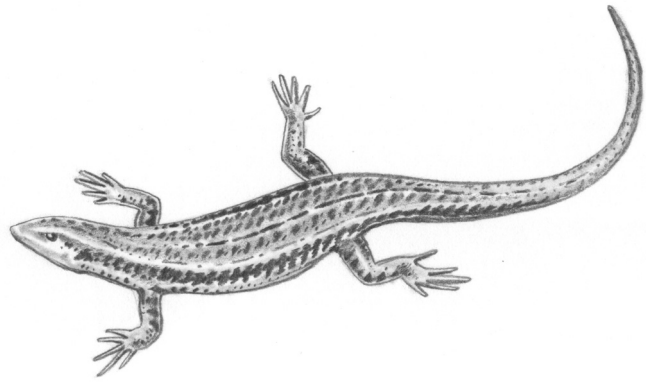
16

Units of Conservation



Grand skink, Section 16.4.2

The zoo directors, curators, geneticists and population biologists who attempt to pursue the elusive goal of preservation of adaptive genetic variation are now considering the question of which gene pools they should strive to preserve.

Oliver A. Ryder (1986)

The choices of what to conserve must often be made with regard to populations that are not separate completely from others, or when information regarding the relationships and degrees of distinction among populations is very incomplete.

Jody Hey et al. (2003)

The identification of appropriate taxonomic and population units for protection and management is essential for the conservation of biological diversity. For species identification and classification, genetic principles and methods are relatively well developed; nonetheless species identification can be controversial. Within species, the identification and protection of genetically distinct local populations should be a major focus in conservation because the conservation of many distinct populations helps maximize evolutionary potential and minimize extinction risks (Hughes et al. 1997; Hilborn et al. 2003; Luck et al. 2003). Furthermore, the local population is often considered the functional unit in ecosystems.

Identification of population units is necessary so that management and monitoring programs can be efficiently targeted toward distinct or independent populations. Biologists and managers must be able to identify populations and geographic boundaries between populations in order to effectively plan harvesting quotas (e.g., to avoid overharvesting) or to devise translocations and reintroductions of individuals (e.g., to avoid mixing of adaptively differentiated populations). In addition, it is sometimes necessary to prioritize which population units (or taxa) to conserve because limited financial resources preclude conservation of all units.

Finally, many governments and agencies have established legislation and policies to protect intraspecific population units. This requires the identification of population units. For example, the ESA (Endangered Species Act of the USA) allows listing and full protection of **distinct population segments** (DPS) of vertebrate species (Example 16.1). Species and subspecies identification is based upon traditional, established taxonomic criteria as well as genetic criteria (although criteria for species identification are sometimes controversial). The choice of criteria to use to delineate intraspecific units for conservation has been highly controversial. Other countries, for example in Europe and Australia, also have laws that depend on the identification of distinct taxa and populations for the protection of species and habits (Example 16.1).

Example 16.1 The US Endangered Species Act (ESA) and conservation units

The ESA of the United States is one of the most powerful pieces of conservation legislation in the world. The ESA has been a major stimulus motivating biologists to develop criteria for identifying population units for conservation. This is because the ESA provides legal protection for subspecies and "distinct population segments" (DPSs) of vertebrates, as if they were full species. According to the ESA:

> The term "species" includes any subspecies of fish or wildlife and plants, and any distinct population segment of any species of vertebrate fish or wildlife which interbreeds when mature.

However, the ESA does not provide criteria or guidelines for delineating DPSs. The identification of intraspecific units for conservation is controversial. This is not surprising given that the definition of a "good species" is controversial (see Section 16.5). Biologists have vigorously debated the criteria for identifying DPSs and other conservation units ever since the US Congress extended full protection of the ESA to "distinct" populations, but did not provide guidelines.

Legislations in other countries around the world have provisions that recognize and protect intraspecific units of conservation. For example, Canada passed the Species at Risk Act (SARA) in 2003. The SARA aims to "prevent wildlife species from becoming extinct, and to secure the necessary actions for their recovery". Under the SARA, "wildlife species" means a species, subspecies, variety, or geographically or genetically distinct population of animal, plant, or other organism, other than a bacterium or virus, which is wild by nature.

In Australia, the Endangered Species Protection Act (ESPA) also allows protection for subspecies and distinct populations. But, like the ESA in the United States, there are problems with defining and identifying intraspecific units (Woinarski and Fisher 1999).

In this chapter, we examine the components of biodiversity and then consider methods to assess taxonomic and population relationships. We discuss the criteria, difficulties, and controversies in the identification of conservation units. We also consider the identification of appropriate population units for legal protection and for management actions (e.g., supplemental translocation of individuals between geographic regions). Recall that in the previous chapter, we considered three spatial scales of genetic population structure for conservation: local, metapopulation, and species.

16.1 What should we try to protect?

Genes, species, and ecosystems are three primary levels of biodiversity (Figure 16.1) recognized by the IUCN. There has been some controversy as to which level should receive priority for conservation efforts (e.g., Bowen 1999). However, it is clear that all three levels must be conserved for successful biodiversity conservation. For example, it is

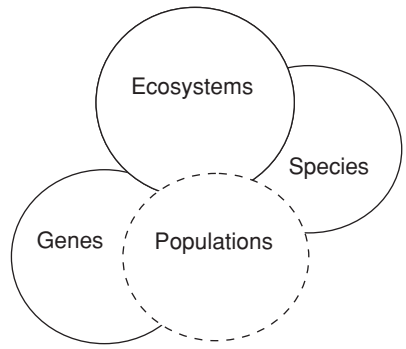


Figure 16.1 Primary levels of biodiversity recognized by the IUCN (solid circles), and a fourth level – populations – recognized as perhaps most crucial for species' long-term persistence (Hughes et al. 1997; Luck et al. 2003). In reality, biodiversity exists across a continuum of many hierarchical levels of organization including genes, genomes (i.e., multilocus genotypes), local populations, communities, ecosystems, and biomes. Additional levels of diversity include metapopulations, subspecies, genera, families, and so on.

as futile to conserve ecosystems without species, as it is to save species without large, healthy ecosystems.

An example of this kind of futility is that of the African rhinoceros, which are being protected mainly in zoos and small nature reserves, but for which little habitat (free from poachers) is currently available. Without conserving vast habitats for future rhino populations, it seems pointless to protect rhinos in small nature reserves surrounded by armed guards and fences.

It is not too late for rhinos. Vast habitats do exist, and rhinos could be successful in these habitats if poaching is eliminated. In addition to conserving rhino species and their habitats, it is also important to conserve genetic variation within rhino species because variation is a prerequisite for long-term adaptive change and the avoidance of fitness decline through inbreeding depression (see Chapter 14). Clearly, it is important to recognize and conserve all levels of biodiversity: ecosystems, species, and genes.

The debate over whether to protect genes, species, or ecosystems is, in a way, a false trichotomy because each level is an important component of biodiversity as a whole. Nonetheless, considering each level separately can help us appreciate the interacting components of biodiversity, and the different ways that genetics can facilitate conservation at different levels. Appreciation of each level also can promote understanding and multidisciplinary collaborations across research domains. Finally, a fourth level of biodiversity – that of genetically distinct local populations – is arguably the most important level for focusing conservation efforts (Figure 16.1). The conservation of multiple, genetically distinct populations is necessary to insure long-term species survival and the functioning of ecosystems, as mentioned above (Luck et al. 2003).

We can also debate which temporal component of biodiversity to prioritize for conservation: past, present, or future biodiversity. All three components are important, although future biodiversity often warrants special concern (Example 16.2).

Example 16.2 Temporal considerations in conservation: past, present, and future

What temporal components of biodiversity do we wish to preserve? Do we want to conserve ancient isolated lineages, current patterns of diversity (ecological and genetic), or the diversity required for future adaptation and for novel diversity to evolve? Most would agree "all of the above". All three temporal components are interrelated and complementary (Figure 16.2). For example, conserving current diversity helps insure future adaptive potential. Similarly, conserving and studying ancient lineages ("living fossils") can help us understand factors important for long-term persistence. Nonetheless, one can argue that the most important temporal component to consider is future biodiversity, i.e., the ability of species and populations to adapt to future environments (e.g., global climate change). If populations do not adapt to future environments then biodiversity will decline – leading to loss of ecosystem functioning and services. Figure 16.2 illustrates how different temporal components of biodiversity (past, present, and future) can be related to different scientific disciplines (systematics, ecology, and evolutionary biology, respectively). These components also are often related to different hierarchical levels of biodiversity: species, ecosystems, and genes, respectively.

