gamma = 1$. Let $\boldsymbol{\Sigma}(t) \equiv (\sum_{k=0}^{t} C_{i,k})$ be the vector of cumulative probabilities of coalescence; then $\boldsymbol{\Sigma}' \equiv \boldsymbol{\Sigma}(t+1)$ is given by

$$\mathbf{1} - \boldsymbol{\Sigma}' = \mathbf{G}\left(\mathbf{1} - \boldsymbol{\Sigma}\right). \tag{4.7}$$

Probabilities of Coalescence

Identity by descent of a pair i of genes can be written as

$$\dot{Q}_i = \sum_{t=1}^{\infty} C_{i,t}\gamma^t \tag{4.8}$$

where $C_{i,t}$ is the probability that two genes coalesce at time t in the past [equation (2.8)]. The i index indicates whether the pair of genes considered are, for example, two homologous genes within a diploid individual, or two genes in different individuals. Equation (4.5) implies that

$$\dot{\mathbf{Q}} = (\mathbf{I} - \gamma\mathbf{G})^{-1}\gamma(\mathbf{I} - \mathbf{G})\mathbf{1}. \tag{4.9}$$

By the Perron-Frobenius theorem for irreducible nonnegative matrices (Horn and Johnson 1985), the eigenvalues λ_j of $\mathbf{G}$ obey $1 > |\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_k|$, where $|\lambda_j|$ is the absolute value (or modulus) of λ_j. We assume that $\mathbf{G}$ is diagonalizable, which is usually the case in practice (see Appendix A for further discussion). Let $\mathbf{e}_j \equiv (e_{j1}, \ldots, e_{jk})^\top$ be a right λ_j eigenvector of $\mathbf{G}$. We can express the vector $(\mathbf{I} - \mathbf{G})\mathbf{1}$ as $\sum_j x_j \mathbf{e}_j$, and using equation (A.10), equation (4.9) can be written

$$\dot{\mathbf{Q}} = \sum_j \frac{\gamma x_j \mathbf{e}_j}{1 - \gamma\lambda_j} = \sum_{t=1}^{\infty} \sum_{j=1}^{k} \gamma^t \lambda_j^{t-1} x_j \mathbf{e}_j. \tag{4.10}$$

Comparison with equation (4.8) shows that probabilities of coalescence are of the form

$$C_{i,t} = \sum_j \lambda_j^{t-1} x_j e_{ji}. \tag{4.11}$$

We also note that all the e_{1j} are nonzero (again a consequence of the Perron-Frobenius theorem), and that for any j, k,

$$\lim_{t\to\infty} C_{j,t}/C_{k,t} = e_{1j}/e_{1k}. \tag{4.12}$$

Hence, the ith element of the first eigenvector is proportional to the ancestral probability that two genes coalesce in the ith demographic class.

With demographic fluctuations (deterministic or random), a more complicated formulation would be required, as the $\mathbf{G}$ matrix would be different from one generation to the next (see chapter 10). However, the main conclusions would be unchanged. The theory of matrix products, particularly the weak ergodic and weak stochastic ergodic theorems (Cohen 1979; Caswell 2001), can be used to show that the ancestral position of gene lineages, and their probability of coalescence in the different classes, becomes independent of their present position. An analog of the larger eigenvalue λ_1, the dominant Lyapunov exponent, can be defined.

INTERPRETATIONS OF F_{ST}

Coalescence before Dispersal

A widely used and simple interpretation of F_{ST} is the probability that two genes, sampled within a deme, coalesce before one ancestral gene

lineage "leaves" the deme when followed backwards in time. This interpretation is intimately tied to its computation as

$$F = (1-m)^2[(1-1/N)F + 1/N] \tag{4.13}$$

[see equation (3.16)]. According to this computation, the only events that contribute to increase the F value are the events of coalescence of genes within demes: whenever two gene lineages come from different demes, the genes are considered unrelated.

Separation of Time Scales

A technical assumption underlies the interpretation of F_{ST} as the probability of coalescence before dispersal. Genes in different demes are independent if mutations occur faster than the coalescence of genes from different demes. For low mutation ($\mu \to 0$), this is obtained by assuming that the number of demes $n_{\mathrm{d}} \to \infty$ and that $n_{\mathrm{d}}\mu \to \infty$. This relative scaling of mutation and coalescence rates is possible when the coalescence of genes occurs at drastically different rates within and between demes: in the island model, genes within demes coalesce at a rate proportional to $1/N$, while genes between demes coalesce at a rate proportional to $1/(Nn_{\mathrm{d}})$. These two rates are quite different when the number of demes is large, a case that can be described as an interdeme separation of time scales. Likewise, an interfamily separation of time scales occurs when the rates of coalescence of genes within and among families scale differently. It justifies the computation of relatednesses from pedigrees in large finite populations. Arguments based on such separations of time scales also provide means to compute the probability of samples of more than two genes by coalescent methods (see chapter 12).

In models with a separation of time scales, the inbreeding or relatedness coefficient F of two genes relative to the total population can be interpreted as a probability of "recent" coalescence. In particular, the probability that two genes are both of a given allelic type k may be written

$$Fp + (1-F)p^2, \tag{4.14}$$

where p is the allele frequency in the total population. Consider, for example, the infinite island model, where this becomes

$$F_{\mathrm{ST}}p + (1-F_{\mathrm{ST}})p^2 \tag{4.15}$$

for genes sampled in different adults within a deme. To obtain this expression, one assumes first that the allele frequency in "recent" generations is identical to the present allele frequency. That is, the effects of drift of the allele frequency in the total population, of mutation, and of selection are neglected on the short time scale. At this scale, only drift within a deme is taken into account. A frequency F_{ST} of pairs of genes have coalesced "recently," and they are both allele k with probability p, hence the $F_{\mathrm{ST}}p$ term. The remaining frequency $1 - F_{\mathrm{ST}}$ of pairs of genes that have not coalesced before dispersal are pairs of "independent" genes, which are both of type k with probability p^2. This accounts for the term $(1 - F_{\mathrm{ST}})p^2$.

An Ancestral Reference Population?

The interpretation of inbreeding coefficients in terms of equation (4.14), where F is independent of the allele frequency and p is the allele frequency in an "ancestral reference population," has an intuitive appeal that has guided early works on the definition of genetic identity (e.g., Cotterman 1940). It has been used for defining relatedness in general, but does not seem to hold without an interfamily or interdeme separation of time scale (Rousset 2002). Arguments based on a separation of time scales do not easily generalize to other models, such as isolation by distance. There are analogous problems with a common definition of the variance effective size, which effectively makes the assumption that two time scales are separated (see chapter 9).

Hence, in the remainder of this book, we will not define inbreeding coefficients and relatedness as coalescence before dispersal of ancestral lineages, or as "recent" coalescence, but from ratios of differences of probabilities of identity. The two types of definitions are equivalent under a separation of time scales. A more general probabilistic interpretation of the definition as a difference of probabilities of identity will now be given.

Differences between Distributions of Coalescence Times

Slatkin (1991) noticed that F-like parameters are related to the expected coalescence times of pairs of genes, as follows. In a finite pop-

ulation, equation (4.8) implies that

$$\dot{Q}_i = \sum_{t=1}^{\infty} C_{i,t}[1 - 2\mu t + O(\mu^2)] = 1 - 2\mu T_i + O(\mu^2), \qquad (4.16)$$

where $T_i \equiv \sum_{t=1}^{\infty} tC_{i,t}$ is the expected coalescence time of a pair of genes of type i. Actually, for other mutation models reviewed in Appendix A, identity in state obeys the same relationship:

$$Q_i = 1 - 2\mu T_i + O(\mu^2). \qquad (4.17)$$

It follows that for any two pairs of genes, indexed w and b, the ratio $F \equiv (Q_{\rm w} - Q_{\rm b})/(1 - Q_{\rm b})$ satisfies

$$\lim_{\mu\to 0} \frac{Q_{\rm w} - Q_{\rm b}}{1 - Q_{\rm b}} = \lim_{\mu\to 0} \frac{1 - 2\mu T_{\rm w} - (1 - 2\mu T_{\rm b})}{1 - (1 - 2\mu T_{\rm b})} = \frac{T_{\rm b} - T_{\rm w}}{T_{\rm b}}. \qquad (4.18)$$

Thus, F is related to the distribution of coalescence events through the expected coalescence times. But F mainly depends on differences between these distributions in recent generations (Rousset 1996, 2002). F approximates how much higher the probability of recent coalescence is for some pair of genes relative to another pair of genes, as shown in figures 4.1 and 4.2. This difference can be described as follows. We can split the area covered by the probability distribution of coalescence times of more related genes (the area delimited by $C_{{\rm w},t}$) into two parts. We take the area below the $C_{{\rm b},t}$ curve (the distribution of coalescence times of less related genes) and consider this surface reduced by the value of the ratio $C_{{\rm w},t}/C_{{\rm b},t}$ for large t. For large t, this reduced area coincides with the area delimited by $C_{{\rm w},t}$; it is shaded in light grey in figure 4.1. The other part is the rest of the area delimited by $C_{{\rm w},t}$; its height is denoted $g(t)$ and it is shaded in darker grey in figure 4.1.

The same graphical comparisons can be made for other population structures, as shown in figure (4.2). The idea of a separation of time scales is recovered in the fact that $C_{{\rm b},t} \to 0$ for all t as $N \to \infty$. That is, the distribution of coalescence times flattens down on the x axis. This is shown in figure (4.2a) in an example where the exact computation is simple, and this more generally underlies the computation of relatedness based on pedigree relationships in a panmictic population. Likewise, a separation of time scales is obtained in the island model, where the probabilities of coalescence of genes in different demes $C_{{\rm b},t} \to 0$ for all t as $n_{\rm d} \to \infty$.

F approximates the area $\sum_{t=1}^{\infty} g(t)$, that is, the increased probability of coalescence in the recent past. This can be more formally shown

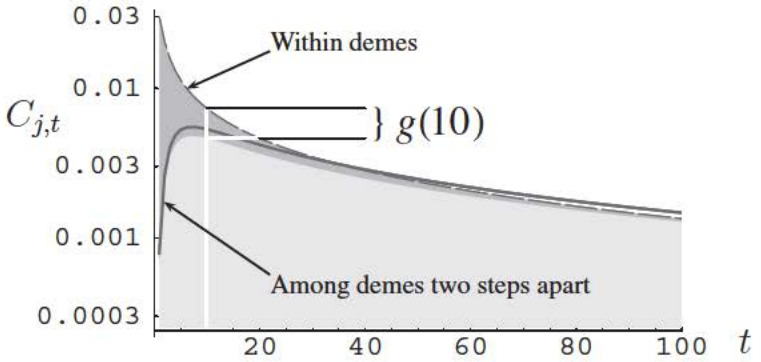

FIGURE 4.1. Differences between distributions of coalescence times. This figure compares distributions of coalescence times of different pairs of genes used to define relatedness. The curves describe two distributions of coalescence times for a one-dimensional stepping stone model (with 100 demes of $N = 10$ haploid individuals and dispersal rate $m = 1/4$). The different areas defined in the text are shown, and $g(t)$ is shown for $t = 10$. Note the logarithmic y-scale.

by the following argument. Let

$$\omega \equiv 1 - \lim_{t\to\infty} C_{\mathrm{w},t}/C_{\mathrm{b},t}. \tag{4.19}$$

Equation (4.12) shows that this limit is well defined for migration matrix models. Then our verbal definition of $g(t)$ becomes

$$g(t) \equiv C_{\mathrm{w},t} - (1-\omega)C_{\mathrm{b},t}. \tag{4.20}$$

As $\sum_{t=1}^{\infty} C_{\mathrm{w},t} = C_{\mathrm{b},t} = 1$ (these are both probability distributions), equation (4.20) implies that $\sum_{t=1}^{\infty} g(t) = \omega$. F differs from ω as a result of mutation, but it slightly differs in the low mutation limit too: equation (4.18) implies

$$\lim_{\mu\to 0} F = \omega - \frac{\sum_1^{\infty} tg(t)}{T_{\mathrm{b}}}. \tag{4.21}$$

For low mutation, the difference between F and ω appears to be of the order of the inverse of the total size N_{T} of the population, and will usually be very small.

In conclusion, $F \approx \sum_{t=1}^{\infty} g(t) = \omega$. That is, F approximates the excess probability of recent coalescence of one pair of genes relative to

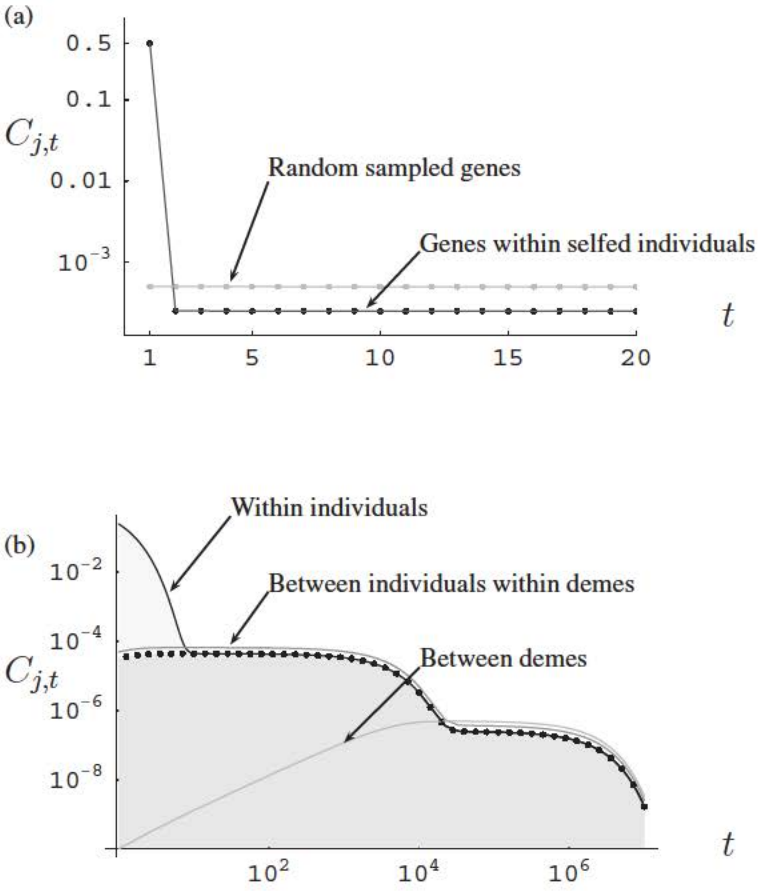

FIGURE 4.2. Probabilities of coalescence in different scenarios as a function of time t. The probabilities $C_{j,t}$ are shown (a) for randomly sampled genes and for genes within selfed individuals in a panmictic, diploid, randomly mating hermaphroditic population of $N = 1{,}000$ individuals, where each offspring may be produced by selfing with probability $1/N$ independently of each other; (b) in an island model with selfing (see Rousset 1996 for details), with 100 demes of $2N = 20{,}000$ genes, a dispersal rate of $m = 1/N$, and a selfing rate of 0.5. $j = 0$: two genes within the same individual; $j = 1$: two genes in different individuals within a deme; $j = 2$: two genes in different demes. Distributions of coalescence times are shown as plain lines. The shaded surface below the dotted line is constructed, as in figure 4.1, as a reduced copy of the area covered by the distribution of coalescence times of genes between individuals. The shaded area above the dotted line is the "initial area" for F_{IS}.

the other. How "recent" is recent depends on the decrease of $g(t)$. If $g(t)$ decreases rapidly enough with t, then the sum $\sum_{t=1}^{\tau} g(t)$ over a short time span τ will be close to ω.

PROPERTIES OF INBREEDING COEFFICIENTS

Arguments based on the comparison of distributions of coalescence times allows us to understand several properties of inbreeding coefficients, as well as the relationship between different definitions of them.

Sensitivity to Mutation and to Past Demographic Events

The decrease of $g(t)$ with t determines the sensitiveness of F values to mutation rates. Given $\sum_{t=1}^{\tau} g(t) \approx \omega$ for some recent time τ, older mutations will have little effect on F. Hence, mutations have notable effect on F only to the extent that mutations are likely to occur within these τ generations.

The decrease of $g(t)$ with t also determines the sensitivity of F values to past fluctuations of population structure. The rate of approach of F to its stationary equilibrium value would be essentially that of $\sum_t g(t)$ to its limit for large t if the identity Q_{b} were not changing with time. More exact results are complicated (e.g., Nagylaki 1983), but the infinite island, infinite allele model is useful for illustration (Takahata 1983; Crow and Aoki 1984; Birky et al. 1989; Whitlock and McCauley 1999). In a population initially at equilibrium, $F = \dot{Q}_{\mathrm{w}}$. The value of F is determined by initial values of N and m. If one or both of them change to new values N^* and m^*, F will converge to a new value $\dot{Q}^*_{\mathrm{w}}$. With sufficiently weak mutation, a little algebra shows that

$$\dot{Q}_{\mathrm{w}}(t+1) - \dot{Q}^*_{\mathrm{w}} = (1-m^*)^2(1-1/N^*)[\dot{Q}_{\mathrm{w}}(t) - \dot{Q}^*_{\mathrm{w}}]. \quad (4.22)$$

Hence, the speed of convergence of F to $\dot{Q}^*_{\mathrm{w}}$ is determined roughly by the larger of $(1-m^*)^2$ and $1-1/N^*$; it is lower for low dispersal rates and high subpopulation sizes. For example, with $m^* = 0.5$ and $N^* = 100$, four generations are enough to go more than 99% of the way to the ultimate value of F. Thus, the equilibrium theory of F-"statistics" is often more relevant for large dispersal rate and low deme size (e.g., Slatkin 1994; Whitlock and McCauley 1999; Rousset 2001).

In this example, F and $\dot{Q}_{\mathrm{w}}$ are equivalent because the identity of genes in different demes is null in all generations. But more generally, the convergence of F_{ST}-like parameters is faster than that of probabilities of identity (Crow and Aoki 1984). This difference is apparent in the comparison of distributions of coalescence times, as the convergence of each identity $\dot{Q}_{\mathrm{w}}$ and $\dot{Q}_{\mathrm{b}}$ will depend on each distribution of coalescence times, while the convergence of F may be seen to depend mainly on recent differences between these two distributions.

Under isolation by distance, the same general trends are observed as in the island model: the approach to equilibrium is faster for low deme size and high dispersal rates. In addition, it is faster for small distances between the populations compared (Slatkin 1993). Conversely, large-scale population structure may be severely affected by recent demographic events such as range expansions (e.g., Austerlitz et al. 2000).

No Mutation

Similar results for the approach to equilibrium have been reached through different definitions of the parameters. If we ignore mutation, identity by descent increases from its value at some initial time t^* in the same way as the probability of coalescence before t^* (we now count time forwards rather than backwards as above). The cumulative probability of coalescence since an initial time t^* has indeed often been considered as a measure of relatedness or "identity by descent" (Cotterman 1940; Malécot 1948; Hill 1972; Chesser et al. 1993; Wang 1997b; Whitlock and Barton 1997). Then, an F_{ST}-like parameter can be defined as a ratio of differences of such "identities":

$$\frac{\sum_{t=1}^{t^*}(C_{\mathrm{w},t} - C_{\mathrm{b},t})}{1 - \sum_{t=1}^{t^*} C_{\mathrm{b},t}}. \tag{4.23}$$

Since $C_{\mathrm{w},t} - C_{\mathrm{b},t} = g(t) - \omega C_{\mathrm{b},t}$, the numerator is $\sum_{t=1}^{t^*}(g(t) - \omega C_{\mathrm{b},t})$, which approaches $\omega(1 - C_{\mathrm{b},t})$ as fast as $\sum_{t=1}^{t^*} g(t)$ approaches ω when t^* increases. Thus, the F_{ST}-like parameter is approximately

$$\frac{\omega(1 - \sum_{t=1}^{t^*} C_{\mathrm{b},t})}{1 - \sum_{t=1}^{t^*} C_{\mathrm{b},t}} = \omega \tag{4.24}$$

and it approaches this limit as fast as $\sum_{t=1}^{t^*} g(t)$ does.

Alternative Measures of Allelic Divergence

The above arguments can be extended to measures of population structure not based on allelic identity. For any measure d of divergence such that $d = 0$ if no mutation occurred, we may compare divergences d_{w} and d_{b} for genes within and among units, as usual. The ratio $[\mathrm{E}(d_{\mathrm{b}}) - \mathrm{E}(d_{\mathrm{w}})]/\mathrm{E}(d_{\mathrm{b}})$ will be approximately equal to $(T_{\mathrm{b}} - T_{\mathrm{w}})/T_{\mathrm{b}}$, provided mutations are unlikely within the "initial" time span. Thus, many different measures of population structure may be defined, with the same properties as F, with respect to relationship with coalescence time and to speed of convergence to stationary values. This holds even if the mutation rate is high, whenever expected divergence d given coalescence t of a pair of genes is a linear function of t. For example, we may consider the sequence divergence (number of nucleotide differences) between two gene copies. Under the simplest models of sequence evolution, the expected number of nucleotide differences is a linear function of their coalescence time (e.g., Hudson 1990; Nordborg 2003). For microsatellites, made of a number of repeats of a short DNA motif, the stepwise mutation model assumes that mutations add or subtract repeats independently of the allele size. Then the squared difference between the lengths of two alleles is a linear measure of their divergence time (Goldstein et al. 1995; Slatkin 1995). In these different cases, the expected divergence for a pair of genes is simply $\mathrm{E}(d) = T\epsilon$, where T is the expected coalescence time of the pair of genes and ϵ is the increase in divergence per time unit.[2]

More or less linear measures of divergence may be defined for various models of evolution of alleles (e.g., Feldman et al. 1997), but the linearity is of course dependent on the validity of such models. Thus, although a measure based on microsatellite allele size may sometimes provide better knowledge of genealogical structure (Balloux et al. 2000), it may also fail. In addition, a high variance of estimators of such measures has been reported, which makes them of little use in various contexts (e.g., Tsitrone et al. 2001; Balloux and Goudet 2002; Leblois et al. 2003).

It may also be assumed that mutation adds a constant amount of

[2]Recursions for expected coalescence times T_i are obtained from recursions such as $\mathbf{Q}' = \gamma(\mathbf{AQ} + \tilde{\mathbf{A}}\mathbf{c})$ by setting $\gamma = 1$, replacing each Q_i with $1 + T_i$ and each Q_i' with T_i' (e.g., Nagylaki 1998a).

additive genetic variance per generation. The random divergence in a quantitative trait is then a function of the coalescence time and of the mutational input in additive variance per generation (Lande 1992). The mathematical analogy has been pushed further by assuming stabilizing selection on the trait, with the approximation that selection removes a constant fraction of genetic variance per generation. Divergence in a quantitative trait can then be analyzed by analogy with identity by descent in neutral models (Nagylaki 1994), with stabilizing selection limiting differentiation in the same way that mutation limits neutral differentiation.

Based on Lande's assumptions, measures of population structure based on the variance of additive effects of genes within and among demes have been defined (e.g., Spitze 1993), and estimators based on phenotypic divergence have been used. The data do not fit this simple theory. Discrepancies are usually attributed to the effects of selection on phenotypic diversity in natural populations, although nonadditive gene effects on phenotype must also have an effect (Merilä and Crnokrak 2001).

CHAPTER FIVE

Evolutionary Dynamics

The analysis of selection through game theoretical approaches seeks the trait values that are most likely to be observed in a population evolving under a recurrent flow of mutations for a given trait. A set of constraints on possible mutations affecting a trait is defined, and the following questions are usually considered. One asks whether there is a trait value such that (1) given continuing mutation, selection will drive the population towards this trait value, starting from any other trait value or at least from a given range of trait values; (2) once it is reached, any mutant allele with a deviant strategy will be selected against. The latter condition corresponds to the definition of an "evolutionarily stable strategy" (Maynard Smith and Price 1973). The former condition (accessibility by evolution from a population expressing other trait values) is known as convergence stability. In this chapter, both of these concepts will be illustrated through a model of competition for resource exploitation in a panmictic population. Next we will consider a range of complications, with a particular focus on those resulting from population structure. This will lead us to consider the probability of fixation of a mutant as a measure of convergence stability.

FITNESS IN A PANMICTIC POPULATION

Example: Resource Competition

Resource competition models investigate the evolution of a population of individuals competing for the exploitation of a range of resources, subject to constraints on the strategy of each individual, that is, on the range of resources that it can exploit. These models show very diverse

dynamics, from stable coexistence of different strategies to cyclic replacement of strategies (e.g., Rummel and Roughgarden 1985; Brown and Vincent 1987; Christiansen 1991; Taper and Case 1992; Day 2001). As such, they provide an elementary illustration of major concepts of evolutionary game theory, and we will consider a simple version of these models.

We consider a panmictic population of N adults. Each individual has access to different types of resources, say prey of size z. The resource of type z has abundance $r(z)$. We assume here that it follows a normal distribution (fig. 5.1),

$$r(z) = e^{-C(z-z_{\mathrm{m}})^2}\sqrt{C/\pi} \tag{5.1}$$

for some constant $C > 0$, and some z_{m} such that resource z_{m} is the most abundant. The smaller C is, the more broadly distributed the resource over different types.

We assume that each individual cannot exploit all resources with equal efficiency. An individual i exploits resource z with an efficiency given by an efficiency function $e_i(z)$ such that individual i gets a fraction $e_i(z)/\sum_{k=1}^{N} e_k(z)$ of resource z, which is the ratio of its own efficiency at exploiting resource z to the sum of efficiencies at exploiting z among all individuals in the population.

Assume that individual i is best at exploiting resource z_i, and that the efficiency function is also normal-shaped (fig. 5.1),

$$e_i(z) = e(z, z_i) = e^{-\alpha(z-z_i)^2} \tag{5.2}$$

for some $\alpha > 0$, identical for all individuals. The smaller α is, the more diverse the resources that an individual may exploit. An individual strategy is therefore characterized by z_i and constrained by the range of resources it can efficiently exploit around z_i.

Finally we assume that the expected number of adult offspring (*fitness*) of an individual is proportional to its share of resource and is

$$w_i = \int_{-\infty}^{+\infty} r(z)\frac{Ne(z, z_i)}{\sum_{k=1}^{N} e(z, z_k)}\, dz. \tag{5.3}$$

Convergence Stability

Intuitively, one expects that individuals will try to exploit the most abundant resource z_{m}. That is, in a population of individuals with strategy $z_a \neq z_{\mathrm{m}}$, we expect that the fitness w_A of deviant individuals with

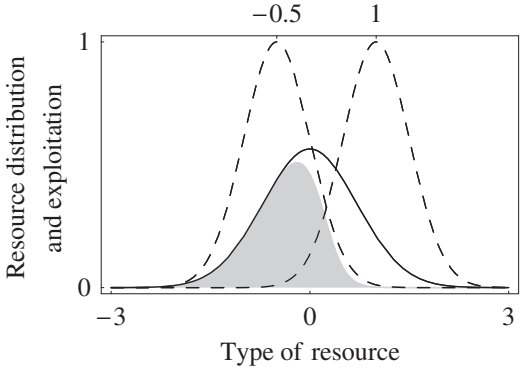

FIGURE 5.1. Shapes of resource distribution and exploitation efficiency functions. This figure shows the distribution of resources (plain line), with $C = 1$, and two different strategies (dashed lines), $z = -0.5$ and $z = 1$, both with $\alpha = 2$. The gray area is the share of resources exploited by the strategy $z = -0.5$ when both strategies are at frequency 0.5. See text for details.

some strategy z_A will be higher than the fitness w_a of individuals with strategy z_a if z_A is closer to z_{m} than z_a is. Thus, we expect $w_A > w_a$ provided $r(z_a) < r(z_A) < r(z_{\mathrm{m}})$. This intuitive expectation is shown on figure 5.2, where for a given strategy z_a, we plot the fitness of deviant strategies near z_a. The first-order derivatives, shown as plain lines in figure 5.2, decrease for increasing z_a. Writing $z_A = z_a + \delta$, we then have

$$\begin{cases} \partial_\delta w_A|_{\delta=0} > 0 & \text{if } z_a < z_{\mathrm{m}}, \\ \partial_\delta w_A|_{\delta=0} < 0 & \text{if } z_a > z_{\mathrm{m}}. \end{cases} \tag{5.4}$$

Under these conditions, we expect the population to evolve under a recurrent flow of mutations of small phenotypic effect until it reaches the point z_{m}, where there is no longer directional selection ($\partial_\delta w_A = 0$). A point where $\partial_\delta w_A = 0$ is called a candidate ESS. When conditions (5.4) are satisfied, z_{m} is said to be convergence stable (Eshel 1996) or attainable (Takada and Kigami 1991; Christiansen 1991). We can also express convergence to z_{m} as $\partial_\delta w_A|_{\delta=0, z_a=z_{\mathrm{m}}} = 0$ and $\partial_{z_a}(\partial_\delta w_A|_{\delta=0})|_{z_a=z_{\mathrm{m}}} < 0$, since these conditions imply the previous inequalities (5.4) near z_{m}. An equivalent mathematical condition may be given in terms of the two strategies z_a and z_A rather than of their difference δ (e.g., Lessard 1990; Eshel 1996; Geritz et al. 1998).

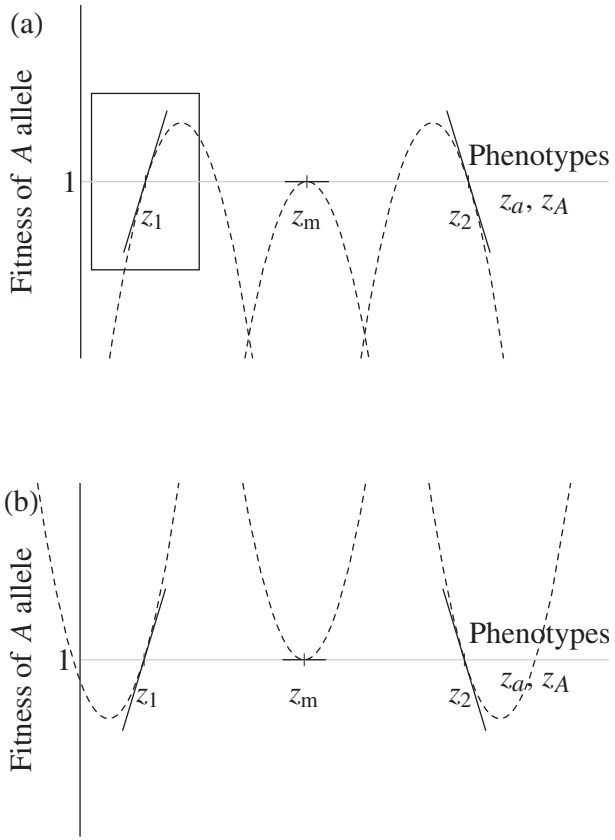

FIGURE 5.2. Fitness of mutants around a candidate ESS. This figure (after Abrams et al. 1993) represents the fitness of a mutant allele A for different combinations of mutant and "resident" trait values, z_A and z_a. In each of the subplots, such as the boxed one, the dashed curve is the fitness of the z_A phenotype in a population of competitors with a fixed z_a phenotype ($z_a = z_1$, z_m, or z_2), and the plain line represents the derivative of fitness $\partial_\delta w_A$ with respect to phenotypic difference δ. In both (a) and (b), the derivatives are positive below z_m and negative above it, which is the condition for convergence stability [equation (5.4)]. In (a), the fitness curves are curved downward [negative second derivative $\partial^2_{\delta,\delta} w_A$, equation (5.9)], which represents evolutionary stability, while in (b) they are curved upward (positive second derivative), which represents invasibility by rare mutants.

We can evaluate $\partial_\delta w_A$ in the competition example. Let n_a and n_A be the number of individuals with some allele a and phenotype z_a, and some allele A with phenotype z_A, respectively. Let $p \equiv n_A/(n_a+n_A)$ be the frequency of the deviant strategy. With the above efficiency and resource abundance functions e and r, the derivative of fitness $\partial_\delta w_A|_{\delta=0}$ is

$$\partial_\delta \int_{-\infty}^{+\infty} r(z) \frac{Ne(z,z_A)}{n_a e(z,z_a)+n_A e(z,z_A)}\, dz = 2\alpha(1-p)(z_{\rm m}-z_a). \tag{5.5}$$

In other words, for small $\delta = z_A - z_a$, the expected number of offspring of an individual with strategy z_A is $1+2\alpha(1-p)(z_{\rm m}-z_a)\delta + O(\delta^2)$ [the "1" follows from the fact that the average fitness $pw_A + (1-p)w_a = 1$ for any p]. Equivalently, the expected change in mutant allele frequency is of the form

$$p\delta\partial_\delta w_A|_{\delta=0} + O(\delta^2) = p(1-p)2\delta\alpha(z_{\rm m}-z_a) + O(\delta^2). \tag{5.6}$$

The derivative is zero at $z_a = z_{\rm m}$, whatever the allele frequency p, and individuals with strategy z_A slightly above z_a will have more offspring when $z_a < z_{\rm m}$ (since the derivative is positive in this case). Conversely, individuals with some strategy z_A slightly below z_a will have more offspring when $z_a > z_{\rm m}$. Hence, $z_{\rm m}$ is convergence stable, as expected.

Evolutionary Stability

Now suppose that the population has reached the point $z^* = z_{\rm m}$, where the first-order derivative vanishes. Will it stay there? Since all individuals maximally exploit the most abundant resource $z_{\rm m}$, it may pay for rare deviants to maximally exploit a less abundant resource $z_A \neq z_{\rm m}$, because there is less competition for this resource. The derivative $\partial_\delta w_A|_{\delta=0, z_a=z_{\rm m}} = 0$, and hence it cannot tell us whether the fitness of a deviant individual is higher or lower than that of $z_{\rm m}$ individuals. Its fitness may be lower (fig. 5.2a) or higher (fig. 5.2b) than that of $z_{\rm m}$ individuals. We see from these figures that these two cases may be distinguished by the curvature of the fitness of "rare" z_A individuals in the $z_{\rm m}$ population. Thus, we need to consider the second derivative $\partial^2_{\delta,\delta} w_A|_{z_a=z_A=z_{\rm m}}$. With the above functions e and a, this is

$$\partial^2_{\delta,\delta} w_A|_{z_a=z_A=z_{\rm m}} = -2\alpha(1-p)\frac{C+(2p-1)\alpha}{C}. \tag{5.7}$$

Hence, the fitness of individuals with strategy $z_{\mathrm{m}} + \delta$ in competition with strategy z_{m} is

$$1 - \alpha(1-p)\frac{C + (2p-1)\alpha}{C}\delta^2 + o(\delta^2) \tag{5.8}$$

where the term $o(\delta^2)$ is small relative to δ^2 (see p. 216 for this notation). Since $\alpha > 0$ and $C > 0$, this fitness is always < 1 near $p = 1$ and is > 1 near $p = 0$ if $\alpha - C > 0$. Then alleles are selected for when rare and against when common.

What does that mean? $\alpha - C < 0$ means that the distribution of resource types is narrower than what an individual may exploit. Thus, individuals with strategy z_{m} already substantially exploit the resource $z \neq z_{\mathrm{m}}$, and a deviant individual with $z_A \neq z_{\mathrm{m}}$ will not gain in terms of lowered competition: any mutant is disfavored. Conversely, $\alpha - C > 0$ means that the individuals with strategy z_{m} do not exploit the resource $z \neq z_{\mathrm{m}}$ substantially, and therefore a deviant individual with $z_A \neq z_{\mathrm{m}}$ will have high fitness when rare. However, it is always disfavored when common.

Thus, a negative second derivative,

$$\partial^2_{\delta,\delta} w_A|_{z_a = z_A = z_{\mathrm{m}}} < 0, \tag{5.9}$$

is interpreted as evidence that the population will stay at z^*. Alternatively, a positive second derivative suggests that a stable polymorphism will initially evolve. This condition ensures that a pair of sufficiently close strategies z_1 and z_2 such that $z_1 < z^* < z_2$ can establish themselves as a polymorphism stabilized by selection. Further, it ensures that for z_1, z_2 sufficiently close to the candidate ESS, a third strategy y will in turn invade if it is also close enough to z^* and $y < z_1$ or $z_2 < y$ (Christiansen 1991; Geritz et al. 1998). Then the intermediate strategy can be eliminated and the two remaining strategies become more and more distinct. This divergence leads the population to a state where its later evolution cannot be predicted from local stability conditions in z^* (Christiansen 1991; Eshel et al. 1997) and may be complex (Matessi and Di Pasquale 1996; Geritz et al. 1998; Kisdi 1999).

The divergence of two lineages is described as "branching"; hence, the point characterized by a positive second derivative is described as a "branching point." Note that in multilocus models, phenotypic selection may be disruptive at a branching point, with a local minimum of fitness for intermediate phenotypes, but branching may not result (see p. 78).

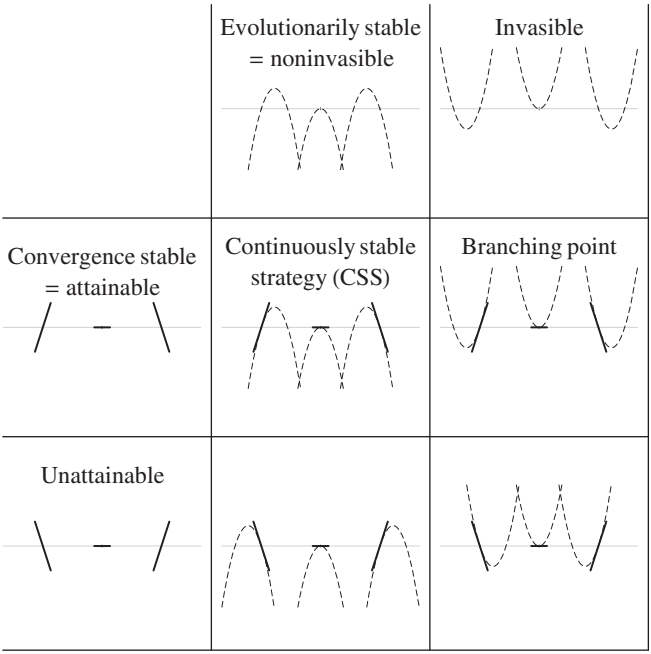

FIGURE 5.3. Classification of candidate ESSs. This figure distinguishes four types of candidate ESSs, whether conditions for convergence and evolutionary stability [equations (5.4) and (5.9)] are satisfied or not. The margins show the graphical translation of each stability condition, whether it is satisfied or not, and the cells show fitness plots as in figure 5.2.

The negative second derivative has been described as the condition for "evolutionary stability" (Maynard Smith and Price 1973; Maynard Smith 1982). This terminology has generated much discussion because this second-order condition does not fully capture concepts of dynamical stability. A strategy may be an "evolutionary stable" in the above sense, yet inaccessible from other states of the population (e.g. Eshel and Motro 1981; Eshel 1983; Nowak 1990), in which case it is an unlikely state for the population. Hence, the convergence stability condition also represents a stability condition from an "evolutionary" point of view (under recurrent mutations).

We summarize in figure 5.3 the different cases and indicate some alternative terminologies. A candidate ESS, where the first-order deriva-

tive vanishes, is sometimes called a singular point (Geritz et al. 1998). It is a branching point, or invasible, if the second derivative is positive; otherwise it is noninvasible, or monomorphic (Christiansen 1991; this is the original concept of "evolutionary stability"). It is attainable if it is convergence stable. Finally, if the stability has been determined only with respect to mutants of small effect, we say it is local. Thus, a strategy z that is both convergence stable and evolutionarily stable by the above mathematical criteria is locally monomorphic attainable. This is also known as a "continuously stable strategy" (CSS, Eshel 1996). Conversely, z_{m} is polymorphic attainable if it is convergence stable but not evolutionarily stable (Christiansen 1991). The concept of "unbeatable strategy" (Hamilton 1967) was not unambiguously described and therefore will not be used in this book. It is not clear whether it was conceived as a local or global stability condition. It differs from the concept of the evolutionarily stable strategy in that an unbeatable strategy could not be "invaded" by any alternative strategy, whatever its initial frequency (Hamilton 1996, p. 373).

Pairwise invasibility plots (van Tienderen and de Jong 1986) allow one to visualize pairs of trait values that can coexist as stable polymorphism. This representation is explained in figure 5.4, and invasibility plots representative of the different cases are shown in figure 5.5.

Applicability of This Framework

In the above example, when the population is away from a candidate ESS, the first-order effect of selection is nonzero; being first order, it dominates higher-order terms (at least for mutations of small effect), and this guarantees that some mutations will be selected in the direction indicated by this first-order effect. One concludes that the population will evolve towards the candidate ESS if the first-order effect is of the appropriate sign. In the candidate ESS, however, the first-order effect vanishes, so near this point the second-order term (or else some higher-order term, see Kisdi 1999) becomes dominant. This is used to determinate the fate of the population near the candidate ESS, namely whether the population will stay monomorphic (ESS) or not (branching). Such approaches have been widely discussed and applied (e.g., Taylor 1989; Lessard 1990; Christiansen 1991; Eshel 1996; Eshel et al. 1997; Geritz et al. 1998; Abrams 2001a). We examine some of the assumptions underlying them.

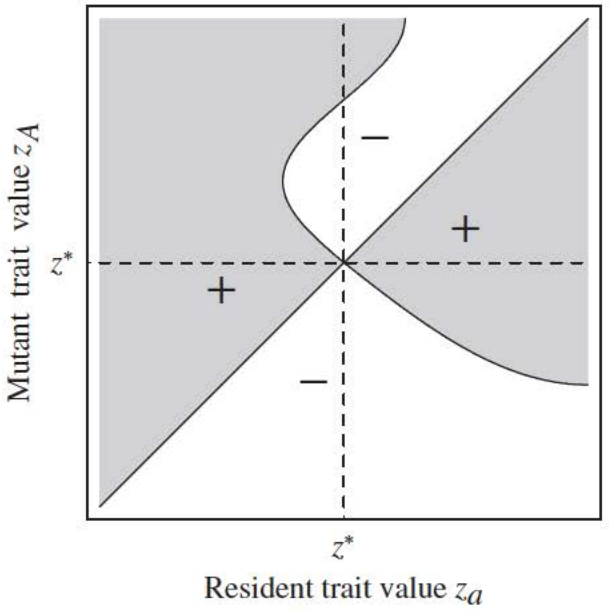

FIGURE 5.4. A pairwise invasibility plot. The shaded areas represent combinations of resident and mutant traits, where mutant fitness is > 1 ("+" signs), while the unshaded ones represent combinations of traits, where mutant fitness is < 1 ("−" signs). Here, for resident $z_a < z^*$, mutant fitness is higher for mutant trait values slightly above the resident one (above the diagonal), while for $z_a > z^*$, mutant fitness is higher for mutant trait values slightly below the resident one (below the diagonal). Hence, z^* is convergence stable. Moreover, mutations with small effect have fitness < 1 when the resident trait value is z^* (along vertical dashed line), so z^* is also evolutionarily stable if only mutants with small effect are considered. However, some mutations with large effect can invade.

ASSUMPTIONS ABOUT GENETIC VARIATION

We have assumed haploid inheritance. In the above example, it can be verified that with diploid inheritance with additive allelic effects on the phenotype, the first derivative (5.5) is unchanged, and the second derivative (5.7) is also unchanged in $p = 0$. Later evolution after branching can be affected. Similar results were obtained by Geritz and Kisdi (2000) for a slightly different model.

Dominance will be discussed below. We made two other impor-

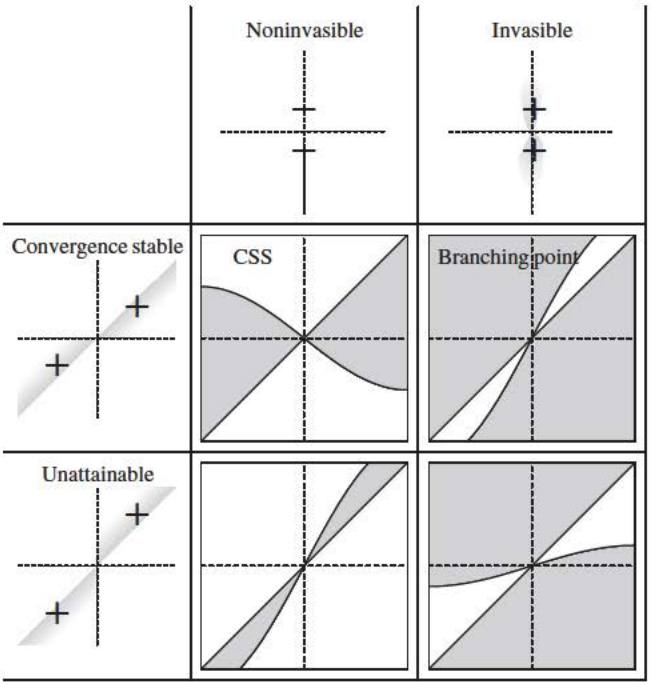

FIGURE 5.5. Pairwise invasibility plots for the different types of candidate ESSs. As in figure 5.3, the margins show the graphical translation of each stability condition: on top, whether mutants near the candidate ESS are selected against (noninvasibility, i.e., evolutionary stability) or for (invasibility), and on the left, whether mutants closer to the candidate ESS are selected for (convergence stability) or not (unattainable point). The cells show invasibility plots representative of the different combinations of cases.

tant assumptions about genetic variation. First, the effects of mutations were small. Second, we considered a one-locus basis for the trait under selection without random environmental effects. The latter assumption clearly does not hold for most traits in natural populations, and the former may not either, although the issue is more complex.

The genetic basis of selected traits in natural populations may not be well described by models ignoring mutations with large effect. On the contrary, the widespread occurrence of genes with major effects is expected from some mechanistic models of metabolic pathways (Bost et al. 1999). Experimental crosses often exhibit genes with major effect

(Tanksley 1993; Bürger 2000; Mackay 2001). However, the type of genetic variation observed between experimental strains of distant origin may not be representative of genetic variation between competing individuals in natural populations. More relevant to this issue, studies of adaptation in natural populations suggest that evolution of new adaptations often occurs through the spread of a few genes with major effect (MacNair 1991; Orr and Coyne 1992; Raymond et al. 2001).

Implications for convergence stability. The mathematical criteria of the previous section are "local," describing selection for "small" δ. They may not account for selective interactions on mutants of large effect. Hence, they are necessary, but sometimes insufficient conditions for global stability (fig. 5.4). Mutations of large effect might affect convergence stability to the extent that they could invade a locally convergence-stable equilibrium. This may occur in particular when there are several candidate ESSs (e.g., Jansen and Mulder 1999; Gandon et al. 2003).

The assumption of a one-locus basis, without random environmental effects, may be defended on the grounds that, waiting long enough, a population will evolve by a sequence of substitutions towards the stable points identified by the single locus analysis (e.g., Hammerstein 1996; Eshel et al. 1998; this idea is of course implicit in much of evolutionary ecology). Yet, for many traits of interest, there is persistent phenotypic variation in natural populations; even recurrent deleterious mutations may alter what is the best (selected) strategy, since the relative fitness of two alleles may depend on the phenotypic effects of all alleles in the population. So it may be worth considering how different types of variation may slow the progress to the point identified by the game theoretical analysis, or even bias it.

In sexual populations, the multilocus model most often considered is the ìnfinitesimal model (Fisher 1918; Bulmer 1971; Bulmer 1980, p. 150; Lynch and Walsh 1998), which remains difficult to establish from first principles (Bürger 2000, p. 189). It assumes that many loci affect the trait under selection, with small, equal, and additive effects. The single-locus and infinitesimal models yield similar conditions for convergence stability, though they are equivalent only under restrictive assumptions (e.g., Charlesworth 1994, p. 185; Abrams et al. 1993; Gomulkiewicz 1998). Differences arise because under the infinitesimal model, the distribution of phenotypes of bearers of any allele is Gaus-

sian, while such a distribution is not implied by the one-locus model. Differences may also arise when random environmental effects are considered, again because the relative fitness of two alleles may depend on the distribution of phenotypes in the population.

Implications for evolutionary stability. Whether divergence actually occurs when selection is disruptive depends on the genetic basis of the trait. In the infinitesimal quantitative genetic model, the distribution of phenotypes in the population is always normal, so branching is excluded, as it would result in a nonnormal distribution of phenotypes. In this model, divergent selection at a branching point may only result in an increased equilibrium phenotypic variance (e.g., Slatkin 1979).

Only a few simulations have investigated the potential for branching in models with several loci (Dieckmann and Doebeli 1999). Stemming from different models, different views have been expressed about the potential for branching in sexual populations (Doebeli and Dieckmann 2000; Abrams 2001b). Whether genetic variation is of a nature promoting branching or not will ultimately be answered by studies of evolution in natural populations. Under multilocus inheritance, branching may require the simultaneous evolution of assortative mating. Various models of sympatric speciation have considered this idea, but very few actually demonstrate it by following the joint evolution of assortative mating and other ecological traits. Doebeli and Dieckmann (2000), and to a much smaller extent Felsenstein (1981), evaluated the potential for such processes, the most recent models emphasizing the fact that, as in the asexual model analyzed in this chapter, the population first evolves to a branching point where extreme resources are not much exploited.

Doebeli and Dieckmann observed a nonmonotonous effect of the number of loci on the expected time for branching; overall, branching was more difficult than in the single-locus model. A full interpretation of these results has not yet been given. Intuitively, one would expect that since branching occurs if rare mutants are favored by escaping competition with common genotypes, it may not occur when no locus has a major effect on the trait relative to other loci, because the distribution of phenotypes associated with different genotypes at one locus is not disjunct enough. However, this argument may make sense only if the distribution of phenotypes in the population is not affected by the number of loci and the distribution of their effects.

Divergence may also be prevented by a large variance of environ-

mental effects on the trait under selection, as again rare mutants do not escape competition with common genotypes as much as in the absence of environmental effects. If branching is a relatively slow process, it leaves time for other species, or simply for other ecomorphs selected in slightly different ecological conditions, to fill the niche created by these unexploited resources. The evolution of different morphs of three-spined sticklebacks (*Gasterosteus aculeatus*), sometimes raised as an example of sympatric speciation, may actually illustrate divergence of two ecomorphs that first evolved in isolation, rather than branching (Taylor and McPhail 2000).

Evolution of Several Traits

When two traits are subject to natural selection, the convergence stability of strategies for these different traits depends on additional conditions beyond those for the convergence stability of a strategy for one trait, assuming that other traits are fixed (Abrams et al. 1993; Motro 1994; Matessi and Di Pasquale 1996; see Leimar, in press, for more than two traits). Abrams et al. (1993) and Matessi and Di Pasquale (1996) describe cases in which convergence stability at either trait is not required for the convergence stability of a joint strategy for the two traits. This may occur when a first trait has convergence-stable strategies when a second trait is fixed, the second trait has only unattainable candidate ESSs when the first is fixed, and the first trait evolves much faster due to limiting mutation on the second trait.

In such arguments it is assumed that one trait is fixed when a mutation appears for the other trait. There are some contexts, such as evolution of mutation or recombination rates, where computation of the ESS for the trait requires the consideration of polymorphism at other loci. In such cases, linkage disequilibria may matter. Consider two loci, each controlling for a different trait. In an infinite panmictic population, no gametic disequilibrium is expected to evolve between neutral alleles. If one assumes weak selection (of order δ) at several loci, it will generate disequilibria of order δ, as given by "quasi-linkage equilibrium" approximations (e.g., Crow and Kimura 1970; Barton and Turelli 1991; Nagylaki 1993; Bürger 2000). The effects of the disequilibria on selection will be of the order of the disequilibria [$O(\delta)$] times the selection coefficients [$O(\delta)$]; hence, they will be only of order δ^2. Thus, convergence stability does not seem to be affected. However, in finite or sub-

divided populations, random gametic disequilibrium may be produced by drift. These are known to matter in the evolution of recombination rates (e.g., Otto and Barton 2001; Otto and Lenormand 2002), and we may expect them to matter in other contexts as well, though what matters in general is not clear. A random disequilibrium generated by drift in a neutral model may imply a selection effect of order δ in the same way that relatedness induces selection effects of order δ in proportion to the neutral value of relatedness. Yet the gametic disequilibrium has zero expectation in the neutral model (see p. 19), so the induced selective effect may also have zero expectation. A selective effect resulting from gametic disequilibrium induced by selection would be of higher order in δ. It would be worth further investigating multilocus evolution in a game theoretical perspective.

Assumptions about Frequency Dependence

An unstated assumption of the methods of analysis described above is that the first-order effect on allele frequency change is frequency independent (or at least of a direction independent of allele frequency). This assumption may be counterintuitive, since these methods are widely applied to models of frequency-dependent selection. However, in the resource competition example, the first-order effect is frequency independent [equation (5.6)]; only higher-order effects are frequency dependent. If the first-order effects were also frequency dependent, one could consider the effects of selection at some "low" allele frequency in order to determine whether an allele will invade or not. But this would in no way tell whether this allele can go to fixation and therefore whether the population would converge towards some point through a sequence of allele substitutions. In other words, it would not be obvious how to define convergence stability or how to define a candidate ESS such that the population would evolve towards it.

Frequency independence of first-order effects is found in a general class of models for haploid panmictic populations (namely, those characterized by fitness functions that are differentiable functions of the individual phenotype and of the mean population phenotype: see p. 95). Models that often serve as examples of frequency dependence, such as the prisoner's dilemma game between two distinct strategies, are part of this class of models when a continuum of strategies is considered (see the appendix to this chapter).

However, there are exceptions, the general importance of which remains to be evaluated. For example, in the model of evolution of body size of Maynard Smith and Brown (1986), the first-order effects are frequency dependent in a way that can generate stable polymorphism. This results from their assumption that the fitness effects are not continuous functions of the difference in phenotype between interacting individuals.

