1

Introduction

What is Ecological Genetics?

Ecological genetics is at the interface of ecology, evolution, and genetics, and thus includes important elements from each of these fields. We can use two closely related definitions to help describe the scope of ecological genetics:

1. Ecological genetics is concerned with the *genetics of ecologically important traits, that is, those traits related to fitness* such as survival and reproduction. Ecology is the study of the distribution and abundance of organisms—in other words, how many individuals there are, where they live, and why. Distribution and abundance are determined by birth rates and death rates, which in turn are determined by interactions with the organism's biotic and abiotic environment. These interactions include predation, competition, and the ability to find mates, food, and shelter. Consider traits that would help an organism deal with each of these interactions. Cryptic coloration could help a beetle avoid being eaten, growing tall could help a plant compete with other plants for light, and a thick coat of fur might help a mouse survive winter cold. These are examples of *ecologically important traits*: those traits that are closely tied to fitness or, in other words, are important in determining an organism's adaptation to its natural environment, both biotic and abiotic.
2. Ecological genetics can also be defined as the *study of the process of phenotypic evolution occurring in present-day natural populations*. Phenotypic evolution can be defined as a change in the mean or variance of a trait

across generations due to changes in allele frequencies. The four processes that can cause evolution are **mutation, genetic drift, migration,** and **natural selection.** All of these processes are described in Chapter 3, and the last three in particular are closely related to ecology and therefore appear throughout the book. Ecological factors can cause population size to decline, and the resulting small population size causes genetic drift. Migration is clearly ecological, but how is natural selection related to ecology? Selection is caused by differences in fitness among organisms in a population, and these fitness differences are caused in part by interactions with the environment as previously mentioned.

Our two definitions are tied together by the concept of *adaptation*, which is the central theme of ecological genetics. An **adaptation** is a phenotypic trait that has evolved to help an organism deal with something in its environment. Like most ecologically important traits, the examples given above are adaptations. Natural selection is special among the four evolutionary processes because it is the only one that leads to adaptation. Mutation, genetic drift, and migration can either speed up or constrain the development of adaptations, but they cannot cause adaptation.

An overview of these ideas is shown in Figure 1.1, which summarizes much of what will be covered in this book. Beginning at the top, ecological factors, both biotic and abiotic, can cause fitness differences among organisms with different phenotypes within the population; this is natural selection. If mutation and recombination create genetic variation for these phenotypic traits, then the selection can act on this variation to change the

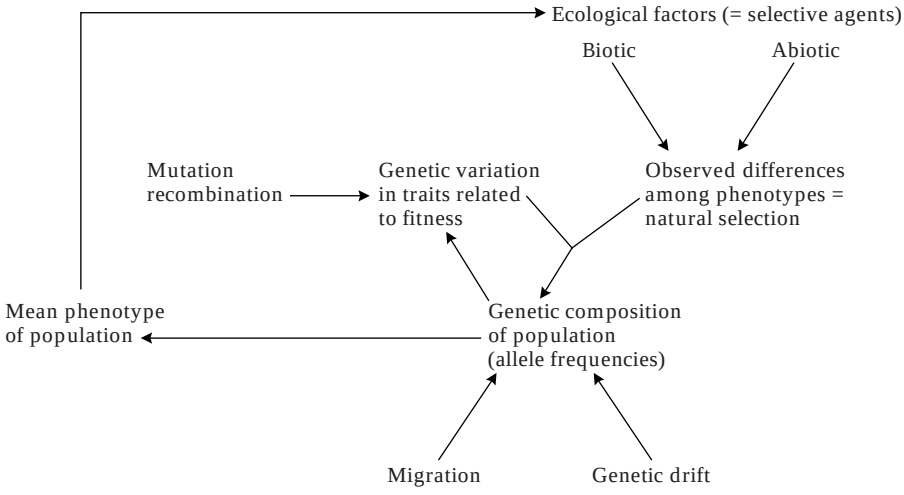

Figure 1.1 A schematic overview of key concepts in ecological genetics.

genetic composition of the population. The genetics of the population can also be affected by gene flow from other populations with different genetic composition, or by genetic drift if the population size is small. All these changes in genetic composition are likely to feed back and affect the genetic variation for the phenotypic traits, as well as change the average phenotype in the population across generations. These phenotypic changes can lead to an improvement in the ability of the population to survive and reproduce in its biotic and abiotic environment; that is, it can lead to adaptation.