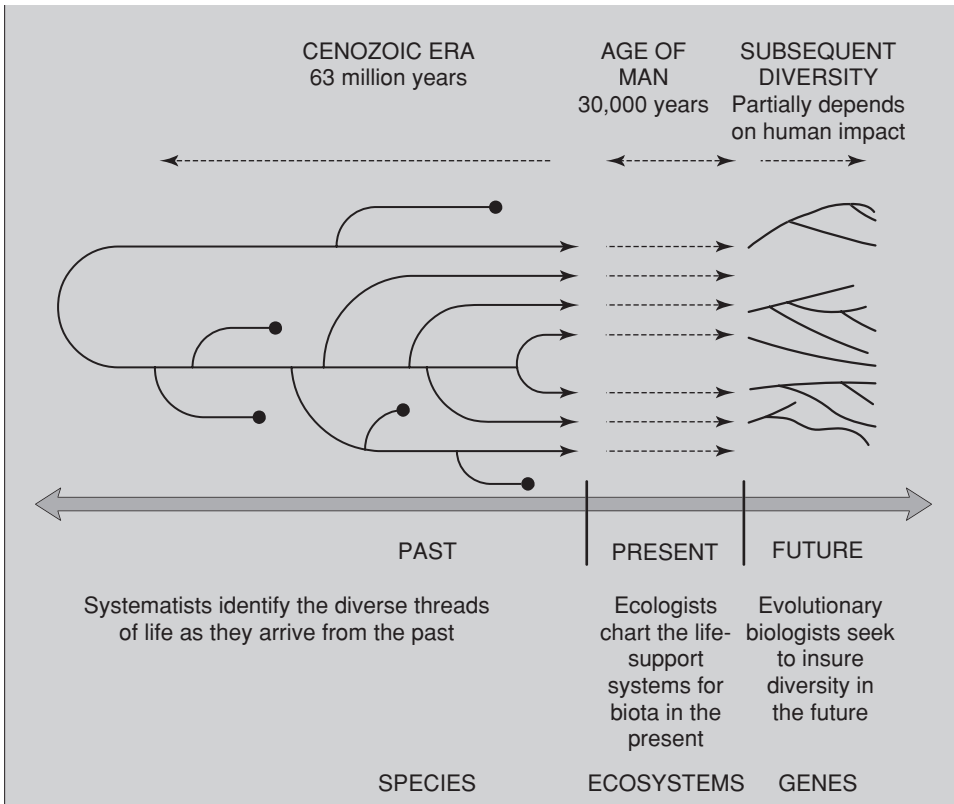


Figure 16.2 The temporal framework (past, present, and future), corresponding disciplines (systematics, ecology, and evolutionary biology), and levels of biodiversity (species, ecosystems, and genes) that are often considered when prioritizing biodiversity for conservation. Modified from Bowen (1999).

Another choice that is often debated is whether we should emphasize protecting the existing **patterns** of diversity or the **processes** that generate diversity (e.g., ecological and evolutionary processes themselves)? Again the answer is, in general, both. It is clear that we should prioritize the preservation of the process of adaptation so that populations and species can continually adapt to future environmental changes. However, one important step toward preserving natural processes is to quantify, monitor, and maintain natural patterns of population subdivision and connectivity (e.g., identify intraspecific population units and boundaries). This, for example, would prevent extreme fragmentation and promote continued natural patterns of gene flow among populations.

How do we conserve the "processes" of evolution, including adaptive evolutionary change? We must first maintain healthy habitats and large wild populations because only in large populations can natural selection proceed efficiently (see Section 8.5). In small populations, genetic drift leads to random genetic change, which is generally nonadaptive. Drift can preclude selection from maintaining beneficial alleles and eliminating deleterious ones. To maintain evolutionary process we also must preserve multiple populations –

ideally from different environments so that selection pressures remain diverse and multilocus genotype diversity remains high. In this scenario, a wide range of local adaptations are preserved within species, as well as some possibility of adaptation to different future environmental challenges.

16.2 Systematics and taxonomy

The description and naming of distinct taxa is essential for most disciplines in biology. In conservation biology, the identification of taxa (taxonomy) and assessing their evolutionary relationships (systematics) is crucial for the design of efficient strategies for biodiversity management and conservation. For example, failing to recognize the existence of a distinct and threatened taxon can lead to insufficient protection and subsequent extinction. Identification of too many taxa (oversplitting) can waste limited conservation resources. The misidentification of a sister taxon could lead to nonideal choice of source populations for supplementing endangered populations.

There are two fundamental aspects of evolution that we must consider: phenotypic change through time (**anagenesis**) and the branching pattern of reproductive relationships among taxa (**cladogenesis**). The two primary taxonomic approaches are based on these two aspects.

Historically, taxonomic classification was based primarily upon phenotypic similarity (**phenetics**), which reflects evolution via anagenesis. That is, groups of organisms that were phenotypically similar were grouped together. This classification is conducted using clustering algorithms (described below) that group organisms based exclusively on "overall similarity" or outward appearance. For example, populations that share similar allele frequencies are grouped together into one species. In this example, the clustering by overall similarity of allele frequencies is phenetic. The resulting diagram (or tree) used to illustrate classification is called a **phenogram**, even if based upon genetic data (e.g., allele frequencies).

A second approach is to classify organisms on the basis of their phylogenetic relationships (**cladistics**). Cladistic methods group together organisms that share **derived** traits (originating in a common ancestor), reflecting cladogenesis. Under cladistic classification, only **monophyletic** groups can be recognized, and only genealogical information is considered. The resulting diagram (or tree) used to illustrate relationships is called a **cladogram** (or sometimes, a **phylogeny**). Phylogenetics is discussed below (see Section 16.3).

Our current system of taxonomy combines cladistics and phenetics, and it is sometimes referred to as evolutionary classification (Mayr 1981). Under evolutionary classification, taxonomic groups are usually classified on the basis of phylogeny. However, groups that are extremely phenotypically divergent are sometimes recognized as separate taxa even though they are phylogenetically related. A good example of this is birds (Figure 16.3). Birds were derived from a dinosaur ancestor, as evidenced from the fossil record showing reptiles with feathers (bird–reptile intermediates). Therefore, birds and dinosaurs are sister groups that should be classified together under a strictly cladistic classification scheme. However, birds underwent rapid evolutionary divergence associated with their development of flight. Therefore, birds are classified as a separate class while dinosaurs are classified as a reptile (class Reptilia).