Another simple example of frequency dependence of first-order effects is given by dominant or recessive alleles in diploid populations. Consider a population with two alleles a and A, and frequency-independent genotypic "fitnesses" (actually relative fecundities) 1, $1+hs$, and $1+s$ for genotypes aa, aA, and AA, respectively. h is a dominance coefficient, and the direction (sign) of selection changes with allele frequency when $h > 1/2$. In a panmictic population, the expected change in allele frequency takes the form

$$\begin{aligned}\Delta p =& \frac{sp(1-p)[h+(1-2h)p]}{1+2p(1-p)sh+p^2 s} \\ =& sp(1-p)(h+(1-2h)p)+O(s^2)\end{aligned} \tag{5.10}$$

instead of the form $sp(1-p)+O(s^2)$ shown in equations (2.1) or (5.6) (e.g. Wright 1969; Hartl and Clark 1997). In particular, it is well known that a heterozygote advantage ($s > 0, h > 1$) will lead to stable polymorphism, and it does so in a way predictable from first-order effects.

However, this example may be of limited interest, as it assumes that mutants with different dominance levels cannot occur. This is not necessarily so (e.g., Bourguet and Raymond 1998). Further, the dominance can evolve through the evolution of modifiers of dominance at a second locus (Mayo and Bürger 1997). Then, dominance may have little consequence on convergence stability, though this has been little considered (see van Dooren 1999 for the evolution of dominance after branching and van Dooren 2000 for the evolution of overdominance).

FITNESS IN A SUBDIVIDED POPULATION

We first examine patterns of frequency dependence in spatially subdivided populations, as they help us to formulate the problem of defining fitness measures for such populations.

Frequency Dependence in Subdivided Populations

In a subdivided population, an allele may be locally common even though globally rare. For example, it may be common among some interacting family members or among members of a single deme in a spatially subdivided population. One may expect different results when selection is considered on only one copy of the allele or, alternatively, when there is a "large" number of copies of it.

We again consider the potential for frequency dependence of first-order effects. Figure 5.6 illustrates its occurrence in the evolution of helping-behavior between neighbors within demes in an island and in a stepping stone model. This model will be considered again in chapter 7. Some details are omitted here, as they are not required for the present argument.

The first-order effects on the average fitness of more helping individuals is shown as a function of their frequency $\bar{p}$ in the total population. That is, we represent $W(\bar{p})$ such that $\mathrm{E}(\bar{p}' - \bar{p}) = W(\bar{p})\bar{p}(1-\bar{p}) + O(\delta^2)$. Under an island model of 200 demes, we see (fig. 5.6a) that fitness is almost independent of $\bar{p}$, except at extreme frequencies. Near $\bar{p} = 0$, fitness varies from its value when there is only one copy of the allele to an approximately constant value when there are many copies of it. Comparing cases with the same number of individuals per deme but variable number of demes, we see that fitness becomes constant for all $0 < \bar{p} < 1$ as the number of demes increases. Hence, the graph of fitness makes a sharper bend (emphasized by thin arrows) near $\bar{p} = 0$ as the number of demes increases.

The constancy of fitness at intermediate frequencies was part of Hamilton's argument for inclusive fitness [see equation (2.4)]. This argument, which will be detailed in chapter 7, was based on the relationship (4.14) between frequencies of identical pairs of genes and relatedness coefficients. It therefore relied on a relationship that may not apply under localized dispersal, as noted in chapter 4.

Indeed, the case of isolation by distance is more complex. In a linear stepping stone model, the first-order effect of selection may not be constant at intermediate frequencies, but as the number of demes increases, selection becomes either positive for all $0 < p < 1$ or negative for all $0 < p < 1$. This result can be intuitively understood by considering the following model. In a population with very limited dispersal,

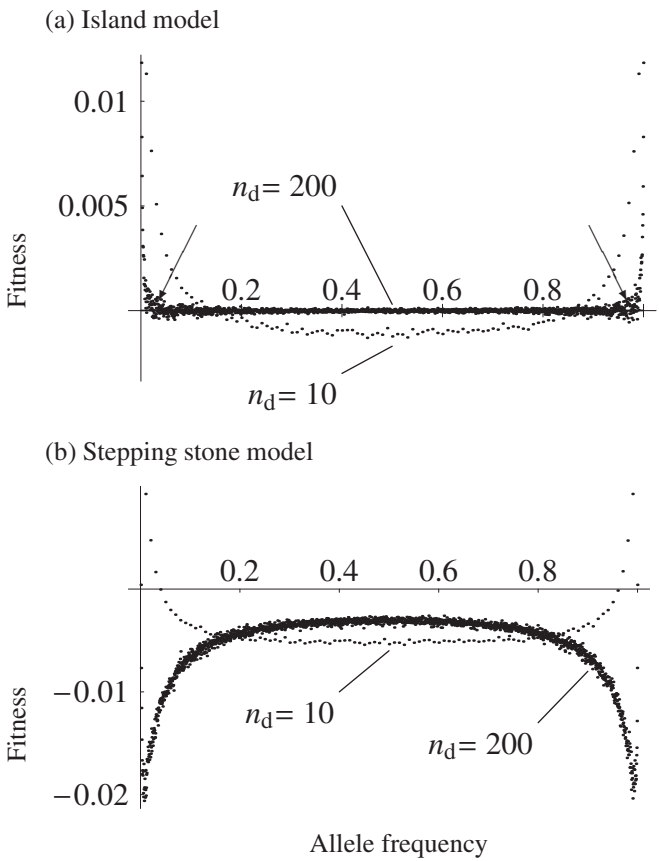

FIGURE 5.6. Frequency dependence of fitness in a structured population. This figure represents average first-order effects on fitness as a function of allele frequency, estimated from simulations (a) in an island model and (b) in a linear stepping stone model. In each case, two values of the number of demes n_d were considered.

a strongly favored allele increases in frequency within a subpopulation faster than it spreads to adjacent subpopulations. Hence, one may view the spread of this allele as the spread of an area almost completely occupied by carriers of the allele. If selection proceeds through replacement of noncarriers with carriers at the boundary of the area, the direction of change in allele frequency will not depend on the size of the area and therefore on $\bar{p}$. The magnitude of this change will be proportional to the length of this boundary, which will not scale as $\bar{p}(1-\bar{p})$.

However, this argument is not conclusive, since it considers a model with deterministic dynamics different from the stochastic model that it attempts to explain. In fact, there is presently no rigorous argument showing that selection is either positive or negative at all intermediate frequencies. The generality of this pattern should depend on the relationship between frequencies of identical pairs of genes and allele frequency in the total population, as will be further discussed in chapter 7 (see p. 108).

Similar patterns of frequency dependence, which are not conducive to stable polymorphisms, may hold for all models of the form considered in chapter 6. These models assume some form of spatial homogeneity. In contrast, an environmental spatial gradient may create variable selection pressures in space. Under isolation by distance, this will lead to frequency-dependent first-order effects conducive to stable polymorphisms in the form of "clines" (e.g., Nagylaki 1975; Barton and Gale 1993).

How to Measure Selection?

Given the above patterns of frequency dependence, one possible conclusion is that the appropriate "invasion" fitness should ignore selection pressures that occur only at extreme allele frequencies. Thus, we may approximate change in allele frequency in a finite island model by change in allele frequency in an infinite island model, the approximation being good except at extreme allele frequencies. This idea, applied to randomly mixing populations, was basic to the original development of the theory of inclusive fitness (Hamilton 1964, 1970; Grafen 1985). According to this theory, inclusive fitness approximates change in allele frequency well, except at extreme frequencies. However, it is not obvious how to apply this idea to the stepping stone example, where the fitness measure should be frequency dependent.

In addition, this idea does not always fit very well with the idea that we should measure selection specifically on "rare alleles." Actually, the latter notion is vague enough to encompass conflicting fitness measures (defining a limit as $\bar{p} \to 0$ is mathematically ambiguous in an infinite population). Fitness measures of "rare" alleles have generally been computed as if the mutant allele were in a unique copy in the population, thereby ignoring inclusive fitness effects (Metz et al.

1992). For example, analyses that have concluded that the optimal rate of dispersal is no dispersal have used fitness measures that ignored kin selection effects (e.g., Holt and McPeek 1996; Parvinen 1999). Yet, in an infinite island model, kin selection effects can be accurately taken into account in an unambiguously defined measure of fitness of rare alleles. This fitness measure is the expected number R_{m} of successful emigrant descendants, over successive generations, of a single deviant immigrant in a deme, computed in the absence of immigration of other deviant individuals in the deme (Metz and Gyllenberg 2001). Kin selection effects are taken into account in this measure by following across generations the number within the deme of descendants of the deviant immigrant, and by accounting for all effects of this number on the immediate number of offspring of each deviant. Inclusive fitness can then be understood as the fitness of a "rare" allele: rare in the total population, though in several copies in some demes.

R_{m} may be seen to depend on the probability that two individuals in a deme descend from the same immigrant (Ajar submitted), which is one interpretation of F_{ST} and similar relatedness coefficients (see previous chapter). As will be detailed in chapter 7, Hamilton's derivation of inclusive fitness was based on the alternative "regression definition" of relatedness, but both definitions lead to identical computations in infinite island models. Hence, to first order, R_{m} is equivalent to Hamilton's inclusive fitness although they were conceived to measure different aspects of selection.

R_{m} allows one to evaluate, numerically at least, both convergence and evolutionary stability in infinite island models. In the next chapter, we will not restrict ourselves to these models. We will avoid ambiguities of the concept of rare allele by deriving exact results for finite populations. There will often be little difference in the outcome of selection in finite and infinite populations: the value of the finite population results is mainly to provide the appropriate framework to construct and validate infinite population models. A drawback is that the results will be exact only to the first order and do not allow one to evaluate evolutionary stability *stricto sensu*. On the other hand, the results allow better analytical understanding of the selective forces involved than numerical approaches.

Convergence Stability

Since measures of convergence stability should indicate to which trait value a population will evolve by a series of allele substitutions, the probability of fixation should be such a measure (Rousset and Billiard 2000). Some trait value z^* is convergence stable if mutants that are closer to z^* have a higher probability of fixation than mutants that are further from z^*. Concretely, we use the fixation probability as follows. Let $\Pi(z_a + \delta, z_a)$ be the fixation probability of an A mutant (with trait $z_A = z_a + \delta$) in a population with "resident" trait z_a. We substitute $\Pi(z_a + \delta, z_a)$ for w_A in equation (5.4), and let a candidate ESS be a trait value z^*, where $\partial_\delta \Pi(z^* + \delta, z^*) = 0$. This approach is practical, as will be demonstrated in subsequent chapters.

The fixation probability is a straightforward way to measure the effects of selection (Haldane 1927; Fisher 1958, chapter 4). It clearly determines the direction of selection when evolution is well described by allele substitutions. If, on the other hand, the first-order effects are frequency dependent in a way that promote stable polymorphisms, probabilities of fixation may be uninformative. Thus, we assume that the frequency dependence shown in figure 5.6 near $p = 0$ and 1 does not strongly support stable polymorphisms. This assumption is supported by the peculiar nature of the observed frequency dependence. Selection changes direction only when there are a few copies of the allele, so it only weakly supports a stable polymorphism and is easily overcome by drift. The effect of drift is reduced by increasing the number of demes, but the pattern of frequency dependence is simultaneously altered so that no stable polymorphism results.

Evolutionary Stability

The resource competition example suggested a simple method to evaluate "evolutionary" stability: only the second-order derivative was frequency dependent, and a stable polymorphism could occur because selection is positive for a whole range of allele frequencies below some $p^* > 0$ and is negative above this p^*. Positive selection for some "vanishing" allele frequency [$p = 0$ in equation (5.7)], or when there is only one copy of the allele, is merely indicative of positive selection over a range of allele frequencies.

In a spatially structured population, one should likewise consider a measure of evolutionary stability indicative of positive selection over a range of allele frequencies. It should take into account the fact that an allele may be locally common even if it is rare in the total population, and the computation should be exact to the second order in δ. Selection depends on the distribution of allele frequency among demes, and the fitness of the A allele can be written as the weighted average

$$w = \sum_{k=1}^{N} \rho_k w_k, \tag{5.11}$$

where N is the deme size, w_k is the fitness of A in a deme with k A copies, and ρ_k is the probability that an A gene finds itself in such a deme (Day 2001). The second-order effect on fitness of phenotypic effect δ is thus

$$\left.\frac{d^2 w}{d\delta^2}\right|_{\delta=0} = \sum_{k=1}^{N} \left(\frac{d^2\rho_k}{d\delta^2} + 2\frac{d\rho_k}{d\delta}\frac{dw_k}{d\delta} + \rho_k \frac{d^2 w_k}{d\delta^2} \right) \tag{5.12}$$

$$= \sum_{k=1}^{N} \left(2\frac{d\rho_k}{d\delta}\frac{dw_k}{d\delta} + \rho_k \frac{d^2 w_k}{d\delta^2} \right) \tag{5.13}$$

because $\sum_{k=1}^{N} \rho_k = 1$. Thus, the second-order effects of selection depend on the first-order effects of selection on the distribution of allele A among demes, as reflected in the $d\rho_k/d\delta$ terms. Day (2001) analyzed evolutionary stability in a case where the coefficients $dw_k/d\delta$ of the derivatives $d\rho_k/d\delta$ are null. In general, however, the first-order effects of selection on the distribution of allele A among demes need to be computed. In the infinite island model, they may be derived from a Markov chain model for the successive numbers over generations of nonemigrant descendants of a mutant immigrant in a deme (Ajar submitted), as already considered in the definition of the fitness measure R_{m} (Metz and Gyllenberg 2001).

CONCLUSION

This chapter has reviewed the diverse concepts of evolutionary stability, focusing on analytical criteria and their application for spatially structured populations. It was suggested that whenever the concept of convergence stability is useful, convergence stability may be evaluated in terms of probabilities of fixation of mutants. This approach will be developed in the following chapters.

APPENDIX: THE PRISONER'S DILEMMA GAME

The "prisoner's dilemma" is a classic conceptual model of game theory. In various extensions, it may also serve as a concrete model for biological interactions (e.g., reviews and further references in Leimar 1997; Dugatkin 1998; see also Turner and Chao 1999). However, it is hard to find explicit analyses of allele frequency changes in this model, as presented below.

In its simplest form (Maynard Smith 1982; Axelrod and Hamilton 1981; Axelrod 1984), a pair of individuals interact. Individuals may either "cooperate" or "defect" and receive "payoffs" depending on the actions of both individuals. In an evolutionary context, these payoffs may be effects on fecundity. If both individuals cooperate, they both get "reward" R of cooperation. If they both defect, they both get "punishment" P. If one cooperates and the other defects, the one that defects receives "temptation" T to defect, and the one that cooperates gets "sucker's payoff" S.

The dilemma itself is the case in which interactors would do better by both cooperating than by both defecting ($R > P$), but it is always better to defect, whatever the other does ($P > S$ and $T > R$). Clearly, cooperation cannot evolve in this case, and many alterations of the rules have been considered, two important ones being the continuous and the iterated versions described below.

Noniterated Game

We first consider the game with arbitrary payoffs R, T, S, P. We assume that individuals pair at random and that they may play "mixed" (Maynard Smith 1982) strategies described by the probability z that an individual cooperates. We compare two strategies z_a and $z_A = z_a + \delta$ in frequencies $1 - p$ and p. The expected payoff received by a "focal" individual with strategy $z_\bullet$ in interaction with an "actor" individual with strategy z_0 is

$$F(z_\bullet, z_0) = z_\bullet z_0 R + z_\bullet(1 - z_0)S + (1 - z_\bullet)z_0 T + (1 - z_\bullet)(1 - z_0)P. \tag{5.14}$$

Therefore, the expected payoff for z_a individuals can be written

$$F_a = \Pr(A|a)F(z_a, z_A) + [1 - \Pr(A|a)]F(z_a, z_a), \tag{5.15}$$

where $\Pr(A|a)$ is the probability that the actor is A, given the focal individual is a (in an infinite panmictic population, this is simply p). As $F(z_a, z_A) = F(z_a, z_a) + \delta\partial F/\partial z_0|_{z_0=z_a} + O(\delta^2)$,

$$F_a = F(z_a, z_a) + \delta\frac{\partial F}{\partial z_0}\Pr(A|a) + O(\delta^2). \tag{5.16}$$

Likewise, the expected payoff for z_A individuals can be written

$$F_A = F(z_a, z_a) + \delta\left(\frac{\partial F}{\partial z_\bullet} + \frac{\partial F}{\partial z_0}\Pr(A|A)\right) + O(\delta^2), \tag{5.17}$$

where $\Pr(A|A)$ is the probability that the actor is A, given the focal is A (in an infinite panmictic population, this is again p).

From these results, one finds that in a panmictic population, the expected frequency in the next generation is

$$p' = p\frac{F_A}{pF_A + (1-p)F_a} = p + \delta p(1-p)\frac{\partial F}{\partial z_\bullet} + O(\delta^2). \tag{5.18}$$

This shows that the first-order effect of selection is not frequency dependent: only higher-order terms are.

Here,

$$\partial F/\partial z_\bullet = S - P + z_a(R - S + P - T). \tag{5.19}$$

Hence, in a population of defectors ($z_a = 0$), some cooperation may invade if $S > P$, which means that it is better to defect when a cooperator loses more than a defector when interacting with a defector. In a population of cooperators ($z_a = 1$), $\partial F/\partial z_\bullet = R - T$, which means that it is better to defect when a cooperator loses more than a defector when they each interact with a cooperator. These conclusions are obvious when considering a single deviant individual, but equation (5.18) shows that in an infinite panmictic population, they extend to any allele frequency, provided the difference between the two strategies is small. If $S > P$, the population converges to the level of cooperation

$$z = \frac{S - P}{R - T - S + P} \tag{5.20}$$

if z is within $[0, 1]$.

Iterated Game

In the iterated prisoner's dilemma, after each interaction, there is a probability w of a further interaction with the same individual. One

may compare the fitness of two strategies, "always defect" (AllD) and "tit for tat" (TFT). The former defects in each interaction in the sequence between the two players. The latter cooperates on the first interaction and then does whatever the other individual did in the previous interaction. Both strategies can be stable against invasion by the other (Maynard Smith 1982; Axelrod 1984).

Now consider again that individuals may play mixed strategies, described by the probability z that an individual plays TFT with another individual. If both individuals play TFT, the expected payoff of the whole sequence is $\rho = R/(1 - \mathsf{w})$. If both individuals play AllD, the expected payoff is $\pi = P/(1 - \mathsf{w})$. In a sequence of interactions between TFT and AllD, the expected payoff of TFT is $\sigma = S + P\mathsf{w}/(1-\mathsf{w})$, and the expected payoff of AllD is $\tau = T + P\mathsf{w}/(1-\mathsf{w})$. If we neglect other strategies (e.g., alternating cooperation and defection: Maynard Smith 1982; Axelrod 1984; "Pavlov": Nowak and Sigmund 1993; and various other stochastic strategies, e.g., Nowak 1990; Brauchli et al. 1999; Killingback and Doebeli 2002), the iterated game defined by R, P, S ,T, and w can be viewed as the noniterated game, with payoff function $F(z_\bullet, z_0)$ [equation (5.17)], with ρ, π, σ, τ in lieu of R, P, S, T. Hence, from equation (5.19), some cooperation invades a population of defectors if $\sigma > \pi$, and the convergence-stable strategy is

$$z = \frac{\sigma - \pi}{\rho - \tau - \sigma + \pi} \tag{5.21}$$

if this is within $[0, 1]$.

Thus, we can seek convergence-stable strategies in both the noniterated and the iterated forms of the prisoner's dilemma game, since first-order effects are not frequency dependent [equation (5.18)].

CHAPTER SIX

Convergence Stability in a Spatially Homogeneous Population

This chapter demonstrates a practical and efficient way to compute the measure of convergence stability proposed in the previous chapter, that is, a first-order effect on the probability of fixation of a mutant. The form of the main result [equation (6.18)] is of particular interest, as it uses only the biologically intuitive concept of fitness (expected number of adult offspring) and the expected coalescence time of pairs of genes. The latter times are themselves directly connected to the (also biologically intuitive) concept of probability of identity between pairs of genes. The derivation is based on a very limited set of mathematical techniques (Rousset 2003c).

Fitness is described by the fitness function, which gives the expected number of adult offspring of a "focal" individual in terms of the phenotypes of all individuals in competition with it. The effects of different individuals on the number of offspring of the focal individual are described by partial derivatives of the fitness function. In line with previous work (e.g., Taylor and Frank 1996), this formalism will be called a *direct fitness* approach. In the next chapter we will discuss how these results relate to the concept of inclusive fitness.

Some models take interactions between individuals into account by following the frequency of demes with different numbers of copies of allele A (e.g., Eshel 1972; Motro 1982; Metz and Gyllenberg 2001). This is feasible under infinite island models, though heavily numerical. The assumptions of the direct fitness method, defined in this chapter (p. 95), ensure that all important features of the distribution of number of A alleles per deme are taken into account by probabilities of identity of gene pairs, thereby simplifying the analysis.

Finally, this chapter illustrates the concrete computation of candidate evolutionarily stable strategies under an island model of dispersal and isolation by distance.

WEAK SELECTION EFFECTS ON PROBABILITY OF FIXATION

We consider a population following the haploid island or isolation by distance models of chapter 3, with n_{d} demes. In these models, all demes are equivalent with respect to their demographic properties, including dispersal. Thus, as defined in chapter 3, they are spatially homogeneous. Spatially inhomogeneous populations will be considered in chapter 9.

As in the previous chapter, we consider two alleles a and A segregating at a locus, with associated phenotypes z_a and $z_A \equiv z_a + \delta$, and the fixation probability $\Pi(z_A, z_a)$ of an A mutant in an a population. We first relate changes over one generation to probabilities of fixation. Using the so-called "direct fitness" method, we describe allele frequency changes over one generation, conditional on the genetic composition of the parental population. This yields expressions in terms of allele frequencies in pairs of individuals. The effects on probability of fixation can then be written in terms of expected allele frequencies in pairs of individuals.

Fixation Probability as Allele Frequency Change

One way to compute the derivative of fixation probability $\phi(z_a) \equiv d\Pi(z_a + \delta, z_a)/d\delta$ is to relate it to the expected change in allele frequency over many generations. With probability Π, a mutant goes to fixation; then the change in frequency is $1 - 1/N_{\mathrm{T}}$ (final frequency minus initial frequency). With probability $1 - \Pi$, the mutant is lost; then its change in frequency is $-1/N_{\mathrm{T}}$.

Let $\mathbf{p}(t) \equiv [p_1(t), \ldots, p_{n_{\mathrm{d}}}(t)]$ be the frequencies of the A allele in the different demes in generation t. Let $\bar{p}(t)$ be the average allele frequency in t. By the above argument, the expected change in frequency is $\mathrm{E}[\bar{p}(\infty) - \bar{p}(0) | \bar{p}(0) = 1/N_{\mathrm{T}}] = \Pi - 1/N_{\mathrm{T}}$, and thus $d\mathrm{E}[\bar{p}(\infty)|\bar{p}(0) = 1/N_{\mathrm{T}}]/d\delta = d\Pi/d\delta$. This expected change is also $\mathrm{E}[\sum_{t=0}^{\infty} \bar{p}(t+1) - \bar{p}(t) | \bar{p}(0) = 1/N_{\mathrm{T}}]$, where the sum is actually until

fixation of either a or A, since there is no frequency change afterwards. We can write

$$\frac{d\Pi}{d\delta} = \frac{d}{d\delta}\sum_{t=0}^{\infty}\sum_{\mathbf{p}(t)}\Pr[\mathbf{p}(t)]\{\mathrm{E}[\bar{p}(t+1)|\mathbf{p}(t)] - \bar{p}(t)\} \tag{6.1}$$

$$= \sum_{t=0}^{\infty}\sum_{\mathbf{p}(t)}\Pr[\mathbf{p}(t);\delta=0]\frac{d}{d\delta}\{\mathrm{E}[\bar{p}(t+1)|\mathbf{p}(t)] - \bar{p}(t)\} + \sum_{t=0}^{\infty}\sum_{\mathbf{p}(t)}\frac{d\Pr[\mathbf{p}(t)]}{d\delta}\underbrace{\{\mathrm{E}[\bar{p}(t+1)|\mathbf{p}(t);\delta=0] - \bar{p}(t)\}}_{0}, \tag{6.2}$$

since expected changes in the neutral process are null. The remaining term depends on the probabilities $\Pr[\mathbf{p}(t);\delta=0]$ in the neutral process and on the selection process (through derivatives with respect to δ). It is the expectation of the given derivative over realizations of $\mathbf{p}(t)$ in the neutral process. We may write it as

$$\frac{d\Pi}{d\delta} = \mathrm{E}^{\circ}\left(\sum_{t=0}^{\infty}\frac{d\{\mathrm{E}[\bar{p}(t+1)|\mathbf{p}(t)] - \bar{p}(t)\}}{d\delta}\right), \tag{6.3}$$

where E° refers to expectations over realizations of $\mathbf{p}(t)$ in the neutral process (i.e., when $z_A = z_a$) and $\mathrm{E}[\bar{p}(t+1)|\mathbf{p}(t)]$ refers to expected changes given $\mathbf{p}(t)$ over one generation in the process with selection. Note that $\mathbf{p}(0)$ itself is random, since the mutation can occur in any of the demes.

Fitness Functions

The expected frequency of A in the total population may be written as the sum over demes of the expected number w_j of adult offspring of an A parent that is in deme j, times the number Np_j of A parents in deme j, divided by the total number of offspring Nn_{d}. Hence,

$$\mathrm{E}(\bar{p}'|\mathbf{p}) = \sum_{j=1}^{n_{\mathrm{d}}} w_j p_j / n_{\mathrm{d}}. \tag{6.4}$$

To evaluate effects of δ on expected allele frequency, we will write w_j as a function of the phenotypes of the different individuals in the population, which are themselves function of allele frequencies and

of δ. As a working example, we will consider the evolution of the dispersal rate in a finite island model, where the phenotypes are the dispersal rate $z_\bullet$ among the focal individual's juveniles and the average dispersal rates $z_1, \ldots, z_{n_\mathrm{d}}$ in the different demes. Since $z_A = z_a + \delta$, $z_k \equiv z_k(\mathbf{p}) = z_a + \delta p_k$.

We assume the following life cycle (consistent with fig. 3.1): (1) Reproduction occurs; an infinite number of juveniles is produced (in other words, J juveniles are produced and we let $J \to \infty$). (2) Juvenile dispersal occurs. (3) Dispersing offspring experience a survival cost c_d, which is the probability that a juvenile dies during dispersal. (4) Adults die. (5) Offspring compete and only N of them survive to adulthood in each deme.

The expected number of adult offspring of an individual is the sum of the expected number of its offspring in each deme, and in each deme these expected numbers can be expressed as the number of adults offspring, N, times the frequency of the focal individual's juveniles among all juveniles that compete for a deme. The latter frequency is simple to compute (Frank 1986). For example, $J(1 - z_\bullet)$ juveniles of the focal individual do not disperse, and for a focal individual in deme j, they are in competition with $JN(1 - z_j)$ juveniles produced in the same deme (including those of the focal individual) and with $JN(1 - c_\mathrm{d})z_k/(n_\mathrm{d} - 1)$ immigrant juveniles from every other deme k. This yields the first term of the fitness function; the other terms are for juveniles that disperse in any deme $i \neq j$. Hence, $w_j = w_j(z_A, \mathbf{z})$, where $\mathbf{z} \equiv (z_1, \ldots, z_{n_\mathrm{d}})$ and

$$w_j(x_\bullet, \mathbf{x}) = \frac{(1 - x_\bullet)}{1 - x_j + (1 - c_\mathrm{d}) \sum_{k \neq j} x_k/(n_\mathrm{d} - 1)} + \sum_{i \neq j} \frac{(1 - c_\mathrm{d})\, x_\bullet/(n_\mathrm{d} - 1)}{1 - x_i + (1 - c_\mathrm{d}) \sum_{k \neq i} x_k/(n_\mathrm{d} - 1)}, \tag{6.5}$$

where $\mathbf{x} \equiv (x_1, \ldots, x_{n_\mathrm{d}})$. The function is here defined independently of the biological significance attached to its variables, so the latter are denoted by $\mathbf{x}$ rather than as phenotypes $\mathbf{z}$. This function is used to evaluate the fitness of an A parent; hence, one evaluates it in $x_\bullet = z_A$

and $x_k = z_a + \delta p_k$. Then

$$\frac{dw_j(z_A, \mathbf{z})}{d\delta} = \sum_{k=\bullet,1}^{n_d} \frac{\partial w_j(x_\bullet, \mathbf{x})}{\partial x_k} \frac{d(z_a + \delta p_k)}{d\delta} \tag{6.6}$$

$$= \frac{\partial w_j(x_\bullet, \mathbf{x})}{\partial x_\bullet} + \sum_{k=1}^{n_d} \frac{\partial w_j(x_\bullet, \mathbf{x})}{\partial x_k} p_k. \tag{6.7}$$

ASSUMPTIONS ABOUT FITNESS

Such linear expressions, in terms of partial derivatives and of allele frequencies, can be written whenever we can describe fitness as follows:

(1) The fitness function, which describes the expected number of adult offspring of an individual as a function of the phenotypes of different individuals, is differentiable with respect to the variables representing these phenotypes in the usual mathematical sense (Courant and John 1989). There may be few counterexamples in the literature (though see Maynard Smith and Brown's model, as noted on p. 81), and their significance could be debated.

(2) The phenotype of each individual is a linear function of (only) one allele frequency, such as the frequency of the allele in its own genome (which is 0 or 1 for haploid organisms: in this case the phenotype is necessarily linear in allele frequency, since there is always a line through two points).

For convenience, the second condition may be relaxed. It is only required that fitness can be described in terms of variables z_k, each of which is, to the first order only, a linear function of some allele frequency p_k:

$$z_k = z_a + \delta p_k + O(\delta^2). \tag{6.8}$$

Then the writing of the fitness function can be simplified. For example, if the fitness of an individual involves multiplicative effects of the phenotypes of different neighbors z_1, z_2, ... [e.g., if it involves products $(1 + z_1)(1 + z_2), \ldots$], then the fitness function can be written in terms of z_1, z_2, ..., satisfying the above requirements (1) and (2). Although formally correct, this is clumsy: it is more convenient to express fitness only in terms of a few average phenotypes, using only (1) and (6.8). Although such a simplification is not exact, it is correct to the first order in δ, which is sufficient for our purposes. For example, $(1+z_1)(1+z_2) =$

$[1+(z_a+p_1\delta)][1+(z_a+p_2\delta)] = 1-{z_a}^2+(1+z_a)(z_1+z_2)+O(\delta^2)$ is, to the first order in δ, a function of the average $(z_1+z_2)/2$. The important point is that all terms involving products p_1p_2 are of the second order in δ.

Thus, the requirement (6.8) is rather weak. Nevertheless, it is not always possible to express fitness in terms of variables with the above property, particularly in the following two cases: (1) When allele A is dominant over a in a diploid population. In this case, the phenotype of an individual is not a linear function of its own "frequency" of A (0, $1/2$ or 1). (2) When some effect on a phenotype z is experienced only if two or more individuals harbor the A allele (say, when $p_1 = p_2 = 1$ for two haploid individuals). Yet the direct fitness formalism may be used to analyze selection on recessive and dominant mutants (chapter 8).

THE SUM OF PARTIAL DERIVATIVES IS NULL

A useful property of fitness functions satisfying the assumptions stated above is that the sum of their partial derivatives is null:

$$\sum_{k=\bullet,1}^{n_{\mathrm{d}}} \frac{\partial w(x_\bullet, x_1, \ldots, x_{n_{\mathrm{d}}})}{\partial x_k} = 0, \tag{6.9}$$

where $\sum_{k=\bullet,0}^{n_{\mathrm{d}}}$ is a sum over all the arguments of w. A proof follows, computing the total derivative $dw(z,\ldots,z)/dz$ in two ways. First, since the total population size is constant across generations and independent of individual phenotypes, the expected number of adult offspring of an individual in a monomorphic population is constant for any z [it is $w(z,\ldots,z) = 1$]. Hence, the total derivative of $w(z,\ldots,z)$ with respect to z is zero. Second, this total derivative is expressed in terms of partial derivatives, as

$$\sum_{k=\bullet,0}^{n_{\mathrm{d}}} \frac{\partial w(x_\bullet, \ldots, x_{n_{\mathrm{d}}})}{\partial x_k} \frac{dx_k}{dz}, \tag{6.10}$$

where all variables $x_k = z$ and so $dx_k/dz = 1$. Therefore, the total derivative, which is known to be null, is the sum of partial derivatives.

Fixation Probability: Direct Fitness Expansion

Equation (6.4) allows one to express the expected allele frequency as

$$\mathrm{E}[\bar{p}'|\mathbf{p}] = \frac{1}{n_\mathrm{d}} \sum_{j=1}^{n_\mathrm{d}} \left(w_j(z_A, \mathbf{z})|_{\delta=0} + \delta \frac{dw_j(z_A, \mathbf{z})}{d\delta} + O(\delta^2) \right) p_j \tag{6.11}$$

$$= \bar{p} + \delta \frac{1}{n_\mathrm{d}} \sum_{j=1}^{n_\mathrm{d}} \frac{dw_j}{d\delta} p_j + O(\delta^2). \tag{6.12}$$

Then, using equation (6.6), the effect of phenotypic difference δ on expected allele frequency is

$$\frac{1}{n_\mathrm{d}} \sum_{j=1}^{n_\mathrm{d}} \frac{dw_j}{d\delta} p_j = \frac{1}{n_\mathrm{d}} \sum_{j=1}^{n_\mathrm{d}} \left(\frac{\partial w_j}{\partial x_\bullet} + \sum_{k=1}^{n_\mathrm{d}} \frac{\partial w_j}{\partial x_k} p_k \right) p_j. \tag{6.13}$$

As shown by equation (6.3), we need to compute the expectation of this expression over the distribution of $\mathbf{p}(t)$ in the neutral process, which is

$$\phi \equiv \frac{d\Pi}{d\delta} = \mathrm{E}^\circ \left[\sum_{t=0}^{\infty} \frac{1}{n_\mathrm{d}} \sum_{j=1}^{n_\mathrm{d}} \left(\frac{\partial w_j}{\partial x_\bullet} + \sum_{k=1}^{n_\mathrm{d}} \frac{\partial w_j}{\partial x_k} p_k \right) p_j \right]. \tag{6.14}$$

Expression in Terms of Parameters of Population Structure

COALESCENCE TIMES

The above expression is a function of the expectation of products $\mathrm{E}^\circ(p_j p_k)$ of allele frequencies over the distribution of $\mathbf{p}(t)$ in a neutral model. It can also be written

$$\phi = \frac{1}{n_\mathrm{d}} \sum_{j=1}^{n_\mathrm{d}} \sum_{k=1}^{n_\mathrm{d}} \frac{\partial w_j}{\partial z_k} \mathrm{E}^\circ \left[\sum_{t=0}^{\infty} (p_k p_j - p_j) \right] \tag{6.15}$$

because $\partial w_j / \partial z_\bullet$ is minus the sum of the other derivatives [equation (6.9)]. All terms $\mathrm{E}^\circ [\sum_{t=0}^{\infty} (p_k p_j - p_j)]$ can be evaluated as follows. First, $\mathrm{E}^\circ[p_j] = \bar{p}(0)$, since expected frequency does not change over generations in the neutral model. Second, in the absence of further mutations since the initial mutation event, two genes are both A only if they have a common ancestor within the t generations following the

mutation. Let $C_{jk}(l)$ be the probability of coalescence of the two lineages in generation l. The probability that the common ancestor was A is $\mathrm{E}^{\circ}[\bar{p}(l)] = \bar{p}(0) = 1/N_{\mathrm{T}}$. Hence, the probability that two genes are both A, t generations after the mutation, may be written

$$\mathrm{E}^{\circ}\left[p_j p_k(t)\right] = \bar{p}(0) \sum_{l=0}^{t} C_{jk}(l) = \frac{1}{N_{\mathrm{T}}} \sum_{l=0}^{t} C_{jk}(l). \tag{6.16}$$

This result implies that $\mathrm{E}^{\circ}[p_j - p_j p_k(t)] = [1 - \sum_{l=0}^{t} C_{jk}(l)]/N_{\mathrm{T}}$, and we evaluate its sum over t using

$$\sum_{t=0}^{\infty} \left(1 - \sum_{l=0}^{t} C_{jk}(l)\right) = \sum_{t=0}^{\infty} \sum_{l=t+1}^{\infty} C_{jk}(l) = \sum_{t=0}^{\infty} t C_{jk,t} \equiv T_{jk}, \tag{6.17}$$

that is, the expected coalescence time of the pair of genes. Thus, equation (6.15) takes the form

$$\phi = -\frac{1}{n_{\mathrm{d}}} \sum_{j=1}^{n_{\mathrm{d}}} \sum_{k=1}^{n_{\mathrm{d}}} \frac{\partial w_j}{\partial z_k} \frac{T_{jk}}{N_{\mathrm{T}}}. \tag{6.18}$$

Probabilities of Identity

The algebra of migration matrix models, described in chapter 4, can be used to compute expected coalescence times. However, equation (6.18) also immediately gives a relationship with stationary probabilities of identity (Q_{jk}) in models with mutation. From equation (4.17),

$$\frac{T_{jk}}{N_{\mathrm{T}}} = \lim_{\mu \to 0} \frac{1 - Q_{jk}}{2 N_{\mathrm{T}} \mu}. \tag{6.19}$$

Hence, equation (6.15) can be written

$$\phi = \lim_{\mu \to 0} \frac{1}{n_{\mathrm{d}}} \sum_{j=1}^{n_{\mathrm{d}}} \sum_{k=1}^{n_{\mathrm{d}}} \frac{\partial w_j}{\partial x_k} \frac{Q_{jk} - 1}{2 N_{\mathrm{T}} \mu} \tag{6.20}$$

Since we considered spatially homogeneous models, the above expression can be simplified. First, although the products $p_j p_k$ differ at any time for different values of j and k, their expectations will depend only on the spatial separation between demes. Hence, it is sufficient to consider probabilities of identity of genes at different distances (Q_k) rather than in specific demes (Q_{jk}). Further, fitness functions w_j have

a common form for all j. As illustrated in the next section, one may express fitness w of a focal individual as a function of its phenotype $z_\bullet$, the average phenotype z_0 of individuals in the focal deme (i.e., the deme of the focal individual) and the average phenotypes z_k in demes at different distances k from the focal one. Equation (6.20) then takes the form

$$\phi = \lim_{\mu\to 0} \sum_{k=1}^{n_\mathrm{d}} \frac{\partial w}{\partial z_k} \frac{Q_k - 1}{2N_\mathrm{T}\mu} = \lim_{\mu\to 0} \frac{S}{2N_\mathrm{T}\mu} \tag{6.21}$$

where

$$S \equiv \frac{\partial w}{\partial z_\bullet} + \sum_{k=1}^{n_\mathrm{d}} \frac{\partial w}{\partial z_k} Q_k. \tag{6.22}$$

Using equation (3.68), this can be written

$$\phi = \lim_{\mu\to 0} \sum_{k=1}^{n_\mathrm{d}} \frac{\partial w}{\partial z_k} \frac{Q_k - 1}{1 - Q_0^\mathrm{D}} = \lim_{\mu\to 0} \frac{S}{1 - Q_0^\mathrm{D}}, \tag{6.23}$$

where Q_0^D is the identity between individuals sampled before population regulation within a deme. The following examples illustrate how this result can be used.

PRACTICAL COMPUTATION OF CONVERGENCE STABILITY

Island Model

Consider the fitness function for the evolution of dispersal, given by equation (6.5). We can gather terms of the fitness function as follows. Since all demes are equivalent, for all demes k different from the focal one, the z_k variables contribute identical terms to the fitness measure. In particular, one may use a single notation Q_1 for the identity of genes between demes. Hence, we write $\sum_k z_k/(n_\mathrm{d} - 1) = \bar{z}$, the average phenotype of parents in demes k different from the focal one; and we write fitness as a function of this variable and the average phenotype z_0^R within the deme of the focal individual,

$$w(z_\bullet, z_0^\mathrm{R}, \bar{z}) = \frac{(1 - z_\bullet)}{(1 - z_0^\mathrm{R}) + (1 - c_\mathrm{d})\,\bar{z}} + \frac{(1 - c_\mathrm{d})\,z_\bullet}{(1 - \bar{z}) + (1 - c_\mathrm{d})\,[(n_\mathrm{d} - 2)\bar{z} + z_0^\mathrm{R}]/(n_\mathrm{d} - 1)}. \tag{6.24}$$

The notation $^{\mathrm{R}}$ is used to emphasize that the deme average z_0 here contains the focal individual. S is then a sum of three terms:

$$S = \frac{\partial w}{\partial z_\bullet} + \frac{\partial w}{\partial z_0^{\mathrm{R}}} Q_0^{\mathrm{R}} + \frac{\partial w}{\partial \bar{z}} Q_1. \tag{6.25}$$

Note that we consider identity Q_0^{R} rather than Q_0^{D}, because Q_0^{R} is the identity between genes sampled with replacement, accounting for the fact that we consider identity between the focal individual and any of the individuals in its deme, including itself. Following equation (6.23), ϕ is the low-mutation limit of

$$\frac{1-Q_1}{1-Q_0^{\mathrm{D}}}\left(\frac{\partial w}{\partial z_\bullet} + \frac{\partial w}{\partial z_0^{\mathrm{R}}}\frac{Q_0^{\mathrm{R}}-Q_1}{1-Q_1}\right) = \frac{1}{1-F_{\mathrm{ST}}}\left(\frac{\partial w}{\partial z_\bullet} + \frac{\partial w}{\partial z_0^{\mathrm{R}}} F_{\mathrm{ST}}^{\mathrm{R}}\right) \tag{6.26}$$

for $F_{\mathrm{ST}}^{\mathrm{R}} \equiv \frac{1}{N} + \frac{N-1}{N} F_{\mathrm{ST}}$. Solving $\phi = 0$ yields z^* in terms of the parameters N, n_{d}, and c_{d} (Gandon and Rousset 1999). The effects of the number of demes on the fitness function, on F_{ST}, and on z^* are generally negligible. In the infinite island model ($n_{\mathrm{d}} \to \infty$),

$$z^* = \frac{F_{\mathrm{ST}}^{\mathrm{R}} - c_{\mathrm{d}}}{F_{\mathrm{ST}}^{\mathrm{R}} - {c_{\mathrm{d}}}^2} \tag{6.27}$$

$$= \frac{1 + 2Nc_{\mathrm{d}} - \sqrt{1 + 4N(N-1){c_{\mathrm{d}}}^2}}{2c_{\mathrm{d}}(1+c_{\mathrm{d}})N}. \tag{6.28}$$

Equation (6.27) was obtained by Frank (1986); equation (6.28) is implicit in Frank (1994b) and related results were given by Taylor (1988) for diploid populations. An important message in such results is that some level of dispersal is always selected for (Hamilton and May 1977). This simple conclusion is very general. Considering the evolution of dispersal from an initial dispersal rate $z = 0$, it is easy to see that an increase in dispersal will always be favored: a nondispersing allele can never end as the common ancestor of individuals in different demes and consequently of the whole population, so it cannot invade and has a null probability of fixation. This conclusion is easily forgotten when one ignores all kin competition effects, including those between juveniles from the same mother.

Isolation by Distance

Assuming isolation by distance in a homogeneous environment, we can gather terms for actors at the same distance from the focal individual.

For example, in a one-dimensional stepping stone model, the evolution of the dispersal rate can be determined as follows. The expected number of offspring of an individual is expressed as a function of the dispersal rate $z_\bullet$ of its juveniles, of the average dispersal rate z_0^{R} of individuals in its deme, of the average dispersal rate $\bar{z}_1$ of individuals in the two adjacent demes, and of the average dispersal rate $\bar{z}_2$ of individuals in the two next-to-adjacent demes. The fitness function, obtained by the same argument as equation (6.5), is

$$\begin{aligned} w\left(z_\bullet, z_0^{\mathrm{R}}, \bar{z}_1, \bar{z}_2\right) =& \frac{1-z_\bullet}{1-z_0^{\mathrm{R}}+(1-c_{\mathrm{d}})\bar{z}_1} \\ &+ \frac{(1-c_{\mathrm{d}})z_\bullet}{1-\bar{z}_1+(1-c_{\mathrm{d}})(z_0^{\mathrm{R}}/2+\bar{z}_2/2)} \end{aligned} \tag{6.29}$$

(Gandon and Rousset 1999). From equation (6.23),

$$\phi = -\frac{\partial w}{\partial z_0^{\mathrm{R}}}\frac{1-Q_0^{\mathrm{R}}}{1-Q_0^{\mathrm{D}}} - \frac{\partial w}{\partial \bar{z}_1}\frac{1-Q_1}{1-Q_0^{\mathrm{D}}} - \frac{\partial w}{\partial \bar{z}_2}\frac{1-Q_2}{1-Q_0^{\mathrm{D}}}, \tag{6.30}$$

where Q_k is the identity between genes at distance k from each other. Here we used Q_0^{R} rather than Q_0^{D} for the same reason as in equation (6.25). The first ratio is $1-1/N$, and the two others are given by equation (3.34) (for $n_{\mathrm{d}} \to \infty$ and $\mu \to 0$):

$$\begin{aligned} \frac{1-Q_1}{1-Q_0^{\mathrm{D}}} =& 1+\frac{1}{N}\left(-1+\frac{1}{2m\sqrt{1-m}}\right), \\ \frac{1-Q_2}{1-Q_0^{\mathrm{D}}} =& 1+\frac{1}{N}\left(-1+\frac{2(1-\sqrt{1-m})}{m^2}\right). \end{aligned} \tag{6.31}$$

Further analysis of this model shows that the selected dispersal rate is approximately the same as in the island model (Gandon and Rousset 1999). Hence, the qualitative conclusions about dispersal in the island model also hold in this model.

CONCLUSIONS

Direct Fitness Method

As described here, allele frequency changes over one generation [equation (6.13)] and the selection measure [equation (6.22)] are of the same form, minimizing risks of error in going from one to the other. This

illustrates the utility of a direct fitness formalism. This formalism, as used here, is based on a minimal number of elementary concepts (number of adult offspring and probability of genetic identity) and is directly applicable to different models, such as finite and infinite populations and localized or "island" dispersal. In particular, elaborate arguments over how to interpret "relatedness" are alleviated. "Relatedness" does not need to be variously interpreted as, for example, the probability of identity or the ratio of such probabilities (see chapter 8 for an application to diploid populations). Rather, the fitness measure ϕ is defined in terms of a single concept of identity. Further, the significance of each term of ϕ is easily grasped: probabilities of identity measure the genetic similarity between an actor and the focal individual and therefore the probability that an actor will have a deviant behavior when the focal individual bears the mutant allele. The partial derivative with respect to the behavior of a category of actors describes the effect of deviant actors' behavior on the focal individual's fitness. Each derivative with respect to a given variable may itself be the sum of several effects, each of which corresponds to an easily identified effect on fitness. For example, in the stepping stone model [equation (6.30)], $\partial w/\partial \bar{z}_1$ involves a term representing the effect on fitness of the focal individual due to the competition of its philopatric juveniles with immigrant juveniles from neighbors in adjacent demes [derivative of the first ratio in expression (6.30)] and a term representing the effect on fitness of the focal individual due to the competition of its emigrant juveniles with philopatric juveniles from neighbors in adjacent demes (derivative of the second ratio).

The interpretation of the direct fitness formalism used here may be more restrictive that in the earlier account of Frank (1998). Here the effects of allele frequencies on fitness are entirely accounted for by the arguments of the fitness function and by its derivatives. If some allele frequency is instead included as a fixed parameter in the definition of the function itself (as in Frank 1998, p. 124), the ensuing computations have no obvious game theoretical significance. Note in addition that it is generally not sufficient to account for dispersal through its effects on relatedness only. The expression for the fitness function must also take dispersal into account in order to represent the expected number of adult offspring under a specified life cycle.

Experience shows that the direct fitness method is counterintuitive

in at least one respect. We describe fitness as function of (average) phenotypes of different types of actors, but we derive the fitness functions with respect to the "phenotypes," ignoring that the phenotypes are random variables. Phenotype as a random variable and "phenotype" as an argument of the fitness function were distinguished as z and x variables in equations (6.5) to (6.20) above, but not afterwards. When using the direct fitness method, it is convenient to ignore this distinction. There is no contradiction here, as the derivative with respect to a variable is well defined independently of the biological significance we may attach to this variable. We are not interested in the fitness of a particular individual, but by the expected effect on probability of fixation. Hence, no random variable can remain in the expression for the fitness measure, which is a parameter.

Fitness Maximization

We did not attempt to define a "fitness" measure that is "maximized" at the ESS, and for good reason. The language of maximization may be confusing when applied to fitness (particularly in class-structured populations, as further discussed on p. 194). Indeed, it is useful to consider fitness maximization if "fitness" has a clearly defined ecological meaning, by which we may compare ecological properties of populations with different trait values z through their fitness. In this sense, the expected number of adult offspring w has no obvious ecological significance, since it is 1 for any population fixed for some trait value z, given the population numbers are regulated (as already noticed on p. 96). On the other hand, it has also long been recognized that ecological concepts of fitness are in general not maximized by selection (e.g., Haldane 1932; Kawecki 1993; Charlesworth 1994).

Fitness w is maximized by evolution only in the following sense. Evolution may reach an evolutionarily stable strategy z^* if it is convergence stable, and individuals expressing an ESS have on average more offspring than deviant individuals competing with them. That is, $w(z+\delta, z)$ is maximized with respect to δ in $z = z^*$, but not with respect to z. However, it is also theoretically possible that evolution leads the population near a branching point, where individuals with deviant strategies have higher fitness.

CHAPTER SEVEN

Inclusive Fitness, Cooperation, and Altruism

The existence of altruistic behaviors, that is, behaviors that decrease the number of offspring of an individual relative to other individuals, has long been (and still is) readily accepted by many biologists and nonbiologists alike. But it has also long been realized that selection should act against altruistic behavior. These opposite conceptions initiated the "group selection" debate (e.g., Trivers 1985, chapter 4; Dawkins 1989, chapter 7).

Genes are transmitted more when their bearers have more offspring, so altruism should be selected against. Thus, one can consider selection of altruistic behavior only in some more restricted sense, particularly when an act is altruistic in some circumstances but not in others. Kin selection theory points out that bearers of genes increasing altruistic behavior also tend to receive more altruistic acts if they interact with genetically related individuals. Hence, although a behavior may be altruistic when a unique individual expresses it, it may be selected for when several altruists interact together. Thus altruism is favored when altruists have on average more offspring over these different cases.

Thus, it is only in the following sense that altruism is considered: it is a behavior that decreases the number of offspring of individuals expressing it when the correlated effects of neighbors on its fitness are ignored. This decrease may be measured by considering the fitness of a unique bearer of the gene for altruism, or, in a mixed population of altruists and nonaltruists, by comparing the fitness of a "focal" individual, whether it behaves altruistically or not, when all other individuals' behavior is kept constant (Grafen 1984). In the latter case, relatedness is neglected, since the other individuals' behavior is not changed when

the focal individual's behavior is changed.

Hamilton's rule (Hamilton 1964) provides a simple condition for the selection of such altruistic behavior: $-c + Rb > 0$, where c is the fitness cost (in terms of number of adult offspring) of the altruistic acts, b is the benefit of altruism received from related neighbors, and R is a relatedness measure between the focal individual and its neighbors. Hamilton's rule was formulated for a family-structured population, and the exact significance and the computation of these terms have been elaborated for this case in many later works (e.g., Michod and Hamilton 1980; Michod 1982; Frank 1998).

However, it was initially not obvious how kin selection theory could be applied to populations with limited dispersal (Hamilton 1964, 1971, 1975). It has since been realized that it could be applied almost without change, at least under an island model of dispersal. Yet the nature of costs, benefits, and relatedness remains a matter of confusion. In this chapter, we detail the close relationship between the fitness measure defined in the previous chapter and inclusive fitness measures. We use several theoretical and empirical examples to provide a better understanding of the components of inclusive fitness expressions in subdivided populations, as well as a basis for evaluating selection for altruism in spatially subdivided populations.

The results plainly support the dour conclusion (Grafen 1984; Maynard Smith 1987; Nunney 1998; Reeve and Keller 1999; Reeve 2000) that much of the debate about "group selection" and "altruism" boils down to an inconsistent use of words. Hence, we will emphasize the consistent use of the above definition of altruism. The issue is more than semantic, since a consistent word usage promotes the proper application of methods and, for example, clearly delineates conditions for the evolution of kin recognition.

WHAT INCLUSIVE FITNESS DOES MEASURE

Inclusive and Direct Fitness

Two ways of accounting for interactions between individuals may be distinguished (fig. 7.1). These are designed to count each interaction once and only once. One way is to consider each individual in turn as an "actor" on itself and on other individuals ("recipients") in the popu-

FIGURE 7.1. Inclusive and direct fitness. The same set of interactions between two individuals (arrows from actor to recipient) is shown on both sides of the figure. On the left, these interactions are grouped by actor (they share a common arrow style for each actor). This is the inclusive fitness approach. On the right, these interactions are grouped by recipient (they share a common arrow style for each recipient). This is the direct fitness approach.

lation. The other is to consider each individual in turn as a "recipient" of acts by individuals ("actors") in the population.

In the former approach, the idea is to count the fitness effects of the focal individual on itself and on other individuals weighted by their relatedness R to the actor, where R is a measure of the extent to which the recipients share the same genes as the actor. Thus, the celebrated formula for inclusive fitness, $-c + Rb$, says that a behavior is selected for if $-c+Rb > 0$, where $-c$ is the effect of the actor's behavior on the actor's fitness, and b is the total effect of the actor's behavior on other recipients in the population. This is the inclusive fitness approach.