As an example, a deer mouse is part of a beetle's biotic environment, and may cause beetles with increased defensive secretions to have higher fitness than those with less secretions, if the secretions deter deer mouse predation. If this phenotypic variation is caused at least in part by underlying genetic variation, then this will cause an increase in the frequency of alleles that increase defensive secretions. (An **allele** is a particular type of a given gene.) This increase in allele frequency across generations may be slowed by random genetic drift, or by gene flow from other beetle populations with low frequencies of the high-secretion alleles (perhaps because there are fewer mice coexisting with those other populations). Selection and drift may decrease genetic variation for secretion quantity, also slowing future evolution of secretions. If the average secretion quantity in the population increases in spite of these constraints, then this may reduce the impact of mouse predation in subsequent generations, increasing adaptation of the beetles to this environmental factor.

Overview of the Book

Chapters 2 and 3 cover the field of population genetics. **Population genetics** is the study of genetic variation within and among populations, focusing on the processes that affect genotypic and allele frequencies at one or a few gene loci. These processes include inbreeding, mutation, migration, drift, and selection; the genotypic and allele frequencies are revealed mainly through molecular markers. Population genetics for the most part does not focus on phenotypes, since the genes and alleles underlying most phenotypic traits are unknown, especially in natural populations. This is because most phenotypic traits are complex, being affected by several to many gene loci and by the environment.

Chapters 4 and 5 cover the field of **quantitative genetics,** which does focus on the phenotype, usually without knowing the genotypes underlying the traits. In the place of genotypic information, statistical abstractions such as variance, correlation, and heritability are used in quantitative genetics to help understand the genetics of complex phenotypes. QTL mapping (covered at the end of Chapter 5) is a marriage of molecular and statistical techniques for studying the genetics of complex phenotypic traits. QTL mapping

is a first step in discovering the genes underlying phenotypic traits in natural populations, bringing together the fields of population and quantitative genetics. This convergence is very likely to lead to fundamental new insights in ecological genetics.

Chapter 6 is on techniques developed from quantitative genetics for studying natural selection on phenotypic traits (rather than on genotypes as in population genetics). These techniques have allowed biologists to measure the strength and direction of selection in natural populations, as well as help determine the ecological causes of the selection. Chapter 6 also synthesizes the quantitative genetic material in Chapters 4 through 6, and shows how short-term evolution can be predicted in natural populations using knowledge of genetic variance and the strength of selection.

Since ecological genetics is at the interface between ecology, evolution and genetics, it is a critical component of all three fields, as well as essential for the study of some of society's problems. In Chapter 7 we will discuss the importance of ecological genetic principles in conservation, the spread of invasive species, the evolution of pesticide, herbicide, and antibiotic resistance, and the environmental effects of genetically modified organisms used in agriculture.

The focus of the book will be on diploid sexual organisms. Most of the concepts covered also apply to asexual and haploid organisms, but there are important differences. Most of our examples will come from studies of plants and animals, because the ecological genetics of most microorganisms and fungi are not as well known.

Basic Genetic Terms

A **gene** is a stretch of **DNA** (deoxyribonucleic acid) coding for a polypeptide chain; one or more polypeptides make up a protein. The genetic information in DNA is coded in the sequence of four nucleotides, abbreviated according to the identity of the nitrogenous **base** that each contains: A (adenine), G (guanine), T (thymine), or C (cytosine). DNA molecules normally consist of two complementary helical strands held together by pairing between the bases: A in one strand is paired with T in the other strand, and G in one strand is paired with C in the other.

The process of creating proteins from the genetic code in DNA is called **gene expression.** The essentials of gene expression in the cells of eukaryotes are outlined in Figure 1.2. The first step is **transcription,** in which the sequence of nucleotides present in one DNA strand of a gene is faithfully copied into the nucleotides of an **RNA** (ribonucleic acid) molecule. As the RNA transcript is synthesized, each base in the DNA undergoes pairing with a base in an RNA nucleotide, which is then added to the growing RNA strand. The base-pairing rules are the same as those in DNA, except that in RNA nucleotides the base U (uracil) is found instead of T (thymine). The second step of gene expression is **RNA processing,** in which intervening sequences or **introns** are removed