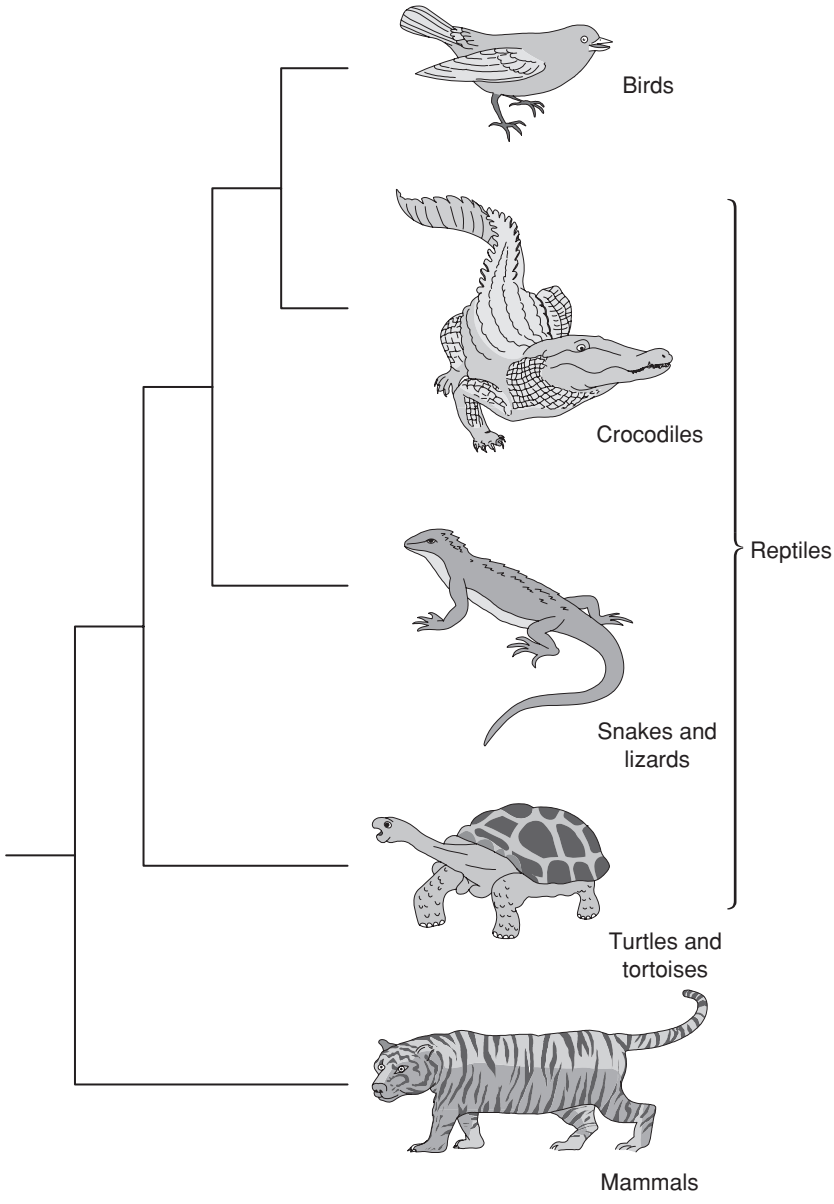


Figure 16.3 Phylogenetic relationships of birds, mammals, and reptiles. Note that crocodiles and birds are more closely related to each other than either is to other reptiles. That is, crocodiles share a more recent common ancestor with birds than they do with snakes, lizards, turtles, and tortoises. Therefore, the classification of the class Reptilia is not monophyletic.

There is a great deal of controversy associated with the "correct" method of classification. We should use all kinds of information available (morphology, physiology, behavior, life history, geography, and genetics) and the strengths of different schools (phenetic and cladistic) when classifying organisms (Mayr 1981; see also Section 16.6).

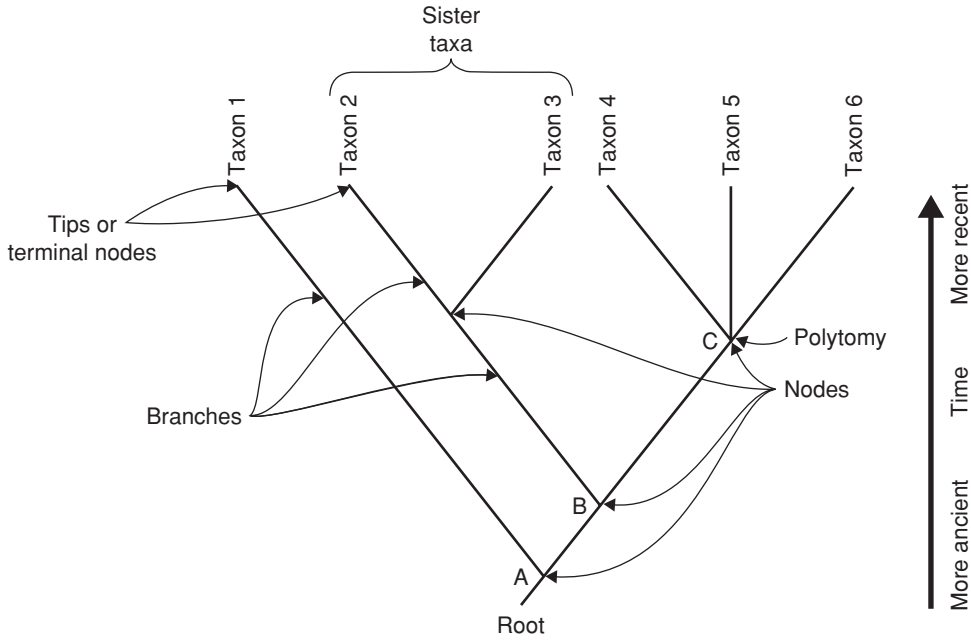


Figure 16.4 A phylogenetic tree (phylogeny). A **polytomy** (node 'C') is when more than two taxa are joined at the same node because data cannot resolve which (two) of the three taxa are most closely related. A widely controversial polytomy 10–20 years ago was that of chimpanzees, gorillas, and humans. However, extensive genetic data now show that chimps and humans are more closely related (i.e., sister taxa). From Freeman and Herron (1998).

16.3 Phylogeny reconstruction

A phylogenetic tree is a pictorial summary that illustrates the pattern and timing of branching events in the evolutionary history of taxa (Figure 16.4). A phylogenetic tree consists of **nodes** for the taxa being considered and branches that connect taxa and show their relationships. Nodes are at the tips of branches and at branching points (representing extinct ancestral taxa). A phylogenetic tree represents a hypothesis about relationships that is open to change as more taxa or characters are added. The same phylogeny can be drawn many different ways. Branches can be rotated at any node without changing the relationship between the taxa, as illustrated in Figure 16.5.

Branch lengths are often proportional to the amount of genetic divergence between taxa. If the amount of divergence is proportional to time, a phylogeny can show time since divergence between taxa. Molecular divergence (through mutation and drift) will be proportional to time if mutation accumulation is stochastically constant (like radioactive decay). The idea that molecular divergence can be constant is called the **molecular clock** concept. In conservation biology, the molecular clock and divergence estimates can help identify distinct populations and prioritize them based on their distinctiveness or divergence times. One serious problem with estimating divergence times is that extreme genetic drift (e.g., bottlenecks and founder events) can greatly inflate estimates of

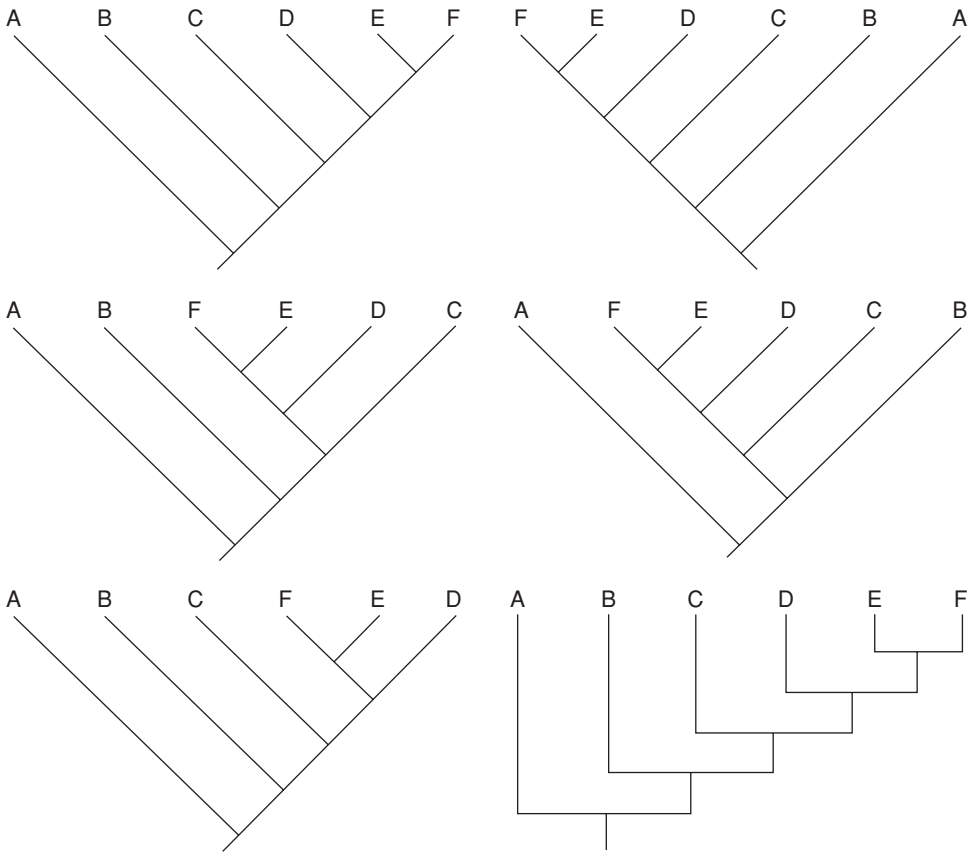


Figure 16.5 Six phylogenetic trees showing identical relationships among taxa. Note that branches can be rotated at the nodes without changing relationships represented on the trees (e.g., E vs. F in the top two trees). From Freeman and Herron (1998).

divergence times leading to long branch lengths and misleading estimates of phylogenetic distinctiveness (see Section 9.7).