In the second approach, the idea is to count the fitness effects of the focal individual on itself and the effects of other individuals weighted by their relatedness R to the focal recipient, where R is a measure of the extent to which the other individuals have a deviant behavior as a result of sharing the same genes as the recipients. This can also be put in the form $-c + bR$, where $-c$ is the effect of the focal individual's behavior on its own fitness, and b is the total effect of the other actors on a focal recipient. This is the "neighbor-modulated" or "personal" or "direct fitness" approach. It simply measures the average fitness of A-bearing individuals (their expected number of adult offspring, averaged over the different possible genotypes of their neighbors) as $1 - c + Rb$.

At first, it was considered that the inclusive fitness approach was easier to develop than the direct fitness one (Hamilton 1964; Maynard Smith 1987), but this has been shown not to be the case (e.g., Taylor and Frank 1996). An old practical problem with inclusive fitness approaches was to count all fitness effects once and only once (Grafen

1984). The direct fitness method automatically solves this problem. If the fitness function describes the expected number of adult offspring, the derivatives of the fitness function count all fitness effects once and only once.

Some authors have noticed that the two approaches could give different results (e.g., Maynard Smith 1982; Frank 1997a; Day and Taylor 1998). The two approaches may actually differ due to additional approximations not inherent in them. For example, conditions for evolutionary stability (*stricto sensu*) derived by either approach may both fail if the effects of selection on the probability that neighbors have a deviant phenotype are not taken into account.

But as presented in figure 7.1, these two methods are merely two different ways of accounting for the same interactions, and therefore they should both yield the same convergence stability condition when properly used. Thus, we make no distinction between direct fitness and inclusive fitness below. That the two are identical confirms that inclusive fitness measures the same quantity as individual fitness, that is, that inclusive fitness techniques indicate that a behavior is favored by selection when the individual fitness of its performers is higher than that of other individuals. Further, if estimation of the inclusive fitness of a gene is of interest, it only requires estimating the individual fitness of the bearers of this gene.

Hamilton's Derivation of Inclusive Fitness

The expected change in allele frequency given $\bar{p}$ (the A frequency in the total population) can be written

$$\mathrm{E}(\bar{p}'|\bar{p})-\bar{p} = \delta\left[\frac{\partial w}{\partial z_\bullet}\bar{p} + \frac{\partial w}{\partial z_0}\mathrm{E}(p_0^2|\bar{p}) + \sum_{k\neq 0}\frac{\partial w}{\partial z_k}\mathrm{E}(p_0 p_k|\bar{p})\right] + O(\delta^2), \tag{7.1}$$

where p_0 is the frequency of allele A in a "focal" deme, and w is the fitness of an A-bearing individual in this deme.

Hamilton (1964, 1970) used an argument that can be stated as follows. It is assumed that

$$\mathrm{E}(p_0|p_\bullet, \bar{p}) = Rp_\bullet + (1-R)\bar{p} \tag{7.2}$$

for some relatedness R, where $p_\bullet$ is the allele frequency in a focal individual and p_0 the frequency in its deme. This assumption corresponds

to a classical argument for defining relatedness, which was detailed in chapter 4. Either genes coalesce in a recent past (with probability R), in which case the allele frequency among within-deme neighbors p_0 equals its frequency in the focal individual $p_\bullet$, or the neighbors and focal individual are "unrelated" (with probability $1-R$), in which case the expected frequency in neighbors is the population frequency $\bar{p}$. In the kin selection literature, R is often referred to as a regression coefficient (Grafen 1985; Frank 1998). The regression of frequency p in the focal deme on $p_\bullet$ is, by definition (e.g., Bulmer 1979; Stuart and Ord 1991), the expectation of p given $p_\bullet$, and the regression coefficient is the coefficient that gives this expectation as a function of $p_\bullet$. According to equation (7.2), the regression is linear in $p_\bullet$ (and in $\bar{p}$): R is independent of $p_\bullet$ and of $\bar{p}$.

Since $\mathrm{E}(p_0^2|\bar{p}) = \mathrm{E}(p_\bullet p_0|\bar{p})$, and given $p_\bullet = 1$ with probability p_0, one then obtains $\mathrm{E}(p_0^2|\bar{p}) = p_0[R + (1-R)\bar{p}] + (1-p_0)(1-R)\bar{p} = R\bar{p} + (1-R)\bar{p}^2$. Likewise, $\mathrm{E}(p_0 p_k|\bar{p}) = \bar{p}^2$ for genes in different demes. Using equation (6.9), equation (7.1) then reduces to

$$\mathrm{E}(\bar{p}'|\bar{p}) - \bar{p} = \delta\left(\frac{\partial w}{\partial z_\bullet} + \frac{\partial w}{\partial z_0}R\right)\bar{p}(1-\bar{p}) + O(\delta^2) \tag{7.3}$$

$$\equiv W_{\mathrm{IF}}\bar{p}(1-\bar{p}) + O(\delta^2). \tag{7.4}$$

Here W_{IF} is the "inclusive fitness"

$$W_{\mathrm{IF}} \equiv \delta\tilde{W}_{\mathrm{IF}} \equiv \delta\left(\frac{\partial w}{\partial z_\bullet} + \frac{\partial w}{\partial z_0}R\right). \tag{7.5}$$

Thus, although fitness is a function of allele frequencies $\mathbf{p}$ in different classes of actors, the change in allele frequency is of the same form as in the case of frequency-independent fecundity effects [equation (2.1)]. This form is more generally obtained, assuming that for all classes k of actors, $\mathrm{E}(p_0 p_k|\bar{p}) = R_c\bar{p} + (1-R_c)\bar{p}^2$ for some R_c's independent of allele frequencies, as expected under an interfamily or interdeme separation of time scales.

Isolation by Distance

A different pattern is observed under isolation by distance, where selection appears frequency dependent, though of constant sign except at extreme allele frequencies (see in particular fig. 5.6b for $n_{\mathrm{d}} = 200$). This

pattern may be deduced from another expression for $\mathrm{E}(p_0 p_k|\bar{p})$. Unpublished simulations such as those reported by Rousset (2002) show that in the two-allele, one-dimensional stepping stone model, the frequency of identical pairs among genes sampled at k steps from each other is of the form

$$\mathrm{E}(p_0 p_k|\bar{p}) \approx \mathrm{E}(p_0 p_k)+\bar{p}-\mathrm{E}(\bar{p})-\frac{x_1}{1-F_k}[\bar{p}-\bar{p}^2-\mathrm{E}(\bar{p}-\bar{p}^2)]+\frac{x_2}{1-F_k} \tag{7.6}$$

where F_k is the pairwise F_{ST} at distance k, and x_1 and x_2 are two constants independent of k. This form implies that the change in allele frequency [equation (7.1)] is of constant sign.

Then it is no longer possible to define an inclusive fitness measure consistent with Hamilton's argument. However, it is still possible to define inclusive fitness from the effects on the fixation probability of mutants, that is, from the convergence stability measures considered in the previous chapters. In particular, we have seen concretely how we could compute convergence stability in a stepping stone model in the previous chapter [equation (6.30)]. This fitness measure is proportional to

$$\frac{\partial w}{\partial z_\bullet}+\frac{\partial w}{\partial z_0^{\mathrm{R}}}\frac{Q_0^{\mathrm{R}}-Q_2}{1-Q_2}+\frac{\partial w}{\partial \bar{z}_1}\frac{Q_1-Q_2}{1-Q_2}. \tag{7.7}$$

By giving unit weight to the cost to the actor, $\partial w/\partial z_\bullet$, this expression is directly comparable to expressions of the form $-c+Rb$. Each of the ratios can be understood as relatedness measures, measuring to what extent genes in the same deme (Q_0^{R}) or adjacent demes (Q_1) are more similar than genes two demes apart (Q_2). Thus, when dispersal is localized, relatedness measures are local and differ from relatedness deduced from one pedigree. Only the relatednesses of parents whose juveniles compete for the next generation matter, and they matter in proportion to the competitive effects with more or less distant neighbors, as measured by the derivatives of the fitness functions with respect to phenotypes of neighbors at different distances.

Discussions of inclusive fitness theory under isolation by distance have suffered from a lack of a consistent definition of relatedness. The above expression provide such definitions. There is no discontinuity here between the island and isolation by distance scenarios. Considering progressively less localized dispersal, competition is with progressively more distant individuals. If dispersal is over large distances

(ideally, following the island model), individuals in different colonies may be considered unrelated. The more precise argument is that pairs of members of adjacent colonies will not be more related than pairs of members of more distant colonies so that only the first relatedness in equation (7.7) remains. But otherwise, the idea that individuals in adjacent demes are unrelated cannot be made very precise. Indeed, the idea that there are both related and unrelated pairs of individuals within the focal individual's deme also cannot be made very precise. Pairs of individuals within demes or from adjacent demes may be viewed as being related by variable pedigrees, and the genetic identity or relatedness parameter will be an average measure over the different pedigrees within *and among* different demes.

ALTRUISM IN SPATIALLY SUBDIVIDED POPULATIONS

Cost, Benefit, and Relatedness

Equation (7.5) is clearly of the form $-c + Rb$, where the meaning of each term is close to those of Hamilton. However, if the fitness function is written as a function of the average phenotype z_0^{R} in the deme of the focal individual, *including* this individual, $\partial w/\partial z_\bullet$ is not exactly the effect of the focal individual on its own fitness. Since the focal individual is involved in z_0, its effect on itself depends on $\partial w/\partial z_0$. To better match Hamilton's concepts of cost and benefit, we write the fitness of the focal individual as a function of the average phenotype z_0 in the deme of the focal individual, *excluding* this individual. The expression for w in terms of the arguments $z_\bullet, z_0^{\mathrm{R}}, z_1$ will be slightly different from the expression for w in terms of the arguments $z_\bullet, z_0, z_1$, and we denote them w^{R} and w^{D}, respectively. It suffices to write $z_0^{\mathrm{R}} = [z_\bullet + (N-1)z_0]/N$ to get $w^{\mathrm{D}}(z_\bullet, z_0, z_1)$ from $w^{\mathrm{R}}(z_\bullet, z_0^{\mathrm{R}}, z_1)$, with the following implications.

(1) The cost $-c \equiv -\delta\tilde{c}$ is the effect of a focal individual on its expected number of adult offspring:

$$-\tilde{c} \equiv \frac{\partial w^{\mathrm{D}}}{\partial z_\bullet} = \frac{\partial w^{\mathrm{R}}}{\partial z_\bullet} + \frac{1}{N}\frac{\partial w^{\mathrm{R}}}{\partial z_0^{\mathrm{R}}}. \tag{7.8}$$

This computation effectively gives the variation in fitness of an individual due to its behavior when all other individuals' behavior is kept constant.

(2) The benefit $b \equiv \delta\tilde{b}$ is the effect of different individuals in the same deme on the expected number of adult offspring of the focal individual:

$$\tilde{b} \equiv \frac{\partial w^{\mathrm{D}}}{\partial z_0} = \frac{N-1}{N}\frac{\partial w^{\mathrm{R}}}{\partial z_0^{\mathrm{R}}}. \tag{7.9}$$

The definition of R must match the definition of w. Thus, if we consider w^{D}, the relatedness is between a focal individual and other individuals in the deme. In the infinite island model, we may compute it as the identity Q_0^{D} between different adults. In the finite island model, we may compute it as $R = \lim_{\mu\to 0}(Q_0^{\mathrm{D}} - Q_1)/(1 - Q_1)$.

This definition of R resolves the inconsistencies, noticed by Seger (1981), that result from equating "identity by descent" and relatedness, even though it can be computed as identity by descent in the infinite island model (chapter 4). We would have similar results for family interactions within a random mixing population (Hamilton's original model). There Q_0 would be the identity between interacting family members, and Q_1 the identity between randomly chosen individuals in the population. R then has essentially the same value as relatedness computed from pedigrees (chapter 4). Thus we recover the concepts of relatedness used by Hamilton.

Helping Neighbors

Is it a good strategy to help neighbors in a subdivided population? The answer is not obvious (Hamilton 1964). To increase one's neighbors' fecundity is to increase the fecundity of individuals more likely to bear the "cooperator" allele, so it should favor the spread of this allele. On the other hand, it increases the number of juveniles competing with the focal individual's juveniles, and this will disfavor the spread of the allele.

A simple model may be used to assess the balance between these two effects (Taylor 1992a). It considers a population with an infinite island structure. There are N adults per deme and the dispersal rate of juveniles is m. The life cycle is the usual one (fig. 3.1), including the following interactions. An individual "actor" bearing allele A engages in some cooperative act that has two effects. First, the fecundity of all neighbors in the deme (including the actor) is increased by B/N (if we

exclude the actor, we have the same model with different notations). Second, this act has a cost C, such that the fecundity of the actor is $1 - C + B/N$ times its fecundity in the absence of the act. A single act thus increases the average fecundity of group members by $(B - C)/N$. For simplicity we assume here additivity of the different costs and benefits, but later results are not affected if we instead assume multiplicativity of some effects on fecundity, for example, when fecundity is $(1 - C)(1 + B/N)$, because there is still additivity of first-order effects (see p. 95).

The case where $B > 0$ and $-C + B/N < 0$ has been described as altruism. However, B and $-C + B/N$ are effects on fecundity, that is, the number of juveniles per adult. These are not fitness costs and benefits, where, as previously, fitness is defined as the expected number of adult offspring of an adult. In general, the direction of selection is measured by following changes in allele frequency over one iteration of the life cycle. Hence, fitness must be defined as an expected number of copies of a parental gene, where this number is counted at the same stage (e.g., adult or juvenile, before or after dispersal) in the life cycle where the parental gene is considered.

We will evaluate the selection between an allele a that does not express the behavior (phenotype $z_a = 0$) and an allele A that engages in it with "intensity" δ (phenotype $z_A = z_a + \delta = \delta$). The costs and benefits are then $-C\delta$ and $B\delta$. Hence, a single act has effect $(B/N - C)\delta$ on the fecundity of the actor and effect $(B - C)\delta/N$ on the average fecundity in its deme. The fitness function then takes the form

$$w = (1 - m)\frac{1 + Bz_0^{\mathrm{R}} - Cz_\bullet}{(1 - m)(1 + (B - C)z_0^{\mathrm{R}}) + m(1 + (B - C)z_1)} + m\frac{1 + Bz_0^{\mathrm{R}} - Cz_\bullet}{1 + (B - C)z_1}. \quad (7.10)$$

The first term is obtained as N times the production of nondispersing juveniles of the focal individual over the production of all juveniles that come into competition in the deme. The focal individual has relative fecundity $1 + Bz_0^{\mathrm{R}} - Cz_\bullet$, because it pays the cost of its own behavior and receives $B\delta/N$ from each cooperator in the deme. Hence, it receives $B\delta$ times the frequency of cooperators in its deme; z_0^{R} is δ times the frequency of cooperators in its deme. $(1 - m)$ of these juveniles stay in the deme, hence the $1 - m$ factor of the first term. The denominator counts competing juveniles as those produced by the deme $(1 + (B - C)z_0^{\mathrm{R}})$ and nondispersing (factor $1 - m$), plus those (fac-

tor m) from other demes, where parents had average relative fecundity $1 + (B - C)z_1$. The second term of w is computed similarly.

To compute inclusive fitness, we need the derivatives

$$\left.\frac{\partial w}{\partial z_\bullet}\right|_{z_a=0} = -C \text{ and } \left.\frac{\partial w}{\partial z_0^{\mathrm{R}}}\right|_{z_a=0} = B - (1-m)^2(B-C) \quad (7.11)$$

and the relatedness Q_0^{R} in the infinite island model. Using $Q_0^{\mathrm{D}} = (1 - m)^2 Q_0^{\mathrm{R}}$ [equation (3.16)], we find that

$$\frac{\partial w}{\partial z_\bullet} + Q_0^{\mathrm{R}} \frac{\partial w}{\partial z_0^{\mathrm{R}}} = (-C + B/N)(1 - Q_0^{\mathrm{D}}).^1 \quad (7.12)$$

Thus, the act is favored only if $-C + B/N > 0$, that is, if the effect of the act on the fecundity of the actor is positive (Taylor 1992a). Both the fitness effects and the identity Q_0^{R} depend on m, but the overall expression is independent of m. It is, of course, possible to modify the assumptions of the model so as to get different results (e.g., Perrin and Lehmann 2001; and models with overlapping generations and juvenile dispersal: Koella 2000; Taylor and Irwin 2000).

What does this result mean in terms of fitness costs and benefits? Following equations (7.8), (7.9) and (7.11), they are

$$\begin{aligned} \tilde{c} &= C - \frac{B - (1-m)^2(B-C)}{N} \\ \text{and} \qquad \tilde{b} &= (N-1)\frac{B - (1-m)^2(B-C)}{N}. \end{aligned} \quad (7.13)$$

In figure 7.2, we follow Hamilton (1964, 1970) in classifying behaviors as altruistic ($c > 0$, $b > 0$), selfish ($c < 0$, $b < 0$), and spiteful ($c > 0$, $b < 0$: the act is deleterious both for the actor and its neighbors). This is consistent with other works explicitly aiming to distinguish between positive and negative effects on the actor's fitness (Axelrod and Hamilton 1981; Trivers 1985; Hamilton 1996, p. 263). Cooperation is the word used in particular by Axelrod and Hamilton (1981) in the remaining case ($c < 0$, $b > 0$: the act is beneficial both for the actor and its neighbors), although it was apparently not defined as explicitly as the other categories. Note that such usages may conflict with the habit to name behaviors as, for example, "cooperative breeding" before knowing whether they are cooperative or altruistic. A perhaps more appropriate, already-used term would be "helping."

[1]Intermediate steps are $-C + Q_0^{\mathrm{R}} B - Q_0^{\mathrm{R}}(1-m)^2(B-C) = -C + Q_0^{\mathrm{R}} B - Q_0^{\mathrm{D}}(B-C) = -C(1-Q_0^{\mathrm{D}}) + B(Q_0^{\mathrm{R}} - Q_0^{\mathrm{D}}) = -C(1-Q_0^{\mathrm{D}}) + B(1-Q_0^{\mathrm{D}})/N$.

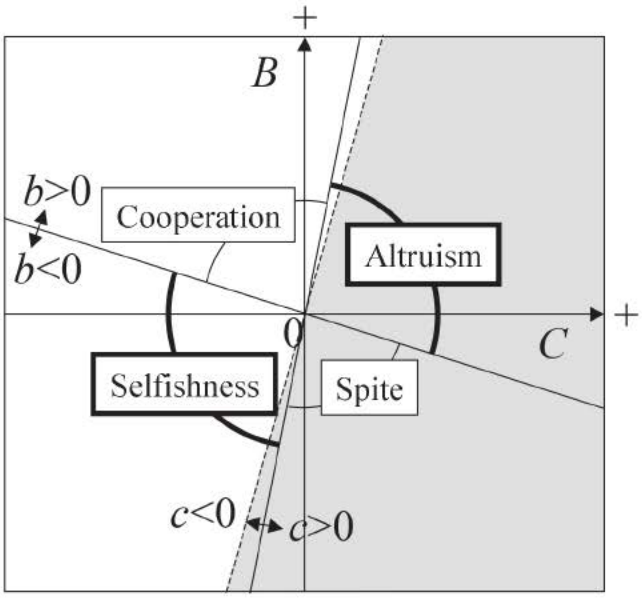

FIGURE 7.2. Selection on altruism and other behaviors in a subdivided population. Quantitative results are represented for the case $N = 4$ and $m = 1/2$. The $c = 0$ ($B = 5C$) and $b = 0$ ($C = -3B$) lines separate the four cases of altruism, cooperation, selfishness, and spite, and they depend on m. The shaded area is the parameter set of behaviors selected against. It is bounded by the dotted line $\phi = 0$ (here $B = 4C$, independent of m).

The line $C - B/N = 0$ and the line $c = 0$ differ (being separated by the narrow area between $B = 4C$ and $B = 5C$ in fig. 7.2). This shows that some acts that increase the fecundity of the actor ($C < 4B$) may reduce its number of offspring when it is the only one to express the behavior ($c > 0$). The reason is that the actor helps its neighbors (which are its main competitors) enough to decrease its own fitness. Such behaviors are altruistic and may yet be selected for. It is therefore important, when testing theoretical predictions about the evolution of altruism in subdivided populations, to determine whether costs and benefits are measured in terms of number of juveniles before dispersal and competition or in terms of adult offspring.

Although the model shows that altruism can be selected for, the area of selected altruism is narrow. It slightly widens when $m \to 0$, because

the $c = 0$ line comes closer to the vertical axis. A similar analysis could be conducted under isolation by distance. The trait is again selected for if $-C + B/N > 0$ [equation (7.21)]. However, for given m and N, the $c = 0$ line may come closer to the vertical axis under isolation by distance than in the island model. Therefore, the direction of selection on a behavior, characterized by given values of C and B, is the same whatever the dispersal rate and dispersal distribution. However, whether the behavior under consideration qualifies as altruistic depends on population structure.

Other Examples

In the dispersal models of chapter 6, dispersal evolves because dispersing juveniles avoid competition with other juveniles born in the same deme, which are likely to be related to them, born either to the same mother or to a related mother. This is a classic example of a behavior subject to kin selection (e.g., Comins et al. 1980; Frank 1986, and many later works). In a population without dispersal ($z = 0$), the fitness cost is $c = c_{\mathrm{d}} - 1/N$, which may be positive or negative. Thus, dispersal may be, but is not necessarily, altruistic when the population is not at the ESS.

In the noniterated prisoner's dilemma (see p. 89), when tit for tat (TFT) is opposed to the all-defect strategy, rare TFT are always selected against, and TFT is thus altruistic. In the iterated version of the dilemma, TFT may be cooperative. Axelrod and Hamilton (1981) thus termed such behavior "reciprocal cooperation," not "reciprocal altruism," as did Trivers (1971) and many later authors (see also Hamilton 1975, p. 151; Hamilton 1996, p. 263). Likewise, in the continuous version of the iterated dilemma, a slight increase in expression of TFT is not always selected against, and thus it is not always altruistic. An intermediate level of expression of the TFT strategy may be convergence stable (see p. 89). This level may increase in subdivided populations (Taylor and Irwin 2000; see also Brauchli et al. 1999 and Killingback et al. 1999 for the noniterated game). This may occur both through effects of relatedness between interactors and through effects of population structure on fitness costs and benefits. The two could be separated by the same techniques as those above.

The Importance of Kin Competition

The importance of kin selection effects seems obvious when one considers family-structured models, but is less obvious with spatially structured models. Where spatial structure differs from family structure is that in many organisms, dispersal is strong enough that relatedness is low. However, it would be misleading to base on this idea a general argument for the weak importance of ensuing kin selection effects. For example, models of split sex ratios in insect colonies have shown that small relatedness asymmetries may have drastic effect on sex ratios (e.g., Boomsma and Grafen 1991; Crozier and Pamilo 1996), which in turn have important consequences on the evolution of other traits. Parallel results may appear as more spatially structured models are investigated.

It is also well known that the importance of kin selection effects depends on the relative magnitude of costs, benefits, and relatedness, and the importance of ecological constraints in shaping costs and benefits has been repeatedly stressed. Slender-tailed meerkats (*Suricata suricatta*) offer an example suggesting that cooperation, rather than kin selection, drives many social behaviors. These mammals live in groups of 2 to 30 adults. All adult group members contribute to guarding and feeding the young (Clutton-Brock et al. 2001b). Meerkat groups include a dominant female and a dominant male; they may also include subordinate females capable of reproduction. However, dominant females may restrict access of subordinate females to unrelated males, and they may kill young born to subordinates. Dominant females can also force subordinates out of the group. They thereby control the reproduction of subordinate females, so that dominant females give birth to the majority of the young. Subordinate reproduction is viewed as a consequence of incomplete control by the dominant female (Clutton-Brock et al. 2001b).

Two additional key features of the meerkat system are (1) expulsion from the group, and more generally emigration from the group, is risky (Doolan and Macdonald 1996); (2) reproductive success is low at very low group size, a form of the Allee effect (Courchamp et al. 1999; Clutton-Brock et al. 2001b). Thus, there are considerable fitness costs in trying to join other colonies or found new ones. These conditions are favorable to the evolution of cooperative (rather than altruistic) so-

cial behavior, because it may be better for subordinates to help rearing the young of the dominant female, so as to maintain a group in which they may eventually have some opportunity to reproduce. Such social behavior may evolve even between unrelated individuals, even though the exact level of helping behavior may be somewhat affected by relatedness. Some data indeed show little effect of relatedness on helping behavior (Clutton-Brock et al. 2001a), although other data suggest effects of relatedness on emigration (Doolan and Macdonald 1996).

Kin Recognition

Kin recognition is the recognition of more related individuals, either through imperfect cues or, more rarely, through the accurate assessment of genetic similarity or of shared genealogy. In some way, a behavior directed only towards neighbors may be viewed as a form of kin recognition; however, we focus here on discrimination between different members of a group of equally distant neighbors.

It is well known that kin recognition is not required for kin selection to operate (Dawkins 1979), and the above models ignored kin recognition. On the other hand, the existence of a kin-discriminating behavior suggests that the behavior has an altruistic component. In the model of fecundity interactions between neighbors, we have seen that the fate of mutants with given effects B and C is independent of the dispersal distribution. Given some tradeoffs between B and C (e.g., Perrin and Lehmann 2001), we expect that a behavior such that $B - C/N = 0$ will evolve. However, as this behavior will be altruistic (as understood above), a kin-discriminating expression of this behavior may be selected for.

Whether acts should be directed only towards closer kin depends on cost-benefit ratios of kin discrimination itself and on the variation in relatedness between an actor and different potential recipients (see Keller 1997, for more discussion). Overall, a kin-discriminating expression of a behavior should be more likely to evolve when the kin selection component of selection on the behavior is larger.

Long-tailed tits (*Aegithalos caudatus*) provide an example of kin recognition in a spatially structured population. They breed in groups of five to eight breeding pairs. All individuals initially breed, but may decide to help another pair following breeding failure. Some group

members are close relatives and others are immigrants. Long-tailed tits almost always help kin rather than nonkin when given the choice, and do not help in the absence of kin (Russell and Hatchwell 2001). Kin recognition is apparently based on learned vocal cues (Hatchwell et al. 2001). These results suggest that helping unrelated individuals is costly, so that according to the previous classification, this "cooperative" breeding is altruism rather than cooperation. In contrast, in meerkat colonies, kin discrimination in helping behavior is not obvious, consistent with the conclusion that helping behavior is cooperative rather than altruistic (Clutton-Brock et al. 2001a). In Seychelles warblers (*Acrocephalus sechellensis*), helping behavior has direct benefits (the helpers often have offspring of their own), yet it is preferentially directed to kin (Richardson et al. 2002). Thus, although kin selection may not have been essential to the evolution of helping in this species, it may have caused the bias towards helping kin rather than nonkin.

Overall, the experimental demonstration of kin selection effects due to spatial structure is not easy. The actual fitness costs and benefits of a deviant behavior are difficult to measure, so most arguments are indirect. Suggestive evidence for the importance of localized dispersal comes from the correlation between behaviors expressed towards relatives, kin recognition, and localized dispersal in some species. Attempts at more quantitative analyses (e.g., Creel and Waser 1994) have used pedigree-based measures of relatedness of individuals within demes. However, these will be biased upwards whenever related competitors are found in different demes [implying, for example, that the Q_2 term in equation (7.7) is nonzero]. A better appreciation of kin selection effects under localized dispersal will be possible when data are analyzed consistently with models accounting for localized dispersal.

IMPLICATIONS FOR MODELING APPROACHES

Inclusive Fitness Theory

This chapter has shown how the precise set of techniques defined in the previous chapter translates into inclusive fitness concepts. We considered only the island model, where an inclusive fitness expression of the form $-c+bR$ can be obtained. With the definition of fitness as number of adult offspring, there is no need to extend inclusive fitness measures

with additional terms.

The analysis presented here answers some objections often raised regarding inclusive fitness measures. Most of them are irrelevant as far as local convergence stability is considered. One of these objections is that such measures are appropriate only for "additive" effects on fitness. Hamilton (1964) initially assumed additive effects on fecundity. For convergence stability, the actual assumptions are different, as stated in chapter 6 (see p. 95). This stems from first-order ("weak selection") effects being usually additive even when effects of different actors are not exactly additive, so the results of exact models converge to those of linearized models under weak selection, as has long been recognized (e.g., Uyenoyama and Feldman 1982; Michod 1982; Grafen 1985). It is sometimes thought that one should consider effects of an altered behavior over several generations. However, this would only contribute terms of order δ^2 to the fitness measure, which again must not be taken into account in a convergence stability measure. It is also sometimes thought that the computation of relatedness should take effects of selection into account. Again, this would only contribute terms of order δ^2; ignoring effects of selection on the genetic identity between individuals is sufficient to compute convergence stability. A final, more subtle objection is that relatedness is computed in a stationary model: it is an average over the distribution of allele frequency in a population, while game theoretical measures of fitness should be measured for "rare" alleles. As discussed in the last two chapters, provided the first-order effects will not sustain a stable equilibrium, a suitable convergence stability measure is given by effects on the probability of fixation, and these effects can be expressed in terms of the stationary probabilities of identity.

Therefore, our analysis emphasizes a restrictive interpretation of the inclusive fitness formalism, where relatedness must be computed in a neutral model and is only determined by genealogical relationships. In contrast, Hamilton (1975, p. 141) noticed that relatedness at a locus under selection could be affected by the selection process, and he proposed to include such effects into the inclusive fitness measure. Frank (1998) has also emphasized that genealogy is not the only possible cause of genetic similarity. For example, interspecific mutualisms have been considered as one case where selection can create correlations between phenotypes in different species (Frank 1994a). However, the

idea is "not terribly successful" (Hamilton 2001, p. 102). Indeed, these correlations cause changes in allele frequency that are only of second order in δ. Hence, they are not sufficient to maintain mutualistic interactions against invasion by selfish mutants, as shown in simulations by Doebeli and Knowlton (1998). They found that population structure within each species, rather than correlations between species, could prevent the invasion of selfish mutants and therefore maintain and even initiate the mutualistic interaction.

In conclusion, the main demonstrated quantitative utility of inclusive fitness techniques is to compute local convergence stability conditions, though some extensions of these techniques to the computation of evolutionary stability are possible (Ajar submitted). Hence, the main limitations of inclusive fitness measures are those of convergence stability measures. On the other hand, this restrictive interpretation allows a number of simplifications in computing inclusive fitness. Similar simplifications are made when computing diffusion approximations for changes in allele frequency, as will be shown in chapter 9.

Other Frameworks

There are many equally valid ways of describing changes in allele frequency (Queller 1992; Frank 1997b; Frank 1998, p. 12). Much previous work (e.g., Hamilton 1970; Seger 1981; Grafen 1985; Wade 1985; Frank 1986, 1998) has used the Price equation (Price 1970, 1972; see Frank 1995, for review). A simple but unusual feature of Price's work was to express allele frequency among offspring in terms of the number of A offspring of each parent, rather than in terms of the number of offspring of each genotype. This has been retained to some extent in equation (6.4). Hierarchical forms of the Price equation, distinguishing changes in allele frequencies within and among groups, have been much discussed, starting with Hamilton (1975). We have not used a hierarchical formalism, but it is easy to produce one as follows. One describes fitness as $w = gf$, where g is a focal group's total number of adult offspring, and f is a focal individual's share of this number of offspring. Therefore, the direction of selection is given by

$$\frac{\partial w}{\partial z_\bullet} + R\frac{\partial w}{\partial z_0} = \left(\frac{\partial f}{\partial z_\bullet} + R\frac{\partial f}{\partial z_0}\right) g^\circ + \left(\frac{\partial g}{\partial z_\bullet} + R\frac{\partial g}{\partial z_0}\right) f^\circ, \quad (7.14)$$

where g° and f° are g and f evaluated when all individuals have the same phenotype z_a (in the previous models, these would simply be N and $1/N$).[2] Hamilton (1975) called effects on f and g "individual" and "group" selection effects, respectively. This usage has been widely adopted, but has been criticized on the ground that "individual selection" is the appropriate term for effects on individual fitness w, not on f (Grafen 1984; Dugatkin and Reeve 1994). Note that when $N = 1$, there is only group selection according to this definition (the ∂f terms are zero). Note also that the group selection term is the sum of the direct effect of the focal individual on g and of the effect of neighbors on g, which is nonzero only if relatedness is nonzero.

Wynne-Edwards initiated a long debate by claims such as whenever "group selection" and "individual selection" conflict, "group selection is bound to win" (Wynne-Edwards 1962, p. 20). It is obvious from equation (7.14) that this is not so, whether we interpret individual selection as effects on f or on w. The direction of selection is given neither by effects on f nor on g only.

Definitions of altruism have generated parallel debates (e.g., Reeve 2000), as altruism is often defined by negative effects on f and by positive effects on g (e.g., Wilson et al. 1992; Taylor 1992a). Likewise, "selfish" intragenomic elements are often defined by their positive effects on f (which represents here their frequency among gametes of an individual) and negative effects on g (which represents here the gamete production by the individual).

Of more interest are ecological conditions that allow effects on g to dominate those on f. This question underlies in one way or another

[2]For comparison, the hierarchical form of the Price equation is as follows. Let f_{ij} and p_{ij} be the f value and allele frequency, respectively, for individual j in deme i. Deme averages are noted $f_{i.}$ and $p_{i.}$. Note that $f_{i.} = 1/N$. Let g_i be the g value for deme i. Then the expected change in allele frequency may be written

$$\mathrm{E}(g_i f_{ij} p_{ij}) - \bar{p} = \frac{1}{n_\mathrm{d}} \sum_i g_i f_{i.} p_{ij} + \frac{1}{N_\mathrm{T}} \sum_i g_i \sum_j (f_{ij} - f_{i.}) p_{ij} - \bar{p} \quad (7.15)$$

$$= \mathrm{E}(g_i p_{i.}) - \bar{p} + \frac{1}{n_\mathrm{d}} \sum_i g_i \mathrm{Cov}(f_{ij}, p_{ij}) \quad (7.16)$$

$$= \mathrm{Cov}(g_i, p_{i.}) + \mathrm{E}[\mathrm{Cov}(w_{ij}, p_{ij})], \quad (7.17)$$

where the latter expectation is the average over demes (Price 1972, equation A.17). ∂g terms arise as first-order effects on $\mathrm{Cov}(g_i, p_{i.})$ and ∂f terms as first-order effects on $\mathrm{E}[\mathrm{Cov}(w_{ij}, p_{ij})]$.

much of the current research on cooperative breeding. The example of meerkats emphasized the importance of Allee effects, which, in one form or another, are an alternative to kin selection as an explanation for various social behaviors (Clutton-Brock 2002). Conversely, we expect little selection for cooperative or altruistic behaviors when there is no effect on g. This occurs in particular if the number of juveniles produced by each deme is independent of its genetic composition through some local regulation of juvenile numbers before dispersal. In a model similar to Taylor's (1992b), but with such local regulation, fitness is in direct proportion to the relative proportion of juveniles in the deme (i.e., f), and it is found that necessarily $c = -b$ [see equation (7.23)]. This result points to the fact that any increase in the number of juveniles of the focal is exactly compensated for by a decrease in the number of juveniles of neighbors. Further, no altruism is selected. This result is consistent with the intuitive expectation that altruism is less favored when there is local regulation of juvenile numbers before dispersal.

In population genetics, local regulation of juvenile numbers has been discussed in the context of models of "hard" and "soft" selection. Soft selection refers to the case where selection has no effect on the number of individuals in each deme right before dispersal, and hard selection is the alternative case where the number of individuals right before dispersal is variable (e.g., Christiansen 1975; Wade 1985; Barton 1993; however, these definitions have not always been used consistently). In this respect, Taylor's model is a hard selection model, since the locus under selection affects the number of individuals right before dispersal. The alternative model with local regulation of juvenile numbers is a soft selection model, and appears less favorable to altruism.

APPENDIX: HELPING NEIGHBORS UNDER ISOLATION BY DISTANCE

Here we consider a more general version of the model of helping behavior between neighbors, considered above. We consider a circular lattice (the extention to two dimensions is only a matter of notation). Actors may have different effects on the fecundity of individuals at different distances. The dispersal rate of juveniles at distance k (for $k = 0, \ldots, n_\mathrm{d} - 1$) is m_k. Equation (7.10) for the fitness function generalizes into

$$w = \sum_{k=0}^{n_\mathrm{d}-1} m_k \frac{1 + \sum_{j=\bullet}^{n_\mathrm{d}-1} z_j d_{-j}}{\sum_{l=0}^{n_\mathrm{d}-1} m_{k-l}(1 + \sum_{j=0}^{n_\mathrm{d}-1} z_j d'_{l-j})}, \tag{7.18}$$

with the following notations: $d_\bullet$ is the specific effect of an A-bearing individual on itself, d_0/N is its effect on each individual in the same deme (including itself, so that the total effect on itself is $d_\bullet + d_0/N$), and its effect on each individual in a deme at j steps is d_j/N. Finally, $d'_j \equiv d_j$ except $d'_0 \equiv d_\bullet + d_0$.

Then

$$S = d_\bullet + \sum_j d_{-j} Q_j^\mathrm{R} - \sum_j d'_j \left(\sum_k \sum_l m_k m_{k-l} Q_{l-j}^\mathrm{R} \right). \tag{7.19}$$

The last summation can be simplified, as equation (3.47) implies that at equilibrium $\sum_k \sum_k m_l m_{k-l} Q_{l-j}^\mathrm{R} = Q_{-j}^\mathrm{D}/\gamma$. Expressing the remaining Q^R's as $Q_0^\mathrm{R} = \gamma[Q_0^\mathrm{D} + (1 - Q_0^\mathrm{D})/N]$ and $Q_k^\mathrm{R} = \gamma Q_k^\mathrm{D}$ otherwise, one obtains

$$\gamma S = d_\bullet(\gamma - Q_0^\mathrm{D}) + \gamma d_0 \frac{1 - Q_0^\mathrm{D}}{N} - (1 - \gamma) \sum_j d_j Q_{-j}^\mathrm{D}. \tag{7.20}$$

Then, using equation (3.68),

$$\phi = \lim_{\mu \to 0} \frac{S}{1 - Q_0^\mathrm{D}} = d_\bullet + \frac{d_0}{N} - \frac{1}{N_\mathrm{T}} \left(d_\bullet + \sum_{j=0}^{n-1} d_j \right). \tag{7.21}$$

Here $d_\bullet + d_0/N$ is the effect of an A individual on its own fecundity. If there are only local interactions ($d_j = 0$ for large j), we obtain in the limit $n_\mathrm{d} \to \infty$ that the direction of selection is determined by this effect. Thus, effects on others' fecundity have no consequence. This result was found by Wilson et al. (1992) in simulations and also by

Taylor (1992b) in the case $N = 1$. On the other hand, if interactions are not local, the effects of actors on others' fecundity matter. For example, if a behavior increases everyone's fecundity by the same factor, then this behavior should be neutral. Indeed, when $d_\bullet = 0$ and $d_j = d$ otherwise, $\phi = 0$.

"Soft" selection. Here we make the same assumptions as those above, except that the number of juveniles produced by each deme is independent of its genetic composition. Using the same notations as in equation (7.18), fitness is

$$w = \frac{1 + \sum_{j=\bullet}^{n_\mathrm{d}-1} z_j d_{-j}}{1 + \sum_{j=0}^{n_\mathrm{d}-1} z_j d'_{-j}}. \tag{7.22}$$

There is no need for distinct terms for migrants in different demes, because the number of competing juveniles in each deme is not affected by the trait values. S immediately simplifies to $d_\bullet(1 - Q_0^\mathrm{R}) = d_\bullet(1 - Q_0^\mathrm{D})(N-1)/N$, and

$$\phi = \frac{N-1}{N} d_\bullet. \tag{7.23}$$

Using the notation $-C = d_\bullet$ for effect on number of juveniles and c, b for effects on fitness, then $c = -b = C(N-1)/N$ and $\phi = -c$. The factor $1 - 1/N$ accounts for the restraint of fitness effects due to regulation before dispersal (when $N = 1$, there is no selection at all).

Maruyama (1974, 1983) found a result that implies $\phi = d_\bullet$, which differs from the two cases considered above. However, his 1974 model has a peculiar life cycle, as best seen from the fact that it has neither selection nor drift when $N = 1$; and his 1983 analysis neglected an effect of order $1/N$, as will be discussed in chapter 9 (p. 151).

CHAPTER EIGHT

Diploidy (and Sex)

The previous chapters have dealt with haploid models, where selection and dispersal concern only a haploid stage. Many organisms have alternating haploid and diploid stages, where in addition the diploid stage may be the longer lasting one. This certainly forces us to consider the consequences of diploidy on genetic structure in general. In this chapter, we will focus on two specific issues without attempting a complete discussion of all consequences of diploidy.

A first specific concern underlying studies of the population structure of diploid individuals is inbreeding depression, that is, a decreased fitness of matings between related individuals. Various traits may have evolved as a means to avoid inbreeding depression (Darwin 1877, 1984; Thornhill 1993). In particular, it selects against self-fertilization (selfing). Patterns of population structure have been used to quantify selfing in natural populations or to demonstrate some of the forces acting on its evolution (Brown 1979; Jarne and Charlesworth 1993; Charlesworth and Pannell 2001). This chapter will reanalyze a model of population structure on which these studies may be based and will consider how to take selfing into account in the quantitative interpretation of spatial patterns.

Most statistical analyses are based on the "hierarchical F-statistics" defined in chapter 2, F_{ST} and F_{IS}, where F_{IS} is used to measure the heterozygote deficit. In particular, they use the relationship $F_{\mathrm{IS}} = s/(2 - s)$, as a function of the selfing rate s. Although alternatives to statistical analysis by F-statistics have been conceived (Nordborg and Donnelly 1997), they rely on the same relationship. $F_{\mathrm{IS}} = s/(2 - s)$ holds for models ignoring spatial population structure, and may be derived as an approximation under broader assumptions. As will be shown, it is nevertheless somewhat inaccurate in populations with highly localized

dispersal, as may occur in some plant populations. This adds to a list of complications that arise in interpreting F_{IS} values.

Next, some specifics of selection in diploid populations will be considered: dominance and parent-offspring conflict. Nothing here is specific to spatially subdivided populations, so we will focus on examples showing how the methods demonstrated in the previous chapters may be safely used for analyzing these phenomena.

POPULATION STRUCTURE OF DIPLOID POPULATIONS

We will consider only models of hermaphrodite (monoecious) populations here. Expressions for differentiation and for heterozygote deficiencies will be given for a model of plant population structure, with pollen dispersal, seed dispersal, and an intermediate level of selfing. Similar models have been considered by Maruyama and Tachida (1992), Tachida and Yoshimaru (1996), and Wang (1997a), who all neglected effects of order $1/N$ on F_{IS} (there are additional discrepancies with Maruyama and Tachida's analysis). The case without pollen dispersal was already considered by Nagylaki (1983). It is applicable to animal or plant populations without dispersal of the haploid stage.

The patterns of differentiation have some intuitive interpretation. The effect of dispersal on differentiation is given by the dispersal rate of a gene lineage through one generation, which is the average of dispersal rates of maternal and paternal genes (Wright 1946). In plant populations, this differs from the average of dispersal rates of pollen and of seeds because the father's genes disperse when the seeds disperse (Crawford 1984). A gene lineage has probability $1/2$ to be transmitted through pollen, so we count half the dispersal rate of pollen. All gene lineages are transmitted through seeds, so we add the dispersal rate of seeds. Thus, given the dispersal rate m_{P} of pollen and the dispersal rate m_{S} of seeds, in the island model, approximate results can be expressed in terms of the overall dispersal rate $m = m_{\mathrm{S}} + m_{\mathrm{P}}/2$. Under isolation by distance, the increase of differentiation with distance is likewise a function of $\sigma^2 = \sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{P}}^2/2$, where σ_{S}^2 and σ_{P}^2 are the mean squared dispersal distances of seeds and pollen, respectively.

Analysis of Pollen and Seed Dispersal

We consider a structured population with demes of N diploid hermaphrodites. Events occur in the following order: mutation, pollen dispersal, fertilization, seed dispersal, and competition among seeds in each deme, such that only N seeds develop into adults.

DIFFERENTIATION

Exact recursions can be constructed as follows. Let Q_{w} be the probability of identity of genes within individuals; let $Q_{\mathbf{r}}$ be the probability of identity of genes in different individuals or seeds at distance $\mathbf{r}$. We denote Q and Q' as the identities between seeds after dispersal and before competition in two successive generations, Q^{J} as the identity between seeds before seed dispersal, and Q^{A} as the identity among adults (as in the haploid life cycle, see fig. 3.1).

Let $p_{\mathrm{S}:\mathbf{r}}$ be the probability of seed dispersal at distance $\mathbf{r}$; let $p_{\mathrm{P}:\mathbf{r}}$ be the probability of pollen dispersal at distance $\mathbf{r}$. Let s be the selfing rate. Note that selfing implies no dispersal of pollen, and hence the probability that pollen does not disperse is equal to s plus a term $p_{\mathrm{P}:\mathbf{0}} - s$ for outcrossed pollen within a deme.

Seed dispersal. For genes in different individuals, identity after and before seed dispersal are related by

$$Q'_{\mathbf{r}} = \sum_{\mathbf{r}_1} \sum_{\mathbf{r}_2} p_{\mathrm{S}:\mathbf{r}_1} p_{\mathrm{S}:\mathbf{r}_2} Q^{\mathrm{J}}_{\mathbf{r}-\mathbf{r}_1+\mathbf{r}_2}. \tag{8.1}$$

This is a general equation for homogeneous dispersal. In the infinite island model, we may adopt the convention that the identity of genes in different demes is zero, and write

$$Q'_{\mathbf{0}} = (1 - m_{\mathrm{S}})^2 Q^{\mathrm{J}}_{\mathbf{0}}. \tag{8.2}$$

This result has a direct parallel under isolation by distance.[1]

[1] Let ψ_{S} be the Fourier transform of the distribution of seed backwards dispersal distance. Let $\mathcal{Q}'$ and $\mathcal{Q}^{\mathrm{J}}$ be Fourier transforms of the Q''s and Q^{J}'s. Then equation (8.1) implies

$$\mathcal{Q}' = \psi_{\mathrm{S}}^2 \mathcal{Q}^{\mathrm{J}}. \tag{8.3}$$

Thus, equations for the infinite island model translate into expressions for isolation by distance by the substitutions $Q'_0 \to \mathcal{Q}'$, $Q^{\mathrm{J}}_{\mathbf{0}} \to \mathcal{Q}^{\mathrm{J}}$, and $1 - m_{\mathrm{S}} \to \psi_{\mathrm{S}}$.

Pollen dispersal. To relate the identity of genes among seeds before seed dispersal to the identity between gametes before pollen dispersal, we consider three cases: both genes are from ovules (and we assume there is no dispersal of ovules at this stage), both from pollen, or one of each. Let $\Delta Q \equiv (1 + Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}$ be the gain in identity resulting when two gene lineages come from the same individual rather than from different individuals in the same deme. Using the notation

$$[\mathbf{j} = \mathbf{l}] \equiv \begin{cases} 1 & \text{if } \mathbf{j} = \mathbf{l} \\ 0 & \text{otherwise,} \end{cases} \tag{8.4}$$

we have

$$\begin{aligned} Q^{\mathrm{J}}_{\mathbf{r}} =& \frac{\gamma}{4}\left(Q_{\mathbf{r}} + [\mathbf{r} = \mathbf{0}]\frac{\Delta Q}{N}\right) \\ &+ \frac{\gamma}{4}\sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2} p_{\mathrm{P}:\mathbf{r}_1} p_{\mathrm{P}:\mathbf{r}_2}\left(Q_{\mathbf{r}-\mathbf{r}_1+\mathbf{r}_2} + [\mathbf{r} - \mathbf{r}_1 + \mathbf{r}_2 = \mathbf{0}]\,\frac{\Delta Q}{N}\right) \\ &+ \frac{\gamma}{2}\sum_{\mathbf{r}_1} p_{\mathrm{P}:\mathbf{r}_1}\left(Q_{\mathbf{r}-\mathbf{r}_1} + [\mathbf{r} - \mathbf{r}_1 = \mathbf{0}]\frac{\Delta Q}{N}\right). \end{aligned} \tag{8.5}$$

The first term is for the case where both genes come from ovules, the second for the case where both come from pollen, and the last is for the case when only one of the two genes comes from pollen. Here the gain in identity ΔQ substitutes for the gain in identity $1 - Q_{\mathbf{0}}$ in the haploid model when two gene lineages coalesce. Computation of Q_{w} is considered on p. 133.

In the infinite island model, this reduces to

$$Q^{\mathrm{J}}_{\mathbf{0}} = \gamma\left(\frac{2 - m_{\mathrm{P}}}{2}\right)^2\left[Q_{\mathbf{0}} + \frac{1}{N}\left(\frac{1 + Q_{\mathrm{w}}}{2} - Q_{\mathbf{0}}\right)\right]. \tag{8.6}$$

Again, there is a parallel result under isolation by distance.[2]

Complete life cycle. From equations (8.2) and (8.6),

$$\mathcal{Q}' = (1 - m_{\mathrm{S}})^2 \mathcal{Q}^{\mathrm{J}}_{\mathbf{0}} \tag{8.8}$$

$$= \gamma(1 - m_{\mathrm{S}})^2\left(1 - \frac{m_{\mathrm{P}}}{2}\right)^2\left[Q_{\mathbf{0}} + \frac{1}{N}\left(\frac{1 + Q_{\mathrm{w}}}{2} - Q_{\mathbf{0}}\right)\right], \tag{8.9}$$

[2]Let ψ_{P} be the Fourier transform of the distribution of pollen dispersal distance. Then

$$\mathcal{Q}^{\mathrm{J}} = \gamma\left(\frac{1 + \psi_{\mathrm{P}}}{2}\right)^2\left[\mathcal{Q} + \frac{1}{N}\left(\frac{1 + Q_{\mathrm{w}}}{2} - Q_{\mathbf{0}}\right)\right]. \tag{8.7}$$

This is equation (8.6), with the substitutions $Q_{\mathbf{0}} \to \mathcal{Q}$, $Q^{\mathrm{J}}_{\mathbf{0}} \to \mathcal{Q}^{\mathrm{J}}$, and $1 - m_{\mathrm{P}} \to \psi_{\mathrm{P}}$.

which yields at equilibrium

$$Q_{\mathbf{0}} = \frac{1}{N}\frac{\gamma\psi^2}{1-\gamma\psi^2}\left(\frac{1+Q_{\mathrm{w}}}{2} - Q_{\mathbf{0}}\right) \tag{8.10}$$

for $\psi \equiv (1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2)$.[3]

QUANTITATIVE INTERPRETATION

These results are consistent with our previous intuitive argument. Results for $F_{\mathrm{ST}\mathbf{r}} = (Q_{\mathbf{0}} - Q_{\mathbf{r}})/(1-Q_{\mathbf{r}})$ in the haploid model translate into results for $\rho_{\mathbf{r}} \equiv (Q_{\mathbf{0}} - Q_{\mathbf{r}})/[(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{r}}]$ (the "correlation between outcrossed mates"; Waller and Knight 1989). In the infinite island model, this yields equation (8.16) below. In a two-dimensional habitat, the increase of

$$\frac{\rho_{\mathbf{r}}}{1-\rho_{\mathbf{r}}} = \frac{Q_{\mathbf{0}} - Q_{\mathbf{r}}}{(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}} \tag{8.12}$$

with logarithm of distance is $1/(2N\pi\sigma^2)$ for $\sigma^2 = \sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{P}}^2/2$.[4] This suggests that we may use $\rho_{\mathbf{r}}/(1-\rho_{\mathbf{r}})$ to estimate $D\sigma^2$ much as we can use $F_{\mathrm{ST}\mathbf{r}}/(1-F_{\mathrm{ST}\mathbf{r}})$ [$\mathtt{a}(\mathbf{r})$ in chapter 3] to estimate $D\sigma^2$ in gametic-dispersal models.

The above results remain valid for $N = 1$ because $Q_{\mathbf{0}}$ is by definition $Q_{\mathbf{0}}^{\mathrm{D}}$, the identity between genes in two seeds competing for the same site. However, for $N = 1$, $Q_{\mathbf{0}}^{\mathrm{D}}$ cannot be estimated unless seeds are sampled before competition. A difficulty then appears if we wish to analyze continuous populations on the basis of the lattice model, as was discussed on p. 42. The same difficulty arises when estimating F_{IS}. In practice, the lattice model is not an exact description of any

[3] Under isolation by distance, one has

$$\mathcal{Q} = \frac{1}{N}\frac{\gamma\psi^2}{1-\gamma\psi^2}\left(\frac{1+Q_{\mathrm{w}}}{2} - Q_{\mathbf{0}}\right) \tag{8.11}$$

for $\psi \equiv \psi_{\mathrm{S}}(1+\psi_{\mathrm{P}})/2$.

[4] $\rho_{\mathbf{r}}/(1-\rho_{\mathbf{r}})$ is given by the expression for $\mathtt{a}(\mathbf{r})$ in the haploid model [equation (3.31)], which is

$$\frac{1}{N}\left[\mathcal{L}_{\mathbf{0}}\left(\frac{\gamma\psi^2}{1-\gamma\psi^2}\right) - \mathcal{L}_{\mathbf{r}}\left(\frac{\gamma\psi^2}{1-\gamma\psi^2}\right)\right], \tag{8.13}$$

again with $\psi = \psi_{\mathrm{S}}(1+\psi_{\mathrm{P}})/2$.

population, but assuming it well describes differentiation in the case in which individuals can settle anywhere in a continuous space, a good substitute for $Q_{\mathbf{0}}$ should be the identity between close seeds. With very limited gametic dispersal, a substitute for $Q_{\mathbf{0}}$ would be the probability of identity between different mates, which is also Q_{w} for outcrossed progeny if these can be identified.