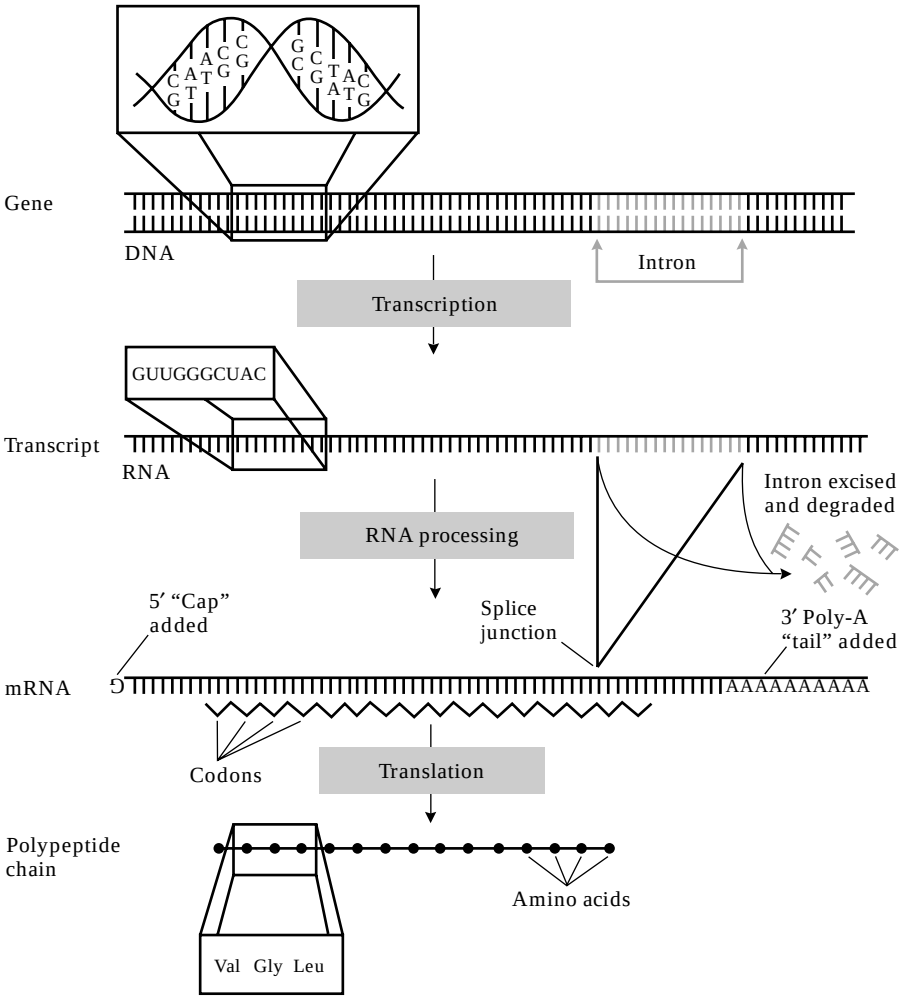

Figure 1.2 Principal processes in gene expression in eukaryotes.

from the RNA transcript by splicing and the ends of the transcript are modified. The regions between the introns that remain in the fully processed RNA are known as **exons;** these are the sequences that actually code for proteins. The fully processed RNA constitutes the **messenger RNA** (mRNA).

The messenger RNA undergoes **translation** on ribosomes in the cytoplasm to produce the polypeptide that is encoded in the sequence of nucleotides. In the translated part of the messenger RNA, each adjacent group of three nucleotides constitutes a coding group or **codon,** which specifies a corresponding amino acid subunit in the polypeptide chain. The stan-

dard **genetic code** showing which codons specify which amino acids is given in Table 1.1. After each three-letter codon are the three- and one-letter designations for the 20 amino acids. The three-letter and one-letter abbreviations are both established conventions. Note that in many cases changes in the third base in the codon do not change the amino acid that is specified; therefore, much variation at this position is not expressed (sometimes called the "silent" position). The codon AUG specifies methionine and also serves as the start codon for polypeptide synthesis. Any of three codons—UAA, UAG, or UGA—specify the end, or termination, of polypeptide synthesis, upon which the completed polypeptide chain is released from the ribosome. The start and stop codons are shaded in Table 1.1.

All the DNA in a cell is collectively called the **genome.** Genome size is typically expressed as the amount of DNA in a reproductive cell (sperm or egg), and it differs greatly among species. For example, the genome of *Arabidopsis thaliana*, a model plant for genetic studies, consists of about 120 million base pairs, whereas the genome of the lily *Fritillaria* is 1000 times as large, about 120 billion base pairs. The human genome is about 3 billion base