16.3.1 Methods

There are two basic steps in phylogeny reconstruction: (1) generate a matrix of character states (e.g., derived versus ancestral states); and (2) build a tree from the matrix. Cladistic methods use only shared derived traits, **synapomorphies**, to infer evolutionary relationships. Phenogram construction is based on overall similarity. Therefore, a phylogenetic tree may have a different topology from a phenogram using the same character state matrix (Example 16.3).

The actual construction of phylogenies is much more complicated than this simple example. It is sometimes difficult to determine the ancestral state of characters. Moreover, the number of possible evolutionary trees to compare rises at an alarming rate. For example, there are nearly 35 million possible rooted, bifurcating trees with just 10 taxa and over 8×10^{21} possible trees with 20 taxa! In addition, there are a variety of other methods

Example 16.3 Phenogram and cladogram of birds, crocodiles, and lizards

As we have seen, birds and crocodiles are sister taxa based upon phylogenetic analysis, but crocodiles are taxonomically classified as reptiles because of their phenetic similarity with snakes, lizards, and turtles. These conclusions are based on a large number of traits. Here we will consider five traits (Table 16.1) to demonstrate how a different phenogram and cladogram can result from the matrix of character states.

Lizards and crocodiles are more phenotypically similar to each other than either is to birds because they share three out of five traits (0.6), while crocodile and birds share just two out of five traits (0.4) (Table 16.2). We can construct a phenogram based upon clustering together the most phenotypically similar groups (Figure 16.6a). The phenotypic similarity of lizards and crocodiles results from their sharing ancestral character states because of the rapid phenotypic changes that occurred in birds associated with adaptation to flight.

Parsimony methods were among the first to be used to infer phylogenies, and they are perhaps the easiest phylogenetic method to explain and understand (Felsenstein 2004, p. 1). There are many possible phylogenies for any group of taxa. Parsimony is the principle that the phylogeny to be preferred is the one that

Table 16.1 Character states for five traits used to construct a phenogram and cladogram of lizards, crocodiles, and birds. Traits: 1, heart (three or four chambered); 2, inner ear bones (present or absent); 3, feathers (present or absent); 4, wings (present or absent); and 5, hollow bones (present or absent).

	Traits*				
Taxon	**1**	**2**	**3**	**4**	**5**
A Lizards	0	0	0	0	0
B Crocodiles	1	1	0	0	0
C Birds	1	1	1	1	1

* 0, ancestral; 1, derived.

Table 16.2 Phenotypic similarity matrix for lizards, crocodiles, and birds based upon the proportion of shared characters states in Table 16.1.

	Lizards	**Crocodiles**	**Birds**
Lizards	1.0		
Crocodiles	0.6	1.0	
Birds	0.0	0.4	1.0

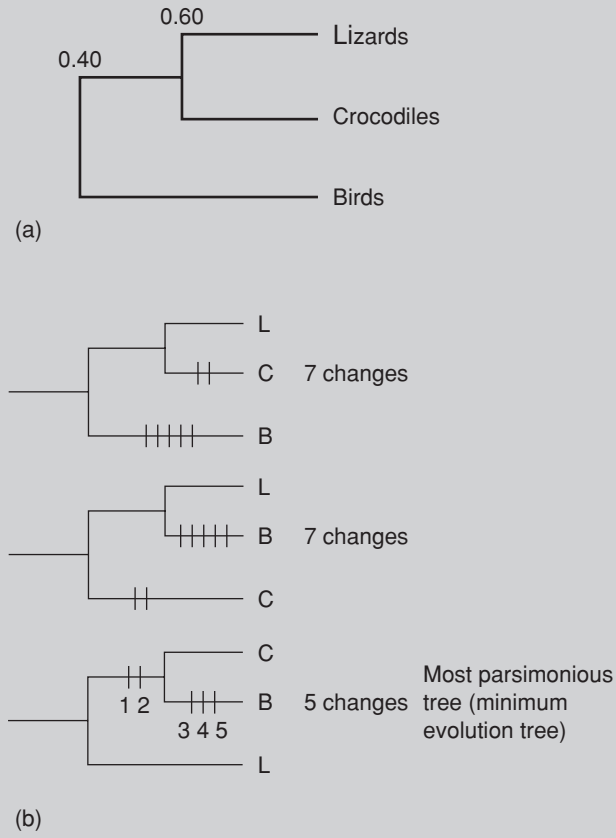


Figure 16.6 (a) Phenogram and (b) cladograms showing phenotypic and evolutionary relationships, respectively, among lizards, crocodile, and birds. Numbers in (a) are genetic distance estimates (e.g., 0.60 distance units between lizards and crocodiles). Vertical slashes in (b) on branches represent changes. Numbers below slashes on the bottom (most parsimonious) tree correspond to the traits (i.e., evolutionary change in traits) listed in Table 16.1.

requires the minimum amount of evolution. To use parsimony, we must search all possible phylogenies and identify the one or ones that minimize the number of evolutionary changes.

There are only three possible bifurcating phylogenies for lizards, crocodiles, and birds. Figure 16.6b shows these trees and the number of evolutionary changes from the ancestral to the derived trait to explain the character state matrix. The upper two phylogenies both require seven changes because certain evolutionary changes had to occur independently in the crocodile and bird branches. The bottom tree requires only five evolutionary changes to explain the character state matrix. Thus, the bottom tree is the most parsimonious. Birds and crocodiles form a monophyletic group because they share two synapomorphies (traits 1 and 2).

besides parsimony for inferring phylogenies (Hall 2004). The field of inferring phylogenies has been marked by more heated controversy than perhaps any other area of evolutionary biology (see Felsenstein 2004).

16.3.2 Gene trees and species trees

It is important to realize that different genes can result in different phylogenies of species, and that gene trees are often different from the true species phylogeny (Nichols 2001). Different gene phylogenies can arise due to four main phenomena: lineage sorting and associated genome sampling error, sampling error of individuals or populations, natural selection, or introgression (following hybridization).

Lineage sorting and sampling error

Ancestral **lineage sorting** occurs when different DNA sequences from a mother taxon are sorted into different daughter species such that lineage divergence times do not reflect population divergence times. For example, two divergent lineages can be sorted into two recently isolated populations, where less-divergent lineages might become fixed in different ancient daughter populations. Lineage sorting makes it important to study many different genes (or independent DNA sequences) to avoid sampling error associated with sampling too few (or an unrepresentative set of) genetic characters (loci).

Sampling error of individuals occurs when too few individuals or nonrepresentative sets of individuals are sampled from a species such that the inferred gene tree differs from the true species tree. For example, many early studies using mtDNA analysis included only a few individuals per geographic location, which could lead to erroneous phylogeny inference. Limited sampling is likely to detect only a subset of local lineages (i.e., alleles), especially for lineages at low frequency.

We can use simple probability to estimate the sample size that we need to detect rare genotypes. For example, how many individuals must we sample to have a greater than 95% chance of detecting an allele with frequency of 0.10 ($p = 0.1$)? Each time we examine one sample, we have a 0.90 chance ($1 - p$) of not detecting the allele in question and a 0.10 chance (p) of detecting it. Using the product rule (see Appendix Section A1), the probability of not detecting an allele at $p = 0.1$ in a sample of size x is $(1 - p)^x$. Therefore, the sample size required to have a 95% chance of sampling an allele with frequency of 0.10 is 29 haploid individuals or 15 diploids for nuclear markers: $(1 - 0.1)^{29} = 0.047$.

Natural selection

Directional selection can cause gene trees to differ from species trees if a rare allele increases rapidly to fixation because of natural selection (**selective sweep**, see Section 10.3.1). For example, a highly divergent (ancient) lineage may be swept to fixation in a recently derived species. Here the ancient age of the lineages would not match the recent age of the newly derived species. In another example, balancing selection could maintain the same lineages in each of two long-isolated species, and lead to erroneous estimation of species divergence, as well as a phylogeny discordant with the actual species phylogeny (and with neutral genes). To avoid selection-induced errors in phylogeny reconstruction, many loci should be used. Analysis of many loci can help identify a locus with unusual

(deviant) phylogenetic patterns due to selection (as in Section 9.6.3). For example, selection might cause rapid divergence at one locus that is not representative of the rest of the genome (or of the true species tree).

Introgression

Introgression also causes gene trees to differ from species trees. For example, hybridization and subsequent backcrossing can cause an allele from species X to introgress into species Y. This has happened between wolves and coyotes that hybridize in northeastern United States, where coyote mtDNA has introgressed into wolf populations. Here, female coyotes hybridize with male wolves, followed by the F_1 hybrids mating with wolves, such that coyote mtDNA introgresses into wolf populations (Roy et al. 1994). This kind of unidirectional introgression of mtDNA (maternally inherited) has been detected in deer, mice, fish, and many other species.