The denominator of $\rho_{\mathbf{r}}/(1-\rho_{\mathbf{r}})$ may be viewed as the gain in identity expected each time two gene lineages occur in the same individual, relative to the identity of genes from different individuals in the same deme. The same was true of the denominator of $\mathsf{a}(\mathbf{r})$, that is, of $F_{\mathrm{ST}\mathbf{r}}/(1-F_{\mathrm{ST}\mathbf{r}})$, assuming the haploid models of chapter 3. As a result, in these different contexts, the parameters yield information about the process of dispersal among demes not confounded by other effects. This approach lends itself to generalizations in models for polyploid populations (Ronfort et al. 1998), "hierarchical" island models (Slatkin and Voelm 1991), or more general kinds of within-deme genetic structure (chapter 9).

Expected values of F_{ST} may be deduced from the relationship

$$\frac{F_{\mathrm{ST}}}{1-F_{\mathrm{ST}}} = \frac{1+F_{\mathrm{IS}}}{2}\frac{\rho_{\mathbf{r}}}{1-\rho_{\mathbf{r}}}, \tag{8.14}$$

which may be checked directly from definitions of the parameters in terms of probabilities of identity (Waller and Knight 1989). Here F_{IS} is the usual measure of heterozygote deficit within demes:

$$F_{\mathrm{IS}} = \frac{Q_{\mathrm{w}}-Q_0}{1-Q_0}. \tag{8.15}$$

The factor $(1+F_{\mathrm{IS}})/2$ in equation (8.14) may be understood as an effect on the rate of coalescence of gene lineages relative to the haploid model with N individuals (equivalently, this rate is increased by $1+F_{\mathrm{IS}}$ relative to a diploid randomly mating population of N individuals).

In the infinite island model, $\rho_{\mathbf{r}} \equiv \rho$ independent of distance, and equation (8.10) implies that

$$\frac{\rho}{1-\rho} = \frac{Q_0}{(1+Q_{\mathrm{w}})/2-Q_0} = \frac{1}{N}\frac{[(1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2)]^2}{1-[(1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2)]^2}. \tag{8.16}$$

For low dispersal rates m_{P} and m_{S}, $(1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2) \approx 1-m$ for $m = m_{\mathrm{S}}+m_{\mathrm{P}}/2$, so $\rho \approx 1/(1+2Nm)$ and $F_{\mathrm{ST}} \approx 1/[1+4mN/(1+F_{\mathrm{IS}})]$ (Nordborg 1997). This result shows that F_{ST} is increased (and within-deme gene diversity decreased) by selfing, both because m is

reduced (there is no dispersal when there is selfing) and because $F_{\rm IS}$ is increased. The reduction of diversity within demes may be viewed as a reduction in effective size (see next chapter) and is a well-known observation (e.g., Brown 1979; Hamrick and Godt 1990, 1996; Charlesworth and Pannell 2001, for sequence data). Likewise, selfing species tends to be more structured (Heywood 1991; Hamrick and Godt 1990, 1996).

Analysis of Heterozygote Deficit

A well-known cause of the heterozygote deficit is the "Wahlund effect" (Wahlund 1928). If the population shows some spatial differentiation in allele frequency, but genotypic frequencies obey Hardy-Weinberg frequencies within each subpopulation, it immediately follows that the probability of identity of genes within individuals will be higher than the probability of identity of genes sampled in different subpopulations. Then, if a sample contains individuals from several subpopulations, $F_{\rm IS}$ will be positive. This effect seems little more than a sampling artifact, although it is sometimes less trivial, a similar effect being caused by variation in allele frequencies among different cohorts in age-structured populations (a temporal Wahlund effect, Christiansen 1988). Other forces that generate positive $F_{\rm IS}$ are selfing and, to some extent, the dispersal of diploid individuals, as the following expressions for $F_{\rm IS}$ show.

With probability s, two genes within an offspring are copies of genes within a selfing parent, in which case they are identical with probability $(1 + Q_{\rm w})/2$. With probability $p_{{\rm P}:\mathbf{0}} - s$, they are copies of genes from different individuals within a deme, in which case they are identical with probability $Q_{\mathbf{0}}$. Finally, dispersal of pollen yields terms $p_{{\rm P}:\mathbf{r}} Q_{\mathbf{r}}$ for $\mathbf{r} \neq \mathbf{0}$. So

$$Q'_{\rm w} = \gamma \left[s \left(\frac{1 + Q_{\rm w}}{2} - Q_{\mathbf{0}} \right) + \sum_{\mathbf{r}=\mathbf{0},\ldots} p_{{\rm P}:\mathbf{r}} Q_{-\mathbf{r}} \right], \qquad (8.17)$$

so that at equilibrium,

$$\frac{Q_{\rm w}}{(1 + Q_{\rm w})/2 - Q_{\mathbf{0}}} = \gamma \left[s + \sum_{\mathbf{r}} p_{{\rm P}:\mathbf{r}} \frac{Q_{-\mathbf{r}}}{(1 + Q_{\rm w})/2 - Q_{\mathbf{0}}} \right]. \qquad (8.18)$$

Then, using equation (8.10), in the infinite island model

$$\frac{2F_{\mathrm{IS}}}{1+F_{\mathrm{IS}}} = \frac{Q_{\mathrm{w}} - Q_{\mathbf{0}}}{(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}} = \gamma s + \frac{1}{N}\frac{\gamma\psi^2[\gamma(1-m_{\mathrm{P}})-1]}{1-\gamma\psi^2}, \tag{8.19}$$

where $\psi \equiv (1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2)$.[5] Henceforth we consider the low-mutation limit ($\gamma \to 1$) and several cases for dispersal, including no pollen dispersal ($m_{\mathrm{P}} = 0$) and more pollen than seed dispersal ($m_{\mathrm{P}} \gg m_{\mathrm{S}}$).

Without pollen dispersal,

$$\frac{2F_{\mathrm{IS}}}{1+F_{\mathrm{IS}}} = s \Leftrightarrow F_{\mathrm{IS}} = \frac{s}{2-s}, \tag{8.23}$$

whatever the distribution of seed dispersal is (i.e., even under isolation by distance). Thus, in this case, it is possible to estimate the selfing rate from F_{IS}, whatever the dispersal rates are between subpopulations.

Equation (8.23) is a well-known result (Wright 1969, p. 195), which can be understood as follows. Either an individual was produced from outcrossing (with probability $1-s$) or it was produced by selfing, in which case its two genes coalesce in their selfed parent (with probability $s/2$) or are copies of the two different genes from the selfed parent (with probability $s/2$). Thus, the probability that the two genes coalesce before an outcrossing event occurs among ancestors is $s/2 + (s/2)^2 + \cdots = (s/2)/(1-s/2) = F_{\mathrm{IS}}$. That is, F_{IS} is the probability of coalescence before the first outcrossing event.

[5] Under isolation by distance, equation (8.18) can be written

$$\frac{Q_{\mathrm{w}}}{(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}} = \gamma s + \frac{1}{N}\mathcal{L}_{\mathbf{0}}\left(\frac{\gamma^2\psi^2\psi_{\mathrm{P}}}{1-\gamma\psi^2}\right). \tag{8.20}$$

To get this result, we interpret Fourier transforms as generating functions, and we use the relationship between convolutions and generating functions (see Appendix A). $\sum_{\mathbf{r}} p_{\mathrm{P}:\mathbf{r}} Q_{-\mathbf{r}}/[(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}]$ is a term of the convolution of the pollen dispersal rates and of the ratios $Q_{-\mathbf{r}}/[(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}]$. The generating function of such convolutions is the product of the generating function ψ_{P} of pollen dispersal rates and of the generating function

$$\frac{1}{N}\frac{\gamma\psi^2}{1-\gamma\psi^2} \tag{8.21}$$

of the latter ratios [see equation (8.11)]. Then equation (8.19) generalizes into

$$\frac{Q_{\mathrm{w}} - Q_{\mathbf{0}}}{(1+Q_{\mathrm{w}})/2 - Q_{\mathbf{0}}} = \gamma s + \frac{1}{N}\mathcal{L}_{\mathbf{0}}\left(\frac{\gamma\psi^2(\gamma\psi_{\mathrm{P}}-1)}{1-\gamma\psi^2}\right). \tag{8.22}$$

Heterozygote excess. From equation (8.10), in the infinite island case equation (8.22) reduces to

$$\frac{2F_{\mathrm{IS}}}{1+F_{\mathrm{IS}}} = s - m_{\mathrm{P}}\frac{\rho}{1-\rho}. \tag{8.24}$$

Therefore, if there is obligate outcrossing ($s = 0$), F_{IS} is slightly negative. Nevertheless, the magnitude of this effect is generally small, of order $1/N$, and not increased by population structure relative to a panmictic population of size N. Here, for $N > 1$, the lowest possible F_{IS} is

$$-1/(2N+1). \tag{8.25}$$

This bound is attained for an obligate outcrossing population in a single deme. It then directly follows from the recursions $Q'_{\mathrm{w}} = \gamma Q_0$ and $Q'_0 = \gamma\{Q_0+[(1+Q_{\mathrm{w}})/2-Q_0]/N\}$. It is also attained in the following conditions in the infinite island model. If the total dispersal is low, so that $(1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2) \approx 1$, and if $m_{\mathrm{P}} \gg m_{\mathrm{S}}$, so that $m_{\mathrm{P}}/\{1-[(1-m_{\mathrm{S}})(1-m_{\mathrm{P}}/2)]^2\} \approx 1$,

$$\frac{2F_{\mathrm{IS}}}{1+F_{\mathrm{IS}}} \approx s - \frac{1}{N} \Leftrightarrow F_{\mathrm{IS}} \approx \frac{s-1/N}{2-(s-1/N)}. \tag{8.26}$$

This yields the lower bound given above. Note that it is attained for $m_{\mathrm{S}} = 0$, $m_{\mathrm{P}} \to 0$, and without selfing, that is, when there is obligate outcrossing within the deme. Large negative F_{IS} values (partially compounded by Wahlund effects) may be attained when all reproduction is by outcrossing with a few close neighbors and when there is very little seed dispersal.[6] These conditions are not unrealistic; the herb *Silene dioica* provides an example (Ingvarsson and Giles 1999). Likewise, some heterozygote excess is expected in mammal populations structured in small social groups (e.g., Wang 1997b; Storz et al. 2001). Nevertheless, although limited dispersal is common in plants (Levin and Kerster 1974), in most cases dispersal may be large enough (relative to density) that little effect of spatial structure on F_{IS} values is expected.

[6]When $N = 1$, there cannot be outcrossing within the deme. F_{IS}, as defined by equation (8.15), can still be computed in terms of the identity $Q_{\mathbf{0}}^{\mathrm{D}}$ between seeds before competition. Low F_{IS} values are again obtained, in the absence of seed dispersal, by outcrossing with the closest possible individuals, that is, in a stepping stone model. For example, in the one-dimensional stepping stone model, evaluation of equation (8.22) shows that the minimum F_{IS} value is $1-\sqrt{2}$, attained for $m_{\mathrm{P}} = 1$.

In a panmictic population with separate sexes, the value of F_{IS} is $-1/(2N-1)$, and this is also a lower bound for a hermaphrodite population with gametic dispersal only (see equation 3 in Rousset 1996). Such results have long been noticed (Kimura and Crow 1963) and somewhat misinterpreted. Random differences in allele frequencies between mothers and fathers tend to create heterozygote excesses in their offspring. The negative F_{IS} within demes is therefore expected as a result of matings with individuals from other demes in a subdivided population, or even in a small panmictic population with separate sexes. The alternative explanation (Kimura and Crow 1963) of this result as a consequence of the "discreteness of the possible numbers of different genotypes in a finite population" is not valid, as it should equally apply to a finite hermaphrodite population with random selfing. In the latter case, however, $F_{\mathrm{IS}} = 0$, as expected from the fact that allele frequencies are identical in mothers and fathers. Nor should the lower bound for F_{IS} be confused with Levene's (1949) result, which gives the expected number of heterozygotes for an allele in a sample, given this allele is in frequency p in the sample, as

$$2p(1-p)\frac{2N}{2N-1}. \tag{8.27}$$

This result actually holds when $F_{\mathrm{IS}} = 0$, since, given p, equation (8.27) implies that $Q_{\mathrm{w}} = Q_0$. They are both $\left[-1 + N(2 - 2p + p^2)\right]/(2N-1)$.

Joint Effects of Selfing and Selection on Population Structure

While the expected effects of selfing on diversity and population structure are qualitatively confirmed by observations from natural populations, it has been argued that there are some consistent quantitative deviations from such predictions. In mostly self-fertilizing species, F_{IS} is often lower than $s/(2-s)$ (Brown 1979). This may in part be explained by pollen dispersal, as shown by equation (8.24), but explanations assuming a role for selection have been considered (Brown 1979). In particular, heterozygotes at one locus tend to be heterozygotes at other loci, because the genealogies at different loci are correlated within an individual (identity disequilibrium: Bennett and Binet 1956; Hartl and Clark 1989). Then selection against deleterious homozygotes at some

loci reduces the frequency of homozygotes at neutral loci. This phenomenon, described as associative overdominance (e.g., David 1998), decreases F_{IS} values at neutral loci.

Recombination is not very efficient in reducing gametic disequilibria in highly selfing populations (Nordborg 2000), so, for example, spatially variable selection, increasing F_{ST} values at a selected locus, will also tend to increase F_{ST} values at linked neutral loci, particularly in selfing species. Selection against deleterious mutations (background selection, Charlesworth et al. 1993) also tends to increase F_{ST} values (Nordborg 1997). This result may be understood as a reduction of effective size (see next chapter) resulting from variation in fitness of individuals with different numbers of deleterious mutations (see also Santiago and Caballero 1998; Stephan et al. 1999). Both effects were considered in simple scenarios by Charlesworth et al. (1997), but their expected magnitude in natural populations is difficult to appreciate.

SELECTION IN SEXUAL DIPLOID POPULATIONS

An obvious complication arising in diploid populations is dominance. Another important consequence is parent-offspring conflict (Trivers 1985, chapter 7; Mock and Parker 1998). It results from sexual reproduction rather than simply from "diploidy" (i.e., a predominantly diploid life cycle). This conflict is characterized by different directions of selection on a trait, whether it is under the control of a parent's or an offspring's genotype. It is a conflict between an offspring and potential future offspring of its parents, and may be seen to result from the difference in relatedness of an offspring with itself versus with its siblings. From a mother's perspective, all offspring may be equivalent, but from an offspring perspective, its own fitness is more important than that of its siblings.

Parent-offspring conflicts, or more general forms of conflicts between relatives within families, have various implications; they have been considered particularly in the context of the evolution of sex ratios and of eusociality in haplodiploid insect colonies, and (legitimately) not in a spatial context (e.g., Bourke and Franks 1995; Crozier and Pamilo 1996; Mock and Parker 1998). These subjects will not be reviewed here. Since conflicts between relatives are not specific to spatially subdivided populations, we treat them briefly, mainly showing

how they can be analyzed by the direct fitness approach.

Parent and Offspring Control

In the following (as throughout this book), we take a gene-centered approach, where each parental gene lineage is taken as the focal "adult": a diploid population is essentially a haploid population with an additional level of structure. Thus, we can use the direct fitness formalism to measure the fitness of a gene as a function of the homologous gene in the same individual and of genes in other individuals. This leads us to consider the identity of genes within and among individuals. In particular, when the actor is the focal parent, the relevant identity coefficient is the average identity between the gene lineage and the two genes of the focal parent; in general, this identity is not 1, in contrast to the haploid models.

CONTROL BY PARENTS

Here we consider a parental control of phenotype and no selection on the gametic stage. As a concrete example, we consider the evolution of seed dispersal under the hermaphroditic plant life cycle previously considered (see p. 129). We further assume that allelic effects are semidominant, that outcrossing is obligate, and that all mating is within the deme, so all dispersal is after mating (no dispersal of pollen).

Consider the fitness function for selection on dispersal in a haploid population [equation (6.24), with $n_{\mathrm{d}} \to \infty$]:

$$w(x_\bullet, x_0, x_1) = \frac{1 - x_\bullet}{1 - x_0 + (1 - c_{\mathrm{d}})x_1} + \frac{x_\bullet(1 - c_{\mathrm{d}})}{1 - c_{\mathrm{d}}x_1} \tag{8.28}$$

and the following variables: $z_\bullet$ for the focal individual's phenotype, z_0^{R} for the average phenotype of individuals in the same patch including the focal individual, and z_1 for a different patch. The fitness of each maternal gene is

$$\frac{1 - z_\bullet}{1 - z_0^{\mathrm{R}} + (1 - c_{\mathrm{d}})z_1} + \frac{z_\bullet(1 - c_{\mathrm{d}})}{1 - c_{\mathrm{d}}z_1} = w(z_\bullet, z_0^{\mathrm{R}}, z_1), \tag{8.29}$$

and the fitness of each paternal gene is

$$\frac{1 - z_{\mathrm{m}}}{1 - z_0^{\mathrm{R}} + (1 - c_{\mathrm{d}})z_1} + \frac{z_{\mathrm{m}}(1 - c_{\mathrm{d}})}{1 - c_{\mathrm{d}}z_1} = w(z_{\mathrm{m}}, z_0^{\mathrm{R}}, z_1), \tag{8.30}$$

where $z_{\rm m}$ is the phenotype of a female mate of the focal male. Since the father has no control, $z_\bullet$ appears nowhere in the fitness of paternal genes.

The fitness function, describing gene transmission through both maternal and paternal ways, may be written as $w = [w(z_\bullet, z_0^{\rm R}, z_1) + w(z_{\rm m}, z_0^{\rm R}, z_1)]/2$. In the notations of p. 129, the identity between a gene lineage and the focal parent (phenotype $z_\bullet$) is $(1 + Q_{\rm w})/2$, the identity between mates (phenotype $z_{\rm m}$) is Q_0 under obligate outcrossing, and phenotype $z_0^{\rm R}$ yields identity

$$Q_0^{\rm R} \equiv \frac{1}{N}\left(\frac{1+Q_{\rm w}}{2} + (N-1)Q_0\right). \tag{8.31}$$

Hence,

$$S = \frac{1}{2}\left[\frac{\partial w}{\partial x_\bullet}\left(\frac{1+Q_{\rm w}}{2} + Q_0\right) + 2\frac{\partial w}{\partial x_0}Q_0^{\rm R}\right] \tag{8.32}$$

in terms of $w(x_\bullet, x_0, x_1)$ as given by equation (8.28).

CONTROL BY OFFSPRING

With an offspring control of the phenotype, fitness can take different values, w_{j1} and w_{j2}, depending on which parental allele was received by an offspring. Then expected allele frequency can be written

$$\mathrm{E}[p'|\mathbf{p}] = \frac{1}{2N_{\rm T}}\sum_{j=1}^{N_{\rm T}}\sum_{l=1,2} p_{jl}w_{jl}, \tag{8.33}$$

which is equation (6.4), except that it is written at the level of the gene lineage jl rather than of the deme. p_{jl} is the allele frequency (0 or 1) for lineage l of individual j.

This expression tells us how to compute relatedness. Consider the evolution of dispersal in the same model as above, but under offspring control. Let $z_\bullet$ be the average phenotype of juveniles bearing a focal parental gene lineage; let z_0' be the average phenotype of all offspring born in the same deme (the offspring of parents in the same deme as the focal gene lineage), including the focal individual; and let z_1 be the average phenotype in a different deme. The same indices will be used for the Q's. The fitness of a parental lineage (maternal or paternal) is

$$\frac{1-z_\bullet}{1-z_0'+(1-c_{\rm d})z_1} + \frac{z_\bullet(1-c_{\rm d})}{1-c_{\rm d}z_1} = w(z_\bullet, z_0', z_1). \tag{8.34}$$

We view juvenile offspring of a focal "parent" (gene lineage) as actors on the fitness of this gene lineage. In particular, the factor of $\partial w/\partial z_\bullet$ must be deduced from the expectation $\mathrm{E}(p_{jl}\bar{p}_\bullet)$ for the focal gene lineage jk and for the frequency $\bar{p}_\bullet$ in the juveniles bearing the focal parental gene lineage. The identity between the two genes within offspring is written Q_w as before. Hence, the identity to be considered for this $\mathrm{E}(p_{jl}\bar{p}_\bullet)$ term is $(1+Q_\mathrm{w})/2$. Likewise, the identity between offspring from parents in the same deme is $\{[(1+Q_\mathrm{w})/2+Q_0]/2+(N-1)Q_0\}/N$ (first term for brothers and sisters that did not disperse, second for nondispersing juveniles of other pairs in the deme). Hence,

$$S=\frac{\partial w}{\partial x_\bullet}\frac{1+Q_\mathrm{w}}{2}+\frac{\partial w}{\partial x_0}\frac{[(1+Q_\mathrm{w})/2+Q_0]/2+(N-1)Q_0}{N} \tag{8.35}$$

in terms of $w(x_\bullet,x_0,x_1)$ as given again by equation (8.28).

Dominance

Consider, for example, parental control of the trait. The expected frequency $\mathrm{E}(\bar{p}')$ in the population in the next generation can be written as $\mathrm{E}(\bar{p}'|\mathbf{p})=\sum_{j=1}^{N_\mathrm{T}}p_jw_j/N_\mathrm{T}$, where p_j is the allele frequency (0, $1/2$, or 1) within individual j, and w_j is its expected number of adult offspring. By the same argument as in chapter 6,

$$\begin{aligned}\frac{d\mathrm{E}(\bar{p}'|\mathbf{p})}{d\delta}&=\frac{1}{N_\mathrm{T}}\sum_{j=1}^{N_\mathrm{T}}p_j\frac{dw_j}{d\delta}\\&=\frac{1}{N_\mathrm{T}}\sum_{j=1}^{N_\mathrm{T}}p_j\sum_{k=1}^{N_\mathrm{T}}\frac{\partial w_j}{\partial z_k}\frac{dz_k}{d\delta},\end{aligned} \tag{8.36}$$

where the z_k's are the phenotypes of all individuals in the population. In haploid models, it was assumed that the phenotype of an individual could be described as $z_a+\delta p+o(\delta)$, where $p=1$ or 0, depending on whether the individual bears the A allele. With diploids, we have three possible phenotypes: z_{aa}, z_{Aa}, and z_{AA}. Writing $\delta\equiv z_{AA}-z_{aa}$, linearity implies the additivity of effect of each gene copy: $z_{Aa}=(z_{AA}-z_{aa})/2$. In contrast, in the case of a completely recessive A allele, $z_{Aa}=(z_{AA}-z_{aa})/2-\delta/2$. Clearly, with dominance, the first-order effects of different alleles are not additive.

We can describe phenotypes in terms of allele frequencies as follows. We distinguish the allele frequencies p_{j1} and p_{j2}, with values 0 or 1, at each of the two homologous genes in individual j. In the case of a recessive A allele, we therefore have $z(p_{j1}, p_{j2}) = z_{aa} + \delta p_{j1} p_{j2}$. More generally, we may write any phenotype in terms of p_{j1} and p_{j2}, of their average p_j, and of the traditional dominance coefficient h. $h = 1$ if the A allele is fully dominant, and 0 if it is fully recessive. For any individual i, $z(p_{j1}, p_{j2}) = z_{aa} + \delta[p_j + (1-2h)(p_{j1}p_{j2} - p_j)]$. Then

$$\frac{d\mathrm{E}(\bar{p}'|\mathbf{p})}{d\delta} = \frac{1}{N_\mathrm{T}} \sum_{j=1}^{N_\mathrm{T}} p_j \sum_{c=1}^{N_\mathrm{T}} \frac{\partial w_j}{\partial z_k} [p_k + (1-2h)(p_{k1}p_{k2} - p_k)]. \quad (8.37)$$

The terms contributed by dominance involve products of three allele frequencies, $p_j p_{k1} p_{k2}$, and are of opposite sign for $h = 1/2 - x$ and $h = 1/2 + x$.

This result can be used to extend the diffusion approximations discussed in the next chapter to dominant mutants, provided expectations of the form $\mathrm{E}(p_j p_{k1} p_{k2}|\bar{p})$, given allele frequency $\bar{p}$ in the total population, are computed (Roze and Rousset 2003).

HIGHLIGHTS

In this chapter we have reviewed some consequences of diploidy on population structure and selection. The example of dispersal rates has shown how different modes of control of the trait could be taken into account by the same methods as those in the haploid models. In the different cases, ϕ is proportional to

$$\frac{\partial w}{\partial x_\bullet} + \frac{\partial w}{\partial x_0} R, \quad (8.38)$$

in terms of the same function w and of different relatedness coefficients. Often R is defined as $2Q_0/(1+Q_\mathrm{w})$ or $2F_\mathrm{ST}/(1+F_\mathrm{IT})$ (e.g., Hamilton 1971; Taylor 1988), but the above examples show that the definition of R depends on, for example, the mode of control of the trait: equation (8.32) implies $R = 2Q_0^\mathrm{R}/[(1+Q_\mathrm{w})/2 + Q_0]$ and equation (8.35) implies $R = \{[(1+Q_\mathrm{w})/2 + Q_0]/2 + (N-1)Q_0\}/[N(1+Q_\mathrm{w})/2]$. In an inclusive fitness perspective (by opposition to the direct fitness perspective, see chapter 7), one interprets the probabilities of identity as measures of "fidelity of transmission," that is, to what extent the genes transmitted by a recipient are related to those of a focal

individual (Frank 1998). Many formulas for R in diploid populations, reviewed in Michod and Hamilton (1980), may be interpreted as ratios of the fidelity of transmission of an allele by the recipient neighbors relative to the fidelity of transmission by the focal individual. This relatedness may be lower or higher than in a corresponding haploid model.

When some selfing is allowed in the above model, the fitness function cannot be expressed simply in terms of function (8.28). Thus, as already noticed on p. 102, the results should not be meant to imply that, in general, models with different life cycles can be treated by different relatedness coefficients for a fixed fitness function.

We did not take into account inbreeding depression as a selective force acting on other traits. There are two main reasons for this neglect. First, it is difficult to predict levels of inbreeding depression in subdivided populations. Some analytical approximations have been suggested by Whitlock (2002) (though see Roze and Rousset 2003). Whitlock et al. (2000) and Theodorou and Couvet (2002) present further numerical analyses. Second, evolution of a trait may feedback on the level of inbreeding depression, but it is unclear how this should be taken in account. A framework for analyzing multilocus evolution in subdivided populations seems required to address these questions. Without such a framework, studies of the evolution of characters that affect levels of inbreeding depression have generally treated inbreeding depression as a fixed factor (e.g. evolution of dispersal: Motro 1991; Gandon 1999; Perrin and Mazalov 1999; evolution of selfing: review in Uyenoyama et al. 1993).

Diploidy versus Other Modes of Inheritance

There are many possible departures from the diploid models considered above. X-linked genes, in humans and many other organisms, are in a haploid state in males and in a diploid state in females. This haplodiploid mode of inheritance may extend to most of the genome, for example, in social Hymenoptera. This mode of inheritance may be complicated by the possibility of asexual reproduction by workers and other phenomena. In such conditions, further analysis generally requires the development of concepts specific to class-structured populations, which are a topic of the next chapter.

CHAPTER NINE

Effective Size: Concepts and Applications to Stable Populations

Attempts to characterize various population structures in terms of a single quantity have led to the concept of "effective size," which gradually emerged from Wright's work (e.g., Wright 1931a, 1938, 1939). Effective size in particular may quantify the extent of drift at the level of the total population, a property important for determining the quantitative balance between drift and selection.

However, different definitions of effective size have been proposed from the consideration of different consequences of drift, making it necessary to compare them and select the best definition (if any). We will see how effective size affects selection. Precise applications of the concept of effective size in selection theory stem from the single-locus diffusion theory, and applications rarely (e.g., Ethier and Nagylaki 1980, 1988; Pollak and Sabran 1992; Caballero and Hill 1992b; Stephan et al. 1999) go beyond the simplest applications of the concept. However, there is renewed effort to construct diffusion approximations for subdivided populations (e.g., Cherry and Wakeley 2003; Whitlock 2003; Wakeley 2003; Roze and Rousset 2003). Here we will see how such approximations can be constructed using the concept of effective size, but also using the direct fitness method to compute expected allele frequency changes. This method leads in a straightforward way to results devoid of the various additional approximations or oversights of some previous analyses.

Many formulas of variable generality have been developed for effective size (e.g., Caballero 1994; Wang and Caballero 1999, for recent reviews and discussion). Rather than detailing all of them, we will focus on the effects of spatial subdivision and on the interpretation of

effective size in terms of elementary events affecting ancestral gene lineages, namely the movements of such a lineage among different categories of individuals and the probabilities of coalescence in each category. We will consider the additional effects of, for example, selfing, age structure, and high variance in number of offspring.

DEFINING EFFECTIVE SIZE

The following definitions have been considered (for references to the early literature, see in particular Ewens 1982, Caballero 1994, and Whitlock and Barton 1997).

Inbreeding effective sizes. In the haploid Wright-Fisher model, the probability of coalescence in the previous generation is $1/N$. Thus, an effective size can be defined as the reciprocal of the probability of coalescence in the previous generation (Ewens 1982). This is generally known as an "inbreeding" effective size. It is instantaneous, in contrast to the asymptotic "inbreeding" effective size, which describes the rate of coalescence of lineages in the distant past, given they have not coalesced before (e.g., Felsenstein 1971; Ewens 1982), and is equivalent to the effective sizes computed by many other authors (Ewens 1982; Whitlock and Barton 1997). For the Wright-Fisher model, the asymptotic effective size is also $1/N$.

The instantaneous effective size defined from the probability of coalescence in the previous generation will generally take a very complicated form and is not further considered. For the asymptotic inbreeding effective size, a unique mathematical definition may be as follows. In a matrix model, using the notation of chapter 4,

$$\frac{1}{N_\mathrm{e}} \equiv \lim_{t\to\infty} \frac{C_{i,t+1}}{1-\sum_{k=1}^{t} C_{i,k}} \tag{9.1}$$

for all pairs i of genes within or among different classes. This corresponds to the verbal definition as the probability of coalescence of genes at time $t+1$ (numerator), given they have not coalesced before (denominator). This limit may be expressed in terms of the largest eigenvalue λ_1 of the $\mathbf{G}$ matrix introduced in equation (4.4), which describes decreases in gene diversity. Using the notations of equation (4.11),

$$\lim_{t\to\infty} \frac{C_{i,t+1}}{\sum_{k=t+1}^{\infty} C_{i,t}} = \frac{\lambda_1^t a_1 r_{1i}}{(\sum_{k=t}^{\infty} \lambda_1^k) a_1 r_{1i}} = (1-\lambda_1). \tag{9.2}$$

Hence, $N_e = 1/(1 - \lambda_1)$. Terms contributed by the other eigenvalues are negligible for large t. Again this holds for all i: one cannot define different effective sizes (e.g., subpopulation and total population size) from different values of i in equation 9.1.

Equivalence to eigenvalue effective size. Considering a complete description of the population (e.g., allele frequencies in all demes), Ewens (1979, 1982) defined the eigenvalue effective size from the largest nonunit eigenvalue of the matrix describing changes in the state of the population. The general theory of Markov chains implies that this eigenvalue determines the asymptotic approach to the equilibrium state of the population (e.g., Ewens 1979, p. 70; this is the same result stated above for probabilities of identity, but applied to a complete description of the population). Since properties of the population determine properties of pairs of genes, the eigenvalue effective size also describes the asymptotic properties of probabilities of identity, and the asymptotic coalescence effective size is equivalent to the eigenvalue effective size in a model without mutation (Ewens 1982; Whitlock and Barton 1997).

Mutation effective size. In the haploid Wright-Fisher model, $\dot{Q} \approx 1/(1 + 2N\mu)$; for more general models, we wish to have $\dot{Q} \approx 1/(1 + 2N_e^{(m)}\mu)$ for some effective size $N_e^{(m)}$ (e.g., Maruyama and Kimura 1980). Being defined through the effects of mutation, this $N_e^{(m)}$ is a "mutation effective size." Depending on its definition, mutation effective size may or may not be exactly related to the eigenvalue effective size. The mutation effective size can be exactly identified to $1-\lambda_1$ only for a particular weighted average of identities. Let $\mathbf{y}_1$ be the leading left eigenvector of $\mathbf{G}$, normalized such that the sum of its elements is 1. Premultiplying equation (4.6) by $\mathbf{y}_1$, one finds that at equilibrium

$$\mathbf{y}_1 \cdot \mathbf{Q} = \gamma \frac{1 - \lambda_1}{1 - \gamma\lambda_1} \tag{9.3}$$

$$= 1 - 2\mu N_e + O(\mu^2), \tag{9.4}$$

where $N_e \equiv (1 - \lambda_1)^{-1}$ is the eigenvalue effective size. In the finite island model of chapter 3, $\mathbf{y}_1 = [O(1/n_d), 1 - O(1/n_d)]$, and hence $\mathbf{y}_1 \cdot \mathbf{Q}$ is approximately the identity of genes in different demes, for which equation (9.4) offers a very crude approximation [and equation (9.3) a better one]. This can be generalized to other models with an "island" mode of dispersal.

Variance effective size. In the haploid Wright-Fisher model, the vari-

ance of allele frequency p' given frequency p one unit time before is $\mathrm{Var}(p') = p(1-p)/N$. This suggests defining an effective size as $\bar{p}(1-\bar{p})/\mathrm{Var}(\bar{p}')$ for some average frequency $\bar{p}$. However, as noticed by Ewens (1979, p. 108; 1982), for two different reasons, the ratio $\bar{p}(1-\bar{p})/\mathrm{Var}(\bar{p}')$ may not be a well-defined parameter. First, there may be no allele frequency whose variance $\mathrm{Var}(\bar{p}')$ at t is determined uniquely by its value $\bar{p}$ at $t-1$ (i.e., no Markov chain on $\bar{p}$). Second, even if there is such a variable, the ratio $\bar{p}(1-\bar{p})/\mathrm{Var}(\bar{p}')$ may not be independent of $\bar{p}$.

In such cases, an effective size can still be defined from the variance of allele frequency changes over one generation by invoking a separation of time scales argument, as was first developed in chapter 4.

For example, consider that genes in deme i are binomially sampled from an infinite number of juveniles. Let $\mathbf{p}$ be the allele frequencies among their parents in all demes, and let $\tilde{p}_i$ be the frequency of allele A among the juveniles. Then

$$\mathrm{Var}(\bar{p}'|\mathbf{p}) = \frac{1}{N_\mathrm{T}}(\bar{p} - \overline{\tilde{p}_i^2}), \tag{9.5}$$

where $\overline{\tilde{p}_i^2}$ is the expected frequency of pairs of A alleles when sampling two different juveniles after dispersal, and also when sampling two different adults. $\overline{\tilde{p}_i^2}$ is determined by the allele frequencies $\mathbf{p}$ in all parental demes, and may differ for different parental populations with the same average frequency $\bar{p}$ but with different values of $\mathbf{p}$. So $\mathrm{Var}(\bar{p}'|\mathbf{p})$ is not a function of $\bar{p}$ only, but also of other features of $\mathbf{p}$.

However, with an increasing number of demes, the frequencies of demes with given numbers of A copies (hence the frequencies of different values of p_i in $\mathbf{p}$) approach their expectations in the infinite island model. Thus, for a given $\bar{p}$, the variance of $\overline{\tilde{p}_i^2}$ vanishes when the number of demes increases. In the infinite island model, $\overline{\tilde{p}_i^2}$ is a function of $\bar{p}$. Following expression (4.15), it may be computed in terms of F_ST as $\bar{p}^2 + \bar{p}(1-\bar{p})F_\mathrm{ST}$. This suggests that

$$\mathrm{Var}(\bar{p}'|\mathbf{p}) = \bar{p}(1-\bar{p})\frac{1-F_\mathrm{ST}}{N_\mathrm{T}}, \tag{9.6}$$

that is, the variance effective size is $N_\mathrm{T}/(1-F_\mathrm{ST})$, as first found by Wright (1943).

However, strictly speaking, we can use expression (4.15) only under an interdeme separation of time scales, that is, when N_T and N_e are

infinite. A more exact statement for a finite number of demes is that

$$\mathrm{Var}(\bar{p}'|\mathbf{p}) = \frac{\bar{p}(1-\bar{p})}{N_\mathrm{e}} + \frac{\epsilon}{n_\mathrm{d}} \tag{9.7}$$

where ϵ is a random (for different values of $\mathbf{p}$ given $\bar{p}$) term whose mean and variance vanish as the number of demes increase. Then the statement that the variance in allele frequency $\bar{p}'$ is of the form

$$\bar{p}(1-\bar{p})/N_\mathrm{e} \text{ for } N_\mathrm{e} \text{ independent of } \bar{p} \tag{9.8}$$

can be understood as a shorthand for the more exact result (9.7).

$N_\mathrm{e} = N_\mathrm{T}/(1-F_\mathrm{ST})$ in the above example, but an effective size satisfying equation (9.7) can be more generally defined from the limit value of $n_\mathrm{d}\mathrm{Var}(\bar{p}'|\bar{p})$ for a large number of demes:

$$\frac{1}{N_\mathrm{e}} \equiv \frac{1}{n_\mathrm{d}} \lim_{n_\mathrm{d}\to\infty} n_\mathrm{d} \frac{\mathrm{Var}(\bar{p}'|\mathbf{p})}{\bar{p}(1-\bar{p})}, \tag{9.9}$$

given there is a separation of time scale in the limit. This N_e becomes equivalent to the eigenvalue effective size in the limit, which is useful for its computation.

Without a separation of time scales, an alternative definition of variance effective size would be required. The de facto definition of variance effective size used in many papers is indeed different: it is based, for example, on an asymptotic rate of increase of $\mathrm{Var}(\bar{p})$ (e.g., Felsenstein 1971), which can be identified with the eigenvalue effective size (Ewens 1982; Whitlock and Barton 1997).

A general concept of total effective size. The above discussion has shown that the eigenvalue effective size describes all asymptotic properties of the population, as well as the instantaneous variance under a separation of time scales. It approximately determines genetic diversity at the level of the total population.

Local effective size. We will distinguish from the above concept a local effective size, which approximately determines population differentiation. The effective size of a total population depends on the number of demes. In contrast, local effective size will be defined so as to be the deme size N in the homogeneous island and isolation by distance models.

APPLICATION IN DIFFUSION APPROXIMATIONS

Diffusion approximations provide relatively simple formulas for various aspects of allele frequency dynamics, such as allele frequency

distributions, fixation probabilities, and fixation time. These formulas involve two main quantities, the expected change in allele frequency p over one generation and its variance. For an allele with fecundity advantage s in a panmictic population, the mean and variance are approximated as

$$\mathrm{E}(p' - p|p) \approx M(p) \equiv sp(1-p) \tag{9.10}$$

[see equation (2.1)], and

$$\mathrm{Var}(p' - p|p) \approx V(p) \equiv p(1-p)/N. \tag{9.11}$$

(e.g., Ewens 1979, pp. 138–139; Karlin and Taylor 1981, p. 180). In the latter expression, any $O(s)$ effect on variance of change in allele frequency is neglected.[1] Thus, this variance is approximated by its value in the neutral Wright-Fisher model [equation (2.3)]. An approximation for the fixation probability of an allele initially in frequency $\tilde{p}$ is then

$$\Pi(\tilde{p}) = \frac{\int_0^{\tilde{p}} G(z)\,dz}{\int_0^1 G(z)\,dz} \tag{9.12}$$

where $G(z) \equiv e^{-\int^z 2M(x)/V(x)\,dx}$. With equations (9.10) and (9.11), this becomes

$$\Pi(\tilde{p}) = \frac{1 - e^{-2Ns\tilde{p}}}{1 - e^{-2Ns}} \tag{9.13}$$

(e.g., Malécot 1952; Kimura 1957; Ewens 1979, pp. 83, 147). It is in general difficult to provide full analytical justification for derivation of the fixation probability or allele frequency distributions from diffusion equations (e.g., Nagylaki 1985; Bürger and Ewens 1995; Watterson 1996, p. 162), and for the diffusion equations in structured populations. Yet diffusion approximations are often very efficient in practice. Here we consider how to compute the diffusion approximation in an island model using arguments based on the separation of time scales (Ethier and Nagylaki 1980, 1988; Wakeley 2003; Roze and Rousset 2003).

First, we use effective size to quantify drift at the level of the total population. We have seen that when the number of demes increases, we

[1]Formally, the diffusion approximation is constructed as a limit case when $N \to \infty, s \to 0$ for Ns fixed, and is a function of the limit values of $N\mathrm{E}(p' - p|p)$ and $N\mathrm{Var}(p' - p|p)$. The $O(s)$ term in the variance has no effect on the latter limit value.

approach an "interdeme" separation of time scales, and we can write the variance of $\bar{p}'$ as $V(\bar{p}) + \epsilon/n_{\mathrm{d}}$ [equation (9.7)], where

$$V(\bar{p}) = \bar{p}(1-\bar{p})/N_{\mathrm{e}} \tag{9.14}$$

and ϵ is a random term that vanishes when $n_{\mathrm{d}} \to \infty$.

Second, we quantify selection. We have seen on p. 109 that in the same condition of separation of time scales, the expected change in average frequency $\mathrm{E}(\bar{p}'|\bar{p}) - \bar{p}$ of an allele with effect δ in a haploid population (or of a semidominant mutant in a diploid population) can be written as $M(\bar{p}) + o(\delta)$, where

$$M(\bar{p}) = W_{\mathrm{IF}}\bar{p}(1-\bar{p}) \tag{9.15}$$

for a selection coefficient W_{IF} of order δ and independent of p. This actually follows from the fact that the expected change in allele frequency can be written $M(\bar{p}) + \delta\epsilon'$, where ϵ' is a random term that vanishes when $n_{\mathrm{d}} \to \infty$. The ϵ' term has the same origin and nature as the ϵ term in the variance, and both can be neglected in the diffusion approximation.[2]

In the above expressions, N_{e} can be computed using equation (9.9), or using results developed later for eigenvalue effective size. W_{IF} was previously recognized as inclusive fitness, which can be computed in a straightforward way by the direct fitness method. Then, given W_{IF} and N_{e}, we can compute diffusion approximations. The probability of fixation of an allele at initial frequency $\tilde{p}$ in the total population is, following equation (9.12),

$$\Pi(\tilde{p}) = \frac{1 - e^{-2N_{\mathrm{e}}W_{\mathrm{IF}}\tilde{p}}}{1 - e^{-2N_{\mathrm{e}}W_{\mathrm{IF}}}}. \tag{9.16}$$

The product $N_{\mathrm{e}}W_{\mathrm{IF}}$ can also be deduced from the first-order effect ϕ on the probability of fixation, considered in chapter 6.[3]

[2]Here the diffusion approximation is constructed as a limit case when $n_{\mathrm{d}} \to \infty, \delta \to 0$ for $n_{\mathrm{d}}\delta$ fixed, and is a function of limit values of $n_{\mathrm{d}}\mathrm{E}(p'-p|p)$ and $n_{\mathrm{d}}\mathrm{Var}(p'-p|p)$. The ϵ and ϵ' terms have no effect on these limit values.

[3]For a single mutant in initial frequency $\tilde{p} = 1/N_{\mathrm{T}}$, and with phenotypic effect δ, W_{IF} can be written as $\delta\tilde{W}_{\mathrm{IF}}$ [equation (7.5)], and equation (9.19) implies

$$\phi = \left.\frac{d\Pi(\tilde{p})}{d\delta}\right|_{\delta=0} = \left(1 - \frac{1}{N_{\mathrm{T}}}\right)\frac{N_{\mathrm{e}}\tilde{W}_{\mathrm{IF}}}{N_{\mathrm{T}}}. \tag{9.17}$$

Hence, neglecting a term of relative magnitude $1/N_{\mathrm{T}}$,

$$\phi \sim N_{\mathrm{e}}\tilde{W}_{\mathrm{IF}}/N_{\mathrm{T}}. \tag{9.18}$$

For an illustration, we will consider fixation probabilities under different conditions. One is the the model of chapter 7, where neighbors may help each other, with fecundity costs and benefits δC and δB, and with hard selection, that is, no regulation of juvenile numbers before dispersal. The diffusion approximation will be compared to simulations under the island and stepping stone model. Second, we will consider the case of soft selection, also previously considered in chapter 7, where there is a fecundity effect s (previously noted δC) of an allele on the relative number of juveniles produced by its bearers in a deme, but where all demes produce the same number of juveniles independently of the genotypes of parents. This is the case most considered in previous discussions of fixation probabilities in subdivided population, to which we will compare our results.

Consider first the hard selection model. With a homogeneous island model of dispersal, effective size is $N_{\mathrm{T}}/(1 - F_{\mathrm{ST}})$ [equation (9.48)], and for hard selection W_{IF} is $\delta(-C + B/N)(1 - F_{\mathrm{ST}})$, as deduced from equation (7.12). Hence,

$$\Pi(\tilde{p}) = \frac{1 - e^{-2N_{\mathrm{T}}\delta(-C+B/N)\tilde{p}}}{1 - e^{-2N_{\mathrm{T}}\delta(-C+B/N)}}. \tag{9.19}$$

Equation (9.19) compares well with simulation results (fig. 9.1). The analytical prediction is good over a wide range of parameter values, but better for weak selection [which means here smaller δB and δC, not only smaller $\delta(-C + B/N)$]. This could be expected, as the diffusion approximation is constructed by assuming a large number of demes and weak selection. It is not required that deme sizes are large, nor that dispersal rates are small, in contrast to the approximations of, for example, Maruyama (1983) and Barton (1993). Rather, it is assumed that transition rates of gene lineages between demes with different allele counts are large relative to selection coefficients, and hence that dispersal is large relative to selection (Wakeley 2003; Roze and Rousset 2003).

Simulations also show that equation (9.19) predicts the probability of fixation in a stepping stone model of dispersal, though somewhat less accurately than in the island model for a given magnitude of the effects on fecundity. This result is not explained by the above argument, as equations (9.15) and (9.14) may not hold, but it may be deduced from a slightly more general argument. Maruyama (1983, p. 533) noticed that $M(\bar{p})$ and $V(\bar{p})$ are proportional to each other even though they may not scale as $\bar{p}(1 - \bar{p})$. Thus, we can compute their ratio for any $\bar{p}$,

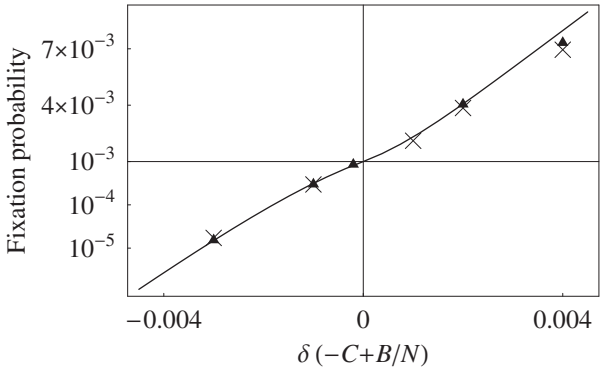

FIGURE 9.1. Fixation probabilities in a subdivided population. Notice the logarithmic y-scale for deleterious mutants. The diffusion approximation (line) and simulation results are shown for the hard selection model. Simulations were run for mutants with $\delta C = 0.02$ and values of δB ranging from 0.08 to 0.12, in island (▲) and stepping stone (×) structures of 200 demes of 5 adults, with dispersal rate $m = 0.2$.

and we can then evaluate the fixation probability using equation (9.12).

Specifically, Maruyama considered the soft selection model with $B = 0$ (no effects of neighbors on fecundity) and fecundity effect $s \equiv -C\delta$. Then $M(\bar{p}) = s[\bar{p} - \mathrm{E}(p_i^2|\bar{p})]$, where p_i is allele frequency among parents in deme i. On the other hand, $V(\bar{p}) = [\bar{p} - \mathrm{E}(\tilde{p}_i^2|\bar{p})]/N_\mathrm{T}$, where $\tilde{p}_i$ is the allele frequency in deme i among juveniles after dispersal [equation (9.5)]. Since $\bar{p} - \mathrm{E}(p_i^2|\bar{p}) = [\bar{p} - \mathrm{E}(\tilde{p}_i^2|\bar{p})](N-1)/N$, then $M(\bar{p})/V(\bar{p}) = N_\mathrm{T}s(N-1)/N$, independent of $\bar{p}$.[4] Maruyama neglected the difference between p_i and $\tilde{p}_i$, and hence he ignored the factor $(N-1)/N$ in $M(\bar{p})$ and $M(\bar{p})/V(\bar{p})$. His result implied that population structure (i.e., N and m) does not affect the probability of fixation of semidominant mutants. Although this is

[4]The expression for $M(\bar{p})/V(\bar{p})$ is also implied by comparison of the different expressions for ϕ. Equations (7.23) and (9.18) imply that

$$N_\mathrm{e}\tilde{W}_\mathrm{IF} = -N_\mathrm{T}C(N-1)/N. \tag{9.20}$$

M/V should be independent of $\bar{p}$ whenever $M(\bar{p})$ can be expressed independently of products of allele frequencies in different demes, and therefore when ϕ can be expressed independently of relatedness coefficients of individuals in different demes, as in equations (7.21) and (7.23).

nearly so in his model, the conclusion is not general, particularly with local extinctions (e.g., Barton 1993; Cherry 2003b).

With dominance, diffusion approximations can also be constructed using the same separation of time scales argument as above. One finds that $M(\bar{p})$ is of the form $\bar{p}(1-\bar{p})(s_1 + s_2\bar{p})$, where s_1 and s_2 are two selection coefficients independent of allele frequency, but dependent on population structure much as W_{IF} is (Cherry 2003a; Roze and Rousset 2003). For moderate dispersal rates, the resulting diffusion approximation is good, but for dispersal rates lower than selection coefficients, the approximations developed by Slatkin (1981) (see also Takahata 1991) are better (Roze and Rousset 2003).

COMPUTATION OF EFFECTIVE SIZE OF A TOTAL POPULATION

Different types of approximations have been developed independently for the eigenvalue or inbreeding effective size. They differ both by assumptions about the magnitudes of different parameters and by the biological significance of the different classes, such as demes (Wright 1943; Maruyama 1972; Nagylaki 1980; Whitlock and Barton 1997) or age classes (Felsenstein 1971; Hill 1972; Charlesworth 1994). These different cases, as well as local and total effective sizes, will here be analyzed in a unified framework. An approximation for λ_1, known as the strong migration limit, will be obtained when the deme sizes N_i are large. More general results will then be discussed.

In natural populations, individuals differ in many ways that will affect their expected number of offspring. This variation may, at least as a first approximation, be described in terms of different classes of individuals. For example, these classes may be habitats with different ecological characteristics, two sexes, or age classes. Effective size may then be computed in terms of the relative genetic contribution of different classes of individuals to the future population. These relative contributions are known as reproductive values and are also important tools for the analysis of selection in class-structured populations, as developed in chapter 11. We first review the concept of reproductive value.

Reproductive Value

CLASS REPRODUCTIVE VALUE

As a simple example, consider a population subdivided into two habitats. Suppose we follow a gene lineage backwards in time and ask what is the probability α_i that this gene lineage is in habitat i. Recursions for the α_i's between times t and $t-1$ follow from the probabilities f_{ij} that a gene lineage in habitat i at t was in habitat j at $t-1$. Note that the f_{ij}'s are the backwards dispersal rates between habitats. Specifically,

$$\begin{aligned}\alpha_j(t-1) &\equiv \Pr[\underline{\text{l}}\text{ineage } \underline{\text{in}} \underline{\text{h}}\text{abitat } j \text{ at time } t-1] \\ &= \sum_i \Pr[\text{l.i.h. } i \text{ at } t] \Pr[\text{l.i.h. } j \text{ at } t-1 | \text{l.i.h. } i \text{ at } t] \\ &= \sum_i \alpha_i(t) f_{ij}.\end{aligned} \tag{9.21}$$

We assume here a constant demography. Hence, at the stationary equilibrium, $\alpha_j(t) = \alpha_j(t-1) = \alpha_j$ implies $\alpha_j = \sum_i \alpha_i f_{ij}$, or in matrix form,

$$\boldsymbol{\alpha} = \boldsymbol{\alpha} \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}. \tag{9.22}$$

The above equation therefore says that $\boldsymbol{\alpha}$ is a left eigenvector of the matrix of backwards dispersal rates between habitats associated with its largest eigenvalue 1. We require in addition that the elements of $\boldsymbol{\alpha}$ are probabilities, which sum to 1 (alternative normalizations are possible). With two classes, $f_{i2} = 1 - f_{i1}$, and

$$\boldsymbol{\alpha} = \left(\frac{f_{21}}{f_{21} + f_{12}}, \frac{f_{12}}{f_{21} + f_{12}} \right). \tag{9.23}$$

Reproductive value may be defined as the asymptotic probability that an ancestral lineage was in some class, or defined as the left eigenvector $\boldsymbol{\alpha}$. The former definition focuses on a long-time (asymptotic) property, while the latter focuses on a short-time property (a recursion over one generation). These two definitions are usually equivalent, since the asymptotic probability can be computed from the recursion.

Individual Reproductive Value

Individual reproductive value v_i is the probability that an ancestral lineage was in a given individual of a given class i. Assuming all individuals within a class are equivalent, this probability is simply the probability that the ancestral lineage was in some class (α_i), divided by the number of individuals in this class N_i: $\alpha_i = N_i v_i$ (we again assume a constant demography, so that $N_i(t) = N_i(t-1) = N_i$ for any i). Hence, $N_j v_j = \sum_i N_i v_i f_{ij} \Leftrightarrow v_j = \sum_i v_i f_{ij} N_i / N_j$. In other words, individual reproductive values are given as elements of a left eigenvector of the matrix with elements $f_{ij} N_i / N_j$. The latter are the expected numbers w_{ij} of offspring in class i per parent in class j. The matrix of such numbers is known as the projection matrix (or Leslie matrix) in demography (Leslie 1945, 1948; Caswell 2001). Projection matrices were originally conceived to describe the growth of populations. They appear here in a more restricted context, since we assumed that class sizes are constant. The largest eigenvalue of the projection matrix is here 1, while there is no such restriction in the original demographic applications.