TABLE 1.1 *The standard genetic code*

First nucleotide in codon (5' end)	Second nucleotide in codon				Third nucleotide in codon (3' end)
	U	C	A	G	
U	UUU Phe/F	UCU Ser/S	UAU Tyr/Y	UGU Cys/C	U
	UUC Phe/F	UCC Ser/S	UAC Tyr/Y	UGC Cys/C	C
	UUA Leu/L	UCA Ser/S	UAA Stop	UGA Stop	A
	UUG Leu/L	UCG Ser/S	UAG Stop	UGG Trp/W	G
C	CUU Leu/L	CCU Pro/P	CAU His/H	CGU Arg/R	U
	CUC Leu/L	CCC Pro/P	CAC His/H	CGC Arg/R	C
	CUA Leu/L	CCA Pro/P	CAA Gln/Q	CGA Arg/R	A
	CUG Leu/L	CCG Pro/P	CAG Gln/Q	CGG Arg/R	G
A	AUU Ile/I	ACU Thr/T	AAU Asn/N	AGU Ser/S	U
	AUC Ile/I	ACC Thr/T	AAC Asn/N	AGC Ser/S	C
	AUA Ile/I	ACA Thr/T	AAA Lys/K	AGA Arg/R	A
	AUG Met/M	ACG Thr/T	AAG Lys/K	AGG Arg/R	G
G	GUU Val/V	GCU Ala/A	GAU Asp/D	GGU Gly/G	U
	GUC Val/V	GCC Ala/A	GAC Asp/D	GGC Gly/G	C
	GUA Val/V	GCA Ala/A	GAA Glu/E	GGA Gly/G	A
	GUG Val/V	GCG Ala/A	GAG Glu/E	GGG Gly/G	G

pairs. Genes are arranged in linear order along microscopic threadlike bodies called **chromosomes.** Each human **gamete** (sperm or egg) contains one complete set of 23 chromosomes; this is the **haploid** chromosome number, designated as *n*. Chromosome number can vary greatly: $n = 2$ in some scorpions and 127 in a species of hermit crab! A typical chromosome contains several thousand genes, in humans averaging approximately 1500 genes per chromosome. The position of a gene along a chromosome is called the **locus** of the gene. Sometimes the words gene and locus are used interchangeably, which can lead to confusion. **Recombination** between loci can occur during meiosis, which creates new combinations of alleles at these different loci. Recombination is rarer between loci that are close together on the chromosome; these loci are said to be genetically **linked.**

In most multicellular organisms, each individual cell contains two copies of each type of chromosome, one inherited from its mother through the egg and one inherited from its father through the sperm (so the **diploid** chromosome number, $2n$, is 46 in humans and 254 in hermit crabs). Note that these two copies of the chromosome are not the two complementary strands of DNA; each chromosome consists of a double-stranded DNA molecule. At any locus, therefore, every diploid individual contains two copies of the gene—one at each corresponding (homologous) position in the maternal and paternal chromosome. These two copies are the alleles of the gene in that individual. If the two alleles at a locus are indistinguishable according to any particular experimental criterion, then the individual is **homozygous** at the locus under consideration. If the two alleles at a locus are distinguishable by means of this criterion, then the individual is **heterozygous** at the locus.

The **genotype** of an individual is the diploid pair of alleles present at a given locus. Therefore, homozygous and heterozygous are the two major categories of genotypes. Typographically, genes are indicated in italics, and alleles are typically distinguished by uppercase or lowercase letters (*A* versus *a*), subscripts (A_1 versus A_2), superscripts (a^+ versus a^-), or sometimes just + and –. Using these symbols, the genotype of homozygous individuals would be portrayed by any of these formulas: *AA*, *aa*, A_1A_1, A_2A_2, a^+a^+, a^-a^-, +/+, or –/–. As in the last two examples, the slash is sometimes used to separate alleles present in homologous chromosomes to avoid ambiguity. The genotype of heterozygous individuals would be portrayed by any of the formulas *Aa*, A_1A_2, a^+a^-, or +/–.

The outward appearance of an organism for a given characteristic is its **phenotype.** Phenotypic traits can be defined at a number of hierarchical levels, each one dependent on a number of traits at lower levels. For example, the form of an enzyme encoded by a gene is a phenotype, as is a physiological function like metabolic rate that depends on a number of enzymes. A number of different physiological functions affect morphological traits like

height, and physiology and morphology together can affect behavioral phenotypes such as courtship. Finally, all these lower level traits can affect life history traits like survival and reproduction, which determine the ultimate trait of individual fitness. The traits that are higher in this hierarchy are more complex and affected by more gene loci. The expression of most phenotypic traits, and especially the higher level ones, are also affected to varying degrees by the environment. This complexity means that the same genotype can produce different phenotypes, through the action of the environment. Conversely, the different genotypes can produce the same phenotypes, again due to the environment and also due to gene interactions. We will discuss complex phenotypic traits and fitness in more detail in Chapters 4 through 6.