MtDNA gene tree versus species tree

An example of a gene tree not equaling the species tree is illustrated in a study of mallard ducks and black ducks (Avise 1990). The black duck apparently recently originated (perhaps via rapid phenotypic evolution) from the more widely distributed mallard duck. This likely occurred when a peripheral mallard population became isolated, evolved into the black duck and became fixed for a single mtDNA lineage (e.g., via lineage sorting or selection). The mallard population is much larger and maintains several divergent mtDNA lineages, including the lineage fixed in the black duck (Figure 16.7). Thus, while the black duck is monophyletic, the mallard is paraphyletic relative to the black duck for mtDNA. Because the black duck mtDNA is common in the mallard, the black duck appears to be part of the mallard species when considering only mtDNA data. However, the black duck has important phenotypic, adaptive, and behavioral differences meriting recognition as a separate species.

This duck example illustrates a problem that is likely to occur when identifying species from molecular data alone (and from only one locus). It shows the importance of considering nonmolecular characteristics (such as life history, morphology, and geography) along with the molecular data (see Section 16.6). This example is analogous to the widely cited example of brown bears that are paraphyletic to polar bears for mtDNA lineages. Despite the lack of **monophyly**, brown bears have important phenotypic, adaptive, and behavioral differences meriting recognition as a separate species apart from polar bears (Paetkau 1999).

16.4 Description of genetic relationships within species

Identifying populations and describing population relationships is crucial for conservation and management (e.g., monitoring population status, measuring gene flow, and planning translocation strategies). Population relationships are generally assessed using multilocus allele frequency data and statistical approaches for clustering individuals or populations with a dendrogram or tree in order to identify genetically similar groups.

Population trees and phylogenetic trees look similar to each other, but they display fundamentally different types of information. Phylogenies show the time since the most recent

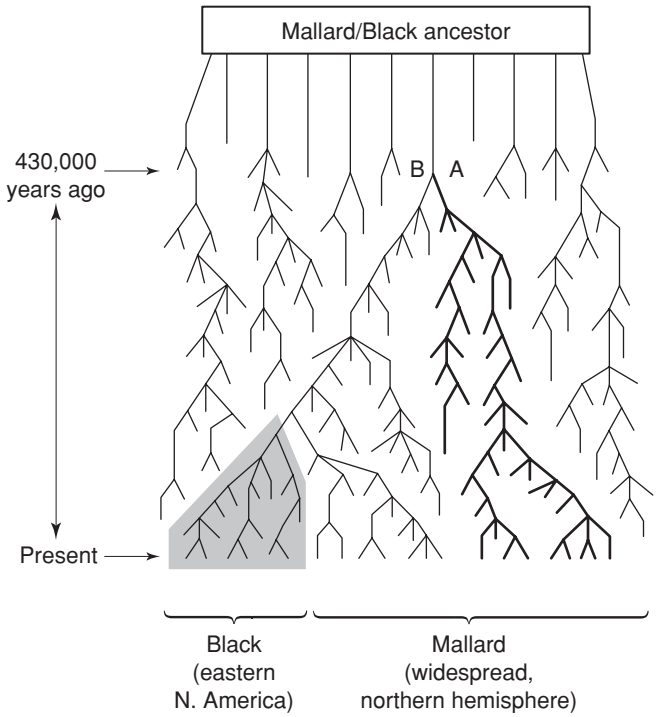


Figure 16.7 Simplified diagrammatic representation of the possible matriarchal ancestry of mallard and black ducks. The mtDNA lineage A is shown in dark lines, and the black duck portion of the phylogeny is shaded. From Avise (1990).

common ancestor (TMRCA) between taxa. Phylogenies represent relationships among taxa that have been reproductively isolated for many generations. A phylogeny identifies monophyletic groups – isolated groups that shared a common ancestor. Phylogenetic trees can be used both for species and for genes (e.g., mtDNA) (Nichols 2001). In the case of species, the branch points represent speciation events; in the case of genes, branch points represent common ancestral genes.

Population trees, in contrast, generally identify groups that have similar allele frequencies because of ongoing genetic exchange (i.e., gene flow). The concept of TMRCA is not meaningful for populations with ongoing gene flow. Populations with high gene flow will have similar allele frequencies and cluster together in population trees.

The differences between population and phylogenetic trees, as described here, are somewhat oversimplified to help explain the differences. In reality there is a continuum in the degree of differentiation among populations in nature. Some populations within the same species may have been reproductively isolated for many generations. In this case, genealogical information and the phylogenetic approach can be used to infer population relationships (see Section 16.4.3).

The description of genetic population structure is the most common topic for a conservation genetics paper in the literature. Individuals from several different geographic locations are genotyped at a number of loci to determine the patterns and amounts of gene flow among populations. This population-based approach assumes that all individuals

sampled from one area were born there and represent a local breeding population. However, new powerful approaches have been developed that allow the description of population structure using an individual-based approach. That is, many individuals are sampled, generally over a wide geographic range, and then placed in population units on the basis of genotypic similarity.

16.4.1 Population-based approaches

A bewildering variety of approaches have been used to describe the genetic relationships among a series of populations. We will discuss several representative approaches.

The initial step in assessing population relationships (after genotyping many individuals) often is to conduct statistical tests for differences in allele frequencies between sampling locations. For example, a chi-square test is used to test for allele frequency differences between samples (e.g., Roff and Bentzen 1989). If two samples are not significantly different, they often are pooled together to represent one population. It can be important to resample from the same geographic location (in different years or seasons) to test for sampling error and for stability of genetic composition through time. After distinct population samples have been identified, the genetic relationships (i.e., genetic similarity) among populations can be inferred.

Population dendrograms

Population relationships are often assessed by constructing a dendrogram based on the genetic similarity of populations. The first step in dendrogram construction is to compute a genetic differentiation statistic (e.g., F_{ST} or Nei's D; see Sections 9.1 and 9.7) between each pair of populations. A genetic distance can be computed using any kind of molecular marker (e.g., allozyme frequencies, DNA haplotypes) and a vast number of metrics (e.g., Cavalli-Sforza's chord distance, Slatkin's R_{ST}, and Wright's F_{ST}; see Section 9.7). This yields a **genetic distance matrix** (Table 16.3).

The second step is to use a clustering algorithm to group populations with similar allele frequencies (e.g., low F_{ST}). The most widely used cluster algorithms are UPGMA

Table 16.3 Genetic distance (D; Nei 1972) matrix based upon allele frequencies at 15 allozyme loci for five populations of a perennial lily. Data from Godt et al. (1997).

	Population				
	FL1	**FL2**	**FL3**	**SC**	**NC**
FL1	–				
FL2	0.001	–			
FL3	0.003*	0.002*	–		
SC	0.029	0.032	0.030	–	
NC	0.059	0.055	0.060	0.062	–

Asterisks and underlining are explained in Example 16.4.

Example 16.4 Dendrogram construction via UPGMA clustering of lily populations

UPGMA (unweighted pair group method with arithmetic averages) clustering was used to assess relationships among five populations of a perennial lily (*Tofieldia racemosa*) from northern Florida (Godt et al. 1997). Allele frequencies from 15 polymorphic allozyme loci were used to construct a genetic distance matrix (Table 16.3) and subsequently a dendrogram using the UPGMA algorithm.

The UPGMA algorithm starts by finding the two populations with the smallest interpopulation distance in the matrix. It then joins the two populations together at an internal node. In our lily example here, population FL1 and FL2 are grouped together first because the distance (0.001) is the smallest (underlined in Table 16.3). Next, the mean distance from FL1 and from FL2 to each other population is used to cluster taxa. The next shortest distance is the mean of FL3 to FL1 and FL3 to FL2 (i.e., the mean of 0.002 and 0.003; see asterisks in Table 16.3); thus FL3 is clustered as the sister group of FL1 and FL2. Next SC is clustered followed by NC (Figure 16.8).

In this example, the genetic distance is correlated with the geographic distance in that SC is geographically and genetically closer to the Florida populations than it is to the North Carolina one. See Guest Box 16 for another example application of dendrogram construction using UPGMA.