Strictly speaking, in an infinite-sized population, individual reproductive value is infinitely small. For this reason, most results below will be expressed in terms of class, rather than individual, reproductive values.

An Example

We illustrate the meaning and computation of differences in reproductive values in a potentially controversial case: differences between social classes in humans. We consider here data from a Norwegian parish (Røskaft et al. 1992). Two social classes were distinguished, $1 =$ low, and $2 =$ high (roughly, $1 =$ employees, and $2 =$ landowners). From church registers, the number of individuals that were born in some class j and first married in class i could be deduced. They are

$$(n_{ij}) = \begin{pmatrix} 230 & 278 \\ 83 & 1069 \end{pmatrix} \tag{9.24}$$

For simplicity, both sexes are pooled, and we make some additional assumptions. We neglect differences in age structure (actually, age structure differs between the two classes: for example, age at marriage was

lower in the high social class). Given the assumption that the class of birth is the class of marriage of parents, the above data give the flow between social classes. Given the assumption that the class structure of the population is stable, the ratio of number of offspring in each class may be considered as the ratio N_2/N_1 of the number of adults in each class (here this is 508/1152), and the number of offspring that marry per parent of each class can be computed as

$$W_j = \frac{n_{1j} + n_{2j}}{n_{j1} + n_{j2}}. \tag{9.25}$$

One finds $W_1 = 0.616$ and $W_2 = 1.169$ for the low and high social class, respectively. Additional data show that this difference results from two causes: the number of births per parent was higher in the high class, and children of high-status parents were more likely to marry. The survival rate of offspring was unrelated to the social status of parents.

What does this imply for reproductive value? Using equation (9.23), one finds that the ratio of individual reproductive values

$$\rho \equiv \frac{\alpha_2/N_2}{\alpha_1/N_1} \tag{9.26}$$

is simply n_{12}/n_{21} (here this is 3.35). It is related to the W_j's through

$$\frac{W_2}{W_1} = \frac{n_{11}n_{22} + \rho n_{21}(n_{11} + \rho n_{21} + n_{22})}{n_{11}n_{22} + n_{21}(n_{11} + n_{21} + n_{22})}. \tag{9.27}$$

Hence, $\rho > 1$ if and only if $W_2 > W_1$, that is, a class has higher reproductive value only if it has higher fitness. As one could expect, reproductive value is proportional to fitness ($\rho = W_2/W_1$) when class is not heritable.[5] Otherwise, differences in reproductive value are larger than differences in fitness ($\rho > W_2/W_1$).

Deme Structure: The Strong Migration Limit

For concreteness, consider two demes with an unequal number of adults N_1 and N_2 in each of them. The matrix $\mathbf{A}$ describing the origin of pairs of genes can be written as follows. We label pairs of genes by a pair

[5]They are proportional when $W_2/W_1 = n_{11}n_{22}/n_{21}^2$, which itself occurs when $n_{11}n_{22} = n_{12}n_{21}$ [from equation (9.25)], that is, when class is not heritable.

of indices ij describing their position, so that we can define a 4×4 matrix with rows and column indices in the order 11, 12, 21, 22. The element $a_{kl,ij}$ (klth row, ijth column) of this matrix gives the probability that parents of a pair of genes in demes k and l were in demes i and j. If juveniles disperse independently from each other, $a_{kl,ij} = f_{ki}f_{lj}$, in terms of the elements f_{ki} of the backwards migration matrix $\mathbf{F}$ for each gene lineage, that is, the matrix giving the probabilities of origin of juveniles.[6] Thus,

$$\begin{aligned}\mathbf{A} &= \begin{pmatrix} f_{11}f_{11} & f_{11}f_{12} & f_{12}f_{11} & f_{12}f_{12} \\ f_{11}f_{21} & f_{11}f_{22} & f_{12}f_{21} & f_{12}f_{22} \\ f_{21}f_{11} & f_{21}f_{12} & f_{22}f_{11} & f_{22}f_{12} \\ f_{21}f_{21} & f_{21}f_{22} & f_{22}f_{21} & f_{22}f_{22} \end{pmatrix} \\ &= \begin{pmatrix} f_{11}\mathbf{F} & f_{12}\mathbf{F} \\ f_{21}\mathbf{F} & f_{22}\mathbf{F} \end{pmatrix} \equiv \mathbf{F} \otimes \mathbf{F}.\end{aligned} \tag{9.29}$$

The $\otimes$ operation is known as the tensor product of two matrices (see Appendix A for further details). Thus, $\mathbf{A}$ is the tensor product of $\mathbf{F}$ with itself.

The probabilities of identity $\dot{Q}_{ij}$ of genes sampled in demes i and j (with $\dot{Q}_{12} = \dot{Q}_{21}$) are then given by

$$\mathbf{Q} \equiv \begin{pmatrix} Q_{11} \\ Q_{12} \\ Q_{21} \\ Q_{22} \end{pmatrix} = \gamma \mathbf{A}(\mathbf{Q} + \mathbf{c}) \text{ for } \mathbf{c} \equiv \begin{pmatrix} \frac{1-Q_{11}}{N_1} \\ 0 \\ 0 \\ \frac{1-Q_{22}}{N_2} \end{pmatrix}. \tag{9.30}$$

The extension of the above example to any number n_d of demes is straightforward, with $\mathbf{A} = \mathbf{F} \otimes \mathbf{F}$ generally.

THE APPROXIMATION FOR EFFECTIVE SIZE

Equation (9.2) showed that effective size can be deduced from the largest eigenvalue λ_1 of the $\mathbf{G}$ matrix, defined in chapter 4 (see p. 55),

[6] For example, if the forward dispersal rate d of juveniles and the fecundity are identical for all parents,

$$\mathbf{F} = \begin{pmatrix} \frac{(1-d)N_1}{(1-d)N_1+dN_2} & \frac{dN_2}{(1-d)N_1+dN_2} \\ \frac{dN_1}{dN_1+(1-d)N_2} & \frac{(1-d)N_2}{dN_1+(1-d)N_2} \end{pmatrix}. \tag{9.28}$$

here describing transition probabilities of pairs of genes between different classes without coalescence. An approximation can be obtained using the perturbation formula (A.6) for the largest eigenvalue. We can view $\mathbf{G}$ as a small perturbation of the migration matrix for pair of genes $\mathbf{A}$. According to the perturbation approximation, effects of order $1/N_i$ on λ_1 are computed in terms of the perturbation $\mathbf{G} - \mathbf{A}$, with elements of order $1/N_i$, and of the leading left and right eigenvectors of the $\mathbf{A}$ matrix (i.e., those associated with eigenvalue $l_1 = 1$ of $\mathbf{A}$).

The leading right eigenvector of $\mathbf{A}$ is the column unit vector. Given the leading left eigenvector $\boldsymbol{\alpha} \equiv (\alpha_i)$ of $\mathbf{F}$, the leading left eigenvector of $\mathbf{A}$ is $\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}$, with elements $\alpha_{ij} = \alpha_i \alpha_j$. The perturbation $\mathbf{G} - \mathbf{A}$ has elements 0 or $a_{kl,ii}/N_i$. Hence, from equation (A.6),

$$\lambda_1 \approx 1 - \sum_i \sum_{kl} \alpha_{kl} \frac{a_{kl,ii}}{N_i} = 1 - \sum_i \frac{\alpha_{ii}}{N_i} = 1 - \sum_i^{n_\mathrm{d}} \frac{\alpha_i^2}{N_i}, \quad (9.31)$$

that is,

$$1/N_\mathrm{e} \approx 1/N_\mathrm{s} \equiv \sum_i^{n_\mathrm{d}} \alpha_i^2 / N_i. \quad (9.32)$$

N_s is lower than the total size N_T, unless the deme reproductive values are proportional to the number of adults, that is, if individual reproductive value is constant among demes. Indeed, N_s can be written as

$$N_\mathrm{T}/[1 + {N_\mathrm{T}}^2 \mathrm{Var}(U_j^2)], \quad (9.33)$$

where U_j is the reproductive value of individual j. The case where individual reproductive values are constant across demes does not seem to correspond to a single easily described condition in terms of dispersal probabilities and (eventual) survival costs of dispersal. Assuming no cost of dispersal and no fecundity differences among adults, this condition holds for any dispersal scheme such that the number of juveniles in any deme after dispersal equals the number of juveniles before dispersal ("conservative migration," Nagylaki 1980).

The assumption of large N_i's was formalized by Nagylaki (1980) as the "strong migration limit," where "migration" is here class transition. Assuming (1) large class sizes ($N_i \to \infty$ for all i), (2) large transition rates between classes, that is, f_{12} and f_{21} fixed (such that, e.g., $N_j f_{ij} \to \infty$), and (3) low mutation rates ($\mu \to 0$, such that $N_i \mu$ is fixed for any i), it is found that

$$1 - Q_{ij} = 2N_\mathrm{s}\mu + O(\mu) \quad (9.34)$$

(Nagylaki 1980). See the appendix to this chapter for a proof.

Expression (9.32) for effective size has an intuitive explanation, relying on the interpretation of reproductive value as the probability that an ancestral lineage was in class i (e.g., Nordborg 1997; Rousset 1999a). If we look far enough in the past, these ancestral probabilities are independent of the present class of the gene lineage. The same argument applies to two independent gene lineages. Therefore, pairs of gene lineages are affected identically by "ancient" events. Only "recent" events may differentially affect them, depending on their current class. In particular, if class 1 is smaller than class 2, with equal backwards dispersal rates in each class, coalescence events will be more common in class 1, so that we expect $Q_{11} > Q_{22}$, particularly for low transition rates between classes. In the limit conditions ($N_i \to \infty$, fixed f_{ij}'s), we ignore such differences, because these conditions make coalescence events unlikely before many transition events between classes have occurred.

The coalescence rate of ancestral lineages can be deduced from this argument. Looking far enough in the past, two gene lineages are in the same class i with probability α_i^2. They can then coalesce with probability $1/N_i$. Hence, $1/N_\mathrm{e}$ is the average ancestral rate of coalescence, as expected.

Differences in mutation rates for different class transitions can also be introduced in the model. One then has

$$1 - Q_{ij} \approx 2N_\mathrm{s}\mu_\mathrm{e} \tag{9.35}$$

for

$$\mu_\mathrm{e} \equiv \sum_i \alpha_i \sum_j f_{ij}\mu_{ij} \tag{9.36}$$

(Rousset 1999a). This result can be understood as follows. If we look far enough in the past, the past class of a gene lineage is independent of its present class. Hence, gene lineages presently in different classes have been accumulating mutations identically at the rate m_e in the distant past. Each term in m_e is made of the probability that the lineage was in class i (α_i) times the probability that it then had its parent in class j (f_{ij}) times the mutation rate given these events (μ_{ij}). Mutation will affect differentially the identity of genes within different classes only to the extent that recent mutations may differentially affect the gene lineages, depending on their class of origin. The assumption of low mutation rates makes mutation events unlikely in a recent past.

FURTHER EXAMPLES

Well-known expressions follow from those above. Consider a diploid population divided in N_f adult females and N_m adult males: the total reproductive value of each sex is 1/2, and hence the rate of coalescence is

$$\frac{1}{N_\mathrm{s}} \equiv \frac{1}{4}\left(\frac{1}{2N_\mathrm{f}} + \frac{1}{2N_\mathrm{m}}\right) \Leftrightarrow N_\mathrm{s} = 2\frac{4N_\mathrm{f}N_\mathrm{m}}{N_\mathrm{f}+N_\mathrm{m}}, \tag{9.37}$$

where the latter ratio is interpreted as an effective number of diploid individuals. This result relies on assumptions about the distribution of number of offspring per parent that are discussed on p. 165.

Another well-known result concerns X-linked genes. In male heterogametic systems, the backwards transition probabilities between the two classes ($1 =$ females; $2 =$ males) is

$$\mathbf{F} = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}, \tag{9.38}$$

which give the reproductive values $\boldsymbol{\alpha} = (2/3, 1/3)$. Hence,

$$\frac{1}{N_\mathrm{s}} \equiv \frac{1}{9}\left(\frac{4}{2N_\mathrm{f}} + \frac{1}{N_\mathrm{m}}\right) \Leftrightarrow N_\mathrm{s} = \frac{9N_\mathrm{f}N_\mathrm{m}}{2N_\mathrm{f}+4N_\mathrm{m}} \tag{9.39}$$

These results are both originally due to Wright (1939, 1969).

A slightly more involved example considers populations with females and hermaphrodites (gynodioecy), or males and hermaphrodites (androdioecy). Since the two cases are symmetrical, we only consider gynodioecy here. We consider an arbitrary pattern of inheritance where a gene is inherited from the mother with probability H and from the father with probability $1 - H$ (i.e., $H = 0$, $1/2$, or 1, respectively, for paternal, Mendelian, and maternal inheritance). We assume the same ploidy x in both sexes. Let N_f and N_h be the number of adult females and hermaphrodites, respectively. The reproductive values of these two classes will depend on the relative contribution of females and hermaphrodites to female gametes. Let f be the frequency of female gametes produced by females rather than by hermaphrodites among all female gametes in the population. If sex is not heritable, the backwards transition probabilities between the two classes ($1 =$ females, and $2 =$ hermaphrodites) form a matrix

$$\mathbf{F} = \begin{pmatrix} fH & 1-fH \\ fH & 1-fH \end{pmatrix}, \tag{9.40}$$

which gives the reproductive values $\boldsymbol{\alpha} = (fH, 1 - fH)$. Hence,

$$\frac{1}{N_\mathrm{e}} \equiv \frac{(fH)^2}{xN_\mathrm{f}} + \frac{(1 - fH)^2}{xN_\mathrm{h}}. \tag{9.41}$$

If sex is heritable (i.e., if females are more likely to have a female mother and hermaphrodites are more likely to have a hermaphrodite mother), this should be taken into account in the transition matrix (see Laporte et al. 2000).

For maternal haploid inheritance, this logic yields $N_\mathrm{e} = N_\mathrm{f}$. This result has been applied to mitochondria in many organisms. Actually, the transmission of mitochondria is generally not really haploid: a population of mitochondria is transmitted (see Birky 1994, 2001, for reviews), which may matter particularly when there is some paternal inheritance (Takahata and Palumbi 1985; Birky et al. 1989; Bromham et al. 2003).

Deme Structure: More Accurate Results

The above results are approximations for large N_i's.[7] The strong migration limit considered in equation (9.34) yields a common approximation for all Q_{ij}'s: it ignores by its nature any spatial pattern when indices i and j stand for different demes. Hence, it may be of limited use for spatially structured populations.

Another type of expression for N_e can be obtained without assuming large N_i's. We consider the recursion for increase in identity in the absence of mutation, $\mathbf{Q}' = \mathbf{A}(\mathbf{Q} + \mathbf{c})$, and premultiply it by the left eigenvector $\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}$ of $\mathbf{A}$. Then $Q'_{\boldsymbol{\alpha}} = Q_{\boldsymbol{\alpha}} + (\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}) \cdot \mathbf{c}$, where $Q_{\boldsymbol{\alpha}} \equiv (\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}) \cdot \mathbf{Q}$. Hence,

$$\frac{1 - Q'_{\boldsymbol{\alpha}}}{1 - Q_{\boldsymbol{\alpha}}} = 1 - \sum_i \frac{\alpha_i^2}{N_i} \frac{1 - Q_{ii}}{1 - Q_{\boldsymbol{\alpha}}}. \tag{9.42}$$

The asymptotic decrease of gene diversity, which is $1 - \lambda_1$, may be written in terms of the F_ST-like parameters,

$$F_i \equiv \lim_{t \to \infty} \frac{Q_{ii} - Q_{\boldsymbol{\alpha}}}{1 - Q_{\boldsymbol{\alpha}}}, \tag{9.43}$$

[7] This is so even in simple cases such as the two-sexes formula (9.37). In this model, the exact expression for N_e follows from $\lambda_1 = [1 - 2/N_\mathrm{s} + \sqrt{1 + (2/N_\mathrm{s})^2}]/2$ (Wright 1969, p. 197; Crow and Kimura 1970, p. 104).

as

$$1 - \lambda_1 = 1 - \sum_i \frac{\alpha_i^2}{N_i}(1 - F_i) \tag{9.44}$$

so that

$$\frac{1}{N_\mathrm{e}} = \sum_i \frac{\alpha_i^2}{N_i}(1 - F_i). \tag{9.45}$$

In the strong migration limit, the F_i's vanish, so we recover equation (9.32) for effective size. Thus, F_i appears as a correction factor, taking into account that the probability that two ancestral lineages are in the same deme, given they have not coalesced before, is $\alpha_i^2(1 - F_i)$, not α_i^2.

Population structure ($F_i > 0$) increases the total effective size. In the absence of mutation, every expression of the form

$$\lim_{t\to\infty} \frac{Q_\mathrm{w} - Q_\mathrm{b}}{1 - Q_\mathrm{b}} \tag{9.46}$$

may be interpreted in terms of cumulative probabilities of coalescence, as

$$\lim_{t\to\infty} \frac{\sum_{l=1}^{t}(C_{\mathrm{w},t} - C_{\mathrm{b},t})}{1 - \sum_{l=1}^{t} C_{\mathrm{b},t}} \tag{9.47}$$

[equation (4.23)], and as equation (4.24) implies, they are usually very close to the other definitions of F_ST parameters discussed in chapter 4. This shows that the (eigenvalue) effective size can be computed in terms of F_ST parameters. Such formulas were derived by Maruyama (1972, equation 4–1) for homogeneous dispersal, and by Whitlock and Barton (1997) for the present models. In the homogeneous island or isolation by distance models, equation (9.45) reduces to

$$N_\mathrm{e} \approx \frac{N n_\mathrm{d}}{1 - F_\mathrm{ST}} \tag{9.48}$$

in terms of a (low mutation limit) F_ST for the total population. Thus, we recover for the eigenvalue effective size the result obtained by Wright for the variance effective size in the infinite island model [see equation (9.6)].

In the island model, equation (9.48) can be derived by the following alternative argument. The probability that two genes in different demes come from a single deme in the previous generation (a_{21} on p. 26) is

$$a_{21} = \frac{1 - (1 - m)^2}{n_\mathrm{d}} + O(1/{n_\mathrm{d}}^2). \tag{9.49}$$

Thus, on a time scale of n_{d} generations, the rate at which genes in different demes come in a single deme is $1-(1-m)^2$. Then the ancestral lineages coalesce immediately on this time scale if they have the same parent (probability: $1/N$), or if they coalesce in this deme in a recent past rather than separate into different demes again [probability: $(1-1/N)F_{\mathrm{ST}}$, since F_{ST} is approximately the probability of recent coalescence of genes within demes]. Then the overall rate of coalescence is

$$\left[1-(1-m)^2\right]\left(F_{\mathrm{ST}}+\frac{1-F_{\mathrm{ST}}}{N}\right) \tag{9.50}$$

per n_{d} generations. From the value of F_{ST}, one finds that this is $(1-F_{\mathrm{ST}})/N$. Thus, the rate of coalescence per generation is $(1-F_{\mathrm{ST}})/(Nn_{\mathrm{d}})$.

More or less crude approximations of (9.48) for the stepping stone model or more general isolation by distance models have been discussed by Wright (1943) and Maruyama (1972). These approximations could be reconsidered in the light of the results of chapter 3, but such technical developments will not be considered here. We only recall that the average F_{ST} tends to be larger in one-dimensional habitats than in two-dimensional ones, and therefore that N_{e} will tend to be larger in one-dimensional habitats for a given N_{T}.

Additional Factors

Here we list some additional factors whose effects on effective size were not apparent in the previous examples. They are not specific to subdivided populations, but must be kept in mind in interpreting effective sizes in general.

SELFING

In hermaphrodite populations, selfing reduces N_{s} by a factor $(1+F_{\mathrm{IS}})$ (Wright 1969). This factor quantifies the increase in the rate of coalescence of genes relative to a randomly mating population (p. 132). For the island and isolation by distance models with selfing (chapter 8),

equation (9.42) generalizes into

$$\frac{1-Q'_{\boldsymbol{\alpha}}}{1-Q_{\boldsymbol{\alpha}}} = 1 - \sum_i \frac{\alpha_i^2}{N_i} \frac{(1+Q_{\mathrm{w}})/2 - Q_{ii}}{1-Q_{\boldsymbol{\alpha}}}, \tag{9.51}$$

where $N_i = N$, the number of diploid adults per deme, and $\alpha_i = 1/n_{\mathrm{d}}$. Given $F_{\mathrm{ST}} = (Q_{ii} - Q_{\boldsymbol{\alpha}})/(1-Q_{\boldsymbol{\alpha}})$ and $F_{\mathrm{IT}} = (Q_{\mathrm{w}} - Q_{\boldsymbol{\alpha}})/(1-Q_{\boldsymbol{\alpha}})$, $[(1+Q_{\mathrm{w}})/2 - Q_{ii}]/(1-Q_{\boldsymbol{\alpha}}) = (1+F_{\mathrm{IT}})/2 - F_{\mathrm{ST}} = (1-F_{\mathrm{ST}})(1+F_{\mathrm{IS}})/2$. Then

$$N_{\mathrm{e}} \approx \frac{2Nn_d}{(1-F_{\mathrm{ST}})(1+F_{\mathrm{IS}})}. \tag{9.52}$$

The relationship between N_{e} and F_{IS} differs from this one under forms of mating between relatives other than selfing (e.g., sib mating: Pollak 1987; Caballero and Hill 1992a). See Wang and Caballero (1999) for references to other models.

Coalescence Not Always Possible within a Class

Expressions for deme-structured populations generally assume that coalescence can occur in any of the classes (i.e., demes), while this is not the case, for example, for age-structured populations. In such populations, the backwards transition matrix for a gene typically takes the form

$$\mathbf{F} = \begin{pmatrix} f_{11} & \cdots & f_{1n_c} \\ 1 & & 0 \\ & 1 \quad 0 & \\ & \ddots & \vdots \\ 0 & 1 & 0 \end{pmatrix}. \tag{9.53}$$

The 1s on the subdiagonal mean that an individual of age $i+1$ necessarily "descends" from individuals of age i (through aging), except for the first age class, which may descend from parents of different ages. Then the backwards transition matrix for two genes may be written $\mathbf{A} = \mathbf{F} \otimes \mathbf{F}$, but the recursion for probabilities of identity is no longer of the form $\mathbf{Q}' = \gamma \mathbf{A}(\mathbf{Q} + \mathbf{c})$, because this would imply, for example, that two gene lineages in individuals of age 2 can coalesce in an individual of age 1 one time unit before. However, at that time the gene lineages were in the same two individuals, aged 1. Such cases may

be taken into account by a recursion of the form $\mathbf{Q}' = \gamma(\mathbf{AQ} + \tilde{\mathbf{A}}\mathbf{c})$, where the elements of $\tilde{\mathbf{A}}$ are $\tilde{a}_{ij,kk} = a_{ij,kk} = f_{ik}f_{jk}$ if coalescence can occur and $\tilde{a}_{ij,kl} = 0$ otherwise (this recursion remains an approximation because it assumes that the realized age structure in the population is constant, while the number of individuals of different age classes will randomly fluctuate). The probability of coalescence within class k is then $\tilde{a}_{ij,kk}/N_k$. Computation of the "strong migration limit" approximation for λ_1 proceeds as before, except that the perturbation $\mathbf{G} - \mathbf{A}$ has elements $\tilde{a}_{kl,ii}/N_i$ rather than $a_{kl,ii}/N_i$. Hence, equation (9.31) generalizes to

$$\lambda_1 \approx 1 - \sum_i \sum_{kl} \alpha_{kl} \frac{\tilde{a}_{kl,ii}}{N_i} \tag{9.54}$$

(Rousset 1999a). Let $\boldsymbol{\kappa} \equiv (\kappa_{11}, \kappa_{12}, \ldots, \kappa_{n_c n_c}) \equiv (\boldsymbol{\alpha} \otimes \boldsymbol{\alpha})\tilde{\mathbf{A}}$. Then equation (9.54) can be written

$$\frac{1}{N_e} \approx \frac{1}{N_s} \equiv \sum_{i=1}^{n_c} \frac{\kappa_{ii}}{N_i}. \tag{9.55}$$

The approximation $1 - Q_{ij} \approx 2N_s\mu$ [equation (9.34)] holds again.

Expression (9.55) for N_s can be understood by the same type of argument as before. Looking far enough in the past, two gene lineages are in classes i and j with probability $\alpha_i\alpha_j$. Then they coalesce one generation before, if (1) they both originate from a single class k and (2) they can coalesce given these transitions. The probability of these events is $\tilde{a}_{ij,kk}/N_k$. Hence, the average coalescence rate in the past is $\sum_{ij} \alpha_i\alpha_j \sum_k \tilde{a}_{ij,kk}/N_k = \sum_k \kappa_{kk}/N_k$, which is $1/N_s$. Since $\kappa_{ii} \leq \alpha_i^2$, equation (9.55) shows that the effective size is increased in comparison to that of equation (9.32). This is expected, since some coalescence events are prevented in comparison to the case considered in (9.32).

The effective size considered in models of age-structured populations (Felsenstein 1971; Hill 1979; Charlesworth 1994) describes the rate of coalescence per "generation," where the generation time is usually the expected interbirth time and depends on the age structure. Thus, for age-structured populations, N_s is equivalent to Felsenstein's (1971) inbreeding effective size, except for a difference in time scale (Rousset 1999a). Age structure increases effective size (i.e., decreases the coalescence rate, measured on the time interval used to construct the matrix model), but it decreases effective size per generation, that is,

increases the rate of coalescence per generation, the latter interpretation being the one usually emphasized. Results for other demographic structures (e.g., populations with seed banks, Kaj et al. 2001; Nordborg and Krone 2002) can be understood by the same arguments.

VARIANCE IN REPRODUCTIVE SUCCESS

In the Wright-Fisher model for a large population of size N, the number of adult offspring of each parent follow a binomial distribution with parameters N and $1/N$. Such a binomial distribution is conveniently approximated by a Poisson distribution with mean 1. In the age-structured model defined by equation (9.53), the variance σ_{g}^2 in lifetime number of offspring is larger than for a Poisson distribution, and N_{e} may be related to this increased variance as

$$N_{\mathrm{e}} \approx N_{\mathrm{s}} = N/\sigma_{\mathrm{g}}^2 \tag{9.56}$$

(Hill 1979, with a different time scale). The exact formula for binomial sampling differs from this Poisson approximation only by a term of relative magnitude $1/N$.

In a hermaphrodite diploid population, the distribution of the number of copies of a parental gene has higher variance than its number of offspring (because of the additional sampling of a parental gamete in each offspring). The effective number of diploid individuals is

$$N_{\mathrm{e}} \approx N_{\mathrm{s}} = 4N_{\mathrm{ad}}/(2 + \sigma_{\mathrm{d}}^2), \tag{9.57}$$

where N_{ad} is the number of diploid adults and σ_{d}^2 is the variance of number of offspring (e.g., Wright 1938 for the exact formula; Hill 1979, again with a different time scale). Indeed, this is equation (9.56), since it gives the effective number of genes $2N_{\mathrm{s}}$ as the number of genes ($2N_{\mathrm{ad}}$) divided by the variance of the number of copies of a gene in the next generation ($1/2 + \sigma_{\mathrm{d}}^2/4$).

Likewise, high variance in the number of offspring will reduce effective size. High variance may result, for example, from correlated events affecting different juveniles of the same family, such as when a clutch of juveniles is eaten by a predator. Likewise, if *expected* contributions vary among different parents, the effective size will also be reduced. In contrast, equation (9.37) assumed that genes of paternal origin in different offspring are independently sampled from all potential fathers with equal probabilities, and likewise for genes of maternal

origin. Hence, expected contributions are all identical, and observed contributions (gene copies) are multinomially distributed. In a model with separate sexes, the covariance between the numbers of male and female offspring of each parent also affects effective size (Latter 1959; Hill 1979).

Most formulas taking into account the variance in offspring number assume that this variance is not due to heritable factors. If there are good and poor habitats with limited dispersal between them, part of the variance in offspring number is due to the environment, here heritable, and the effective size is given by equation (9.33). Its denominator $1 + N^2\mathrm{Var}(U^2)$ is higher than the variance σ_g^2 of the number of copies in offspring of each parental gene (which depends on the variation in habitat quality), so the effective size is smaller than would be predicted by simple observation of σ_g^2. How much? This has been investigated in the case where part of the variance in offspring number is due to genetic differences between individuals. The resulting effect on effective size can again be interpreted as an effect on the variance of the ultimate contribution of a gene copy (Wray and Thompson 1990), that is, on the variance of reproductive value. Assuming random monogamous mating, the variance in family size is approximately $\sigma_d^2 + 4\sigma_f^2$, where $\sigma_d^2 \approx 2$ under binomial sampling and where σ_f^2 is the variance in family fitness. Assuming that fitness differences are due to additive effects at many loci unlinked to the neutral locus for which we compute effective size, a gene in an offspring inherits half of these effects (the usual parent-offspring regression in fitness). The cumulated effects on the ultimate contribution of a gene copy sum as $1 + 1/2 + 1/4 + 1/8... = 2$, and the variance in ultimate contribution is increased fourfold (Robertson 1961). Then the effective size is approximately

$$\frac{4N_{\mathrm{ad}}}{2 + \sigma_d^2 + 16\sigma_f^2} = \frac{N_{\mathrm{ad}}}{1 + 4\sigma_f^2}. \tag{9.58}$$

Santiago and Caballero (1995) provide a detailed derivation and generalizations of this result; Santiago and Caballero (1998) and Campbell (1999) further discuss the effects of linkage between the different loci considered.

CLONALITY

Diploid models with random selfing may be treated as haploid models, because the identity within individuals and between individuals of the same class is the same. This is no longer so if there is clonal reproduction of diploid individuals. Clonal reproduction in plants raises other complications for the computation of effective size (Orive 1993). Some vegetative forms in plants may persist over several time intervals and may also reproduce identically. For example, stolons produce other stolons; thus, transitions from "lineage in stolon" to "lineage in stolon" may occur both through reproduction or through survival. Coalescence will occur in one case and not in the other.

LOCAL EFFECTIVE SIZE

The previous results do not give approximations for F_{ST} or similar population structure parameters when there is some structure at the level of local demes. This section will describe such approximations. They are obtained assuming large transition rates between different demographic classes (e.g., age classes). However, in contrast to the "strong migration limit" previously considered, no such assumption is made for spatial dispersal. We will see that the results of the homogeneous island and isolation by distance models, in terms of local size N, can be generalized in terms of a local effective size of the same form as N_{s}, the effective size in the strong migration limit.

We consider different demes on a homogeneous lattice, each with several classes of individuals. The different classes within a deme may actually be two spatially distinct subpopulations, but the rate of exchange between them may be different from the rate of exchange between adjacent demes, and we consider that such subpopulations are at the same position on the lattice. Dispersal may be spatially localized or not. The latter case (a generalized island model), with two classes of individuals in each deme, is detailed in the appendix to this chapter. See Rousset (1999a) for a more general analysis.

Provided transition rates are higher than dispersal rates, one finds that

$$F_{\mathrm{ST}} \approx \frac{1}{1 + 2N_{\mathrm{e}}m_{\mathrm{e}}} \tag{9.59}$$

where m_e is an "effective" dispersal rate and N_e is an effective size of a deme, not of the total population. In fact, the definition of the appropriate local effective size N_e will be formally identical to the definition of the total effective size in the spatially unstructured case [equation (9.55)]. The "effective" migration rate is

$$m_e \equiv \sum_i \alpha_i \sum_j v_{ij} m_{ij}, \tag{9.60}$$

where v_{ij} is the probability that a lineage in class i was in class j one time unit before, and m_{ij} is the backwards dispersal probability, conditional on each possible transition from class j to i. This definition can be understood in the same way as the definition of μ_e [equation (9.36)], as an average migration rate of an ancestral gene lineage. Likewise, under isolation by distance, the results of pp. 39 and 40 generalize in terms of the "effective σ^2"

$$\sigma_e^2 \equiv \sum_i \alpha_i \sum_j v_{ij} \sigma_{ij}^2 \tag{9.61}$$

in terms of the second moments σ_{ij}^2 of the dispersal distributions, given each possible transition from class j to i.

The approximation (9.59) for F_{ST} may be rather poor, as it depends on transition rates being larger than dispersal rates. Otherwise, differentiation within and between classes may depart from this value in ways that can be predicted from the relative diversities in each class (Rousset 1999b). These diversities may themselves be complex functions of several demographic parameters. They bring little information about the relative values of the different transition rates between classes. In particular, without further assumptions they cannot be used to identify demographic "sources" and "sinks."

CONCLUDING REMARKS

So far, the mathematical developments seem to answer all problems. We have a fairly unitary concept of effective size, since the approximate identity between several definitions of effective size have been shown, with a notable exception for the instantaneous variance effective size. We could use effective size to quantify aspects of allele frequency dynamics such as fixation probabilities and to deepen some of

the conclusions of chapter 7. For example, the difference between fixation probabilities under "hard" and "soft" selection appear inversely related to the deme size. Finally, we were able to express effective size in terms of characteristics of the life cycle (e.g., age structure, selfing rates) and of $F_{\rm ST}$ parameters. Such results show that limited dispersal increases the global effective size, but variation in reproductive value decreases effective size.

But serious difficulties await us. First, taking all factors into account, any realistic formula will be fairly complicated. Rough approximations have been sought. A reduction of effective size by a factor of two has been suggested as a quick approximation for many of the complications arising from age structure and sex-specific effects (Nunney 1991). Waite and Parker (1996) suggest that this rule apply only to organisms with a short prereproductive period relative to reproductive period. Remember that the effective size considered in such discussions describes the rate of coalescence per generation (expected interbirth time).

Second, the exact definition of the $F_{\rm ST}$ considered in expressions for eigenvalue effective size has important implications, as it is a global $F_{\rm ST}$ for the total population. The global $F_{\rm ST}$ measured from genetic identity is usually reduced by mutation (see p. 41). The global $F_{\rm ST}$ affecting effective size would be a low mutation limit. It is essential to keep in mind that this $F_{\rm ST}$ is usually not the one that can be measured from gene diversity data. This problem may be alleviated by considering sequence divergence rather than gene identity, as sequence divergence may accumulate linearly in time, but the amount of sequence data required to have accurate estimates might be considerable. Further, estimating a global $F_{\rm ST}$ would require a sampling design giving appropriate weights to all subpopulations of a species. Another crucial issue is that whatever kind of data considered, the global $F_{\rm ST}$ estimate will be affected by past demographic fluctuations such as range expansions and bottlenecks.

Indeed, it may seem desirable to have expressions for effective size taking demographic fluctuations into account. The latter could be used to predict genetic diversity. However, it would not necessarily describe an effective size relevant for evaluating the strength of selection in current populations: for example, a bottleneck in the distant past does not affect the probability of fixation of a new mutant. From this per-

spective, we must rather focus on the consequences of recurrent demographic fluctuations at a short time scale. This is the topic of the next chapter.

APPENDIX

Dispersal and Class Transitions

We consider an infinite island model, with two classes of individuals per deme, which may be, for example, age classes or more and less fit individuals. There are N_1 and N_2 adults of classes 1 and 2 in each deme. Exchanges between classes are class-transition events, and exchanges between demes are dispersal events. We describe transition events by the matrix $\mathbf{V}$ of backwards transition rates between classes, and dispersal events by the backwards probabilities m_{ij} of dispersal between demes, given that a transition between i and j occurs. For example, the probability that a gene in class 1 had its parent in the same deme in the same class is v_{11}[for probability of parent in class 1]$\times(1-m_{11})$[for no dispersal given parent in class 1]. Likewise, the probability that a gene in class 1 had its parent in class 2 in another deme is $v_{12}m_{12}$. The probabilities $v_{ij}(1-m_{ij})$ of class transition without dispersal form a 2×2 matrix $\mathbf{F}$.

We use the conventions of the infinite island model: probabilities of identities of genes in different demes are considered null. Then the equilibrium relationships for probabilities of identity Q_{ij} of genes in classes i and j in the same deme are of the form

$$\mathbf{Q}\equiv\begin{pmatrix}Q_{11}\\Q_{12}\\Q_{21}\\Q_{22}\end{pmatrix}=\gamma(\mathbf{AQ}+\tilde{\mathbf{A}}\mathbf{c})\text{ for }\mathbf{c}\equiv\begin{pmatrix}c_{11}\\c_{12}\\c_{21}\\c_{22}\end{pmatrix}\equiv\begin{pmatrix}\frac{1-Q_{11}}{N_1}\\0\\0\\\frac{1-Q_{22}}{N_2}\end{pmatrix}, \tag{9.62}$$

where $\mathbf{A}$ is the tensor product $\mathbf{F}\otimes\mathbf{F}$ of the backwards dispersal matrix $\mathbf{F}$ with itself. $\tilde{\mathbf{A}}$ has elements $\tilde{a}_{kl,ij}=a_{kl,ij}$ if the two genes in classes i and j can coalesce, and $\tilde{a}_{kl,ij}=0$ if they cannot. That is, it is constructed from $\mathbf{A}$ in the same way as it would be without spatial structure: see age-structured models on p. 163. As noted there, it is only an approximation. For an age-structured population with juvenile dispersal, this matrix is

$$\tilde{\mathbf{A}}=\begin{pmatrix}a_{11,11}&0&0&a_{11,22}\\a_{12,11}&0&0&0\\a_{21,11}&0&0&0\\0&0&0&0\end{pmatrix}. \tag{9.63}$$

Here elements $a_{11,11}$ and $a_{11,22}$ correspond to birth events from two age 1 or two age 2 parents, and elements $a_{12,11}$ and $a_{21,11}$ correspond to one birth event and one survival event. Coalescence can occur in this case when the age 2 individual was the parent of the age 1 individual. In all other cases, $a_{ij,kl} = 0$, or coalescence events are impossible.

Equation (9.62) has solution

$$\mathbf{Q} = \gamma(\mathbf{I} - \gamma\mathbf{A})^{-1}\tilde{\mathbf{A}}\mathbf{c}. \tag{9.64}$$

We deduce an approximation for probabilities of identity by a perturbation analysis, assuming low dispersal rates m_{ij}. To that aim, we use a spectral representation of $\mathbf{F}$ [see equation (A.11)]: $\mathbf{F} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^{-1}$. From the properties of the spectral representation, the columns of $\mathbf{H}$ are right eigenvectors of $\mathbf{F}$, and the rows of $\mathbf{H}^{-1}$ are left eigenvectors of $\mathbf{F}$. In particular, in the absence of dispersal (when all $m_{ij} \to 0$, which we write simply as "$m \to 0$"), $\mathbf{F}$ reduces to the matrix of class transition rates $\mathbf{V}$. Then the first row of $\mathbf{H}^{-1}$ reduces to the reproductive value vector $\boldsymbol{\alpha}$, and the first column of $\mathbf{H}$ reduces to the unit vector $\mathbf{1} = (1,1)^\top$. For the largest eigenvalue ℓ_1 of $\mathbf{F}$, from equation (A.6),

$$\left.\frac{d\ell_1}{dm_{ij}}\right|_{m_{ij}=0} = \boldsymbol{\alpha}\left.\frac{d\mathbf{F}}{dm_{ij}}\right|_{m_{ij}=0}\mathbf{1} = -\alpha_i v_{ij}. \tag{9.65}$$

Hence, neglecting higher-order terms in the m_{ij}'s, $\ell_1 \approx 1 - m_{\mathrm{e}}$ for

$$m_{\mathrm{e}} \equiv -\sum_{i,j} m_{ij}\frac{d\ell_1}{dm_{ij}} = \sum_{i,j}\alpha_i v_{ij} m_{ij}. \tag{9.66}$$

For any matrix $\mathbf{M}$, let $\mathbf{M}^{\otimes 2} \equiv \mathbf{M} \otimes \mathbf{M}$. The spectral representation of $\mathbf{A} = \mathbf{F}^{\otimes 2}$ is $\mathbf{H}^{\otimes 2}\boldsymbol{\Lambda}^{\otimes 2}(\mathbf{H}^{-1})^{\otimes 2}$, and that of $(\mathbf{I} - \gamma\mathbf{A})^{-1}$ follows, so that equation (9.64) takes the form

$$\mathbf{Q} = \mathbf{H}^{\otimes 2}(\mathbf{I} - \gamma\boldsymbol{\Lambda}^{\otimes 2})^{-1}(\mathbf{H}^{-1})^{\otimes 2}\gamma\tilde{\mathbf{A}}\mathbf{c}. \tag{9.67}$$

The first row of $\mathbf{X}^{-1} \equiv \lim_{m\to 0}(\mathbf{H}^{-1})^{\otimes 2}$ is $\boldsymbol{\alpha}^{\otimes 2}$, and the first column of $\mathbf{X} \equiv \lim_{m\to 0}\mathbf{H}^{\otimes 2}$ is the vector $(1,1,1,1)^\top$. The largest eigenvalue of $\mathbf{I} - \gamma\boldsymbol{\Lambda}^{\otimes 2}$ is $1 - 2\mu - 2m_{\mathrm{e}} +$ (higher-order terms in μ and in the m_{ij}'s). The other eigenvalue ℓ_2 of $\mathbf{F}$ is bounded below 1 in the low migration limit. Hence,

$$(\mathbf{I} - \gamma\boldsymbol{\Lambda}^{\otimes 2})^{-1} = \mathbf{D} + \mathbf{O}(1), \tag{9.68}$$

where $\mathbf{D}$ is a diagonal matrix

$$\mathbf{D} = \mathrm{diag}\left(\frac{1}{2(\mu + m_{\mathrm{e}})}, 0, 0, 0\right) \tag{9.69}$$

and $\mathbf{O}(1)$ is a matrix whose elements are all bounded as $m \to 0$ and $\mu \to 0$. Further,

$$\mathbf{Q} = [\mathbf{X}\mathbf{D}\mathbf{X}^{-1} + \mathbf{O}(1)]\tilde{\mathbf{A}}\mathbf{c} \equiv [\mathbf{\Xi} + \mathbf{O}(1)]\mathbf{c}, \tag{9.70}$$

where $\mathbf{\Xi} \equiv (\mathbf{X}\mathbf{D}\mathbf{X}^{-1}) \lim_{m\to 0} \tilde{\mathbf{A}}$. Therefore, $Q_{kl} \approx (\mathbf{\Xi}\mathbf{c})_{kl}$ where

$$\begin{aligned}
(\mathbf{\Xi}\mathbf{c})_{kl} &= \sum_{i=1,2}\sum_{j=1,2} \xi_{kl,ij} c_{ij} \\
&= \sum_{i} \xi_{kl,ii} c_{ii} \qquad (\text{since } c_{ij} = 0 \text{ for } j \neq i) \\
&= \sum_{i} \frac{(\mathbf{X})_{kl,11}}{2(\mu + m_{\mathrm{e}})} \sum_{k'=1,2}\sum_{l'=1,2} (\mathbf{X}^{-1})_{11,k'l'} \left(\lim_{m\to 0} \tilde{a}_{k'l',ii} \right) c_{ii} \\
&= \frac{1}{2(\mu + m_{\mathrm{e}})} \sum_{i} \kappa_{ii} c_{ii} \quad \text{for} \quad \boldsymbol{\kappa} \equiv \boldsymbol{\alpha}^{\otimes 2} \lim_{m\to 0} \tilde{\mathbf{A}}.
\end{aligned} \tag{9.71}$$

Let

$$w_i \equiv \frac{\kappa_{ii}/N_i}{\sum_j \kappa_{jj}/N_j} = N_{\mathrm{s}} \frac{\kappa_{ii}}{N_i}. \tag{9.72}$$

Let $Q_{\mathrm{w}} \equiv \sum_i w_i Q_{ii}$. Then equation (9.71) implies

$$Q_{kl} \approx \frac{1}{2(\mu + m_{\mathrm{e}})} \frac{1 - Q_{\mathrm{w}}}{N_{\mathrm{s}}} \tag{9.73}$$

for N_{s} defined by equation (9.55), the error in the approximation being given by the remaining $\mathbf{O}(1)\mathbf{c}$ term in equation (9.70). Since the approximation holds for all identities within demes, Q_{kl}, it holds for Q_{w} and finally for any k, l:

$$Q_{kl} = \frac{1}{1 + 2N_{\mathrm{s}}(\mu + m_{\mathrm{e}})} + O\left(\frac{1}{N_{\mathrm{s}}}\right). \tag{9.74}$$

However, the $O(1/N_{\mathrm{s}})$ term may be large relative to the other one, particularly when the assumptions of the analysis are not well verified, that is, when transition rates between classes are not large relative to dispersal rates (Rousset 1999b). Note that, as usual for infinite island models, the within-deme identity by descent also gives F_{ST} values.

Strong Migration Limit

The "strong migration limit" between two classes of individuals can be viewed as the previous model with strong *transition* rates between

classes relative to rates of coalescence. Transition plays here the role of migration, and no dispersal between different locations is considered. Hence, we set all $m_{ij} = 0$ and $m_{\mathrm{e}} = 0$. Then equation (9.74) reduces to

$$Q_{kl} = \frac{1}{1 + 2N_{\mathrm{s}}\mu} + O\left(\frac{1}{N_{\mathrm{s}}}\right). \tag{9.75}$$

CHAPTER TEN

Fluctuating Demography: Neutral Models

In this chapter we formulate models of genetic structure in populations with fluctuations of deme size. In particular, we focus on the application of the concept of effective size in such a case. We have seen that population structure, in the sense of limited dispersal, increases effective size. But we have also seen that effective size is reduced when reproductive value differs among different demes. We will see that local demographic fluctuations also decrease effective size, but the cause of this reduction is not simply a variation in reproductive value. This effect is better described in terms of the movements of a pair of gene lineages rather than in terms of the movements of a single lineage (which determine reproductive values). When both limited dispersal and fluctuating demography are taken into account, effective size may be much lower than actual total population size (Maruyama and Kimura 1980; Hedrick and Gilpin 1997; Whitlock and Barton 1997). Some models have been considered to get a feeling of the magnitude of this reduction, which we will reconsider here. We first consider how to extend the concept of effective size in fluctuating populations (structured or not).

EFFECTIVE SIZE OF AN ISOLATED DEME

Since effective size is defined so as to describe an asymptotic rate of coalescence, we may intuitively expect it will be an average over different demographic states of the population, of the rate of coalescence.

Defining Effective Size

Consider a single population of size N following the Wright-Fisher model, except that $N = N(t)$ randomly fluctuates from generation to generation. In each generation, gene diversity decreases as $1 - Q' = [1 - 1/N(t)](1 - Q)$, so that the total decrease in gene diversity over t generations can be written as $\prod_{i=0}^{t-1}[1 - 1/N(i)]$. Wright (1938) somewhat implicitly defined effective size so that this product is $(1-1/N_{\mathrm{e}})^t$ (see also Wright 1969, p. 210). He further suggested that N_{e} could be approximated by the harmonic mean of $N(i)$, that is,

$$\frac{1}{N_{\mathrm{e}}} \approx \frac{1}{t}\sum_{i=0}^{t-1}\frac{1}{N(i)}. \tag{10.1}$$

This approximation is rather straightforward if variation in $N(i)$ is cyclic. It leads to the simple conclusion that N_{e} is lower than the average (arithmetic mean) of population size (Wright 1969).

On the other hand, if the sequence of population sizes $N(i)$ is random, a parametric definition of N_{e} must consider all possible sequences of $N(i)$. One possible definition of effective size is given by

$$1 - \frac{1}{N_{\mathrm{e}}} \equiv \lim_{t\to\infty} \mathrm{E}\left[\left(\prod_{i=0}^{t-1}[1 - 1/N(i)]\right)^{1/t}\right] \tag{10.2}$$

and another one is

$$1 - \frac{1}{N_{\mathrm{e}}} \equiv \lim_{t\to\infty} \left[\mathrm{E}\left(\prod_{i=0}^{t-1}[1 - 1/N(i)]\right)\right]^{1/t}. \tag{10.3}$$

These two definitions differ, and the harmonic mean is a better approximation for the former than for the latter. In particular, the latter may strongly depart from the harmonic mean when population size changes slowly (Iizuka 2001), while in such cases the harmonic mean is still a good approximation for the former definition (10.2) of effective size.

Which effective size do we need? The simple model detailed below shows that the effective size defined by equation (10.3) gives the asymptotic decrease in gene diversity. It is also the mutation effective size (see p. 145) such that expected gene identity Q is approximately $1/(1+2N_{\mathrm{e}}\mu)$ as considered by Motro and Thomson (1982), and Iizuka et al. (2002).

An Example

Consider the simple case where the population size follows a Markov chain with two states N_1 and N_2 (Iizuka 2001). The transition probabilities between the two states may be written

$$\begin{pmatrix} 1-q_1 & q_1 \\ q_2 & 1-q_2 \end{pmatrix}, \tag{10.4}$$

where q_j is the probability that a population of size N_j switches to the other size in the next generation. Thus, the stationary probabilities of N_1 and N_2 are

$$(\pi_1, \pi_2) = \left(\frac{q_2}{q_1+q_2}, \frac{q_1}{q_1+q_2} \right) \tag{10.5}$$

and the harmonic mean size is

$$\frac{q_2/N_1 + q_1/N_2}{q_1+q_2}. \tag{10.6}$$

In the absence of mutation, the expected gene diversity after t generations is

$$1 - Q(t) = \mathrm{E}\left(\prod_{i=0}^{t-1} [1 - 1/N(i)] \right) [1 - Q(0)]. \tag{10.7}$$

The asymptotic decrease of the expected product in this equation corresponds to definition (10.3) of effective size. It can be computed by considering recursions for the identities Q_j, given that present population size is N_j. $1 - Q_j(t)$ can be written $Z_j(t)[1 - Q(0)]$, where $Z_j(t) \equiv \mathrm{E}\left\{ \prod_{i=0}^{t-1}[1 - 1/N(i)] \middle| N(t-1) = N_j \right\}$. The recursion for Z_1 is

$$\begin{aligned} Z_1(t+1) =& \frac{\pi_1(1-q_1)(1-1/N_1)Z_1(t) + \pi_2 q_2 (1-1/N_2) Z_2(t)}{\pi_1(1-q_1) + \pi_2 q_2} \\ =& (1-q_1)(1-1/N_1)Z_1(t) + q_1(1-1/N_2)Z_2(t). \end{aligned} \tag{10.8}$$

This recursion and the similar one for $Z_2(t+1)$ can be written in matrix form, $\mathbf{Z}' \equiv \mathbf{Z}(t+1) = \mathbf{G}\mathbf{Z}$ for $\mathbf{Z} \equiv (Z_1, Z_2)^\top$ and for

$$\mathbf{G} \equiv \begin{pmatrix} (1-q_1)(1-1/N_1) & q_1(1-1/N_2) \\ q_2(1-1/N_1) & (1-q_2)(1-1/N_2) \end{pmatrix}. \tag{10.9}$$

Then the largest eigenvalue λ_1 of $\mathbf{G}$ describes the asymptotic decrease in gene diversity. Viewing $\mathbf{G}$ as a small perturbation (for large N_i's)

of the transition matrix (10.4) between population sizes, we obtain a perturbation approximation for λ_1 [using equation (A.6)]. The corresponding approximation for effective size is the harmonic mean

$$\frac{1}{N_e} \approx \sum_j \frac{\Pr(N(t) = N_j)}{N_j}. \tag{10.10}$$

The harmonic mean is thus expected to be a good approximation for λ_1 when deme sizes are large relative to the transition rates [large $N_i(q_1 + q_2)$]. On the other hand, if $q_1 + q_2$ is low relative to the $1/N_i$'s, N_e may approach the largest of the N_i's (Iizuka 2001).

With mutation, we can write the recursion for probabilities of identity in the form of equation (4.6):

$$\gamma \mathbf{1} - \mathbf{Q}' = \gamma \mathbf{G} (\mathbf{1} - \mathbf{Q}). \tag{10.11}$$

Premultiplying both sides of this equation by the left λ_1 eigenvector $\mathbf{y}_1$ of $\mathbf{G}$, we obtain the same result (9.3) as in models without fluctuations. This shows that λ_1 is the mutation effective size.

To better understand the significance of these results, consider a realization of the above model in the case $N_1 = 10$, $N_2 = 1{,}000$, and $q_j = 1/N_j$ (fig. 10.1). This figure shows the expected diversity in each generation, that is the expected diversity $H(t) \equiv 1 - Q(t)$, given the past sequence of population sizes. Equation (10.11) only gives the average values H_j of $H(t)$, given the present population size $N(t) = N_j$. In the present case, the population experiences bottlenecks ($N_1 = 10$) for short periods with an expected length of 10 generations. Hence, $H(t)$ fluctuates widely, as it quickly drops during each bottleneck. Recovery of diversity after each bottleneck depends on the mutation rate, here $\mu = 10^{-3}$. In a given generation, the realized gene diversity in the population would further fluctuate around the average H_t as a result of the randomness, for a given history of population sizes, of the mutation events and of the genealogical tree of the population.

STRUCTURED POPULATIONS

Two kinds of demographic fluctuations are often distinguished (e.g., Caswell 2001), environmental stochasticity, which refers to the effects of demographic forces that affect several individuals at a time through a common environment, and demographic stochasticity, which refers

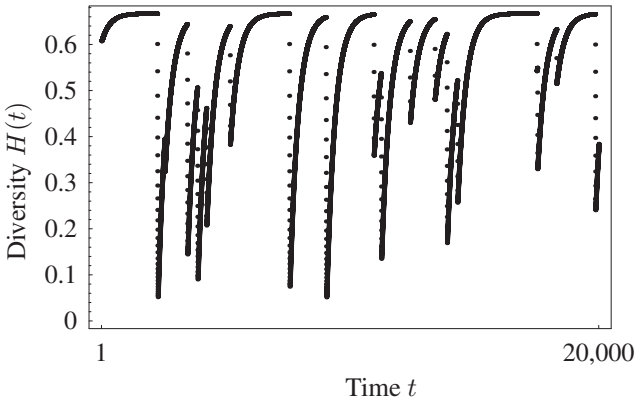

FIGURE 10.1. Gene diversity under random fluctuations of population size. This figure represents 20,000 generations of evolution of expected gene diversity in the two-state model with $N_1 = 10$, $N_2 = 1{,}000$, $q_j = 1/N_j$, and $\mu = 10^{-3}$. Diversity drops quickly when $N = 10$ and recovers more slowly when $N = 1{,}000$. The eigenvalue effective size is 657 and the harmonic effective size is 505.

to the effects of demographic forces assumed to affect each individual independently, for example, when the death of different individuals is assumed independent. In this section, we derive expressions for a "metapopulation" model with both forms of stochasticity. It allows the investigation of the consequences of different modes of population regulation. This model is very close to the one of Whitlock and Barton (1997).