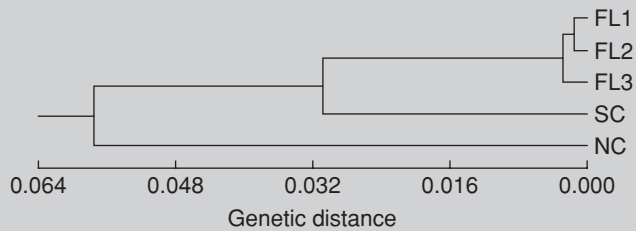


Figure 16.8 Dendrogram generated using the UPGMA clustering algorithm and the genetic distance matrix from Table 16.3. FL is Florida; SC and NC are South Carolina and North Carolina, respectively. From Godt et al. (1997).

(unweighted pair group method with arithmetic averages) and neighbor-joining (Salemi and Vandamme 2003). UPGMA clustering for dendrogram construction is illustrated by a study assessing population relationships of a perennial lily from Florida (Example 16.4).

Neighbor-joining is one of the most widely used algorithms for constructing dendrograms from a distance matrix (Salemi and Vandamme 2003). Neighbor-joining is different from UPGMA in that the branch lengths for sister taxa (e.g., FL1 and FL2, Table 16.3) can be different, and thus can provide additional information on relationships between populations. For example, FL1 is more distant from FL3 than FL2 is from FL3 (Table 16.3). This is not evident in the UPGMA dendrogram (Figure 16.8), but would be in a neighbor-joining tree. It follows that neighbor-joining is especially useful for data sets with lineages evolving at substantially different rates. Other advantages include that neighbor-joining is

fast and thus useful for large data sets and for bootstrap analysis (see next paragraph), which involves the construction of hundreds of replicate trees. It also permits correction for multiple character changes when computing distances between taxa. Disadvantages are it gives only one possible tree and it depends on the model of evolution used.

Bootstrap analysis is a widely used sampling technique for assessing the statistical error when the underlying sampling distribution is unknown. In dendrogram construction, we can bootstrap resample across loci from the original data set, meaning that we sample with replacement from our set of loci until we obtain a new set of loci, called a "bootstrap replicate". For example, if we have genotyped 12 loci, we randomly draw 12 numbers from 1 and 12 and these numbers (loci) become our bootstrap replicate data set. We repeat this procedure 100 times to obtain 100 data sets (and 100 dendrograms). The proportion of the random dendrograms with the same cluster (i.e., branch group) will be the bootstrap support for the cluster (see Figure 16.9a).

Multidimensional representation of relationships among populations

Dendrograms cannot illustrate complex relationships among multiple populations because they consist of a one-dimensional branching diagram. Thus dendrograms can oversimplify and obscure relationships among populations. Note that this is not a limitation in using dendrograms to represent phylogenic relationships, as these can be represented by a one-dimensional branching diagram as long as there has not been secondary contact following speciation.

There are a variety of multivariate statistical techniques (e.g., principal component analysis, PCA) that summarize and can be used to visualize complex data sets with multiple dimensions (e.g., many loci and alleles) so that most of the variability in allele frequencies can be extracted and visualized on a two- or three-dimensional plot (Example 16.5). Related multivariate statistical techniques include PCoA (principal coordinates analysis), FCA (frequency correspondence analysis), and MDS (multidimensional scaling).

Example 16.5 How many species of tuatara are there?

We saw in Example 1.1 that tuatara on North Brother Island in Cook Strait, New Zealand, were described as a separate species primarily on the basis of variation at allozyme loci (Daugherty et al. 1990). A neighbor-joining dendrogram based on allele frequencies at 23 allozyme loci suggested that the North Brother tuatara population is highly distinct because it is separated on a long branch (Figure 16.9a).

More recent molecular genetic data, however, have raised some important questions about this conclusion. Analysis of mtDNA sequence data indicates that tuatara from North Brother and three other islands in Cook Strait are similar to each other, and that they are all distinct form the northern tuatara populations (Hay et al. 2003; Hay et al., in preparation). Allele frequencies at microsatellite loci also support the grouping of tuatara from Cook Strait (Hay et al., in preparation).

Principal component analysis (PCA) of the allozyme data supports the similarity of the Cook Strait tuatara populations. Three major population groupings are

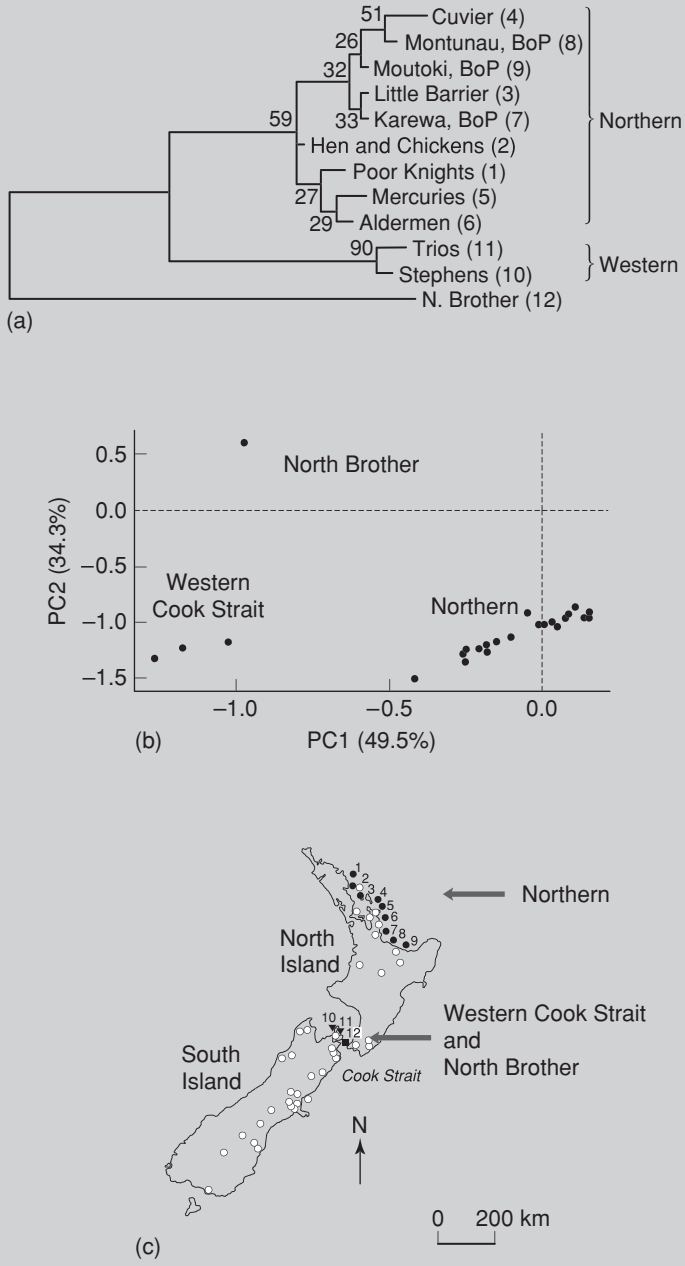


Figure 16.9 (a) Neighbor-joining dendrogram (the numbers on the branches are the bootstrap values) and (b) principal component analysis based on allele frequencies at 23 allozyme loci. (c) The map of New Zealand shows the geographic locations of populations sampled. Open circles indicate where fossil remains have been found.

apparent in the plot of the first two components of the PCA analysis (Figure 16.9b). The first component distinguishes between the northern group and the Cook Strait populations; the North Brother population clusters closely with the Western Cook Strait populations on this axis. PC2 separates the North Brother population from the other populations. The North Brother population clusters with the other Cook Strait populations on PC1, which explains nearly 50% of the variation. The North Brother population is distinct only for the second main variance component (PC2), which explains 34% of the variation. These results suggest that the Cook Strait populations are much more genetically similar to each other than they are to the northern populations.

North Brother Island is very small and the tuatara on this island have substantially less genetic variation at microsatellite loci. Thus, genetic distinctiveness of the North Brother tuatara is likely due to a small population and rapid genetic drift rather than long-term isolation that might warrant species status.

This example illustrates the limitations of one-dimensional tree diagrams and the possible loss (or oversimplification) of information when data are collapsed into one dimension.

16.4.2 Individual-based methods

Individual-based approaches are used to assess population relationships through first identifying populations by delineating genetically similar clusters of individuals. Clusters of genetically similar individuals are often identified by building a dendrogram in which each branch tip is an individual. Second, we quantify genetic relationships among the clusters (putative demes).

Individual-based methods for assessing population relationships make no a priori assumptions about how many populations exist or where boundaries between populations occur on the landscape. If individual-based methods are not used, we risk wrongly grouping individuals into populations based on somewhat arbitrary traits (e.g., color) or an assumed geographic barrier (a river) identified by humans subjectively.

One example of erroneous a priori grouping would be migratory birds that we sample on migration routes or on overwintering grounds. Here, we might wrongly group together individuals from different breeding populations, because we sampled them together at the same geographic location. A similar potential problem could exist in migratory butterflies, salmon, and whales, for example, if we sample mixtures containing individuals from different breeding groups with different geographic origins.

An individual-based approach was used by Pritchard et al. (2000) to assess relationships among populations of the endangered Taita thrush in Africa. The authors built a tree of individuals based on pairwise genetic distance between individuals. Each individual was genotyped at seven microsatellite loci (Galbusera et al. 2000). The genetic similarity index (Nei's genetic distance) between each pair of individuals was computed, and then a clustering algorithm (neighbor-joining) was used to group similar individuals together on branches. The geographic location of origin of individuals was also plotted on the branch

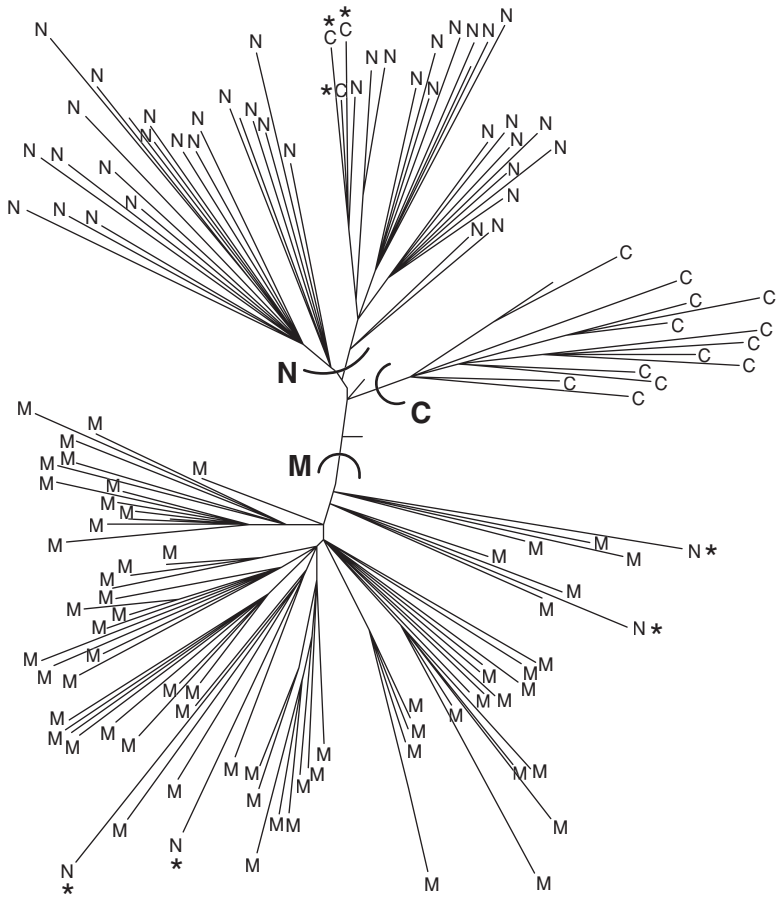


Figure 16.10 Tree of individuals (Taita thrush) constructed using the genetic distance between individuals and the neighbor-joining tree building algorithm (Chawia, 17 individuals; Ngangao, 54 individuals; Mbololo, 80 individuals). The three curved slashes (N, M, and C) across the branches identify the three population clusters. The letters on branch tips are sampling locations (i.e., population names); asterisks on branch tips represent putative immigrants (e.g., three migrants from Ngangao into Chawia (top of figure). Modified from Pritchard et al. (2000).

tips to help identify population units. The analysis revealed three distinct populations represented by three discrete clusters of individuals (Figure 16.10).

This example illustrates a strength of the individual-based approach: the ability to identify migrants. Individuals (i.e., branches) labeled with "N" and an asterisk (bottom of tree, Figure 16.10) were sampled from the "N" location (Ngangao) but cluster genetically with Mbololo (labeled "M"). This suggests these individuals are migrants from Mbololo into the Ngangao population.

An individual-based and model-based approach to identify populations as clusters of individuals was introduced by Pritchard et al. (2000). "Model based" refers to the use of a model with k populations (demes) that are assumed to be in Hardy–Weinberg (HW) and

gametic equilibrium. This approach first tests if our data fit a model with $k = 1$, 2, 3, or more populations. The method uses a computer algorithm to search for the set (k) of individuals that minimizes the amount of HW and gametic disequilibrium in the data. Many or all possible sets of individuals are tested. Once k is inferred (step 1), the algorithm estimates, for each individual, the (posterior) probability (Q) of the individual's genotype originating from each population (step 2). If an individual is equally likely to have originated from population X and Y, then Q will be 0.50 for each population.

For example, Berry et al. (2005) used 15 microsatellites and Pritchard's model-based approach to study dispersal and the affects of agricultural land conversion on the connectivity of insular populations of the grand skink from New Zealand. The skink lives in small populations (approximately 20 individuals) on rock outcrops separated by 50–150 m of inhospitable vegetation (native tussocks grassland or exotic pasture). A total of 261 skinks were genotyped from 12 rock outcrops. The number of dispersers inferred from Pritchard's cluster analysis was lower for the exotic pasture than for the native grassland habitat. For example, nine known dispersers were detected among rock outcrops within the native grassland site T1, versus only one disperser within the exotic pasture site P1 (Figure 16.11; see open squares above bar graphs representing dispersers). This study suggests that exotic pasture fragments populations; this likely increases population extinction risks by increasing genetic and demographic stochasticity (see Chapter 14).

Individual-based analyses can also be conducted with many multivariate statistical methods (e.g., PCA) if individuals are used as the operational unit (instead of populations). These multivariate approaches make no prior assumptions about the population structure model, e.g., HW and gametic equilibrium are not assumed.

Individual-based methods are useful to identify cryptic subpopulations and localize population boundaries on the landscape. Once genetic boundaries are located, we can test if the boundaries are concordant with some environmental gradient or some ecological or landscape feature (e.g., a river or temperature gradient). This approach of associating population genetic "boundaries" with landscape or environmental features has been called **landscape genetics** (Manel et al. 2003).

A final strength of individual-based methods is that they can help evaluate data quality by detecting human errors in sampling; for example, a sample with the wrong population label. Such mislabeled samples would show up as outliers (or candidate "migrants") from a different population (Figure 16.11).

A disadvantage of individual-based methods is that they often require the analysis of many individuals (hundreds), sampled across relatively evenly spaced locations. In a continuous population, we might wrongly infer a genetic discontinuity (barrier) between sampling locations if clusters of individuals are sampled from locations far apart. For example, in an isolation by distance scenario (see Figure 9.8), we could infer different (discrete) populations by sampling distant locations with no individuals in between the locations. However, this problem could arise even with the classic population-based methods (see Section 16.4.1).

A potential problem with individual-based methods is that they still can yield uncertain results if genetic differentiation among populations is not substantial. Plus the performance and reliability of individual-based methods has not been thoroughly evaluated (but see Evanno et al. 2005 for a performance evaluation of the individual-based clustering method of Pritchard et al. 2000). Thus it seems useful and prudent to use both individual-based and population-based methods.