Intuitively, we expect effective size to be expressed in terms of the probabilities that gene lineages are in some demes in an ancestral generation (i.e., reproductive values), times the probability of coalescence given the position of gene lineages. However, a complication appears, as the probability that two lineages are in the same deme is not the square of reproductive values, contrary to the models of chapter 9.

A Metapopulation Model

We consider a finite number n_{d} of demes with different numbers of haploid adults, $(N_1, \ldots, N_{n_{\mathrm{d}}})$, which collectively define the demographic state $\mathbf{N}$ of the population. This state is random, following

a Markov chain defined by the transition probabilities $\Pr(\mathbf{N}'|\mathbf{N})$ between states at generations t and $t+1$.

Let $f_{ij}(\mathbf{N},\mathbf{N}')$ be the probability that an adult sampled in deme i had its parent in deme j, given the demographic states $\mathbf{N}$ and $\mathbf{N}'$. If we assume Poisson distributions of juveniles, and that juveniles disperse independently from each other, finite fecundity can be taken into account, and the f_{ij}'s are actually independent of the demographic state $\mathbf{N}'$. Thus, they form a matrix $\mathbf{F}(\mathbf{N})$. Further, the probability that a pair of genes in demes i and k had their parent(s) in demes j and l is the product of probabilities f_{ij} and f_{kl} of the origin of each gene. In other words, juveniles are sampled with replacement with respect to their origin.

Conditional on the parental sizes $\mathbf{N}$ and parental identities $\mathbf{Q} \equiv (Q_{11}, Q_{12}, \ldots, Q_{n_\mathrm{d} n_\mathrm{d}})^\top$,

$$\mathbf{Q}' = \gamma \mathbf{A}(\mathbf{N})(\mathbf{Q} + \mathbf{c}), \tag{10.12}$$

where $\mathbf{c}$ is a column vector with elements $(1-Q_{ii})/N_i$ if $N_i > 0$, and 0 otherwise, and $\mathbf{A}(\mathbf{N})$ is a matrix with elements $a_{ik,jl} = f_{ij}(\mathbf{N}) f_{kl}(\mathbf{N})$.

Then the equilibrium probabilities of identity $\mathbf{Q}(\mathbf{N})$, given the actual deme sizes $\mathbf{N}$, are given by the set of recursions

$$\mathbf{Q}(\mathbf{N}') = \sum_{\mathbf{N}} \Pr(\mathbf{N}|\mathbf{N}')\mathbf{A}(\mathbf{N})[\mathbf{Q}(\mathbf{N}) + \mathbf{c}(\mathbf{N})], \tag{10.13}$$

where $\Pr(\mathbf{N}|\mathbf{N}')$ is the probability of demographic state $\mathbf{N}$ in the parental population given the demographic state $\mathbf{N}'$ in the offspring generation.

As shown in chapter 9, we can deduce the effective size by considering a weighted average of the probabilities of identity, where the weights are defined so that a simple expression results for the decrease in weighted identity over one generation. Here an appropriate definition of weights involves the set of vectors $\boldsymbol{\kappa}(\mathbf{N}) \equiv [\kappa_{ij}(\mathbf{N})]$ solutions of the recursions

$$\boldsymbol{\kappa}(\mathbf{N}) = \sum_{\mathbf{N}'} \Pr(\mathbf{N}'|\mathbf{N})\boldsymbol{\kappa}(\mathbf{N}')\mathbf{A}(\mathbf{N}). \tag{10.14}$$

$\mathbf{A}(\mathbf{N})$ describes the backwards movements of a pair of gene lineages, and hence the elements ij of the $\boldsymbol{\kappa}(\mathbf{N})$ vectors may be understood as the probabilities, given $\mathbf{N}$, that two gene lineages are in demes i and j, ignoring the possibility of coalescence of gene lineages. By the same argument, the reproductive values [i.e., the probabilities $\alpha_i(\mathbf{N})$ that an

ancestral lineage was in deme i, given $\mathbf{N}$], form vectors $\boldsymbol{\alpha}(\mathbf{N})$, which are solutions of the recursions

$$\boldsymbol{\alpha}(\mathbf{N}) = \sum_{\mathbf{N}'} \Pr(\mathbf{N}'|\mathbf{N})\boldsymbol{\alpha}(\mathbf{N}')\mathbf{F}(\mathbf{N}). \tag{10.15}$$

Note that in general, $\kappa_{ii}(\mathbf{N}) \neq \alpha_i^2(\mathbf{N})$.

The appropriate weights for the probabilities of identity $\mathbf{Q}(\mathbf{N})$ are $\Pr(\mathbf{N})\boldsymbol{\kappa}(\mathbf{N})$, because these probabilities obey a recursion in terms of successive values of the weighted average $Q_{\boldsymbol{\kappa}} \equiv \sum_{\mathbf{N}} \Pr(\mathbf{N})\boldsymbol{\kappa}(\mathbf{N}) \cdot \mathbf{Q}(\mathbf{N})$:

$$Q'_{\boldsymbol{\kappa}} = \sum_{\mathbf{N}'} \Pr(\mathbf{N}')\boldsymbol{\kappa}(\mathbf{N}') \cdot \sum_{\mathbf{N}} \Pr(\mathbf{N}|\mathbf{N}')\mathbf{A}(\mathbf{N})[\mathbf{Q}(\mathbf{N}) + \mathbf{c}(\mathbf{N})] \tag{10.16}$$

$$= \sum_{\mathbf{N}} \Pr(\mathbf{N})\boldsymbol{\kappa}(\mathbf{N}) \cdot [\mathbf{Q}(\mathbf{N}) + \mathbf{c}(\mathbf{N})] \tag{10.17}$$

$$= Q_{\boldsymbol{\kappa}} + \sum_{\mathbf{N}} \Pr(\mathbf{N})\boldsymbol{\kappa}(\mathbf{N}) \cdot \mathbf{c}(\mathbf{N}). \tag{10.18}$$

Here each element $\kappa_{ii}(\mathbf{N})$ of $\boldsymbol{\kappa}(\mathbf{N})$ is the probability that the two ancestral lineages are in gametes produced by adults in deme i, and each element $c_{ii}(\mathbf{N})$ of $\mathbf{c}(\mathbf{N})$ is the gain in genetic identity when these two gametes were produced by the same adult. With respect to demes with only one adult, remember that recursions such as (10.12) are most generally valid in terms of the probability of identity after dispersal but before population regulation ($Q \equiv Q^{\mathrm{D}}$, see p. 24). Thus, Q_1 is meaningful: there may be several juveniles in a deme where there will be only one adult.

Then we can generalize equation (9.44) as

$$\lambda_1 = \frac{1 - Q'_{\boldsymbol{\kappa}}}{1 - Q_{\boldsymbol{\kappa}}} = 1 - \sum_{\mathbf{N}} \Pr(\mathbf{N}) \sum_i \frac{\kappa_{ii}(\mathbf{N})}{N_i} \frac{1 - Q_{ii}(\mathbf{N})}{1 - Q_{\boldsymbol{\kappa}}} \tag{10.19}$$

which yields

$$\frac{1}{N_{\mathrm{e}}} = \sum_{\mathbf{N}} \Pr(\mathbf{N}) \sum_i \frac{\kappa_{ii}(\mathbf{N})}{N_i}(1 - F_i) \tag{10.20}$$

for $(1 - F_i) \equiv \lim_{t\to\infty}[1 - Q_{ii}(\mathbf{N})]/(1 - Q_{\boldsymbol{\kappa}})$. The F_i's are F_{ST}-like parameters with the same properties as the F_i's considered in populations with stable demography (p. 160). In particular, in infinite island models, they can be assimilated to probabilities of identity.

APPROXIMATIONS

In a metapopulation, the number of demographic states $\mathbf{N}$ will usually be far too large for expression (10.20) to be useful for computation. However, we can approximate the weights κ_{ii} and the spatial structure parameters F_i using infinite island approximations. The results of this section are equivalent to those of Whitlock and Barton (1997), except for their ν_{ij} terms, which should not appear in the equations (this is most easily seen by considering the island model with constant deme size as a special case of the metapopulation model).

The recursion for identities (10.13) simplifies as follows. The probabilities of identity by descent Q_n for demes of size n are given by the recursion

$$\mathbf{Q} = \tilde{\mathbf{A}}(\mathbf{Q} + \mathbf{c}), \tag{10.21}$$

where $\mathbf{Q}$ is the column vector (Q_n), $\mathbf{c}$ is a vector with elements $(1 - Q_n)/n$ if $n > 0$ and 0 otherwise, and $\tilde{\mathbf{A}}$ is a matrix with elements $\tilde{a}_{n',n}$ obtained as the backward transition probability $v(n|n')$ that a deme of size n' derives from a deme of size n, times the probability that both genes are descendants of parents in the deme, given n. The latter probability may be written $(1 - m_n)^2$, where m_n is the immigration rate (probability that individuals at $t + 1$ are of immigrant origin) in a deme that was of size n at t.

The expression (10.20) for effective size is approximated as follows. For a deme i of size n, F_i is given by Q_n. The probability $\kappa_{ii}(\mathbf{N})$ that two gene lineages are in the same deme given $\mathbf{N}$ is of the order of $1/{n_\mathrm{d}}^2$. Hence, a nontrivial limit when $n_\mathrm{d} \to \infty$ is obtained only for values scaled to ${n_\mathrm{d}}^2$. Below we consider the limit ρ_n of ${n_\mathrm{d}}^2\kappa_{ii}$ for a deme i of size n, so that the effective size can be written

$$\frac{1}{N_\mathrm{e}} \sim \frac{1}{n_\mathrm{d}} \sum_n \Pr(n)\rho_n \frac{1 - Q_n}{n}, \tag{10.22}$$

where $\sim$ means that higher-order terms in $1/n_\mathrm{d}$ are neglected and $\Pr(n)$ is the equilibrium probability that a deme is of size n. In this expression, $\Pr(n)\rho_n/n_\mathrm{d}$ is the probability that two lineages are both in anyone of the demes of size n.

Recursions for ρ_n can be deduced in a relatively intuitive and terse way from those for the total reproductive values α_n in demes of size n. As usual, these reproductive values are obtained as a left eigenvector of

the matrix of backward transition probabilities, denoted $l_{n'n}$, of gene lineages between successive deme sizes n and n'. When $n_{\mathrm{d}} \to \infty$, these transition probabilities may be written

$$l_{n'n} \sim v(n|n')(1-m_n) + \sum_q v(q|n')m_{q,n} \tag{10.23}$$

where $m_{q,n}$ is the probability that a lineage sampled in a deme formerly of size q is an immigrant from a deme formerly of size n (see Rousset and Ronce in press for a more concrete computation). In particular, given a deme size n' at $t+1$, its size at t was n with probability $v(n|n')$, and the probability that a gene is not immigrant is then $(1-m_n)$. In the recursion for α_n between t and $t+1$, nondispersed genes therefore contribute a term $(1-m_n)\sum_{n'} v(n|n')\alpha_{n'}$.

By definition of κ_{ii}, the computation of ρ_n ignores coalescence events. Two independently evolving ancestral lineages would be in the same deme, of size n at t, with probability $\sim \{\alpha_n/[n_{\mathrm{d}}\Pr(n)]\}^2$, the square of the probability that each is in a particular deme of size n. However, when two lineages are in the same deme, they are not changing deme type independently: if neither was immigrant, the deme size in the previous generation was the same for both lineages. Hence, the recursion for ρ_n differs from that for $[\alpha_n/\Pr(n)]^2$ as follows.

As noted above, in the recursion for α_n between t and $t+1$, nondispersed genes contribute a term $(1-m_n)\sum_{n'} v(n|n')\alpha_{n'}$. In the recursion for $\alpha_n/\Pr(n)$, this contributes a term

$$(1-m_n)\sum_{n'} v(n|n')\frac{\alpha_{n'}}{\Pr(n)} = (1-m_n)\sum_{n'} u(n'|n)\frac{\alpha_{n'}}{\Pr(n')} \tag{10.24}$$

where $u(n'|n) = v(n|n')\Pr(n')/\Pr(n)$ is the forward transition probability between demes sizes n and n'. In a recursion for $[\alpha_n/\Pr(n)]^2$, this would contribute the square of this term, which needs to be replaced in the recursion for ρ_n by $(1-m_n)^2\sum_{n'} u(n'|n)\rho_{n'}$.[1] Other terms in the recursion for $[\alpha_n/\Pr(n)]^2$ are also affected by the nonindependence of gene lineages, but the effects are neglected because they are of higher order in $1/n_{\mathrm{d}}$. Hence, taking the difference between

[1]This term could also be obtained by considering the contribution to the probability $\sim \Pr(n)\rho_n/n_{\mathrm{d}}$ that two lineages are both in some deme of size n, by gene lineages which are both in some deme of size n' in the next generation [probability $\sim \Pr(n')\rho_{n'}/n_{\mathrm{d}}$]. This contribution occurs when this deme was of size n in the previous generation [probability $v(n|n')$] and when the gene lineages were nonimmigrants.

recursions for ρ_n and for $[\alpha_n/\Pr(n)]^2$, we find that the ρ_n's are the solutions of a system of linear equations

$$\rho_n \sim \left(\frac{\alpha_n}{\Pr(n)}\right)^2 + (1-m_n)^2\left[\sum_{n'} u(n'|n)\rho_{n'} - \left(\sum_{n'} u(n'|n)\frac{\alpha_{n'}}{\Pr(n')}\right)^2\right]. \tag{10.25}$$

Alternatively, one can compute the products $\rho_n(1-Q_n)$, which are the scaled probabilities that two ancestral lineages, considered before population regulation, are distinct and both in a given deme of size n after regulation. The recursion for these products must take coalescence into account, and is given by the above recursion modified by adding coefficients $(1-1/n')$ to $\rho_{n'}$.

These results allow evaluation of effective size under different demographic scenarios. For example, we may consider the effect of fecundity (number of juveniles produced before population regulation) and dispersal rate on the ratio $N_{\mathrm{e}}/N_{\mathrm{T}}$ of effective size to actual population size. Higher fecundity, by allowing faster growth of a deme colonized by a single individual, is expected to reduce $N_{\mathrm{e}}/N_{\mathrm{T}}$. But at the same time it allows larger numbers of colonizers, which may increase the effective size. Thus, a low $N_{\mathrm{e}}/N_{\mathrm{T}}$ is expected when backwards dispersal rates are low and fecundity is high. Numerical analysis of equation (10.22) confirms such trends (fig. 10.2). Whitlock and Barton (1997, fig. 4) also present cases of tenfold-times reduction or increase of $N_{\mathrm{e}}/N_{\mathrm{T}}$, depending on whether fecundity is high or low.

Propagule Models

In the above model, colonizers were independently sampled from all other demes in the total population. Some models have been formulated to examine the effects of different origins of colonizers (Slatkin 1977; Wade and McCauley 1988). They ignore demographic stochasticity, but incorporate the possibility that most colonizers come from a single deme. In the "migrant pool" model, recolonizers come from the whole metapopulation: in the "propagule pool" model, they preferen-

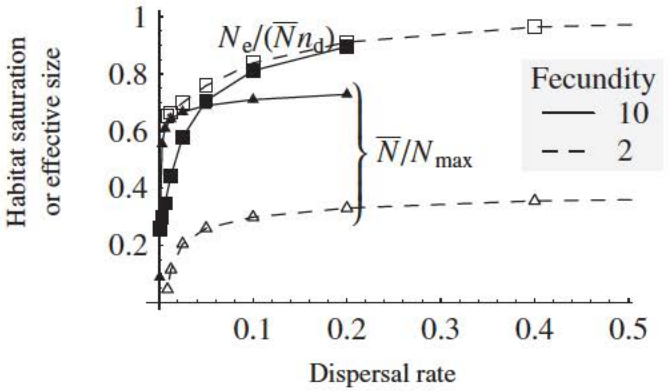

FIGURE 10.2. Habitat saturation and effective size. The demographic model considered in these computations assumed a probability 0.1 of environmental local extinction, a maximum number of adults per patch $N_{max} = 35$, a Poisson distributed fecundity with mean 2 or 10 juveniles produced, and no survival cost of dispersal. The number of adults was determined from the number of juveniles j competing for a deme by assuming that each juvenile independently had a probability $1/(1 + 0.03j)$ of survival. The resulting number of adults was set to N_{max} if it exceeded it. Habitat saturation (△, ▲) is defined as mean deme size $\bar{N}$ relative to maximum deme size N_{max}. Effective size (□, ■) is given relative to the actual total size $\bar{N}n_d$.

tially come from a a pool of individuals formed in a single population after the dispersal of juveniles.

In the absence of extinction, events occur in the following order: gamete production, dispersal, and population regulation, where N adult offspring survive. In each generation, any deme can independently become extinct with some probability e, and is immediately recolonized so that there are always N adults in all demes. Thus, there are two rounds of dispersal and reproduction within the time only one round is considered for demes that do not become extinct. This assumption was made to simplify the mathematics, but could be relaxed (Whitlock et al. 1993; Rousset 2003a).

By allowing an effectively infinite fecundity and a low number of recolonizers, these models allow for extreme reduction in effective size. If the number of colonizers k is small, the identity among their N descendants may be higher than the identity among adults in other demes. In this case, extinction has an additional effect of reducing diversity

within demes and increasing differentiation (Slatkin 1977; Wade and McCauley 1988). The simplest way to derive effective size here is to generalize the argument developed on p. 161. Accordingly, we compute effective size as the product of the probability that two lineages in different demes originate from gametes from the same deme in the previous generation, times the probability that lineages in such gametes coalesce rather than separate again in different demes. As in the case without extinctions, the latter probability can be approximated by the identity among gametes produced by adults within a deme in the infinite island model, $Q^{\mathrm{R}} \equiv 1/N + (N-1)F_{\mathrm{ST}}/N$. The former probability is computed as follows. With local extinctions, the two demes where the genes are sampled may have both become extinct in the previous generation (probability e^2) or neither became extinct [probability $(1-e)^2$], or only one did [probability $2e(1-e)$]. If neither became extinct, the probability that the two lineages come from a single deme in the previous generation is

$$\frac{1-(1-m)^2}{(1-e)n_{\mathrm{d}}} + O(1/{n_{\mathrm{d}}}^2), \tag{10.26}$$

where $(1-e)n_{\mathrm{d}}$ is the number of parental demes that contribute to the next generation. If one or both demes became extinct, the probability is simply $1/[(1-e)n_{\mathrm{d}}]$.

Taking the different cases into account, the overall rate of coalescence is

$$\frac{1}{N_{\mathrm{e}}} \sim \frac{1}{n_{\mathrm{d}}}\left((1-e)[1-(1-m)^2] + 2e + \frac{e^2}{1-e}\right) Q^{\mathrm{R}} \tag{10.27}$$

$$\frac{1}{N_{\mathrm{e}}} \sim \frac{1}{n_{\mathrm{d}}}\frac{1-(1-m)^2(1-e)^2}{1-e} Q^{\mathrm{R}}, \tag{10.28}$$

where $\sim$ means that the computation is correct to leading order in $1/n_{\mathrm{d}}$.

This result only slightly differs from the previous analysis of effective size in such models by Pannell and Charlesworth (1999), but differs more from that of Whitlock and Barton (1997), as discussed in Rousset (2003a). For $n_{\mathrm{d}} \to \infty$, Q^{R} can be computed, for example, as in Wade and McCauley (1988), or can be deduced from the recursion

$$Q' = (1-e)(1-m)^2 Q^{\mathrm{R}} + e\left[\frac{1}{k} + \left(1-\frac{1}{k}\right)\phi Q^{\mathrm{R}}\right], \tag{10.29}$$

where $\phi = 0$ for the migrant pool model and $\phi = (1-m)^2$. It yields

$$Q^{\mathrm{R}} = \frac{1/N + e/k - e/(kN)}{1-(1-1/N)[(1-m)^2(1-e) + e\phi(1-1/k)]} \tag{10.30}$$

at equilibrium. These results show that propagule size k and probability of common origin ϕ affect effective size only through their effects on Q^{R}, that is, on F_{ST}. As could be intuitively expected, lower k and higher ϕ reduce the effective size.

The quantitative implications of these results may be appreciated by applying them to two sets of estimates of demographic parameters from the literature (Whitlock 1992a; Ingvarsson et al. 1997). In both cases, estimates of F_{ST} from genetic data are roughly consistent with expectations from estimates of demographic parameters, and the ratios $N_{\mathrm{e}}/(Nn_{\mathrm{d}})$ are 0.67 and 0.35 (Rousset 2003a). Thus, the overall effect of population structure seems to be a moderate reduction of effective size. These examples show moderate bottlenecks at recolonization ($k/N = 0.49$ and 0.36), and substantially larger reduction in effective size may occur for lower numbers k of colonizers relative to N.

INFERENCES

The results of this chapter confirm the intuition that demographic fluctuations reduce the metapopulation effective size and demonstrate relatively simple coalescent arguments that confirm this conclusion. Reanalysis of some data suggest that the reduction in not necessarily marked.

An observer contemplating an isolated deme evolving as shown in figure 10.1 will find that the eigenvalue effective size is not predictive of the observed gene diversity after a rare bottleneck. This is to be expected, since the eigenvalue effective size is computed by taking into account all possible histories of the population. This observer might conclude that some other definition of effective size would be more appropriate. Nevertheless, expected diversity after a bottleneck will depend on the probabilities of different past histories of the population, and this is what is taken into account by N_{e}. The problem with using *some* effective size here is that this concept appears useful mainly when one can distinguish fast processes, such as fast fluctuations of deme size, from slow processes, such as the slow rate of coalescence measured by N_{e}. Thus, figure 10.1, where fluctuations of deme size are not fast, may actually illustrate cases where the concept of effective size is not useful for predicting diversity in a given population. This further implies that it is practically impossible to estimate the effec-

tive size of a population evolving as in figure 10.1 from observations over a few generations. Gene diversity will reflect recent fluctuations in population size, and different loci will be identically affected by them.

In the metapopulation example, the fast process is the transition of gene lineages between different deme sizes through dispersal of the lineage or through change in deme size; the slow process is the drift process at the level of the total population. Slow fluctuations in the total number of demes would restrict the appropriateness of N_{e} for the analysis of expected levels of gene diversity in the total population, but N_{e} would remain appropriate for the analysis of processes that depend on recent events (e.g., current adaptive processes).

For a single deme with fast fluctuations, the harmonic mean is a good approximation for effective size. Vucetich et al. (1997) estimated the ratio of harmonic effective size to arithmetic mean effective size for diverse data sets. Values ranged from 7.5×10^{-4} to 0.98, with larger values for vertebrate than invertebrate populations. An important factor reducing this ratio is high fecundity, as it allows large fluctuations of population size from generation to generation. The metapopulation effective size is not simply comparable to a harmonic effective size. It depends on the number of colonizers and on the rate of local growth of deme size, and therefore on an interaction between dispersal rate and fecundity. The metapopulation $N_{\mathrm{e}}/N_{\mathrm{T}}$ ratio should therefore also show taxon-specific trends, but its magnitude in realistic settings remains to be evaluated.

These ratios are relative to the number of adults. One can obtain impressively lower ratios when comparing to the number of juveniles, but these are somehow meaningless; for example, in Wright's island model the ratio of N_{e} to the number of juveniles is essentially zero, since the number of juveniles is "infinite."

CHAPTER ELEVEN

Selection in Class-Structured Populations

In a class-structured population, it is clearly better to have one offspring with a high probability to be the ancestor of the future population than one offspring with low potential to be the ancestor. Then a measure of the effects of selection on a trait may be given as a weighted average of effects on the number of offspring in different classes. The weights are the reproductive values defined in chapter 9. This result has long been recognized but little applied in a spatial context (Houston and McNamara 1992; Kawecki and Stearns 1993), much as it has been little used in conjunction with kin selection theory (Frank 1998).

Expressions for fitness measures in class-structured populations with stable demography have essentially identical forms whether the class structure represents a fine grained spatial heterogeneity or has no spatial component. In this chapter, we present a slight generalization that applies to more general migration patterns than infinite island models. We build on chapter 6 to derive the fitness measures concisely: an expression for allele frequency change over one generation directly translates into an expression for the fitness measure. We also review what is known about spatial variation in reproductive values in natural populations.

These first results assume a stable demography. However, local extinctions and recolonizations have important consequences on phenotypic evolution. For example, it has long been recognized that local extinctions promote the evolution of dispersal (e.g., van Valen 1971; Gadgil 1971; Levin et al. 1984). They should also be important for the evolution of other life-history traits that have been considered more recently (e.g., McCauley et al. 1990; Pannell 2000; de Jong et al. 2000;

Ronce et al. 2000; Ronce and Olivieri in press). Some numerical methods have been developed to compute fitness measures in metapopulations (Metz and Gyllenberg 2001). Noting that metapopulations may also be viewed as populations structured in different classes of demes, we extend to them the fitness measure for class-structured populations. We consider in particular the case, which has not been previously analyzed, where demographic fluctuations depend on the trait under selection (Rousset and Ronce in press). These results are compared to analytical approaches for fitness in class-structured populations.

The fitness measures considered in this chapter are both more easily computed and more relevant when there is some large-scale spatial homogeneity. Although the results given are correct under more general conditions, they are not useful when there are large-scale spatial inhomogeneities of selection pressures, resulting, for example, in spatial frequency clines of selected alleles.

FITNESS MEASURES

As in chapter 5, we seek effects on the probability of fixation of mutants. There we used the fact that, in a neutral model, expected allele frequency in the offspring generation is allele frequency in the parental generation [equation (6.2)]. In a class-structured population, the allele frequency with the same property is a weighted average of allele frequencies in the different classes. The appropriate weights are given by class reproductive values, as the following shows.

Stable Demography

First, we write expected allele frequencies in the offspring generation as

$$\mathrm{E}(\mathbf{p}'|\mathbf{p}) = \mathbf{F}\mathbf{p} \tag{11.1}$$

for some matrix $\mathbf{F} = (f_{ij})$. For equation (11.1) to be a correct description of expected allele frequency, $f_{ij}p_j$ must be the probability that a gene in class i is a copy of a gene from any of the A parents in class j. This probability can be expressed as a function of the expected number of i offspring of an A-bearing parent of class j, that is, in terms of a

fitness function $w_{ij}(z_\bullet, \mathbf{z})$, where $z_\bullet$ is the focal individual fitness and $\mathbf{z} \equiv (z_1, \ldots, z_{n_c})$ are the average phenotypes in the different classes:

$$f_{ij}(z_\bullet, \mathbf{z}) = \frac{N_j}{N_i} w_{ij}(z_\bullet, \mathbf{z}). \tag{11.2}$$

f_{ij} is N_j times the probability that a gene in class i comes from a given A adult in class j.

We assume that the $\mathbf{F}$ matrix is constant through time. Let $^\circ$ generically denote variables and expectations evaluated in the neutral model. $\mathbf{F}^\circ$ is the backwards transition matrix between classes, so the class reproductive values are a left eigenvector of $\mathbf{F}^\circ$: $\boldsymbol{\alpha} = \boldsymbol{\alpha}\mathbf{F}^\circ$. We consider the weighted allele frequency, $\boldsymbol{\alpha}\cdot\mathbf{p} \equiv \sum_{i=1}^{n_c} \alpha_i p_i$. This weighted allele frequency does not change in expectation in the neutral model:

$$\mathrm{E}^\circ(\boldsymbol{\alpha}\cdot\mathbf{p}'|\mathbf{p}) = \boldsymbol{\alpha}\mathbf{F}^\circ\mathbf{p} = \boldsymbol{\alpha}\cdot\mathbf{p}. \tag{11.3}$$

Then we can extend equation (6.3) as

$$\phi \equiv \frac{d\Pi}{d\delta} = \mathrm{E}^\circ\left(\sum_{t=0}^{\infty} \boldsymbol{\alpha}\cdot\frac{d\{\mathrm{E}[\mathbf{p}(t+1)|\mathbf{p}(t)] - \mathbf{p}(t)\}}{d\delta}\right), \tag{11.4}$$

where the sum is over generations, starting with the occurrence of the mutation ($t = 0$). From equations (11.1) and (11.3) this can be written

$$\phi = \mathrm{E}^\circ\left(\sum_{t=0}^{\infty} \boldsymbol{\alpha}\cdot\frac{d\mathbf{F}}{d\delta}\mathbf{p}(t)\right), \tag{11.5}$$

and by "direct fitness" expansion of each element of $d\mathbf{F}/d\delta$, this is

$$\phi = \mathrm{E}^\circ\left[\sum_{t=0}^{\infty}\sum_{i=1}^{n_c}\alpha_i\sum_{j=1}^{n_c}\left(\frac{\partial f_{ij}}{\partial z_\bullet} + \sum_{k=1}^{n_c}\frac{\partial f_{ij}}{\partial z_k}p_k\right)p_j\right]. \tag{11.6}$$

Following the same argument as in chapter 6, this can be expressed in terms of coalescence times:

$$\phi = -\sum_{i=1}^{n_c}\alpha_i\sum_{j=1}^{n_c}\sum_{k=1}^{n_c}\frac{\partial f_{ij}}{\partial z_k}\frac{T_{jk}}{N_\mathrm{T}}, \tag{11.7}$$

where T_{jk} is the expected coalescence time of genes sampled in parental class j and in actor category k, and N_T is the total population size $\sum_i N_i$. Hence, ϕ can be computed as $\lim_{\mu\to 0} S/(2N_\mathrm{T}\mu)$, where

$$S \equiv \sum_{i=1}^{n_c}\alpha_i\sum_{j=1}^{n_c}\left(\frac{\partial f_{ij}}{\partial z_\bullet} + \sum_{k=1}^{n_c}\frac{\partial f_{ij}}{\partial z_k}Q_{jk}\right), \tag{11.8}$$

(Leturque and Rousset 2002). Here Q_{jk} is the probability of identity of genes sampled in parental class j and in actor category k.

This result is formally similar to previous ones for various slightly different biological scenarios (e.g., Taylor 1990; Taylor and Frank 1996; Frank 1998). Rewriting f_{ij} as $N_j w_{ij}/N_i$, we see that it weights effects on fitness w_{ij} by the number of parents N_j subject to such effects and by the individual reproductive value α_i/N_i of the type of offspring considered. This result also shows that the fitness measure will be independent of the reproductive values if for any class j of parents, $\sum_{k=\bullet,1}^{n_c} \partial f_{ij}/\partial z_k Q_{jk}$ is independent of i. That is, when the distribution of offspring reproductive value is not affected by the trait under selection ($\partial f_{ij}/\partial z_k$ independent of i), fitness effects depend only on the effects on the total number of offspring, independently of the reproductive value of different classes of offspring (Taylor 1990).

Fluctuating Demography

Demographic fluctuations imply that the functions f_{ij}, standing for the probability that a lineage in deme i originates from deme j, may differ from generation to generation. The demographic fluctuations may themselves depend on the trait under selection, for example, when this trait is the dispersal rate and the recolonization probability of an extinct deme depends on the immigration rate. To see what it entails for fitness measures, we reconsider the metapopulation model of chapter 10, where deme sizes may fluctuate. Then $\mathbf{F} \equiv (f_{ij})$ is a function $\mathbf{F}(\mathbf{N})$ of the different deme sizes $\mathbf{N} \equiv (N_1, \ldots, N_{n_d})$ (we use the same notation as in chapter 10). Equation (11.2) generalizes into

$$f_{ij}(z_\bullet, \mathbf{z}, \mathbf{N}) = \frac{N_j}{N_i'} w_{ij}(z_\bullet, \mathbf{z}, \mathbf{N}, \mathbf{N}'), \tag{11.9}$$

where N_j and N_i' are the size of the demes j and i, considered in the parental and offspring generation, respectively. The fitness function $w_{ij}(z_\bullet, \mathbf{z}, \mathbf{N}, \mathbf{N}')$ is now defined to describe the expected number of i offspring of a parent in deme j, given the demographic states in successive generations $\mathbf{N}, \mathbf{N}'$. Given $\mathbf{N}$, we consider the average allele frequency $\boldsymbol{\alpha}(\mathbf{N}) \cdot \mathbf{p}$, where the vectors $\boldsymbol{\alpha}(\mathbf{N})$ are reproductive values defined as solutions of equation (10.15). Expected allele frequency in the offspring generation can be expressed in terms of the different pos-

sible demographies $\mathbf{N}'$ in the offspring generation, as

$$\mathrm{E}[\boldsymbol{\alpha}(\mathbf{N}') \cdot \mathbf{p}'|\mathbf{p}, \mathbf{N}] = \sum_{\mathbf{N}'} \boldsymbol{\alpha}(\mathbf{N}') \Pr(\mathbf{N}'|\mathbf{N}) \mathbf{F}(\mathbf{N}) \mathbf{p}. \tag{11.10}$$

In the neutral model, this simplifies as

$$\mathrm{E}^{\circ}[\boldsymbol{\alpha}(\mathbf{N}') \cdot \mathbf{p}'|\mathbf{p}, \mathbf{N}] = \boldsymbol{\alpha}(\mathbf{N}) \cdot \mathbf{p} \tag{11.11}$$

by definition of $\boldsymbol{\alpha}(\mathbf{N})$. Thus, the expected weighted allele frequency does not change over generations.

As before, we obtain the fitness measure by summing first-order effects of selection on allele frequency change over generations, starting from the occurrence of a mutant. The first-order effects are deduced from equation (11.10) for allele frequency change, which is an expectation conditional not only on allele frequencies, but also on the demographic state $\mathbf{N}$ in the parental generation. Hence, the expected sum of derivatives is obtained as an expectation over the distribution of $\mathbf{N}$. If demographic fluctuations depend on the trait under selection, $\Pr(\mathbf{N}'|\mathbf{N})$ depends on phenotypes z_k. Hence, first-order effects are given by the derivatives of the products $\Pr(\mathbf{N}'|\mathbf{N})\mathbf{F}(\mathbf{N})$, rather than simply of $\mathbf{F}$ as previously considered. After some algebra, one obtains $\phi = \lim_{\mu \to 0} S/[2\mathrm{E}(N_{\mathrm{T}})\mu]$, where

$$S = \sum_{\mathbf{N}} \Pr(\mathbf{N}) \sum_{\mathbf{N}'} \sum_{i} \alpha_i(\mathbf{N}') \sum_{j} \sum_{k} \frac{\partial \Pr(\mathbf{N}'|\mathbf{N}) f_{ij}(\mathbf{N})}{\partial z_k} Q_{jk|\mathbf{N}}, \tag{11.12}$$

where $\Pr(\mathbf{N})$ is the probability of demographic state $\mathbf{N}$, and $Q_{jk|\mathbf{N}}$ is the probability of identity of genes sampled in parental class j and in actor category k, given $\mathbf{N}$. This expression can be simplified when dispersal follows an infinite island model (Rousset and Ronce in press), assuming that a stationary distribution of deme size is attained in this case.

S may be decomposed as

$$S = S_f + S_{\Pr}, \tag{11.13}$$

where S_f involves derivatives of the f_{ij}'s and $S_{\Pr}$ involves derivatives of the $\Pr(\mathbf{N}'|\mathbf{N})$'s. The expression for S_f has the same interpretation as the expression for S in models with stable demography [equation (11.8)]. $S_{\Pr}$ describes the additional selective effects, due to the effects on an individual's offspring that result from changes in the probabilities of different future demographic states $\mathbf{N}'$ of the population. A

higher probability of some demographic state matters, because the reproductive value of offspring depends on $\mathbf{N}'$. Hence, the variations of $\Pr(\mathbf{N}'|\mathbf{N})$ are weighted by reproductive values.

Fitness as Eigenvalue

Many works compute fitness as follows. One writes the recursion for the relative numbers of A copies in different classes as

$$\mathbf{v}' = \mathbf{S}\mathbf{v} \tag{11.14}$$

for some matrix $\mathbf{S}$ and for $\mathbf{v} = (v_1, \ldots, v_{n_c})^\top$ being the vector of the relative numbers of A copies in the different classes. Strictly speaking, $\mathbf{S}\mathbf{v}$ may be only the expectation of $\mathbf{v}'$ at t as a function of the values $\mathbf{v}$ at $t-1$, but stochastic fluctuations are neglected.

If, in addition, $\mathbf{S}$ is constant across generations, then equation (A.9) implies that after a large number of generations,

$$\mathbf{v}(t) = \mathbf{S}^t\mathbf{v}(0) \approx x\lambda^t\mathbf{e}, \tag{11.15}$$

where x depends on $\mathbf{v}(0)$ but is independent of t, λ is the largest eigenvalue of $\mathbf{S}$, and $\mathbf{e}$ is a right λ eigenvector of $\mathbf{S}$. Then λ, described as an "asymptotic growth rate," is an intuitive measure of fitness. A candidate ESS may be sought as the value of the resident strategy z_a, where $d\lambda/d\delta|_{z_a} = 0$. This can be done using a perturbation approximation for λ [see equation (A.6)]:

$$d\lambda = \frac{\mathbf{y}(d\mathbf{S})\mathbf{e}}{\mathbf{y}\cdot\mathbf{e}}, \tag{11.16}$$

where $\mathbf{y}$ is a left λ eigenvector of the matrix $\mathbf{S}|_{\delta=0}$ associated with the largest eigenvalue (typically the vector $\boldsymbol{\alpha}$ of reproductive values). There are some variations on this framework, but the majority of authors have used equation (11.16). If $\mathbf{S} = \mathbf{F}$, the matrix describing changes in allele frequency, $d\lambda/d\delta$ is formally similar to the fitness measure S [equation (11.8)]. Kin selection effects ignored, it has been abundantly used in previous analyses of class-structured populations (e.g., Charlesworth 1994; Frank 1998; Caswell 2001).

The idea of fitness maximization has been quite confusing in this context. Various works have sought the resident trait value that supposedly maximizes λ, or else the "net reproductive rate" or "lifetime

reproductive success" R, which is simply the lifetime number of adult offspring of an individual (e.g., references in Mylius and Diekmann 1995). There are two issues here.

The first issue is whether $dR/d\delta$ or $d\lambda/d\delta$ is the appropriate measure to determine the direction of selection. In some conditions, dR and $d\lambda$ are equivalent (Charlesworth 1994, p. 179), but more generally $d\lambda$ should be considered, particularly when all offspring are not equivalent in terms of fitness.

The second issue is the distinction between two concepts of maximization, one which is finding the resident trait value z_a such that $d\lambda/d\delta|_{z_a} = 0$ as we have just considered, and the other which is maximization with respect to the resident trait value, that is, finding z_a such that $d\lambda/dz_a = 0$ (see p. 103). In general, to determine the effects of selection, one must seek effects on a mutant's fitness in a context set in part by the resident strategy, which is the first concept. In an age-structured "density-independent population," selection also increases λ (Charlesworth 1994, p. 178), so its direction is also determined by $d\lambda/dz_a$. But density-dependent regulation must eventually occur, and then maximization with respect to resident trait value is meaningless. Evolution may rather result in an increase in the size of some age class (Charlesworth 1994, p. 183), or in the increase of R for an individual in an unoccupied habitat, which is no longer the same as the increase of λ in the same conditions (Mylius and Diekmann 1995).

For fluctuating demography, the generalization of eigenvalue as a measure of fitness is the dominant Lyapunov exponent, which generalizes eigenvalue for random matrix products (e.g., Tuljapurkar 1989, 1990; Caswell 2001; and related invasion exponents, Rand et al. 1994; Dieckmann and Law 1996). Tuljapurkar's results are not easily compared to those of the present chapter. For example, he defined reproductive value as a function of the whole past sequence of demographic fluctuations, rather than as a function of the present demographic state (Tuljapurkar 1990; Caswell 2001, p. 384). The results of this chapter should amount to the computation of perturbations of the dominant Lyapunov exponent, provided the matrices considered have exactly the same meaning. Further comparison is not yet possible, because the practical computation of S has been achieved only in the infinite island model, to which Tuljapurkar's results have not been applied.

Diffusion Approximations

Diffusion approximations for class-structured populations following an island model can be constructed by straightforward generalization of the argument previously given for homogeneously structured populations in chapter 9. Effective size is used to quantify drift in the total population. Expected change in weighted allele frequency takes the form $\bar{p}(1-\bar{p})W_{\mathrm{IF}}$, where W_{IF} has the same form as S [equations (11.8) and (11.12)], where genetic identities are interpreted as relatedness or as F-statistics. Note that in a class-structured population, $\bar{p}$ should be defined as the average weighted according to reproductive value. Otherwise, the change in allele frequency in a neutral model may involve terms of order $1/N_{\mathrm{T}}$, and required conditions for diffusion approximations (Ethier and Nagylaki 1980, equation 1.1; Wakeley 2003, equation A5) will not hold.

Such diffusion approximations will of course be as complex as the expression for S and effective size. Simpler expressions may be obtained for, for example, fertility selection in a metapopulation following the propagule models described in the previous chapter (Roze and Rousset 2003). As expected, a reduction in effective size due to local extinctions entails a reduction in the strength of selection, in the sense that fixation probabilities of selected alleles relative to neutral alleles are less decreased for deleterious alleles, and less increased for favorable alleles (see also Barton 1993; Cherry 2003b).

INFERENCES

Some of the expressions reviewed in this chapter are cumbersome, but they allow numerical analysis of complex models and give a framework within which additional approximations can be developed and evaluated. They also show that a range of problems can be analyzed by essentially the same techniques as in previous chapters. It remains to be seen how much such results increase our ability to understand evolution under complex demographic scenarios, rather than following it in simulations or in numerical studies. Several qualitative conclusions may be noted.

First, although selection on a mutant is evaluated "in the context set by the resident strategy," the mutant affects its own environment

through the correlated effects of neighbors on the number of offspring (as accounted for by kin selection terms), but also through the correlated effects of itself and its neighbors on the demographic conditions experienced by its offspring, hence on their reproductive value.

Second, effects of individual strategies on the demography of a local group of individuals (i.e., the S_{Pr} term) may be viewed as a component of group selection that was ignored in simpler models with fixed demography. Much as with stable demography, the magnitude of these effects decreases with increasing deme sizes, because S_{Pr} involves relatedness coefficients as factors in all of its terms.

As far as conceptual debates about group selection are concerned, there is little to add to the discussion of simpler models (chapter 7): the present results show that the effects of demographic fluctuations on selection can be understood using the same concepts as used previously. Allowing group size to depend on the trait under selection does not ensure evolution of group beneficial traits. On the contrary, it makes selection-driven extinction possible, as may occur through the evolution of too low dispersal rates (Rousset and Ronce in press; see also Gyllenberg et al. 2002).

Spatial Variation in Reproductive Value

The results also show that the importance of demographic structure depends on the extent of variation in reproductive value among different types of demes. Is such variation expected and observed?

We intuitively expect that organisms engage in behaviors that "maximize" their contribution to the future gene pool. Hence, it is clearly better to produce offspring with higher individual reproductive value. This does not mean that a class that produces less offspring than another one can compensate for this decreased fitness with a higher reproductive value of its offspring, as the class with higher fitness also has higher reproductive value [see equation (9.27)]. Rather, this means that within a class, it may pay for individuals to produce less offspring of higher quality.

Thus, there is selection for dispersal towards habitats with high reproductive value. This, of course, increases the number of competitors in habitats that have higher value. Each of them will be less likely to become an adult, which reduces its expected reproductive value. Thus,

selection on habitat choice tends to reduce the variation in the expected reproductive value of juveniles to the point where they become equal, a case known as the ideal free distribution (IFD, Fretwell and Lucas 1970).

This argument ignores many factors, often discussed under the heading of "source-sink" demography. Persistent variation of reproductive value in a temporally constant environment may result from asymmetries in competitive abilities among individuals. This has initially been modeled in some extreme models (Fretwell and Lucas 1970). The difficulty of developing more general mechanistic models of asymmetric competition may have led to a relative neglect of such models, to a theoretical bias towards the IFD, and to an underappreciation of variation in reproductive value among habitats. Passive dispersal, a lack of cues for individuals to assess habitat quality, or maladaptation to a constantly changing environment may contribute to variation in reproductive value (Pulliam 1996; Holt 1997; Diffendorfer 1998; Abrams 2000). Thus, larger variations are expected in species with passive dispersal, such as plants, and in metapopulations with recurrent local extinctions. The example of Norwegian parishes (p. 154) may rather illustrate the results of asymmetric competition.

Kin selection effects may also select for dispersal behaviors that allow persistent spatial variation in reproductive value, because individuals may reduce competition with related individuals by going into habitats with lower individual reproductive value (Leturque and Rousset 2002). However, this may be a minor reason for heterogeneity of reproductive value in natural populations, relative to the other factors.

CHAPTER TWELVE

Overview and Perspectives

In this book we have reviewed models of genetic structure and their uses in analyzing selection in subdivided populations. We have related different approaches such as game theoretical concepts of stability, inclusive fitness, effective size, and applications of diffusion approximations. All these developments are based on simple models. However, there is increased pressure for more quantitative analyses of natural populations. Much current development in population genetics is focused on statistical analyses of genetic data. So the question is how useful such models may be in the latter perspective.

The utility of simple models is well demonstrated by the errors committed when their conclusions are ignored. Simple models argue against a number of erroneous conclusions, some of which will be listed below, often drawn from statistical analyses of genetic diversity in natural populations, and indeed supposedly on the basis of these models. But simple models are in general not sufficient for quantitative analysis of either selection or neutral genetic differentiation in natural populations. In this chapter, we discuss, sometimes rather speculatively, the limitations of current theory and issues facing the further development both of statistical analyses of genetic structure and of selection modeling.

STATISTICAL ANALYSES OF GENETIC STRUCTURE

All Those Data

Various methods of analysis are routinely applied to genetic data from structured populations (e.g., Slatkin 1994; Excoffier 2003; Rousset 2003b; Arbogast et al. 2002), and new ones are proposed virtually ev-

ery day. There are at least two reasons to try to infer the parameters of population structure from genetic data. First, in comparison with other estimates of demographic parameters, it allows an evaluation of the ability of simple models to capture the main quantitative determinants of genetic structure. Second, an appreciation of demographic parameters of population structure should help in guiding theoretical work towards more reasonable evolutionary scenarios of adaptation in natural populations and in the design of experiments to test them. Genetic approaches may allow estimation of parameters that are often very difficult to estimate by other methods.

Slatkin (1994) has argued that there is often qualitative agreement between genetic estimates of gene flow and demographic estimates of dispersal (see also Hastings and Harrison 1994; Ward et al. 1994), with discrepancies revealing the impact of past demographic processes. Yet it is difficult to draw more quantitative conclusions and statistical criteria to demonstrate such an impact in a given species.

Further, genetic analyses have often been misguided in a way that suggested discrepancies between genetics and demography. Consider, for example, the main malaria vector, the mosquito *Anopheles gambiae*. Its genetic structure has been the object of many studies (see Donnelly et al. 2002 for references). F_{ST} estimates of, say, 0.04, have been obtained between populations distant from each other by several hundreds, or even thousands, of kilometers. Such results are quite compatible with most (or even all) dispersal occurring within a few kilometers (chapter 3). However, the result $F_{\mathrm{ST}} \approx 1/(1 + 4Nm)$ has been widely abused to infer that dispersal between distant sampled populations was of the order of several individuals per generation, even though what was a "population" was not clear and even though the inference is not valid under the island model itself (the dispersal rate between any two populations being rather Nm/n_{d}). Such studies confuse our understanding of the actual demography of natural populations. Overall, dispersal may be much more restricted than has often been concluded from analyses of genetic structure.

Estimation Methods

MOMENT VERSUS LIKELIHOOD APPROACHES

In essence, there are two main statistical approaches to estimation, moment and maximum likelihood methods (for general statistical background, see, for example, Cox and Hinkley 1974; Bulmer 1979; Stuart and Ord 1991). The simplest moment method consists of estimating functions of probabilities of identity by the corresponding functions of frequencies of identical pairs of genes in a sample. Thus, $F_{\mathrm{ST}} = (Q_1 - Q_2)/(1 - Q_2)$ can be estimated by $(\hat{Q}_1 - \hat{Q}_2)/(1 - \hat{Q}_2)$, where $\hat{Q}_1$ is the frequency of identical pairs of genes among different individuals within demes, and $\hat{Q}_2$ is the frequency of identical pairs of genes among individuals in different demes. It is little realized that the most widely used estimator of F_{ST} (Weir and Cockerham 1984) is of this form (Rousset 2003b). Refinements are possible, particularly in specific contexts assuming an interfamily or interdeme structure with a separation of time scales. There have been various attempts in this direction and some evaluation of their performance (e.g., Raufaste and Bonhomme 2000 for F_{ST} estimators; Lynch and Ritland 1999; van de Casteele et al. 2001; Wang 2002; Milligan 2003 for other relatedness estimators).

Maximum likelihood approaches require either explicit analytical expressions for the probability of a sample as a function of model parameters, which are rarely available, or various simulation algorithms. There are hopes that these methods will replace moment methods (e.g., Beerli and Felsenstein 2001). Most currently developed likelihood methods use coalescent approaches to compute the probability of the sample. Common to the different coalescent approaches described below is the assumption of neutrality of mutants, so that selection does not affect the genealogy of a sample. Hence, samples can be generated and their probabilities computed by first looking backwards at the coalescence events defining the genealogical relationships between individuals, independent of mutation, then going down the genealogical tree and adding mutations.

THE n-COALESCENT

The genealogy of a sample of n genes may be described by the sets of genes that have a common ancestor at different times in the past. The n-coalescent (e.g., Kingman 1982; Hudson 1990; Nordborg 2003) is a process describing the sequence of such sets through time for a neutral Wright-Fisher model with large N (formally, in the limit $N \to \infty$). As such, it is just another way of looking at the neutral Wright-Fisher model, but it aims to describe properties of a sample rather than of the total population, and it does so by considering the genealogy of the sample rather than of the total population. Of course, the latter feature traces back to Malécot's work for pairs of genes, but analyses of the n-coalescent have provided many results for larger samples (Tavaré 1984), and it is worth recognizing the argument in its generality.

STRUCTURED COALESCENTS

A variety of generalizations of the n-coalescent have been considered, the most notable in the present context being "structured coalescents" with two time scales (Fu 1997; Nordborg 1997; Nordborg and Krone 2002), which are characterized by the separation of time scales considered in chapter 4. Thus, infinite island models (Takahata 1991; Wakeley 1999; Wakeley and Aliacar 2001) and random mixing populations with selfing (Nordborg and Donnelly 1997; Möhle 1998) can be described by structured coalescents. The separation of time scales simplifies the computation of the probability of samples. The analysis of pairs of genes remains useful for such purposes (e.g., Nordborg 1997).

COALESCENT ALGORITHMS FOR MORE GENERAL MODELS

Not every model of population structure has a hierarchical structure with two time scales. Although "structured coalescents" can be defined for other models (e.g., Notohara 1993), they have not brought many analytical results beyond those already known for pairs of genes. Rather, simulation algorithms [often described as Markov chain Monte Carlo (MCMC) or importance sampling (IS) algorithms] have been defined to estimate the probability of a sample (e.g., Griffiths and Tavaré 1994;

Beerli and Felsenstein 2001). See the appendix to this chapter for a description of some of these algorithms. So far, development of these methods has been focused on difficult computational issues. Beerli and Felsenstein (1999, 2001) have tested their performance in some very simple settings.

Robustness

With data from natural populations, many factors are uncontrolled. Therefore, inferences from such data may not be as reliable as those from experimental data. Further, inferences are based on highly simplified models. Attempts to evaluate more complex demographic scenarios will face considerable difficulties, as the number of parameters required to describe the history of more than two populations may be huge, making estimation of each of them less reliable. Hence, a major concern is the robustness of the analyses based on a given model when some of its assumptions do not hold. Robustness issues determine what can be estimated. If a method is sensitive both to recent and more distant events, it can be used to estimate something only when it is much more sensitive to either distant or recent events. For approaches based on F_{ST}'s, the most clear-cut case is that of analyses at a local spatial scale, which depend mainly on recent events (chapter 4). Essentially nothing is currently known about the robustness the coalescence-based maximum likelihood methods.

Estimation of Effective Size

The models predict some correlation between genetic diversity in a species and total effective size. Correlations between genetic diversity and population size are observed (e.g., Frankham 1996), though the strength of the correlation is not one that would allow accurate estimation of population size from genetic diversity. There are several reasons for this lack of accuracy. First, data are often some measure of local genetic diversity and some local population sizes, rather than the measures considered in the models, which refer to the total population. Second, the relationship between effective and actual size will differ among species. Third, genetic diversity is the result of processes

acting on a long time scale, and hence methods used to estimate "effective sizes" from observed genetic diversity will be sensitive to the assumption of long-term demographic stability. Thus, the assumption of long-term demographic stability is more than a purely technical device, and estimating effective size by genetic methods is more problematic than analyzing spatial structure. The conclusion that the effective size estimated by such approaches is some sort of long-term average may not be very useful when the nature of this long-term average is not well understood. Genetic methods will also be sensitive to selection events that may have occurred over a long time scale at the locus considered or at linked loci.

Estimates of effective size based on short-term fluctuations of genetic diversity should be less affected by such problems, but are not devoid of other difficulties. The range of models considered to define estimators has been restricted to panmictic populations (e.g., Wang 2001 and references therein), with few exceptions (Vitalis and Couvet 2001; Wang and Whitlock 2003). It is not clear what is estimated from data from subdivided populations by methods based on models of panmictic populations: the estimated values may be intermediate between local and global effective sizes, or even show a more complex behavior.

With the exceptions of the analyses discussed on p. 187, the current models of effective size in metapopulations have not been applied to real metapopulations. In particular, expressions for effective size are in terms of properties of a set of connected populations, and hence they may be difficult to translate into properties observable from a sampling of a limited subset of these populations.

Some Easy Improvements

Given the "black box" character of and unknown practical performance of currently developed coalescent methods, moment methods should remain useful at least for some time. Leaving aside elaborate theoretical considerations, we may notice that such analyses of genetic structure would be greatly improved if a series of common flaws were eradicated, in light of the following conclusions:

- Analyses of pairs of samples do not yield estimates of dispersal between pairs of populations. Under essentially no model, a

pairwise F_{ST} provides an estimate of the number of migrants between two populations. Under the island model, it provides only an estimate of the number of immigrants from all other populations, assuming the term "population" has a clear meaning.

- "$Nm > 1$" (or $F_{\mathrm{ST}} < 0.2$) says nothing about long-distance dispersal. It is expected over large areas under equilibrium models of isolation by distance, even with with small deme size and very localized dispersal (e.g., fig. 3.3).
- There is often low power to detect isolation by distance, given typical sample sizes. Absence of significant evidence for isolation by distance is therefore an uninformative result. It would be much better to provide confidence intervals for a measure of isolation by distance that can be related to density and dispersal parameters (see Leblois et al. 2003 for an attempt in this direction).
- Theory does not support estimating total effective size from gene diversities in the way it is often done. Usually such estimates are derived from a geographically localized sampling. But none of the theoretical results for local genetic diversity [e.g., equation (3.68); or more accurate formulas for local diversity under island and stepping stone models] relate it to the total effective size in a useful way.

There may be some hope that these problems will be overcome by application of maximum likelihood coalescent approaches, but it is an easy bet that the latter will be as misused and misinterpreted as the moment methods. For example, they do not guard against the confusion of the number of samples with the number of populations in the statistical model, or from the lack of a clear definition of any "population." Again, understanding the important assumptions underlying such methods will be necessary before drawing conclusions based on them.

PROSPECTS FOR THE ANALYSIS OF SELECTION

Current State of Theory

Some common patterns emerge from the models reviewed in this book.

In a panmictic population, as illustrated by the Wright-Fisher models of chapters 2 and 5, the expected frequency change of an allele with phenotypic effect δ, in frequency p, may be written as

$$\Delta p = \delta A p(1-p) + \delta^2 B(p) + o(\delta^2), \tag{12.1}$$

where given fairly weak assumptions (but excluding dominance), A does not depend on p, while $B(p)$ is generally frequency dependent. In a game theoretical perspective, this allows the determination of candidate ESSs as trait values such that $A = 0$ and of convergence to this trait value z^* from the sign of A near z^*. The sign of $B(p)$ is used to check whether this candidate ESS is really an ESS or a branching point. In diffusion approximations, we use only the simpler result $\Delta p = \delta A p(1-p) + o(\delta)$ and the expression for the variance of change in allele frequency $V(p) = p(1-p)/N$ to deduce dynamic properties of allele frequency changes in terms of A and N.

In a random mixing population, such as populations structured by selfing considered by various authors, or the family-structured models considered by Hamilton (1964) to formulate the theory of inclusive fitness, equation (12.1), where p is the allele frequency in the total population, holds as an approximation for large population size. In the same conditions, the variance of change in allele frequency takes the form $V(p) = p(1-p)/N_{\mathrm{e}}$ for some effective size N_{e}. All further analyses then proceed as in the panmictic case, except that there has been little conclusive discussion of how to compute $B(p)$.

In island models, equation (12.1) holds as an approximation for a large number of demes, and variance can again approximately be written as $V(p) = p(1-p)/N_{\mathrm{e}}$. The case is quite comparable to the random mixing case. In both cases, there is a separation of time scales, which has many implications. It allows the consideration that coalescence events of gene lineages within demes occur at a faster time scale than the encounter in the same deme of gene lineages from different demes. This was used to compute effective sizes in chapters 9 and 10. Further, it allows the consideration of "fast events" affecting each deme, such as random fluctuations of allele frequencies, while keeping al-

lele frequency in the total population fixed. This underlies Wright's diffusion approximations for structured populations, as well as more recent works on diffusion approximations and structured coalescents. A separation of time scales underlies classical definitions or properties of relatedness and inbreeding coefficients. This argument is intuitive enough that these definitions have been widely accepted despite often little effort to make them more rigorous (chapter 4). The result that A is independent of p also follows from the properties of relatedness (chapter 7) and from fairly weak assumptions on fitness, made clear in chapter 6. Finally, evolutionary stability *stricto sensu* can be evaluated, at least numerically, though this has been little explored so far.

Under localized dispersal (isolation by distance), most of these arguments break down. A must be considered a function of p. Simulations and heuristic arguments (chapter 5) suggest that its sign is independent of p when the number of demes is large, and we can determine convergence to a candidate ESS from an average of A across different value of p. This average may be computed in terms of coalescence times of pairs of genes, or more commonly in terms of measures of relatedness that bear simple relationship to coalescence times (chapter 6).

These results leave unsolved a number of theoretical issues that would require further analysis. These include analytical methods for practically assessing branching and phenotypic divergence; specific aspects of multilocus evolution in subdivided populations, including effects of inbreeding; and a simultaneous account of fluctuating demography and of spatially localized dispersal. On a more empirical side, we still lack a general perspective on the effects of population structure on social behaviors in natural populations.

There are also a number of topics that can in principle be analyzed using current methods, though this has not been detailed in this book. Some selective factors, such as dominance, can be taken into account by considering genetic relationships between more than two genes (Roze and Rousset 2003). These relationships have attracted little attention, except in the context of maximum likelihood analyses of population structure (see above). They have also been used to compute the variance of gene diversity in a population (e.g., Takahata 1983). In a more ecological perspective, theoretical consequences of kin recognition have been little investigated, despite its widespread expression in several groups of organisms (but see Lehmann and Perrin

2002, in press). Also, different traits do not evolve independently, so it is important to consider their joint evolution. Phenotypic plasticity may be viewed in this perspective, as the variable expression of a trait in different conditions may be viewed as the expression of different traits. Though there is lasting and increasing interest in, for example, condition-dependent sex ratios or density-dependent dispersal, not very many models have dealt with such joint evolution. Modeling behaviors dependent on kin recognition or on local conditions is difficult because these factors induce a class structure as considered in chapter 11.

Avenues for Simplifications

The methods reviewed in this book are approximate. They employ simplified biological assumptions and additional assumptions are made for the purpose of mathematical analysis (e.g., weak selection approximations). Within this context, they are arguably the simplest available methods. In particular, the direct fitness method has proven efficient for computing expected allele frequency changes in more or less complex biological scenarios. This technique is useful even when kin selection effects are considered "weak" in some sense.

However, fitness measures get complicated when kin selection and fluctuating demography are taken into account. The balance between the different forces acting on a trait is no longer easily deduced from the fitness measure, so that even qualitative effects of changing one parameter are not easily predicted. Simulation studies, although useful, are hardly a reduction in complexity. The main question is therefore how to simplify the theory further.

An approximate analysis is useful when it can be more easily interpreted than a more exact one, while at the same time it reaches qualitatively the same conclusion as the latter. How to best simplify the analysis of specific models has not been systematically considered. The practical assessment of the usefulness of any method of approximation requires both an understanding of the theoretical conditions where it may be expected to perform well and an evaluation of its practical relevance by a thorough understanding of several natural systems. Long-term studies of natural populations (Thomas and Hanski 1997) should be particularly important in this respect.

Several types of approximations have been considered. For exam-

ple, deme size is sometimes treated as a continuous variable, but it is unclear to what extent this is a robust approximation and even whether it brings any simplification. "Moment closure" or "pair approximation" methods have also been proposed (e.g., Nakamaru et al. 1997). These methods start from the principle that fitness will depend, for example, on the frequencies not only of different pairs of alleles among interacting individuals, but also of triplets of alleles, and so on indefinitely. Then, various approximations are considered to reduce the system to a small number of variables. Many approaches have also ignored kin selection effects. Kin selection effects due to limited dispersal often appear to be less important when deme sizes are large. Hence, it is logical to consider approximations that neglect relatedness, or at least some aspects of it. But it is difficult to draw a boundary between cases when such approximations will be reliable and cases when they will not. In particular, in a metapopulation, the number of recolonizers after an extinction is an important determinant of population structure. Even with high emigration rates, the number of recolonizers may be low, and hence relatedness may be high (e.g., *Melitaea* butterfly populations studied by Heino and Hanski 2001). Thus, it is of dubious value to ignore relatedness between offspring of the same parent in metapopulations (Ronce and Olivieri in press). With overlapping generations, competition effects between, for example, mothers and daughters may be a major selective force on many traits (e.g., dispersal: de Fraipont et al. 2000; sex ratio: Johnson et al. 2001), but they would be ignored if all kin selection effects were ignored.

CONCLUSION

Many of the issues discussed in this chapter are not unique to the genetics of subdivided populations. The field of population dynamics faces them too: a common problem is how to make theoretical models simpler (e.g., Levin and Pacala 1997), while models seem too simple when it comes to statistical analyses of natural populations. Current models of population structure, even the most elaborate ones, remain of limited use for reliable statistical analysis. A better understanding of more complex models will be needed if more ambitious and more general statistical methods are to be developed. Nevertheless, in some cases, approximate quantitative fit between predictions of simple mod-

els and observations is possible, and a range of other questions can be addressed using a few relatively simple concepts and methods.

APPENDIX: ALGORITHMS FOR LIKELIHOOD ESTIMATION

Griffiths and Tavaré's (1994) algorithm was originally defined for panmictic populations and has been extended to subdivided populations by Nath and Griffiths (1996) and Bahlo and Griffiths (2000). The main idea of this algorithm can be described as follows.

We first consider recursions between probabilities of samples at successive generations (Kingman 1980, chapter 3). Following a typical coalescent argument, we first consider the genealogical relationships between individuals independently of the nature of alleles, then add mutations on a genealogical tree. To obtain the recursions, (1) we consider the coalescence events in the previous generation (in structured populations, we also consider dispersal events: all properties of ancestral lineages, except allelic states, are included in the definition of the genealogy); (2) we consider the rest of the genealogy until the common ancestor; (3) we add mutations on the tree considered in (2); and (4) we add mutations in the latest generation, considered in (1).

In a subdivided population, a sample may be characterized by the number of genes of different allelic types in each deme sampled (these numbers are collectively denoted $\mathbf{N}'$), and the sample "size" $\mathbf{n}'$ is a vector of sample sizes for different demes. One generation earlier, the parental lineages form an ancestral sample, noted $\mathbf{N}$, of $\mathbf{n}$ genes. Summing over the different possible ancestral states $\mathbf{N}$, one can write a recursion for the probability $q(\mathbf{N}')$ of a sample by the straightforward use of conditional probabilities:

$$q(\mathbf{N}') = \sum_{\mathbf{N}} \Pr(\mathbf{n}) q(\mathbf{N}) \Pr(\mathbf{N}'|\mathbf{N}), \tag{12.2}$$

where $q(\mathbf{N}')$ is the stationary probability of sample $\mathbf{N}'$, given sample size $\mathbf{n}'$; $\Pr(\mathbf{n})$ is the probability (considered independent of $\mathbf{N}'$ by the coalescent argument) of genealogical events in the latest generation such that the parental lineages form a sample of $\mathbf{n}$ genes; $q(\mathbf{N})$ is the stationary probability of sample $\mathbf{N}$ given parental "sample size" $\mathbf{n}$, determined by events in steps (2) and (3); and $\Pr(\mathbf{N}'|\mathbf{N})$ is the probability (given $\mathbf{n}'$) of mutation events in the latest generation leading to $\mathbf{N}'$ in step (4).

If the term for $\mathbf{N}' = \mathbf{N}$ is singled out on the right-hand side, the latter equation can be written in the form

$$q(\mathbf{N}') = \sum_{\mathbf{N} \neq \mathbf{N}'} f(\mathbf{N}') p(\mathbf{N}|\mathbf{N}') q(\mathbf{N}), \tag{12.3}$$

where

$$f(\mathbf{N}') \equiv \frac{\sum_{\mathbf{N}\neq\mathbf{N}'} \Pr(\mathbf{n}) \Pr(\mathbf{N}'|\mathbf{N})}{1 - \sum_{\mathbf{N}\neq\mathbf{N}'} \Pr(\mathbf{n}) \Pr(\mathbf{N}'|\mathbf{N})} \tag{12.4}$$

and

$$p(\mathbf{N}|\mathbf{N}') \equiv \frac{\Pr(\mathbf{n}) \Pr(\mathbf{N}'|\mathbf{N})}{\sum_{\mathbf{N}\neq\mathbf{N}'} \Pr(\mathbf{n}) \Pr(\mathbf{N}'|\mathbf{N})}. \tag{12.5}$$

The p's are thus transition probabilities defining a Markov chain between different samples ($\mathbf{N} \neq \mathbf{N}'$), which means that recursion (12.3) may be obtained as a recursion over the first time that some event, such as mutation, dispersal, or coalescence, affects some ancestral lineage(s).

In practice, approximations are required for $\Pr(\mathbf{N}'|\mathbf{N})$ and $\Pr(\mathbf{n})$ to take simple forms. It is assumed that only one event (e.g., mutation, dispersal, coalescence) can occur at one time, so these probabilities are computed as approximations for large deme size and small mutation and dispersal rates. For example with a single haploid population of N individuals, where n' have been sampled, the probability that a coalescence event occurs among the n' lineages over one generation is $\Pr(n = n' - 1) \approx n'(n' - 1)/2N$ for large N. The probabilities of the mutation and migration events may be determined by equally simple algebra, but will depend on the details of the models assumed (see references above).

Now index the successive samples as $\mathbf{N}(t)$, where $t = 0$ for the actual sample. Iterating the recursion (12.3), we see that

$$\begin{aligned} &q[\mathbf{N}(0)] \\ &\quad = f[\mathbf{N}(0)] \sum_{\mathbf{N}(1)} f[\mathbf{N}(1)] \sum_{\mathbf{N}(2)} q[\mathbf{N}(2)] p[\mathbf{N}(2)|\mathbf{N}(1)] p[\mathbf{N}(1)|\mathbf{N}(0)] \end{aligned} \tag{12.6}$$

$$= \prod_{t=0}^{\tau} f[\mathbf{N}(t)] \sum_{\mathbf{N}(1)} \cdots \sum_{\mathbf{N}(\tau)} \prod_{t=0}^{\tau-1} p[\mathbf{N}(t+1)|\mathbf{N}(t)], \tag{12.7}$$

where τ is the (random) number of events until there is a single ancestor, and $f[\mathbf{N}(\tau)]$ is the stationary probability of $\mathbf{N}(\tau)$, that is, of the allelic type of the common ancestor. The right-hand product is the probability of the ancestral history, that is, of the sequence of states until the common ancestor, given the transition probabilities $p(\mathbf{N}|\mathbf{N}')$.

Thus, the above equation is

$$q[\mathbf{N}(0)] = \mathrm{E}\left(\prod_{t=0}^{\tau} f[\mathbf{N}(t)]\right), \tag{12.8}$$

where the expectation is taken over realizations of the Markov chain. We can estimate it by computing the average value of $\prod_{t=0}^{\tau} f[\mathbf{N}(t)]$ over a large number of realizations of the Markov chain.

This algorithm is not very efficient, because it will average over ancestral histories that may be very unlikely under the model considered. Improvements have been considered by Stephens and Donnelly (2000) and de Iorio and Griffiths (submitted). In essence, they rewrite equation (12.2) as

$$q(\mathbf{N}') = \sum_{\mathbf{N}\neq\mathbf{N}'} \hat{f}(\mathbf{N}', \mathbf{N})\hat{p}(\mathbf{N}|\mathbf{N}')q(\mathbf{N}) \tag{12.9}$$

for some transition probabilities $\hat{p}(\mathbf{N}|\mathbf{N}')$ forming a Markov transition matrix, and for

$$\hat{f}(\mathbf{N}', \mathbf{N}) \equiv \frac{f(\mathbf{N}')}{\hat{p}(\mathbf{N}|\mathbf{N}')}. \tag{12.10}$$

Then

$$q[\mathbf{N}(0)] = \mathrm{E}\left(\prod_{t=0}^{\tau} \hat{f}[\mathbf{N}(t-1), \mathbf{N}(t)]\right) \tag{12.11}$$

over realizations of the Markov chain defined by $\hat{p}$. Thus, the new algorithm effectively replaces $f(\mathbf{N}')$ with $\hat{f}(\mathbf{N}', \mathbf{N})$ and p with $\hat{p}$. With a suitable choice of the transition probabilities $\hat{p}$, faster estimation of $q[\mathbf{N}(0)]$ is possible. This may be viewed as an application of importance sampling (e.g., Hastings 1970). In the language of importance-sampling literature, the proposal distribution is the distribution of ancestral histories generated by $\hat{p}$, and the importance sampling weights are the products in equation (12.11).

The algorithms of Beerli and Felsenstein (1999, 2001) are also importance-sampling algorithms, but they are based on a recurrent Markov chain visiting the different possible ancestral genealogies of a given sample, rather than on an absorbing Markov chain visiting successive ancestral samples until the common ancestor.

APPENDIX A

Mathematical Appendix

NOTATION

Some notation used throughout the book is summarized in Table A.1.

Vectors are usually denoted as bold lowercase characters ($\mathbf{v}$) and matrices as bold uppercase characters ($\mathbf{S}$). However, bold uppercase symbols are also used for vectors of variables that have been denoted using uppercase characters (e.g., deme size N or genetic identity Q). The transpose of a vector $\mathbf{v}$ is denoted $\mathbf{v}^\top$. Thus, (v_1, v_2) is a row vector and $(v_1, v_2)^\top$ is a column vector. Likewise, $\mathbf{S}^\top = (s_{ji})$ is the transpose of a matrix $\mathbf{S} = (s_{ij})$, where rows are exchanged for columns and vice versa. The scalar product $\sum_i u_i v_i$ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is denoted $\mathbf{u} \cdot \mathbf{v}$. $\mathbf{I}$ is the identity matrix of appropriate dimension, for example,

$$\mathbf{I} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{A.1}$$

with 1s on the diagonal and 0s elsewhere. $\mathbf{0}$ is the (row or column, depending on context) zero vector of appropriate dimension, $(0, \ldots, 0)$. Likewise, $\mathbf{1}$ is the row or column unit vector $(1, \ldots, 1)$.

$\mathrm{E}(X)$ denotes the expectation of variable X, and $\mathrm{E}(X|Y = y)$ denotes conditional expectation, given that variable Y has value y. $\equiv$ is the equivalence symbol, used for definitions. In recursions over successive times, primes ($'$) are used to denote any variable considered at the later time.

A function $f(x)$ is $O[g(x)]$ near $x = x_0$ if $f(x)$ does not become infinitely greater than $g(x)$ for x near x_0 (the exact meaning being that there is some constant K such that $|f(x)/g(x)| < K$ in a neighborhood of x_0). It usually means that $f(x)$ can be neglected if $g(x)$ can, and that $f(x)$ may not be negligible if $g(x)$ is not negligible.

A function $f(x)$ is $o[g(x)]$ near x_0 if $\lim_{x\to x_0} f(x)/g(x) = 0$. It usually means that $f(x)$ is negligible relative to $g(x)$ near x_0.

TABLE A.1. Notation Used Consistently

$\mathbf{A} \equiv (a_{ij})$	Backwards dispersal or transition matrix for pairs of genes
$C_t, C_{i,t}$	Probability of coalescence
c	Fitness cost to an actor
c_{d}	Survival cost of dispersal
$\mathbf{c} \equiv (c_i)$	Vector of gains in identity through coalescence
$\mathbf{e}_i$	Right eigenvectors of a matrix
$\mathbf{F} \equiv (f_{ij})$	Backwards dispersal or transition matrix for single genes
$\mathbf{G}$	Matrix of probabilities of noncoalescence of pairs of genes
$\imath$ (dotless i)	$\sqrt{-1}$
l_i	Eigenvalues of migration matrix $\mathbf{A}$
ℓ_i	Eigenvalues of various matrices
m	Backwards dispersal rate
n_{c}	Number of classes of individuals
n_{d}	Number of demes
N	Deme size
N_{e}	Effective population size
N_{s}	Effective population size (strong migration limit)
N_{T}	Total population size
p_k	Frequency of allele k
Q_j	Probability of identity in state for a pair of genes specified by index j
$\dot{Q}_j$	Probability of identity by descent,[a] for a pair of genes specified by index j
Q_{w}, Q_{b}	w and b, generic indices for identity of pairs of genes compared *within* a group and *between* two demographic groups
Q^{A}	Identity right after a competition stage (e.g., between adults)
Q^{D}	Identity right before a competition stage (e.g., between juveniles after dispersal)

Continued on next page

Table A.1 (*continued*)

Q^{J}	Identity among juveniles before dispersal
Q^{R}	Same as Q^{A} but with resampling
r	Euclidian distance
$\mathbf{r}$	Vectorial distance
w	Fitness, that is, expected number of adult offspring
$\mathbf{y}_i$	Left eigenvectors of a matrix
z_A, z_a	Phenotypes of individuals bearing A or a alleles, respectively
$\boldsymbol{\alpha} \equiv (\alpha_i)$	Class reproductive values
γ	$(1-\mu)^2$, probability that neither of two genes mutates
δ	Phenotypic difference $z_A - z_a$
λ_j	Eigenvalues of $\mathbf{G}$ matrix
μ	Mutation rate
Π	Fixation probability
$\phi \equiv \partial\Pi/\partial\delta$	Weak selection effect on probability of fixation

Notes: [a] The notation $\dot{Q}$ is used only when identity by descent and identity in state are contrasted. Otherwise, the notation Q is used for identity by descent, viewed as identity in state under the infinite allele model.

MATRIX ALGEBRA

Eigenvalues and Eigenvectors

DEFINITIONS

A column vector $\mathbf{e}$ is a right ℓ eigenvector of matrix $\mathbf{S}$ if $\mathbf{Se} = \ell\mathbf{e}$. ℓ is then an eigenvalue of $\mathbf{S}$. Let $\bar{x}$ be the complex conjugate of x (for two real numbers a, b and $x = a + \imath b$, $\bar{x} = a - \imath b$). A row vector $\mathbf{y}$ is a left ℓ eigenvector of matrix $\mathbf{S}$ if $\bar{\mathbf{y}}\mathbf{S} = \ell\bar{\mathbf{y}}$ (Horn and Johnson 1985).

VECTOR BASES

Any vector $\mathbf{v}$ may be represented in a basis of linearly independent vectors $\mathbf{w}_i$, as $\mathbf{v} = \sum_i x_i \mathbf{w}_i$. Two nonzero vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ are orthogonal if and only if $\mathbf{e}_i \cdot \bar{\mathbf{e}}_j = 0$. In a basis of orthogonal vectors, the above x_i's may be obtained as $\mathbf{e}_i \cdot \bar{\mathbf{x}} / \mathbf{e}_i \cdot \bar{\mathbf{e}}_i$. Hence, in an orthogonal

basis of eigenvectors,

$$\mathbf{v} = \sum_i \frac{\mathbf{e}_i \cdot \bar{\mathbf{x}}}{\mathbf{e}_i \cdot \bar{\mathbf{e}}_i} \mathbf{e}_i. \tag{A.2}$$

In chapter 3, we consider the basis of eigenvectors

$$\mathbf{e}_j \equiv (1, \ldots, e^{\imath 2\pi jk/n}, \ldots, e^{\imath 2\pi j(n-1)/n}) \tag{A.3}$$

for $j = (0, \ldots, n-1)$. These vectors are orthogonal, since for any j, l,

$$\mathbf{e}_j \cdot \bar{\mathbf{e}}_l = \sum_{k=0}^{n-1} e^{\imath \frac{2\pi kj}{n}} e^{-\imath \frac{2\pi kl}{n}} = n[j = l] \tag{A.4}$$

using notation (8.4). To obtain this result, we note that $\sum_0^{n-1} z^k = (1 - z^n)/(1 - z)$ and that $e^{\imath\pi} = -1 \Rightarrow e^{\imath 2\pi k} = 1$ for any k. Then

$$\begin{aligned} \sum_{k=0}^{n-1} e^{\imath \frac{2\pi kj}{n}} e^{-\imath \frac{2\pi kl}{n}} &= \sum_{k=0}^{n-1} e^{\imath \frac{2\pi k(j-l)}{n}} \\ &= \begin{cases} \sum_{k=0}^{n-1} 1 = n & \text{if } j - l = 0 \\ \frac{1 - e^{\imath 2\pi k(j-l)}}{1 - e^{\imath 2\pi k(j-l)/n}} = 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{A.5}$$

THE LARGEST EIGENVALUE

A formula for perturbations of the largest eigenvalue ℓ_1 of a matrix, in terms of perturbations of its elements, is

$$d\ell_1 = \frac{\mathbf{y}^\circ (d\mathbf{S}) \mathbf{e}^\circ}{\mathbf{y}^\circ \cdot \mathbf{e}^\circ}, \tag{A.6}$$

where $^\circ$ denotes any variable considered in the unperturbed case, and $\mathbf{y}$ and $\mathbf{e}$ are ℓ_1 left and right eigenvectors (Horn and Johnson 1985, theorem 6.3.12).[1]

[1]This result holds in particular for matrices that have only one linearly independent right ℓ_1 eigenvector [in particular, "irreducible" matrices with nonnegative elements, which have real $\mathbf{y}$ and $\mathbf{e}$ (Horn and Johnson 1985, section 8.4)]. To get equation (A.6), write the differential of the equation $\mathbf{S}\mathbf{e} = \ell_1 \mathbf{e}$,

$$\mathbf{S}^\circ d\mathbf{e} + (d\mathbf{S})\mathbf{e}^\circ = \ell_1^\circ d\mathbf{e} + (d\ell_1)\mathbf{e}^\circ \tag{A.7}$$

and premultiply (A.7) by $\mathbf{y}^\circ$:

$$\mathbf{y}^\circ \mathbf{S}^\circ d\mathbf{e} + \mathbf{y}^\circ (d\mathbf{S})\mathbf{e}^\circ = \ell_1^\circ \mathbf{y}^\circ \cdot d\mathbf{e} + d\ell_1 \mathbf{y}^\circ \cdot \mathbf{e}^\circ. \tag{A.8}$$

Since $\mathbf{y}^\circ \mathbf{S} = \ell_1 \mathbf{y}^\circ$, the first terms on the left-hand and right-hand sides cancel, which yields equation (A.6).

TENSOR PRODUCT

The tensor (or direct or Kronecker) product or two matrices $\mathbf{S} \equiv (s_{ik})$ and $\mathbf{S}' \equiv (s'_{jl})$, of dimensions $m \times n$ and $m' \times n'$, is a matrix of dimensions $mm' \times nn'$ with ij, klth element $s_{ik}s'_{jl}$ [see e.g., equation (9.29)]. It is denoted $\mathbf{S} \otimes \mathbf{S}'$. The eigenvalues of $\mathbf{S} \otimes \mathbf{S}'$ are the products of eigenvalues of the two matrices $\mathbf{S}$ and $\mathbf{S}'$, and its eigenvectors are the tensor products of the eigenvectors of these two matrices (Horn and Johnson 1991, theorem 4.2.12).

Diagonalizable Matrices

A diagonalizable $n \times n$ matrix has n linearly independent right eigenvectors $\mathbf{e}_i$ with associated eigenvalues ℓ_i. Thus, any vector $\mathbf{v}$ may be represented in such a basis as $\mathbf{v} = \sum_i x_i \mathbf{e}_i$. This representation directly yields $\mathbf{S}\mathbf{v} = \sum_i \ell_i x_i \mathbf{e}_i$, and by induction

$$\mathbf{S}^k \mathbf{v} = \sum_i \ell_i^k x_i \mathbf{e}_i. \tag{A.9}$$

This further gives results for any function of $\mathbf{S}$ that can be represented as a power series. For example, for any γ, $(1 - \gamma\mathbf{S})^{-1} = \sum_{k=0}^{\infty} \gamma^k \mathbf{S}^k$, provided that the inverse exists, and then

$$(\mathbf{I} - \gamma\mathbf{S})^{-1}\mathbf{v} = \sum_i \frac{1}{1 - \gamma\ell_i} x_i \mathbf{e}_i. \tag{A.10}$$

Diagonalizable matrices can be represented using a "spectral" representation (Karlin and Taylor 1975):

$$\mathbf{S} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}^{-1}, \tag{A.11}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues of $\mathbf{S}$, and $\mathbf{H}$ is a matrix whose columns are the right eigenvectors of $\mathbf{S}$. Then $\mathbf{H}^{-1}$ is a matrix whose rows are the left eigenvectors of $\mathbf{S}$. The left eigenvectors may be chosen to form an orthogonal basis, and then the right eigenvectors will also form an orthogonal basis.

For example, if $\mathbf{S}$ is a 2×2 matrix, then, using the notation $Z \equiv \sqrt{{s_{11}}^2 + 4s_{12}s_{21} - 2s_{11}s_{22} + {s_{22}}^2}$,

$$\mathbf{\Lambda} \equiv \begin{pmatrix} l_1 = \frac{Z + s_{11} + s_{22}}{2} & 0 \\ 0 & l_2 = \frac{-Z + s_{11} + s_{22}}{2} \end{pmatrix} \tag{A.12}$$

is the diagonal matrix of eigenvalues of $\mathbf{S}$, and

$$\mathbf{H} \equiv \begin{pmatrix} \frac{(Z+s_{11}-s_{22})(Z-s_{11}+2s_{21}+s_{22})}{4Zs_{21}} & \frac{Z-s_{11}+s_{22}}{2s_{21}} \\ \frac{-(Z-s_{11}+2s_{21}+s_{22})}{2Z} & 1 \end{pmatrix}. \tag{A.13}$$

The spectral representation, or simply equation (A.9), is used to show that for large k, the terms ℓ_i^k contributed by the subdominant eigenvalues are negligible with respect to the term ℓ_1^k contributed by the largest eigenvalue. When the matrix is not diagonalizable, a generalization of the spectral representation, known as the "Jordan" representation (e.g., Horn and Johnson 1985), can be used to the same effect (e.g., Cox and Miller 1965). Since the Jordan representation is cumbersome, we ignore it in this book.

GENERATING FUNCTIONS

Given a sequence of numbers $\{p_k\} \equiv p_0, p_1, \ldots$, the generating function of this sequence is a function of the variable z,

$$G(z) \equiv \sum_{k=0}^{\infty} p_k z^k. \tag{A.14}$$

For example, identity by descent can be viewed as the generating function of coalescence times, $\sum_t C_t z^t$, evaluated in $z = (1-\mu)^2$. A discrete Fourier transform, $\sum_k p_k e^{\imath x k}$, is $G(z)$ evaluated in $z = e^{\imath x}$ (here the sum over k may extend over negative integers). The Fourier transform of a probability distribution is known as its characteristic function.

Generating functions are convenient tools for dealing with sums of independent random variables (e.g., Feller 1968; Grimmett and Stirzaker 1992). We note only one of their properties here. Given two independent random variables X and Y, with generating functions $G_1(z)$ and $G_2(z)$, the generating function of their sum is $G_1(z)G_2(z)$:

$$\begin{aligned} \sum_k \Pr(X+Y=k) z^k &= \sum_k \left(\sum_l \Pr(Y=l) \Pr(X=k-l) \right) z^k \\ &= \sum_l \Pr(Y=l) z^l \sum_k \Pr(X=k-l) z^{k-l} \\ &= G_1(z) G_2(z). \end{aligned} \tag{A.15}$$

The sequence of sums $\{\sum_l \Pr(Y = l) \Pr(X = k - l)\}$ is also known as the convolution of the sequences $\{\Pr(X = k)\}$ and $\{\Pr(Y = k)\}$.

COMPUTATION OF IDENTITY IN STATE

The computation of identity in state under different mutation models offers a good illustration of several of the techniques presented in the previous sections. The following models are usually considered.

Mutation Models

The infinite allele model (IAM) assumes that every allele created by mutation is different from preexisting ones. Identity in state is then identical to identity by descent $\dot{Q}$, whose computation has been described under various models throughout this book.

The K-allele model (KAM) assumes that there are K possible alleles and that the mutation rates from one type to another are identical. Thus, given the total mutation probability μ, the probability of mutation from one type to another is $\mu/(K - 1)$.

The stepwise mutation model (SMM) was introduced by Ohta and Kimura (1973) to describe the evolution of alleles characterized by their electrophoretic mobility, and assumes that mutation reduces or increases this mobility by some unit (e.g., by one repeat of a microsatellite motif). Generalized versions of this model consider that mutation may reduce or increase this value by several units (e.g., Griffiths 1980; Di Rienzo et al. 1994). The distribution of mutation effects may be arbitrary, but it is assumed independent of the "value" of the mutating allele.

A more general model simply considers any arbitrary mutation pattern between alleles. It is a Markov mutation model, where the Markovian variable is the allele type. This model is characterized by the probability transition matrix between alleles. The techniques demonstrated below apply to this class of models. We do not consider models where the mutation process affecting one gene would depend, for example, on the allelic type of the homologous gene copy in a diploid through a gene conversion event or through some other interchromosomal interaction.

Identity in the Different Mutation Models

The probability of identity in state between two genes may be written in the form

$$Q = \sum_t C_t Q_{|t} \tag{A.16}$$

in terms of the probability of coalescence C_t at time t and the probability $Q_{|t}$ of identity, given that coalescence occurs at t. For identity by descent, the conditional value is $Q_{|t} = (1-\mu)^{2t}$ in terms of the mutation rate μ, and the unconditional one is $\sum_t C_t(1-\mu)^{2t} \equiv \dot{Q}[(1-\mu)^2]$. More generally, the conditional probability is given by a Markov chain model. For example, for a two-allele model with mutation rate μ, the transition probabilities between the two allelic states $1, 2$ along a gene lineage are described by the matrix

$$\mathbf{U} \equiv (u_{ij}) = \begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix}. \tag{A.17}$$

Thus, if the gene was initially of type 1, the probabilities that its descendant t generations later is of type 1 or 2 are the elements of the vector

$$\mathbf{U}^t \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{A.18}$$

If we follow two gene lineages, we have four possible states, the allelic states of two genes: 11, 12, 21, or 22. The transition probabilities between these four states are given by the 4×4 matrix

$$\begin{aligned} \begin{pmatrix} u_{11}u_{11} & u_{11}u_{12} & u_{12}u_{11} & u_{12}u_{12} \\ u_{11}u_{21} & u_{11}u_{22} & u_{12}u_{21} & u_{12}u_{22} \\ u_{21}u_{11} & u_{21}u_{12} & u_{22}u_{11} & u_{22}u_{12} \\ u_{21}u_{21} & u_{21}u_{22} & u_{22}u_{21} & u_{22}u_{22} \end{pmatrix} &= \begin{pmatrix} u_{11}\mathbf{U} & u_{12}\mathbf{U} \\ u_{21}\mathbf{U} & u_{22}\mathbf{U} \end{pmatrix} \\ &= \mathbf{U} \otimes \mathbf{U}, \end{aligned} \tag{A.19}$$

which is the tensor product of $\mathbf{U}$ with itself [this analysis is comparable to the one leading to equation (9.28)].

Thus, if the common ancestor was initially of type 2, the probabilities that the two descendants t generations later are both of type 1 is

element 11 of the vector

$$(\mathbf{U} \otimes \mathbf{U})^t \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{A.20}$$

More generally, the ancestral state can be represented by a vector δ_{ij} having all elements being 0 except element ij being 1, where ij is the allelic state of the ancestral lineage(s) of the pair of genes. Then the probabilities of the different allelic states t generations later are given by $(\mathbf{U} \otimes \mathbf{U})^t \boldsymbol{\delta}_{22}$.

These expressions are conveniently expressed in terms of the eigenvalues ℓ_{ij} and eigenvectors $\mathbf{e}_{ij}$ of the matrix $\mathbf{U} \otimes \mathbf{U}$. Assuming the matrix to be diagonalizable, we express any vector $\mathbf{v}$ in the basis of eigenvectors $\mathbf{e}_{ij}$ as $\mathbf{v} = \sum_{ij} a_{ij}(\mathbf{v})\mathbf{e}_{ij}$. Then

$$(\mathbf{U} \otimes \mathbf{U})^t \mathbf{v} = \sum_{ij} \ell_{ij}^t a_{ij}(\mathbf{v})\mathbf{e}_{ij}. \tag{A.21}$$

Hence, the vector of probabilities of allelic states, given that coalescence occurred t generations ago and that the common ancestor was of state 2, is given by $\sum_{ij} \ell_{ij}^t a_{ij}(\boldsymbol{\delta}_{22})\mathbf{e}_{ij}$. The probabilities of identity not conditional on coalescence time t may be computed without any explicit knowledge of the conditional probabilities, given t. For example, the probability that two genes are both 1, given that their common ancestor was 2 is obtained by summing the above expression over the distribution of coalescence times so that it is element 11 of the same vector $\sum_{ij} \ell_{ij}^t a_{ij}(\boldsymbol{\delta}_{22})\mathbf{e}_{ij}$. This can be written

$$\sum_{ij} \left(\sum_t C_t \ell_{ij}^t \right) a_{ij}\left(\boldsymbol{\delta}_{22}\right) \mathbf{e}_{ij} = \sum_{ij} \dot{Q}(\ell_{ij}) a_{ij}(\boldsymbol{\delta}_{22})\mathbf{e}_{ij}. \tag{A.22}$$

Thus, such probabilities can be expressed as a function of probabilities of identity by descent for various mutation rates. In practice, one only needs to find $\dot{Q}$ as a function of $(1-\mu)^2$ for the given demographic model and to find the eigenvectors and eigenvalues of the mutation matrix $\mathbf{U}$. This is the separation of the mating and mutation systems noticed by Tachida (1985).

Summing over possible ancestral states l, each with probability $\pi(l)$, the probability that two genes are both of type 1 is element 11 of

$$\sum_{ij} \dot{Q}(\ell_{ij}) a_{ij} \left(\sum_l \pi(l) \boldsymbol{\delta}_{ll} \right) \mathbf{e}_{ij} \tag{A.23}$$

and the probability that two genes are identical is the sum of elements kk of this vector.

Computations are made easier by the fact that the eigenvalues ℓ_{ij} and eigenvectors $\mathbf{e}_{ij}$ are the products $\ell_i\ell_j$ and tensor products $\mathbf{e}_i \otimes \mathbf{e}_j$ of the eigenvalues and eigenvectors of $\mathbf{U}$. For example, for the two-allele version of the KAM, $\ell_1 = 1$, $\ell_2 = 1 - 2\mu$; $\mathbf{e}_1 = (1/2, 1/2)$ and $\mathbf{e}_2 = (1/2, -1/2)$. Thus, $\ell_{12} = 1 - 2\mu$, $\mathbf{e}_{12} = [1/2(1/2, -1/2), 1/2(1/2, -1/2)]$, and so on. By applying the above formulas, one can check that

$$Q = \frac{1}{2}\dot{Q}(1) + \frac{1}{2}\dot{Q}[(1-2\mu)^2] = \frac{1}{2}\{1 + \dot{Q}[(1-2\mu)^2]\}, \quad \text{(A.24)}$$

the more general formula for a symmetric KAM being

$$Q = \frac{1}{K} + \frac{K-1}{K}\dot{Q}\left[\left(1 - \frac{K\mu}{K-1}\right)^2\right] \quad \text{(A.25)}$$

(this is not the simplest way to get this particular result; see, for example, Crow and Aoki 1984).

Now consider a four-allele stepwise mutation model with transition matrix

$$\mathbf{U} = \begin{pmatrix} 1-\mu & \mu/2 & 0 & \mu/2 \\ \mu/2 & 1-\mu & \mu/2 & 0 \\ 0 & \mu/2 & 1-\mu & \mu/2 \\ \mu/2 & 0 & \mu/2 & 1-\mu \end{pmatrix}, \quad \text{(A.26)}$$

that is, mutation is possible towards each of the two "neighbouring" states 4 and 2 for allele 1, 1 and 3 for allele 2, and so on. This transition matrix has the same structure as the migration matrix for a stepping stone model with four demes on a circle [equation (3.19)]. Since this matrix is a circulant matrix (i.e., where each row is the previous one cycled forward one step), the general formulas above reduce to the formulas of discrete Fourier analysis. We directly generalize to K alleles. The eigenvectors are $\mathbf{e}_j = (1, \ldots, e^{\imath 2\pi jk/K}, \ldots, e^{\imath 2\pi j(K-1)/K})$ and the eigenvalues are $\ell_j = \psi(e^{\imath 2\pi j/K})$, where $\psi(e^{ix}) = 1 - \mu[1 - \cos(x)]$ (e.g., they are 1, $1-\mu$, $1-2\mu$ and $1-\mu$ in the present example). Thus, the probability that two genes are identical is

$$Q = \frac{1}{K}\sum_{k=0}^{K-1} \dot{Q}[\psi^2(e^{\imath 2\pi k/K})]. \quad \text{(A.27)}$$

One can recognize here a Fourier inverse of the characteristic function $\dot{Q}(\psi^2)$ (see p. 46). This illustrates the more general result that if ψ^2 is

the characteristic function describing the effects of mutation over one generation on the difference in size between two alleles, and $\dot{Q}(\gamma)$ is the probability of identity by descent of two sampled genes, then $\dot{Q}(\psi^2)$ is the characteristic function of the difference in size of these two gene copies (e.g., Moran 1975). The probability that the size of two gene copies differs by j units is therefore

$$\frac{1}{K}\sum_{k=0}^{K-1}\dot{Q}[\psi^2(e^{\imath 2\pi k/K})]e^{-\imath 2\pi kj/K}. \tag{A.28}$$

Likewise, the result for the unbounded stepwise mutation model is the limit when $K \to \infty$ of the above expression:

$$\frac{1}{\pi}\int_0^{\pi}\dot{Q}[\psi^2(e^{ix})]\cos(jx)\,dx. \tag{A.29}$$

These formulas hold for more general unbounded stepwise mutation models: see Rousset (1996) for examples. Slightly different formulas were used by, for example, Li (1976). Given that coalescence occurred t generations in the past, the Fourier transform of the distribution of difference in size between two alleles is ψ^{2t}. Hence, the probability that the size of two gene copies differs by j units is obtained by Fourier inversion as

$$\frac{1}{\pi}\int_0^{\pi}\psi^{2t}(e^{ix})\cos(jx)\,dx. \tag{A.30}$$

For the single-step SMM, this is approximately $e^{-2\mu t}I_j(2\mu t)$ in terms of the modified Bessel function of the first kind $I_j(x)$. Summing over the distribution of coalescence time, the probability that the "size" of two genes differs by j units is therefore approximately

$$\sum_t C_t e^{-2\mu t}I_j(2\mu t). \tag{A.31}$$

C_t must then be evaluated separately.

Very similar techniques have been used to describe sequence polymorphism under a simple model of sequence divergence (the infinite site model: Kimura 1969). Given the generating function ψ of the number of mutations per generation on a sequence and given the probability of IBD $\dot{Q}(\gamma)$, $\dot{Q}(\psi^2)$ is the generating function of the distribution of the number of nucleotide differences between two sequences (e.g., Griffiths 1981; Tachida 1985).

References

Page numbers of reference citations are given in square brackets.

Abramovitz, M., and I. A. Stegun, eds. 1972. *Handbook of mathematical functions*. New York: Dover. [39]

Abrams, P. A. 2000. The impact of habitat selection on the spatial heterogeneity of resources in varying environments. *Ecology* 81:2902–2913. [198]

Abrams, P. A. 2001a. Adaptive dynamics: neither F nor G. *Evolutionary Ecology Research* 3:369–373. [74]

Abrams, P. A. 2001b. Modelling the adaptive dynamics of traits involved in inter- and intraspecific interactions: an assessment of three methods. *Ecology Letters* 4:166–175. [78]

Abrams, P. A., H. Matsuda, and Y. Harada. 1993. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. *Evolutionary Ecology* 7:465–487. [70, 77, 79]

Ajar, É. submitted. Analysis of disruptive selection in subdivided populations. [85, 87, 121]

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* 33:707–740. [199]

Austerlitz, F., S. Mariette, N. Machon, P.-H. Gouyon, and B. Godelle. 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* 154:1309–1321. [63]

Axelrod, R. 1984. *The evolution of cooperation*. New York: Basic Books. [88, 90]

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. *Science* 211:1390–1396. Reprinted in Hamilton (2001), pp. 142–159. [88, 114, 116]

Bahlo, M., and R. C. Griffiths. 2000. Inference from gene trees in a subdivided population. *Theoretical Population Biology* 57:79–95. [211]

Balloux, F., H. Brünner, N. Lugon-Moulin, J. Hausser, and J. Goudet. 2000. Microsatellites can be misleading: an empirical and simulation study. *Evolution* 54:1414–1422. [64]

Balloux, F., and J. Goudet. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology* 11:771–783. [64]

Barton, N. H. 1993. The probability of fixation of a favoured allele in a subdivided population. *Genetical Research* 62:149–157. [11, 123, 150, 152, 196]

Barton, N. H., F. Depaulis, and A. M. Etheridge. 2002. Neutral evolution in spatially continuous populations. *Theoretical Population Biology* 61:31–48. [43]

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. In R. G. Harrison, ed., *Hybrid zones and the evolutionary process*, 13–45. Oxford: Oxford University Press. [84]

Barton, N. H., and M. Turelli. 1991. Natural and sexual selection on many loci. *Genetics* 127:229–255. [79]

Bateman, A. J. 1950. Is gene dispersion normal? *Heredity* 4:353–363. [31]

Beattie, A. J., and D. C. Culver. 1979. Neighborhood size in *Viola*. *Evolution* 33:1226–1229. [31]

Beerli, P., and J. Felsenstein. 1999. Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–773. [203, 213]

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the U.S.A.* 98:4563–4568. [201, 202, 203, 213]

Bennett, J. H., and F. E. Binet. 1956. Association between Mendelian factors with mixed selfing and random mating. *Heredity* 10:51–55. [136]

Birky, C. W. 1994. Relaxed and stringent genomes: why cytoplasmic genes don't obey Mendel's laws. *Journal of Heredity* 85:355–365. [160]

Birky, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms and models. *Annual Review of Genetics* 35:125–148. [160]

Birky, C. W., P. Fuerst, and T. Maruyama. 1989. Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121:613–627. [62, 160]

Boomsma, J. J., and A. Grafen. 1991. Colony-level sex ratio selection in the social Hymenoptera. *Journal of Evolutionary Biology* 3:383–407. [117]

Boshier, D. H., M. R. Chase, and K. S. Bawa. 1995. Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood, and population substructure. *American Journal of Botany* 82:484–490. [31]

Bost, B., C. Dillmann, and D. de Vienne. 1999. Fluxes and metabolic pools as model traits for quantitative genetics. I. The L-shaped distribution of gene effects. *Genetics* 153:2001–2012. [76]

Bourguet, D., and M. Raymond. 1998. The molecular basis of dominance relationships: the case of some recent adaptive genes. *Journal of Evolutionary Biology* 11:103–122. [81]

Bourke, A. F. G., and N. R. Franks. 1995. *Social evolution in ants*. Princeton: Princeton University Press. [137]

Brauchli, K., T. Killingback, and M. Doebeli. 1999. Evolution of cooperation in spatially structured populations. *Journal of Theoretical Biology* 200:405–417. [90, 116]

Bromham, L., A. Eyre-Walker, N. H. Smith, and J. Maynard Smith. 2003. Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends in Ecology and Evolution* 18:2–4. [160]

Brown, A. H. D. 1979. Enzyme polymorphism in plant populations. *Theoretical Population Biology* 15:1–42. [127, 133, 136]

Brown, J. S., and T. L. Vincent. 1987. Coevolution as an evolutionary game. *Evolution* 41:66–79. [68]

Bulmer, M. G. 1971. The effects of selection on genetic variability. *American Naturalist* 105:201–211. [77]

Bulmer, M. G. 1979. *Principles of statistics*. New York: Dover. [109, 201]

Bulmer, M. G. 1980. *The mathematical theory of quantitative genetics*. Oxford: Clarendon Press. [77]

Bürger, R. 2000. *The mathematical theory of selection, recombination, and mutation*. Chichester, U.K.: Wiley. [77, 79]

Bürger, R., and W. J. Ewens. 1995. Fixation probabilities of additive alleles in diploid populations. *Journal of Mathematical Biology* 33:557–575. [148]

Caballero, A. 1994. Developments in the prediction of effective population size. *Heredity* 73:657–679. [143, 144]

Caballero, A., and W. G. Hill. 1992a. Effective size of nonrandom mating populations. *Genetics* 130:909–916. [163]

Caballero, A., and W. G. Hill. 1992b. Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* 131:493–507. [11, 143]

Campbell, R. B. 1999. The coalescent time in the presence of background fertility selection. *Theoretical Population Biology* 55:260–269. [166]

Caswell, H. 2001. *Matrix population models*. Sunderland, Mass.: Sinauer. [xvi, 56, 154, 178, 194, 195]

Charlesworth, B. 1994. *Evolution in age-structured populations*. 2d ed. Cambridge: Cambridge University Press. [77, 103, 152, 164, 194, 195]

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303. [137]

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* 70:155–174. [137]

Charlesworth, D., and J. R. Pannell. 2001. Mating system and population genetic structure in the light of coalescent theory. In J. Silvertown and J. Antonovics, eds., *Integrating ecology and evolution in a spatial context*, 73–95. Oxford: Blackwell. [127, 133]

Cherry, J. L. 2003a. Selection in a subdivided population with dominance or local frequency dependence. *Genetics* 163:1511–1518. [152]

Cherry, J. L. 2003b. Selection in a subdivided population with local extinction and recolonization. *Genetics* 164:789–795. [152, 196]

Cherry, J. L., and J. Wakeley. 2003. A diffusion approximation for selection and drift in a subdivided population. *Genetics* 163:421–428. [11, 143]

Chesser, R. K., O. E. Rhodes, D. W. Sugg, and A. Schnabel. 1993. Effective sizes for subdivided populations. *Genetics* 135:1221–1232. [63]

Christiansen, F. B. 1975. Hard and soft selection in a subdivided population. *American Naturalist* 109:11–16. [123]

Christiansen, F. B. 1988. The Wahlund effect with overlapping generations. *American Naturalist* 131:149–156. [133]

Christiansen, F. B. 1991. On conditions for evolutionary stability for a continuously varying character. *American Naturalist* 138:37–50. [68, 69, 72, 74]

Clark, J. S., M. Lewis, and L. Horváth. 2001. Invasion by extremes: population spread with variation in dispersal and reproduction. *American Naturalist* 157:537–554. [42]

Clark, J. S., M. Silman, R. Kern, E. Macklin, and J. HilleRisLambers. 1999. Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology* 80:1475–1494. [31, 32]

Clutton-Brock, T. 2002. Breeding together: kin selection and mutualism in cooperative vertebrates. *Science* 296:69–72. [123]

Clutton-Brock, T. H., P. N. M. Brotherton, M. J. O'Riain, A. S. Griffin, D. Gaynor, R. Kansky, L. Sharpe, and G. M. McIlrath. 2001a. Contributions to cooperative rearing in meerkats. *Animal Behaviour* 61:705–710. [118, 119]

Clutton-Brock, T. H., P. N. M. Brotherton, A. F. Russell, M. J. O'Riain, D. Gaynor, R. Kansky, A. Griffin, M. Manser, L. Sharpe, G. M. McIlrath, T. Small, A. Moss, and S. Monfort. 2001b. Cooperation, control and concession in meerkat groups. *Science* 291:478–481. [117]

Cockerham, C. C., and B. S. Weir. 1986. Estimation of inbreeding parameters in stratified populations. *Annals of Human Genetics* 50:271–281. [19]

Cockerham, C. C., and B. S. Weir. 1987. Correlations, descent measures: drift with migration and mutation. *Proceedings of the National Academy of Sciences of the U.S.A.* 84:8512–8514. [18, 19]

Cockerham, C. C., and B. S. Weir. 1993. Estimation of gene flow from F-statistics. *Evolution* 47:855–863. [19]

Cohen, J. E. 1979. Ergodic theorems in demography. *Bulletin of the American Mathematical Society* 1:275–295. [56]

Comins, H., W. D. Hamilton, and R. May. 1980. Evolutionarily stable dispersal strategies. *Journal of Theoretical Biology* 82:205–230. Reprinted in Hamilton (1996), pp. 505–528. [116]

Cotterman, C. W. 1940. A calculus for statistico-genetics. Ph.D. thesis, Ohio State University, Columbus. Reprinted in 1974 in P. Ballonoff, ed., *Genetics and social structure*, 157–272. Stroudsburg, Pa.: Dowden, Hutchinson & Ross. [58, 63]

Courant, R., and F. John. 1989. *Introduction to calculus and analysis*. Vol. 2. New York: Springer-Verlag. [95]

Courchamp, F., T. Clutton-Brock, and B. Grenfell. 1999. Inverse density dependence and the Allee effect. *Trends in Ecology and Evolution* 14:405–410. [117]

Cox, D. R., and D. V. Hinkley. 1974. *Theoretical statistics*. London: Chapman & Hall. [201]

Cox, D. R., and H. D. Miller. 1965. *The theory of stochastic processes*. London: Chapman & Hall. [220]

Coyne, J. A., N. H. Barton, and M. Turelli. 1997. A critique of Sewall Wright's shifting balance theory of evolution. *Evolution* 51:643–671. [2]

Crawford, T. 1984. The estimation of neighborhood parameters for plant populations. *Heredity* 52:273–283. [31, 128]

Creel, S. R., and P. M. Waser. 1994. Inclusive fitness and reproductive strategies in dwarf mongooses. *Behavioral Ecology* 5:339–348. [119]

Crow, J. F. 1954. Breeding structure of populations. II. Effective population number. In O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush, eds., *Statistics and mathematics in biology*, 543–556. Ames: Iowa State University Press. [15]

Crow, J. F., and K. Aoki. 1984. Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences of the U.S.A.* 81:6073–6077. [62, 63, 224]

Crow, J. F., and M. Kimura. 1970. *An introduction to population genetics theory*. New York: Harper & Row. [79, 160]

Crow, J. F., and T. Maruyama. 1971. The number of neutral alleles maintained in a finite, geographically structured population. *Theoretical Population Biology* 2:437–453. [50]

Crozier, R. H., and P. Pamilo. 1996. *Evolution of social insect colonies*. Oxford: Oxford University Press. [117, 137]

Darwin, C. R. 1877. *The different forms of flowers on plants of the same species*. London: Murray. Electronic text available on the *Online Books Page* website, http://onlinebooks.library.upenn.edu. [127]

Darwin, C. R. 1984. *The various contrivances by which orchids are fertilised by insects*. 2d ed. Chicago: University of Chicago Press. [127]

David, P. 1998. Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* 80:531–537. [137]

Dawkins, R. 1979. Twelve misunderstandings of kin selection. *Zeitschrift für Tierpsychologie* 51:184–200. [118]

Dawkins, R. 1989. *The selfish gene*. 2d ed. Oxford: Oxford University Press. [105]

Day, T. 2001. Population structure inhibits evolutionary diversification under competition for resources. *Genetica* 112-113:71–86. [68, 87]

Day, T., and P. D. Taylor. 1998. Unifying genetic and game theoretic models of kin selection for continuous traits. *Journal of Theoretical Biology* 194:391–407. [108]

de Fraipont, M., J. Clobert, H. John-Alder, and S. Meylan. 2000. Increased pre-natal maternal corticosterone promotes philopatry of offspring in common lizards *Lacerta vivipara*. *Journal of Animal Ecology* 69:404–413. [209]

de Iorio, M., and R. C. Griffiths. submitted. Importance sampling on coalescent histories. [213]

de Jong, T. J., P. G. L. Klinkhamer, and J. L. H. de Heiden. 2000. The evolution of generation time in metapopulations of monocarpic perennial plants: some theoretical considerations and the example of the rare thistle *Carlina vulgaris*. *Evolutionary Ecology* 14:213–231. [189]

Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the U.S.A.* 91:3166–3170. [221]

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. *Nature* 400:354–357. [78]

Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of Mathematical Biology* 34:579–612. [195]

Diffendorfer, J. E. 1998. Testing models of source-sink dynamics and balanced dispersal. *Oikos* 81:417–433. [198]

Dobzhansky, T., and S. Wright. 1941. Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. *Genetics* 26:23–51. [2, 12]

Doebeli, M. 1999. The marginal value of relations (book review). *Journal of Evolutionary Biology* 12:837–838. [4]

Doebeli, M., and U. Dieckmann. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *American Naturalist* 156:S77–S101. [78]

Doebeli, M., and N. Knowlton. 1998. The evolution of interspecific mutualisms. *Proceedings of the National Academy of Sciences of the U.S.A.* 95:8676–8680. [121]

Doncaster, C. P., J. Clobert, B. Doligez, L. Gustafsson, and É. Danchin. 1997. Balanced dispersal between spatially varying local populations: an alternative to the source-sink model. *American Naturalist* 150:425–445. [30]

Donnelly, M. J., F. Simard, and T. Lehmann. 2002. Evolutionary studies of malaria vectors. *Trends in Parasitology* 18:75–80. [200]

Doolan, S. P., and D. W. Macdonald. 1996. Dispersal and extraterritorial prospecting by slender-tailed meerkats (*Suricata suricatta*) in the south-western Kalahari. *Journal of Zoology* 240:59–73. [117, 118]

Dugatkin, L. A. 1998. Game theory and cooperation. In L. A. Dugatkin and H. K. Reeve, eds., *Game theory and animal behavior*, 38–63. Oxford: Oxford University Press. [88]

Dugatkin, L. A., and H. K. Reeve. 1994. Behavioral ecology and levels of selection: dissolving the group selection controversy. *Advances in the Study of Behavior* 23:101–133. [122]

Dutech, C., J. Seiter, P. Petronelli, H. I. Joly, and P. Jarne. 2002. Evidence of low gene flow in a neotropical clustered tree species in two

rainforest stands of French Guyana. *Molecular Ecology* 11:725–738. [31]

Eanes, W. F. 1999. Analysis of selection on enzyme polymorphisms. *Annual Review of Ecology and Systematics* 30:301–326. [3]

Eguiarte, L. E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán, and D. Piñero. 1993. Direct and indirect estimates of neighborhood and effective population size in a tropical palm, *Astrocaryum mexicanum*. *Evolution* 47:75–87. [31]

Endler, J. A. 1977. *Geographical variation, speciation, and clines*. Princeton: Princeton University Press. [30, 31]

Eshel, I. 1972. On the neighbor effect and the evolution of altruistic traits. *Theoretical Population Biology* 3:258–277. [91]

Eshel, I. 1983. Evolutionary and continuous stability. *Journal of Theoretical Biology* 103:99–111. [4, 73]

Eshel, I. 1996. On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *Journal of Mathematical Biology* 34:485–510. [69, 74]

Eshel, I., M. W. Feldman, and A. Bergman. 1998. Long-term evolution, short-term evolution, and population genetic theory. *Journal of Theoretical Biology* 191:391–396. [77]

Eshel, I., and U. Motro. 1981. Kin selection and strong evolutionary stability of mutual help. *Theoretical Population Biology* 19:420–433. [73]

Eshel, I., U. Motro, and E. Sansone. 1997. Continuous stability and evolutionary convergence. *Journal of Theoretical Biology* 185:333–343. [72, 74]

Ethier, S. N., and T. Nagylaki. 1980. Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Advances in Applied Probability* 12:14–49. [11, 143, 148, 196]

Ethier, S. N., and T. Nagylaki. 1988. Diffusion approximations of Markov chains with two time scales and applications to population genetics, II. *Advances in Applied Probability* 20:525–545. [11, 143, 148]

Ewens, W. J. 1979. *Mathematical population genetics*. Berlin: Springer Verlag. [10, 11, 145, 146, 148]

Ewens, W. J. 1982. On the concept of the effective population size. *Theoretical Population Biology* 21:373–378. [144, 145, 146, 147]

Excoffier, L. 2003. Analysis of population subdivision. In D. J. Balding, M. Bishop, and C. Cannings, eds., *Handbook of statistical genetics*, 745–750. Chichester, U.K.: Wiley, second ed. [199]

Feder, J. L., J. B. Roethele, B. Wlazlo, and S. H. Berlocher. 1997. Selective maintenance of allozyme differences between sympatric host races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the U.S.A.* 94:11417–11421. [3]

Feldman, M. W., A. Bergman, D. D. Pollock, and D. B. Goldstein. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* 145:207–216. [64]

Feller, W. 1968. *An introduction to probability theory and its applications*. 3d ed. Vol. 1. New York: Wiley. [xvi, 54, 220]

Felsenstein, J. 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68:581–597. [144, 147, 152, 164]

Felsenstein, J. 1975. A pain in the torus: some difficulties with models of isolation by distance. *American Naturalist* 109:359–368. [43]

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* 10:253–280. [43]

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* 35:124–138. [78]

Fenster, C. B. 1991. Gene flow in *Chamaecrista fasciculata* (Leguminosae) I. Gene dispersal. *Evolution* 45:398–409. [31]

Fenster, C. B., X. Vekemans, and O. J. Hardy. 2003. Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution* 57:995–1007. [44]

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52:399–433. Available on the *Collected Papers of R.A. Fisher* website, http://www.library.adelaide.edu.au/digitised/fisher/index.html. [12, 77]

Fisher, R. A. 1922. On the dominance ratio. *Transactions of the Royal Society of Edinburgh* 42:321–341. [10]

Fisher, R. A. 1958. *The genetical theory of natural selection*. 2d ed. New York: Dover. [10, 86]

Frank, S. A. 1986. Dispersal polymorphism in subdivided populations. *Journal of Theoretical Biology* 122:303–309. [94, 100, 116, 121]

Frank, S. A. 1994a. Genetics of mutualism: the evolution of altruism between species. *Journal of Theoretical Biology* 170:393–400. [120]

Frank, S. A. 1994b. Kin selection and virulence in the evolution of protocells and parasites. *Proceedings of the Royal Society (London) B* 258:153–161. [100]

Frank, S. A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* 377:520–522. [121]

Frank, S. A. 1997a. Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *Journal of Theoretical Biology* 189:307–316. [108]

Frank, S. A. 1997b. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* 51:1712–1729. [121]

Frank, S. A. 1998. *Foundations of social evolution*. Princeton: Princeton University Press. [102, 106, 109, 120, 121, 142, 189, 192, 194]

Frankham, R. 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology* 10:1500–1508. [203]

Fretwell, S. D., and H. L. Lucas. 1970. On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica* 19:16–36. [198]

Friesen, V., W. Montevecchi, A. Gaston, R. Barrett, and W. Davidson. 1996. Molecular evidence for kin groups in the absence of large scale genetic differentiation in a migratory bird. *Evolution* 50:924–930. [30]

Fu, Y.-X. 1997. Coalescent theory for partially selfing populations. *Genetics* 146:1489–1499. [202]

Gadgil, M. 1971. Dispersal: population consequences and evolution. *Ecology* 52:253–260. [189]

Gale, J. S. 1990. *Theoretical population genetics*. London: Unwin Hyman. [10]

Gandon, S. 1999. Kin competition, the cost of inbreeding and the evolution of dispersal. *Journal of Theoretical Biology* 200:345–364. [142]

Gandon, S., M. Mackinnon, S. Nee, and A. Read. 2003. Imperfect vaccination: some epidemiological and evolutionary consequences. *Proceedings of the Royal Society (London) B* 270:1129–1136. [77]

Gandon, S., and F. Rousset. 1999. Evolution of stepping stone dispersal rates. *Proceedings of the Royal Society (London) B* 266:2507–2513. [100, 101]

Geritz, S. A. H., and É. Kisdi. 2000. Adaptive dynamics in diploid, sexual populations and the evolution of reproductive isolation. *Proceedings of the Royal Society (London) B* 267:1671–1678. [75]

Geritz, S. A. H., É. Kisdi, G. Meszéna, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* 12:35–57. [69, 72, 74]

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471. [64]

Gomulkiewicz, R. 1998. Game theory, optimization, and quantitative genetics. In L. A. Dugatkin and H. K. Reeve, eds., *Game theory and animal behavior*, 283–303. Oxford: Oxford University Press. [77]

Grafen, A. 1984. Natural selection, kin selection and group selection. In J. R. Krebs and N. B. Davies, eds., *Behavioural ecology. An evolutionary approach*, 62–84. London: Blackwell. [105, 106, 107, 122]

Grafen, A. 1985. A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology* 2:28–89. [84, 109, 120, 121]

Greene, D. R., and E. A. Johnson. 1989. A model of wind dispersal of winged or plumed seeds. *Ecology* 70:339–347. [32]

Griffiths, R. C. 1980. Allele frequencies in multidimensional Wright-Fisher models with a general symmetric mutation structure. *Theoretical Population Biology* 17:51–70. [221]

Griffiths, R. C. 1981. Neutral two-locus multiple allele models with recombination. *Theoretical Population Biology* 19:169–186. [225]

Griffiths, R. C., and S. Tavaré. 1994. Simulating probability distributions in the coalescent. *Theoretical Population Biology* 46:131–159. [202, 211]

Grimmett, G. R., and D. R. Stirzaker. 1992. *Probability and random processes*. 2d ed. Oxford: Oxford University Press. [xvi, 220]

Gyllenberg, M., K. Parvinen, and U. Dieckmann. 2002. Evolutionary suicide and evolution of dispersal in structured metapopulations. *Journal of Mathematical Biology* 45:79–105. [197]

Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Transactions of the Cambridge Philosophical Society* 28:838–844. [10, 86]

Haldane, J. B. S. 1932. *The causes of evolution*. London: Longmans Green. Reprinted by Princeton University Press (1990). [103]

Hamilton, W. D. 1963. The evolution of altruistic behavior. *American Naturalist* 97:354–356. Reprinted in Hamilton (1996), pp. 6–8. [3]

Hamilton, W. D. 1964. The genetical evolution of social behavior. I. *Journal of Theoretical Biology* 7:1–16. Reprinted in Hamilton (1996), pp. 31–46. [3, 12, 84, 106, 107, 108, 112, 114, 120, 206]

Hamilton, W. D. 1967. Extraordinary sex ratios. *Science* 156:477–488. Reprinted in Hamilton (1996), pp. 143–169. [4, 74]

Hamilton, W. D. 1970. Selfish and spiteful behaviour in an evolutionary model. *Nature* 228:1218–1220. Reprinted in Hamilton (1996), pp. 177-182. [84, 108, 114, 121]

Hamilton, W. D. 1971. Selection of selfish and altruistic behavior in some extreme models. In J. F. Eisenberg and W. S. Dillon, eds., *Man and Beast: comparative social behavior*, 58–91. Washington, D.C.: Smithsonian Institution Press. Reprinted in Hamilton (1996), pp. 198–227. [12, 106, 141]

Hamilton, W. D. 1975. Innate social aptitudes in man, an approach from evolutionary genetics. In R. Fox, ed., *Biosocial anthropology*, 133–157. London: Malaby Press. Reprinted in Hamilton (1996), pp. 329–361. [106, 116, 120, 121, 122]

Hamilton, W. D. 1996. *Narrow roads to gene land*. Vol. 1. Oxford: Freeman. [74, 114, 116, 232, 239]

Hamilton, W. D. 2001. *Narrow roads to gene land*. Vol. 2. Oxford: Oxford University Press. [121, 227]

Hamilton, W. D., and R. May. 1977. Dispersal in stable habitats. *Nature* 269:578–581. Reprinted in Hamilton (1996), pp. 377–385. [100]

Hammerstein, P. 1996. Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology* 34:511–532. [77]

Hamrick, J. L., and M. J. W. Godt. 1990. Allozyme diversity in plant species. In A. H. D. Brown, M. T. Clegg, A. L. Kahler, and B. S.

Weir, eds., *Plant population genetics, breeding, and genetic resources*, 43–63. Sunderland, Mass.: Sinauer. [133]

Hamrick, J. L., and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society (London) B* 351:1291–1298. [133]

Hardy, O. J., and X. Vekemans. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* 83:145–154. [36]

Hartl, D. L., and A. G. Clark. 1989. *Principles of population genetics*. second ed. Sunderland, Mass.: Sinauer. [136]

Hartl, D. L., and A. G. Clark. 1997. *Principles of population genetics*. third ed. Sunderland, Mass.: Sinauer. [12, 81]

Hastings, A., and S. Harrison. 1994. Metapopulation dynamics and genetics. *Annual Review of Ecology and Systematics* 25:167–188. [200]

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109. [213]

Hatchwell, B. J., D. J. Ross, M. K. Fowlie, and A. McGowan. 2001. Kin discrimination in cooperatively breeding long-tailed tits. *Proceedings of the Royal Society (London) B* 268:885–890. [119]

Hedrick, P. W. 1999. Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313–318. [41]

Hedrick, P. W., and M. E. Gilpin. 1997. Genetic effective size of a metapopulation. In I. A. Hanski and M. E. Gilpin, eds., *Metapopulation biology: ecology, genetics and evolution*, 165–181. San Diego: Academic Press. [175]

Heino, M., and I. Hanski. 2001. Evolution of migration rate in a spatially realistic metapopulation model. *American Naturalist* 157:495–511. [209]

Heywood, J. S. 1991. Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics* 22:335–355. [133]

Hill, W. G. 1972. Effective size of populations with overlapping generations. *Theoretical Population Biology* 3:278–289. [55, 63, 152]

Hill, W. G. 1979. A note on effective population size with overlapping generations. *Genetics* 92:317–322. [164, 165, 166]

Holt, R. D. 1997. On the evolutionary stability of sink populations. *Evolutionary Ecology* 11:723–731. [198]

Holt, R. D., and M. A. McPeek. 1996. Chaotic population dynamics favors the evolution of dispersal. *American Naturalist* 148:709–718. [85]

Horn, R. A., and C. R. Johnson. 1985. *Matrix analysis*. Cambridge: Cambridge University Press. [xvi, 56, 217, 218, 220]

Horn, R. A., and C. R. Johnson. 1991. *Topics in matrix analysis*. Cambridge: Cambridge University Press. [219]

Houston, A. I., and J. M. McNamara. 1992. Phenotypic plasticity as a state-dependent life-history decision. *Evolutionary Ecology* 6:243–253. [189]

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* 7:1–44. [64, 202]

Iizuka, M. 2001. The effective size of fluctuating populations. *Theoretical Population Biology* 59:281–286. [176, 177, 178]

Iizuka, M., H. Tachida, and H. Matsuda. 2002. A neutral model with fluctuating population size and its effective size. *Genetics* 161:381–388. [176]

Ingvarsson, P. K., and B. Giles. 1999. Kin-structured colonization and small-scale genetic differentiation in *Silene dioica*. *Evolution* 53:605–611. [135]

Ingvarsson, P. K., K. Olsson, and L. Ericson. 1997. Extinction-recolonization dynamics in the mycophagous beetle *Phalarus substriatus*. *Evolution* 51:187–195. [44, 187]

Jansen, V. A. A., and G. S. E. E. Mulder. 1999. Evolving biodiversity. *Ecology Letters* 2:379–386. [77]

Jarne, P., and D. Charlesworth. 1993. The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Annual Review of Ecology and Systematics* 24:441–466. [127]

Johnson, C. N., M. Clinchy, A. C. Taylor, C. J. Krebs, P. J. Jarman, A. Payne, and E. G. Ritchie. 2001. Adjustment of offspring sex ratios in relation to the availability of resources for philopatric offspring in the common brushtail possum. *Proceedings of the Royal Society (London) B* 268:2001–2005. [209]

Kaj, I., S. M. Krone, and M. Lascoux. 2001. Coalescent theory for seed bank models. *Journal of Applied Probability* 38:285–301. [165]

Karlin, S., and H. M. Taylor. 1975. *A first course in stochastic processes*. 2d ed. San Diego: Academic Press. [219]

Karlin, S., and H. M. Taylor. 1981. *A second course in stochastic processes*. San Diego: Academic Press. [148]

Kawecki, T. J. 1993. Age and size at maturity in a patchy environment: fitness maximization versus evolutionary stability. *Oikos* 66:309–317. [103]

Kawecki, T. J., and S. C. Stearns. 1993. The evolution of life histories in spatially heterogeneous environments: optimal reaction norms revisited. *Evolutionary Ecology* 7:155–174. [189]

Keller, L. 1997. Indiscriminate altruism: unduly nice parents and siblings. *Trends in Ecology and Evolution* 12:99–103. [118]

Kempthorne, O. 1954. The correlation between relatives in a random mating population. *Proceedings of the Royal Society (London) B* 143:103–113. [15]

Killingback, T., and M. Doebeli. 2002. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *American Naturalist* 160:421–438. [90]

Killingback, T., M. Doebeli, and N. Knowlton. 1999. Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *Proceedings of the Royal Society (London) B* 266:1723–1728. [116]

Kimura, M. 1957. Some problems of stochastic processes in genetics. *Annals of Mathematical Statistics* 28:882–901. [148]

Kimura, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61:893–903. [225]

Kimura, M., and J. F. Crow. 1963. The measurement of effective population number. *Evolution* 17:279–288. [136]

Kimura, M., and T. Maruyama. 1971. Pattern of neutral polymorphism in a geographically structured population. *Genetical Research* 18:125–131. [37]

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. [35]

Kingman, J. 1980. *Mathematics of genetic diversity*. Philadelphia: Society for Industrial and Applied Mathematics. [211]

Kingman, J. F. C. 1982. On the genealogy of large populations. *Journal of Applied Probability* 19A:27–43. [202]

Kisdi, É. 1999. Evolutionary branching under asymmetric competition. *Journal of Theoretical Biology* 197:149–162. [72, 74]

Koella, J. C. 2000. The spatial spread of altruism versus the evolutionary response of egoists. *Proceedings of the Royal Society (London) B* 267:1979–1985. [114]

Kreitman, M., and H. Akashi. 1995. Molecular evidence for natural selection. *Annual Review of Ecology and Systematics* 26:403–422. [2]

Lande, R. 1992. Neutral model of quantitative genetic variance in an island model with local extinction and recolonization. *Evolution* 46:381–389. [65]

Laporte, V., J. Cuguen, and D. Couvet. 2000. Effective population sizes for cytoplasmic and nuclear genes in a gynodioecious species: the role of the sex determination system. *Genetics* 154:447–458. [160]

Latter, B. D. H. 1959. Genetic sampling in a random mating population of constant size and sex ratio. *Australian Journal of Biological Sciences* 12:500–505. [166]

Leblois, R., A. Estoup, and F. Rousset. 2003. Influence of mutational and sampling factors on the estimation of demographic parameters in a "continuous" population under isolation by distance. *Molecular Biology and Evolution* 20:491–502. [64, 205]

Lehmann, L., and N. Perrin. 2002. Altruism, dispersal, and phenotype-matching kin recognition. *American Naturalist* 159:451–468. [207]

Lehmann, L., and N. Perrin. in press. Inbreeding avoidance through kin recognition: choosy females boost male dispersal. *American Naturalist*. [208]

Leimar, O. 1997. Reciprocity and communication of partner quality. *Proceedings of the Royal Society (London) B* 264:1209–1215. [88]

Leimar, O. in press. Multi-dimensional convergence stability and the canonical adaptive dynamics. In U. Dieckmann and J. A. J. Metz, eds., *Elements of adaptive dynamics*. Cambridge: Cambridge University Press. [79]

Lenormand, T., D. Bourguet, T. Guillemaud, and M. Raymond. 1999. Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* 400:861–864. [30]

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. *Biometrika* 33:183–212. [154]

Leslie, P. H. 1948. Some further notes on the use of matrices in population mathematics. *Biometrika* 35:213–245. [154]

Lessard, S. 1990. Evolutionary stability: one concept, several meanings. *Theoretical Population Biology* 37:159–170. [69, 74]

Leturque, H., and F. Rousset. 2002. Convergence stability, probabilities of fixation, and the ideal free distribution in a class-structured population. *Theoretical Population Biology* 62:169–180. [192, 198]

Levene, H. 1949. On a matching problem arising in genetics. *Annals of Mathematical Statistics* 20:91–94. [136]

Levin, D. A., and H. W. Kerster. 1974. Gene flow in seed plants. *Evolutionary Biology* 7:139–220. [135]

Levin, S. A., D. Cohen, and A. Hastings. 1984. Dispersal strategies in patchy environments. *Theoretical Population Biology* 26:165–191. [189]

Levin, S. A., and S. W. Pacala. 1997. Theories of simplification and scaling of spatially distributed processes. In D. Tilman and P. Kareiva, eds., *Spatial ecology*, 271–295. Princeton: Princeton University Press. [209]

Lewis, O. T., C. D. Thomas, J. K. Hill, M. I. Brookes, T. P. R. Crane, Y. A. Graneau, J. L. B. Mallet, and O. C. Rose. 1997. Three ways of assessing metapopulation structure in the butterfly *Plebejus argus*. *Ecological Entomology* 22:283–293. [44]

Lewontin, R. 1974. *The genetic basis of evolutionary change*. New York: Columbia University Press. [2]

Li, W.-H. 1976. Electrophoretic identity of proteins in a finite population and genetic distance between taxa. *Genetical Research* 28:119–127. [225]

Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* 152:1753–1766. [201]

Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sunderland, Mass.: Sinauer. [12, 77]

Mackay, T. F. C. 2001. The genetic architecture of quantitative traits. *Annual Review of Genetics* 35:303–339. [77]

MacNair, M. R. 1991. Why the evolution of resistance to anthropogenic toxins normally involves major gene changes: the limits to natural selection. *Genetica* 84:213–219. [77]

Malécot, G. 1948. *Les mathématiques de l'hérédité*. Paris: Masson. [15, 16, 28, 63]

Malécot, G. 1950. Quelques schémas probabilistes sur la variabilité des populations naturelles. *Annales de l'Université de Lyon A* 13:37–60. [28, 38]

Malécot, G. 1951. Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en génétique de population. *Annales de l'Université de Lyon A* 14:79–117. [38, 46]

Malécot, G. 1952. Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques. *Publications de l'Institut de Statistique de l'Université de Paris* 1:1–16. [10, 148]

Malécot, G. 1967. Identical loci and relationship, In L. M. Le Cam and J. Neyman, eds., *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 4, 317–332, Berkeley: University of California Press. [38, 43]

Malécot, G. 1975. Heterozygosity and relationship in regularly subdivided populations. *Theoretical Population Biology* 8:212–241. [15, 38, 43, 46, 48]

Maruyama, K., and H. Tachida. 1992. Genetic variability and geographical structure in partially selfing populations. *Japanese Journal of Genetics* 67:39–51. [128]

Maruyama, T. 1970a. Effective number of alleles in a subdivided population. *Theoretical Population Biology* 1:273–306. [50]

Maruyama, T. 1970b. Stepping stone models of finite length. *Advances in Applied Probability* 2:229–258. [33, 46]

Maruyama, T. 1972. Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. *Genetics* 70:639–651. [38, 43, 152, 161, 162]

Maruyama, T. 1974. A simple proof that certain quantities are independent of the geographical structure of population. *Theoretical Population Biology* 5:148–154. [125]

Maruyama, T. 1983. Stochastic theory of population genetics. *Bulletin of Mathematical Biology* 45:521–554. [11, 125, 150]

Maruyama, T., and M. Kimura. 1980. Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences of the U.S.A.* 77:6710–6714. [145, 175]

Matessi, C., and C. Di Pasquale. 1996. Long-term evolution of multilocus traits. *Journal of Mathematical Biology* 34:613–653. [72, 79]

Maynard Smith, J. 1964. Group selection and kin selection. *Nature* 201:1145–1147. [3]

Maynard Smith, J. 1982. *Evolution and the theory of games*. Cambridge: Cambridge University Press. [4, 73, 88, 90, 108]

Maynard Smith, J. 1987. How to model evolution. In J. Dupré, ed., *The latest on the best*, 119–131. Cambridge: MIT Press. [106, 107]

Maynard Smith, J., and R. L. W. Brown. 1986. Competition and body size. *Theoretical Population Biology* 30:166–179. [81, 95]

Maynard Smith, J., and G. Price. 1973. The logic of animal conflict. *Nature* 246:15–18. [67, 73]

Mayo, O., and R. Bürger. 1997. The evolution of dominance: a theory whose time has passed? *Biological Reviews* 72:97–110. [81]

McCauley, D. E., M. S. Olson, and D. R. Taylor. 1990. An association between chloroplast DNA haplotype and gender in a plant metapopulation. *Evolutionary Ecology* 14:181–194. [189]

Merilä, J., and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology* 14:892–903. [65]

Metz, J. A. J., and M. Gyllenberg. 2001. How should we define fitness in structured metapopulation models? Including an application to the calculation of evolutionarily stable dispersal strategies. *Proceedings of the Royal Society (London) B* 268:499–508. [85, 87, 91, 190]

Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz. 1992. How should we define 'fitness' for general ecological scenarios? *Trends in Ecology and Evolution* 7:198–202. [84]

Michod, R. E. 1982. The theory of kin selection. *Annual Review of Ecology and Systematics* 13:23–55. [106, 120]

Michod, R. E., and W. D. Hamilton. 1980. Coefficients of relatedness in sociobiology. *Nature* 288:694–697. [12, 106, 142]

Milligan, B. G. 2003. Maximum-likelihood estimation of relatedness. *Genetics* 163:1153–1167. [201]

Mock, D. W., and G. A. Parker. 1998. *The evolution of sibling rivalry*. Oxford: Oxford University Press. [137]

Möhle, M. 1998. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Advances in Applied Probability* 30:493–512. [202]

Moran, P. A. P. 1975. Wandering distributions and the electrophoretic profile. *Theoretical Population Biology* 8:318–330. [225]

Motro, U. 1982. Optimal rates of dispersal. I. Haploid populations. *Theoretical Population Biology* 21:394–411. [91]

Motro, U. 1991. Avoiding inbreeding and sibling competition: the evolution of sexual dimorphism for dispersal. *American Naturalist* 137:108–115. [142]

Motro, U. 1994. Evolutionary and continuous stability in asymmetric games with continuous strategy sets: the parental investment conflict as an example. *American Naturalist* 144:229–241. [79]

Motro, U., and G. Thomson. 1982. On heterozygosity and the effective size of populations subject to size changes. *Evolution* 36:1059–1066. [176]

Mylius, S. D., and O. Diekmann. 1995. On evolutionarily stable life histories, optimization and the need to be specific about density dependence. *Oikos* 74:218–224. [195]

Nagylaki, T. 1974. The decay of genetic variability in geographically structured populations. *Proceedings of the National Academy of Sciences of the U.S.A.* 71:2932–2936. [38, 43]

Nagylaki, T. 1975. Conditions for the existence of clines. *Genetics* 80:595–615. [84]

Nagylaki, T. 1976. The decay of genetic variability in geographically structured populations. II. *Theoretical Population Biology* 10:70–82. [38]

Nagylaki, T. 1980. The strong migration limit in geographically structured populations. *Journal of Mathematical Biology* 9:101–114. [152, 157, 158]

Nagylaki, T. 1982. Geographical invariance in population genetics. *Journal of Theoretical Biology* 99:159–172. [50]

Nagylaki, T. 1983. The robustness of neutral models of geographical variation. *Theoretical Population Biology* 24:268–294. [26, 49, 62, 128]

Nagylaki, T. 1985. Biased intrachromosomal gene conversion in a chromosome lineage. *Journal of Mathematical Biology* 21:215–235. [148]

Nagylaki, T. 1989. Gustave Malécot and the transition from classical to modern population genetics. *Genetics* 122:253–268. [15, 38]

Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics* 134:627–647. [79]

Nagylaki, T. 1994. Geographical variation in a quantitative character. *Genetics* 136:361–381. [65]

Nagylaki, T. 1998a. The expected number of heterozygous sites in a subdivided population. *Genetics* 149:1599–1604. [64]

Nagylaki, T. 1998b. Fixation indices in subdivided populations. *Genetics* 148:1325–1332. [41]

Nakamaru, M., H. Matsuda, and Y. Iwasa. 1997. The evolution of cooperation in a lattice-structured population. *Journal of Theoretical Biology* 184:65–81. [209]

Nath, H. B., and R. C. Griffiths. 1996. Estimation in an island model using simulation. *Theoretical Population Biology* 50:227–253. [211]

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the U.S.A.* 70:3321–3323. [19]

Nei, M. 1986. Definition and estimation of fixation indices. *Evolution* 40:643–645. [19]

Neubert, M. G., M. Kot, and M. A. Lewis. 1995. Dispersal and pattern formation in a discrete-time predator-prey model. *Theoretical Population Biology* 48:7–43. [32]

Nordborg, M. 1997. Structured coalescent processes on different time scales. *Genetics* 146:1501–1514. [132, 137, 158, 202]

Nordborg, M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154:923–929. [137]

Nordborg, M. 2003. Coalescent theory. In D. J. Balding, M. Bishop, and C. Cannings, eds., *Handbook of statistical genetics*, 631–635. Chichester, U.K.: Wiley, second ed. [64, 202]

Nordborg, M., and P. Donnelly. 1997. The coalescent process with selfing. *Genetics* 146:1185–1195. [127, 202]

Nordborg, M., and S. M. Krone. 2002. Separation of time scales and convergence to the coalescent in structured populations. In M. Slatkin and M. Veuille, eds., *Modern developments in theoretical population genetics*, 194–232. Oxford: Oxford University Press. [165, 202]

Notohara, M. 1993. The strong-migration limit for the genealogical process in geographically structured populations. *Journal of Mathematical Biology* 31:115–122. [202]

Nowak, M. 1990. An evolutionarily stable strategy may be inaccessible. *Journal of Theoretical Biology* 142:237–241. [73, 90]

Nowak, M., and K. Sigmund. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's dilemma game. *Nature* 364:56–58. [90]

Nunney, L. 1991. The influence of age structure and fecundity on effective population size. *Proceedings of the Royal Society (London) B* 246:71–76. [169]

Nunney, L. 1998. Are we selfish, are we nice, or are we nice because we are selfish? (book review). *Science* 281:1620–1621. [106]

Oddou-Muratorio, S., R. J. Petit, B. Le Guerroué, D. Guesnet, and B. Démesure. 2001. Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution* 55:1123–1135. [31]

Ohta, T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences of the U.S.A.* 79:1940–1944. [19]

Ohta, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* 22:201–204. [221]

Orive, M. E. 1993. Effective population size in organisms with complex life-histories. *Theoretical Population Biology* 44:316–340. [167]

Orr, H. A., and J. A. Coyne. 1992. The genetics of adaptation: a reassessment. *American Naturalist* 140:725–742. [77]

Otto, S. P., and N. H. Barton. 2001. Selection for recombination in small populations. *Evolution* 55:1921–1931. [80]

Otto, S. P., and T. Lenormand. 2002. Resolving the paradox of sex and recombination. *Nature Reviews Genetics* 3:252–261. [80]

Packer, C., and A. Pusey. 1993. Dispersal, kinship, and inbreeding in African lions. In N. W. Thornhill, ed., *The natural history of inbreeding and outbreeding*, 375–391. Chicago: University of Chicago Press. [30]

Pannell, J. R. 2000. A hypothesis for the evolution of androdioecy: the joint influence of reproductive assurance and local mate competition in a metapopulation. *Evolutionary Ecology* 14:195–211. [189]

Pannell, J. R., and B. Charlesworth. 1999. Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53:664–676. [186]

Parvinen, K. 1999. Evolution of migration in a metapopulation. *Bulletin of Mathematical Biology* 61:531–550. [85]

Perrin, N., and L. Lehmann. 2001. Is sociality driven by the costs of dispersal or the benefits of philopatry? A role for kin-discrimination mechanisms. *American Naturalist* 158:471–483. [114, 118]

Perrin, N., and V. Mazalov. 1999. Dispersal and inbreeding avoidance. *American Naturalist* 154:282–292. [142]

Pollak, E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117:353–360. [163]

Pollak, E., and M. Sabran. 1992. On the theory of partially inbreeding finite populations. III. Fixation probabilities under partial selfing when heterozygotes are intermediate in viability. *Genetics* 131:979–985. [11, 143]

Porter, A. H., R. Wenger, H. Geiger, A. Scholl, and A. M. Shapiro. 1997. The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. *Evolution* 51:1561–1573. [30]

Portnoy, S., and M. F. Willson. 1993. Seed dispersal curves: behavior of the tails of the distribution. *Evolutionary Ecology* 7:25–44. [31]

Price, G. 1970. Selection and covariance. *Nature* 227:520–521. [121]

Price, G. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics* 35:485. [121, 122]

Provine, W. B. 1986. *Sewall Wright and evolutionary biology*. Chicago: University of Chicago Press. [1, 52]

Pulliam, H. R. 1996. Sources and sinks: empirical evidence and population consequences. In O. E. Rhodes, R. K. Chesser, and M. H. Smith, eds., *Population dynamics in ecological space and time*, 45–69. Chicago: University of Chicago Press. [198]

Queller, D. C. 1992. Quantitative genetics, inclusive fitness, and group selection. *American Naturalist* 139:540–558. [121]

Rand, D. A., H. B. Wilson, and J. M. McGlade. 1994. Dynamics and evolution: evolutionarily stable attractors, invasion exponents and phenotype dynamics. *Philosophical Transactions of the Royal Society (London) B* 343:261–283. [195]

Raufaste, N., and F. Bonhomme. 2000. Properties of bias and variance of two multiallelic estimators of F_{ST}. *Theoretical Population Biology* 57:285–296. [201]

Raymond, M., C. Berticat, M. Weill, N. Pasteur, and C. Chevillon. 2001. Insecticide resistance in the mosquito *Culex pipiens*: what

have we learned about adaptation? *Genetica* 112–113:287–293. [77]

Reeve, H. K. 2000. Multi-level selection and human cooperation (book review). *Evolution and Human Behavior* 21:65–72. [106, 122]

Reeve, H. K., and L. Keller. 1999. Levels of selection: burying the units-of-selection debate and unearthing the crucial new issues. In L. Keller, ed., *Levels of selection in evolution*, 3–14. Princeton: Princeton University Press. [106]

Richardson, D. S., T. Burke, and J. Komdeur. 2002. Direct benefits and the evolution of female-biased cooperative breeding in Seychelles warblers. *Evolution* 56:2313–2321. [119]

Robertson, A. 1961. Inbreeding in artificial selection programmes. *Genetical Research* 2:189–194. [166]

Ronce, O., and I. Olivieri. in press. What makes life in a metapopulation so special?. In I. Hanski and O. Gaggiotti, eds., *Ecology, genetics, and evolution of metapopulations*. San Diego: Academic Press. [190, 209]

Ronce, O., F. Perret, and I. Olivieri. 2000. Landscape dynamics and evolution of colonizer syndromes: interactions between reproductive effort and dispersal in a metapopulation. *Evolutionary Ecology* 14:233–260. [189]

Ronfort, J., É. Jenczewski, T. Bataillon, and F. Rousset. 1998. Analysis of population structure in autotetraploid species. *Genetics* 150:921–930. [132]

Røskaft, E., A. Wara, and Å. Viken. 1992. Reproductive success in relation to resource-access and parental age in a small Norwegian farming parish during the period 1700–1900. *Ethology and Sociobiology* 13:443–461. [154]

Rousset, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* 142:1357–1362. [59, 61, 136, 225]

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228. [40, 41, 44]

Rousset, F. 1999a. Genetic differentiation in populations with different classes of individuals. *Theoretical Population Biology* 55:297–308. [158, 164, 167]

Rousset, F. 1999b. Genetic differentiation within and between two habitats. *Genetics* 151:397–407. [40, 41, 168, 173]

Rousset, F. 2000. Genetic differentiation between individuals. *Journal of Evolutionary Biology* 13:58–62. [43, 44]

Rousset, F. 2001. Genetic approaches to the estimation of dispersal rates. In J. Clobert, É. Danchin, A. A. Dhondt, and J. D. Nichols, eds., *Dispersal*, 18–28. Oxford: Oxford University Press. [42, 62]

Rousset, F. 2002. Inbreeding and relatedness coefficients: what do they measure? *Heredity* 88:371–380. [58, 59, 110]

Rousset, F. 2003a. Effective size in simple metapopulation models. *Heredity* 91:107–111. [185, 186, 187]

Rousset, F. 2003b. Inferences from spatial population genetics. In D. J. Balding, M. Bishop, and C. Cannings, eds., *Handbook of statistical genetics*, 681–712. Chichester, U.K.: Wiley, second ed. [19, 44, 199, 201]

Rousset, F. 2003c. A minimal derivation of convergence stability measures. *Journal of Theoretical Biology* 221:665–668. [91]

Rousset, F., and S. Billiard. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *Journal of Evolutionary Biology* 13:814–825. [86]

Rousset, F., and O. Ronce. in press. Inclusive fitness for traits affecting metapopulation demography. *Theoretical Population Biology*. [183, 190, 193, 197]

Roze, D., and F. Rousset. 2003. Diffusion approximations for selection and drift in subdivided populations: a straightforward method and examples involving dominance, selfing and local extinctions. *Genetics* in press. [141, 142, 143, 148, 150, 152, 196, 207]

Ruckelshaus, M. H. 1996. Estimation of genetic neighborhood parameters from pollen and seed dispersal in the marine angiosperm *Zostera marina* L. *Evolution* 50:856–864. [31]

Rummel, J. D., and J. Roughgarden. 1985. A theory of faunal buildup for competition communities. *Evolution* 39:1009–1033. [68]

Russell, A. F., and B. J. Hatchwell. 2001. Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proceedings of the Royal Society (London) B* 268:2169–2174. [119]

Santiago, E., and A. Caballero. 1995. Effective size of populations under selection. *Genetics* 139:1013–1030. [166]

Santiago, E., and A. Caballero. 1998. Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* 149:2105–2117. [137, 166]

Sawyer, S. 1976. Results for the stepping stone model for migration in population genetics. *Annals of Probability* 4:699–728. [38]

Sawyer, S. 1977. Asymptotic properties of the equilibrium probability of identity in a geographically structured population. *Advances in Applied Probability* 9:268–282. [38, 39, 43]

Seger, J. 1981. Kinship and covariance. *Journal of Theoretical Biology* 91:191–213. [14, 112, 121]

Slatkin, M. 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* 12:253–262. [184, 186]

Slatkin, M. 1979. Frequency- and density-dependent selection on a quantitative character. *Genetics* 93:755–771. [78]

Slatkin, M. 1981. Fixation probabilities and fixation times in a subdivided population. *Evolution* 35:477–488. [11, 152]

Slatkin, M. 1987. The average number of sites separating DNA sequences drawn from a subdivided population. *Theoretical Population Biology* 32:42–49. [2, 49]

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genetical Research* 58:167–175. [15, 58]

Slatkin, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47:264–279. [63]

Slatkin, M. 1994. Gene flow and population structure. In L. A. Real, ed., *Ecological Genetics*, 3–17. Princeton: Princeton University Press. [62, 199, 200]

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462. [64]

Slatkin, M., and N. H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–1368. [43]

Slatkin, M., and L. Voelm. 1991. F_{ST} in a hierarchical island model. *Genetics* 127:627–629. [132]

Spitze, K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* 135:367–374. [65]

Spong, G., and S. Creel. 2001. Deriving dispersal distance from genetic data. *Proceedings of the Royal Society (London) B* 268:2571–2574. [30]

Stephan, W., B. Charlesworth, and G. McVean. 1999. The effect of background selection at a single locus on weakly selected, partially linked variants. *Genetical Research* 73:133–146. [137, 143]

Stephens, M., and P. Donnelly. 2000. Inference in molecular population genetics (with discussion). *Journal of the Royal Statistical Society* 62:605–655. [213]

Storz, J. F., H. R. Bhat, and T. H. Kunz. 2001. Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*. I. Inbreeding, outbreeding, and population subdivision. *Evolution* 55:1215–1223. [135]

Stoyan, D., and S. Wagner. 2001. Estimating the fruit dispersion of anemochorous forest trees. *Ecological Modelling* 145:35–47. [32]

Strobeck, C. 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* 117:149–153. [49]

Stuart, A., and J. K. Ord. 1991. *Kendall's advanced theory of statistics*. 5th ed. Vol. 2. London: Edward Arnold. [109, 201]

Sumner, J., A. Estoup, F. Rousset, and C. Moritz. 2001. 'Neighborhood' size, dispersal and density estimates in the prickly forest skink (*Gnypetoscincus queenslandiae*) using individual genetic and demographic methods. *Molecular Ecology* 10:1917–1927. [44]

Tachida, H. 1985. Joint frequencies of alleles determined by separate formulations for the mating and mutation systems. *Genetics* 111:963–974. [223, 225]

Tachida, H., and C. C. Cockerham. 1986. Analysis of linkage disequilibrium in an island model. *Theoretical Population Biology* 29:161–197. [19]

Tachida, H., and H. Yoshimaru. 1996. Genetic diversity in partially selfing populations with the stepping-stone structure. *Heredity* 77:469–475. [128]

Takada, T., and J. Kigami. 1991. The dynamical attainability of ESS in evolutionary games. *Journal of Mathematical Biology* 29:513–529. [69]

Takahata, N. 1983. Gene identity and genetic differentiation of populations in the finite island model. *Genetics* 104:497–512. [62, 207]

Takahata, N. 1991. Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics* 129:585–595. [152, 202]

Takahata, N., and M. Nei. 1984. F_{ST} and G_{ST} statistics in the finite island model. *Genetics* 107:501–504. [19]

Takahata, N., and S. R. Palumbi. 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics* 109:441–457. [160]

Tanksley, S. D. 1993. Mapping polygenes. *Annual Review of Genetics* 27:205–233. [77]

Taper, M. L., and T. J. Case. 1992. Models of character displacement and the theoretical robustness of taxon cycles. *Evolution* 46:317–333. [68]

Tavaré, S. 1984. Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Population Biology* 26:119–164. [202]

Taylor, E. B., and J. D. McPhail. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society (London) B* 267:2375–2384. [79]

Taylor, P. D. 1988. An inclusive fitness model for dispersal of offspring. *Journal of Theoretical Biology* 130:363–378. [100, 141]

Taylor, P. D. 1989. Evolutionary stability in one-parameter models under weak selection. *Theoretical Population Biology* 36:125–143. [74]

Taylor, P. D. 1990. Allele-frequency change in a class-structured population. *American Naturalist* 135:95–106. [192]

Taylor, P. D. 1992a. Altruism in viscous populations – an inclusive fitness model. *Evolutionary Ecology* 6:352–356. [112, 114, 122]

Taylor, P. D. 1992b. Inclusive fitness in a homogeneous environment. *Proceedings of the Royal Society (London) B* 249:299–302. [123, 125]

Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. *Journal of Theoretical Biology* 180:27–37. [91, 107, 192]

Taylor, P. D., and A. J. Irwin. 2000. Overlapping generations can promote altruistic behaviour. *Evolution* 54:1135–1141. [114, 116]

Theodorou, K., and D. Couvet. 2002. Inbreeding depression and heterosis in a subdivided population: influence of the mating system. *Genetical Research* 80:107–116. [142]

Thomas, C. D., and I. Hanski. 1997. Butterfly metapopulations. In I. Hanski and M. E. Gilpin, eds., *Metapopulation biology: ecology, genetics, and evolution*, 359–386. London: Academic Press. [208]

Thornhill, N. W., ed. 1993. *The natural history of inbreeding and outbreeding*. Chicago: University of Chicago Press. [127]

Trivers, R. 1985. *Social evolution*. Menlo Park, Calif.: Benjamin/Cummings. [105, 114, 137]

Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [116]

Tsitrone, A., F. Rousset, and P. David. 2001. Heterosis, marker mutational processes, and population inbreeding history. *Genetics* 159:1845–1859. [64]

Tuljapurkar, S. 1989. An uncertain life: demography in random environments. *Theoretical Population Biology* 35:227–294. [195]

Tuljapurkar, S. 1990. *Population dynamics in variable environments*. New York: Springer-Verlag. [195]

Turchin, P. 1998. *Quantitative analysis of movement*. Sunderland, Mass.: Sinauer. [31]

Turner, P. E., and L. Chao. 1999. Prisoner's dilemma in an RNA virus. *Nature* 398:441–443. [88]

Uyenoyama, M. K., and M. Feldman. 1982. Population genetic theory of kin selection. II. The multiplicative model. *American Naturalist* 120:614–627. [120]

Uyenoyama, M. K., K. E. Holsinger, and D. M. Waller. 1993. Ecological and genetic factors directing the evolution of self-fertilization. *Oxford Surveys in Evolutionary Biology* 9:327–381. [142]

van de Casteele, T., P. Galbusera, and E. Matthysen. 2001. Comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology* 10:1539–1549. [201]

van Dooren, T. J. M. 1999. The evolutionary ecology of dominance-recessivity. *Journal of Theoretical Biology* 198:519–532. [81]

van Dooren, T. J. M. 2000. The evolutionary dynamics of direct phenotypic overdominance: emergence possible, loss probable. *Evolution* 54:1899–1914. [81]

van Tienderen, P. H., and G. de Jong. 1986. Sex ratio under the haystack model: polymorphism may occur. *Journal of Theoretical Biology* 122:69–81. [74]

van Valen, L. 1971. Group selection and the evolution of dispersal. *Evolution* 25:591–598. [189]

Vitalis, R., and D. Couvet. 2001. Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* 157:911–925. [204]

Vucetich, J. A., T. A. Waite, and L. Nunney. 1997. Fluctuating population size and the ratio of effective to census population size. *Evolution* 51:2017–2021. [188]

Wade, M., and D. McCauley. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* 42:995–1005. [184, 186]

Wade, M. J. 1985. Soft selection, hard selection, kin selection, and group selection. *American Naturalist* 125:61–73. [121, 123]

Wahlund, S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus Betrachtet. *Hereditas* 11:65–106. [133]

Waite, T. A., and P. G. Parker. 1996. Dimensionless life histories and effective population size. *Conservation Biology* 10:1456–1462. [169]

Wakeley, J. 1999. Nonequilibrium migration in human evolution. *Genetics* 153:1863–1871. [202]

Wakeley, J. 2003. Polymorphism and divergence for island-mainland species. *Genetics* 163:411–420. [143, 148, 150, 196]

Wakeley, J., and N. Aliacar. 2001. Gene genealogies in a metapopulation. *Genetics* 159:893–905. Correction in *Genetics* 160:1263. [202]

Waller, D. M., and S. E. Knight. 1989. Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. II. Outcrossing rates and genotypic correlations. *Evolution* 43:860–869. [131, 132]

Wang, J. 1997a. Effective size and F-statistics of subdivided populations. I. Monoecious species with partial selfing. *Genetics* 146:1453–1463. [128]

Wang, J. 1997b. Effective size and F-statistics of subdivided populations. II. Dioecious species. *Genetics* 146:1465–1474. [63, 135]

Wang, J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetical Research* 78:243–257. [204]

Wang, J. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics* 160:1203–1215. [201]

Wang, J., and A. Caballero. 1999. Developments in predicting the effective size of subdivided populations. *Heredity* 82:212–226. [143, 163]

Wang, J., and M. C. Whitlock. 2003. Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163:429–446. [204]

Ward, R. D., M. Woodward, and D. O. F. Skibinski. 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology* 44:213–232. [200]

Watterson, G. A. 1996. Motoo Kimura's use of diffusion theory in population genetics. *Theoretical Population Biology* 49:154–188. [148]

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370. [201]

Weiss, G. H., and M. Kimura. 1965. A mathematical analysis of the stepping stone model of population genetics. *Journal of Applied Probability* 2:129–149. [35]

Whitlock, M. C. 1992a. Nonequilibrium population structure in forked fungus beetles: extinction, colonization, and the genetic variance among populations. *American Naturalist* 139:952–970. [44, 187]

Whitlock, M. C. 1992b. Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* 46:608–615. [24]

Whitlock, M. C. 2002. Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160:1191–1202. [142]

Whitlock, M. C. 2003. Fixation probability and time in subdivided populations. *Genetics* 164:767–779. [11, 143]

Whitlock, M. C., and N. H. Barton. 1997. The effective size of a subdivided population. *Genetics* 146:427–441. [63, 144, 145, 147, 152, 161, 175, 179, 182, 184, 186]

Whitlock, M. C., P. K. Ingvarsson, and T. Hatfield. 2000. Local drift load and the heterosis of interconnected populations. *Heredity* 84:452–457. [142]

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{\rm ST} \neq 1/(4Nm + 1)$. *Heredity* 82:117–125. [2, 62]

Whitlock, M. C., P. C. Phillips, and M. J. Wade. 1993. Gene interaction affects the additive genetic variance in subdivided populations with migration and extinction. *Evolution* 47:1758–1769. [185]

Williams, G. C. 1966. *Adaptation and Natural Selection*. Princeton: Princeton University Press. [3]

Wilson, D. S., G. B. Pollock, and L. A. Dugatkin. 1992. Can altruism evolve in purely viscous populations? *Evolutionary Ecology* 6:331–341. [122, 124]

Wood, J. W., P. E. Smouse, and J. C. Long. 1985. Sex-specific dispersal patterns in two human populations of highland New Guinea. *American Naturalist* 125:747–768. [30]

Wray, N. R., and R. Thompson. 1990. Prediction of rates of inbreeding in selected populations. *Genetical Research* 55:41–54. [166]

Wright, S. 1922. Coefficients of inbreeding and relationship. *American Naturalist* 56:330–338. Reprinted in Wright (1986), pp. 13–21. [12]

Wright, S. 1931a. Evolution in Mendelian populations. *Genetics* 16:97–159. Reprinted in Wright (1986), pp. 98–160. [10, 11, 13, 16, 143]

Wright, S. 1931b. Statistical theory of evolution. *Journal of the American Statistical Society* 26 (Suppl.):201–208. Reprinted in Wright (1986), pp. 89–96. [1]

Wright, S. 1937. The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the U.S.A.* 23:307–320. Reprinted in Wright (1986), pp. 236–249. [13]

Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* 87:430–431. Reprinted in Wright (1986), pp. 251–252. [143, 165, 176]

Wright, S. 1939. *Statistical genetics in relation to evolution*. Vol. 802 of *Actualités scientifiques et industrielles*. Paris: Hermann et Cie. Reprinted in Wright (1986), pp. 283–341. [143, 159]

Wright, S. 1942. Statistical genetics and evolution. *Bulletin of the American Mathematical Society* 48:223–246. Reprinted in Wright (1986), pp. 465–489. [10]

Wright, S. 1943. Isolation by distance. *Genetics* 28:114–138. Reprinted in Wright (1986), pp. 401–425. [12, 13, 14, 28, 43, 146, 152, 162]

Wright, S. 1946. Isolation by distance under diverse systems of mating. *Genetics* 31:39–59. Reprinted in Wright (1986), pp. 444–464. [28, 41, 43, 128]

Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323–354. Reprinted in Wright (1986), pp. 580–611. [12, 17]

Wright, S. 1969. *Evolution and the genetics of populations. II. The theory of gene frequencies*. Chicago: University of Chicago Press. [12, 17, 81, 134, 159, 160, 162, 176]

Wright, S. 1986. *Evolution: selected papers*. Chicago: University of Chicago Press. [259, 260]

Wynne-Edwards, V. C. 1962. *Animal dispersion in relation to social behaviour*. Edinburgh: Oliver and Boyd. [122]

Index