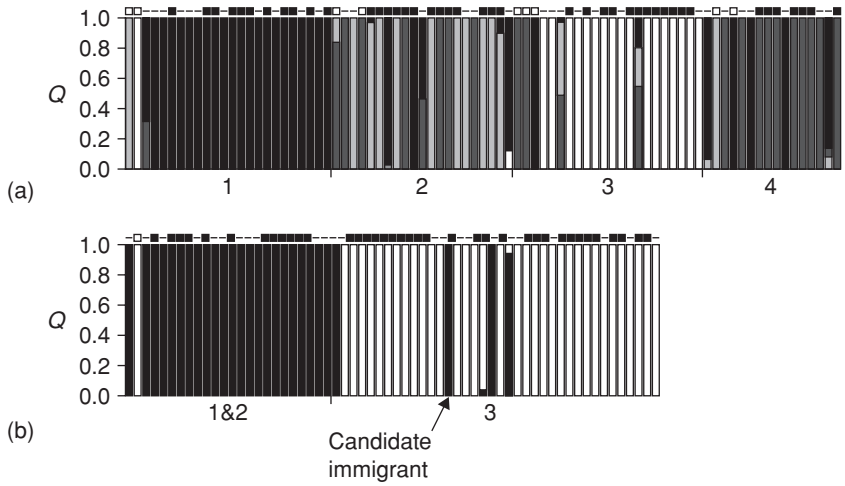


Figure 16.11 Bayesian clustering of individual skink genotypes. Each site is shown separately: (a) site T1 (native tussocks) and (b) site P1 (exotic pasture). Individuals are represented across the *x*-axis by a vertical bar that may be divided into shaded segments that represent the individual's probability of originating (*Q*) from each of the rocks at a study site (computed using STRUCTURE 2.0; Pritchard et al. 2000). Skinks are also grouped across the *x*-axis according to the rock they were captured on (e.g., 1, 2, 3, or 4). Filled squares above an individual indicate that the natal rock was known and the individual did not disperse, open squares indicate that the individual was a known disperser (from mark–recapture data), and dashes indicate that the natal rock was not known for that individual. The arrow points toward one (of several) putative immigrants. From Berry et al. (2005).

16.4.3 Phylogeography

Phylogeography is the assessment of the correspondence between phylogeny and geography (Avise 2000b). We expect to find phylogeographic structuring among populations with long-term isolation. Isolation for hundreds of generations is generally required for new mutations to arise locally, and to preclude their spread across populations. Phylogeographic structure is expected in species with limited dispersal capabilities, with **philopatry**, or with distributions that span strong barriers to gene flow (e.g., mountains, rivers, roads, and human development). In conservation biology, detecting phylogeographic structuring is important because it helps identify long-isolated populations that might have distinct gene pools and local adaptations. Long-term reproductive isolation is one major criterion widely used to identify population units for conservation (see Section 16.5.2).

Intraspecific phylogeography was pioneered initially by J. C. Avise and colleagues (Avise et al. 1987). In a classic example, Avise et al. (1979a) analyzed mtDNA from 87 pocket gophers from across their range in southeastern United States. The study revealed 23 different mtDNA genotypes, most of which were localized geographically (Figure 16.12). A major discontinuity in the maternal phylogeny clearly distinguished eastern and western populations. A potential conservation application of such results is that eastern and

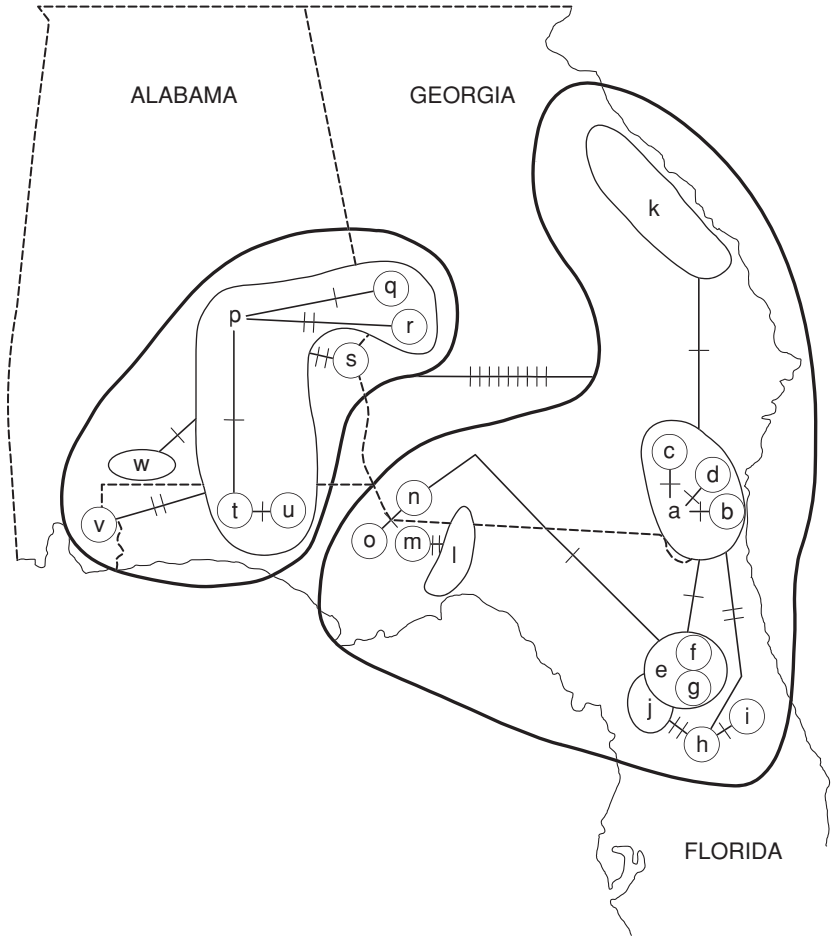


Figure 16.12 Mitochondrial DNA phylogenetic network for 87 pocket gophers; mtDNA genotypes are represented by lower case letters and are connected by branches in a parsimony network. Slashes across branches are substitutions. Nine substitutions separate the two major mtDNA clades encircled by heavy lines. From Avise (1994).

western populations of pocket gopher appear to be highly divergent with long-term isolation and thus potentially adaptive differences; this could warrant management as separate units. However additional data (including nuclear loci and nongenetic information) should be considered before making conservation management decisions (e.g., Section 16.6 and Guest Box 4).

Phylogeographic studies can help identify **biogeographic** provinces containing distinct flora and fauna worth conserving as separate geographic units in nature reserves. For example, multispecies phylogeographic studies in the southwest United States (Avise 1992) and northwest Australia (Moritz and Faith 1998) have revealed remarkably concordant phylogeographic patterns across multiple different species. Such multispecies concordance can be used to identify major biogeographic areas that can be prioritized as separate conservation units and to identify locations to create nature reserves (Figure 16.13).

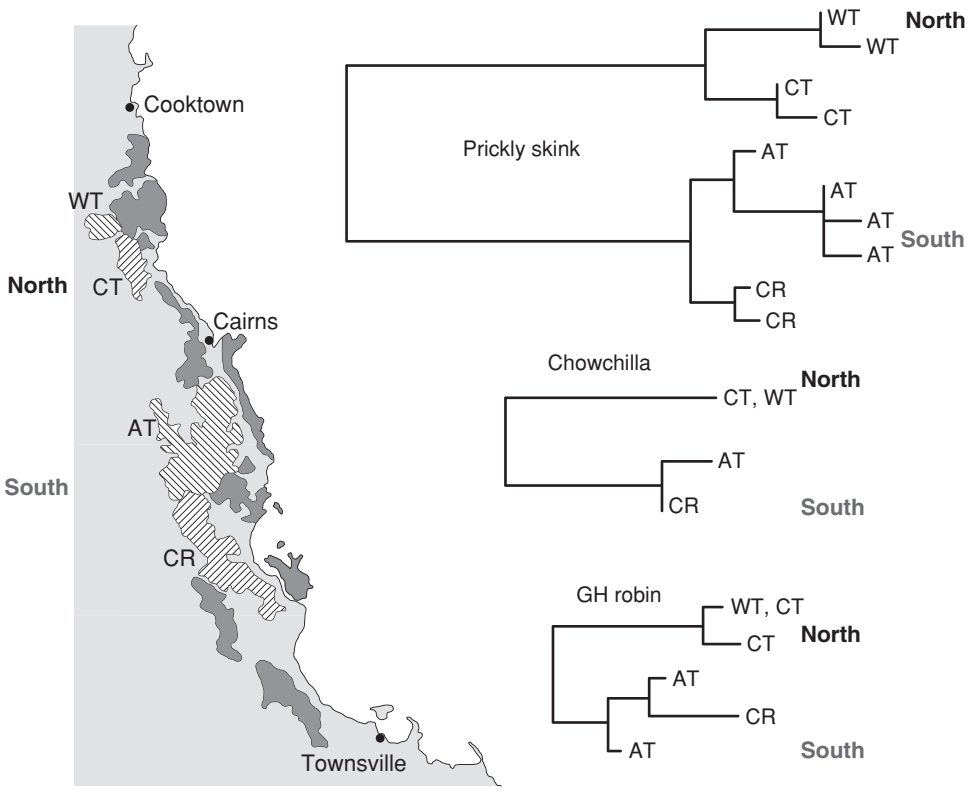


Figure 16.13 Phylogeographic analysis for three species sampled from each of four geographic areas from the tropical rain forests of northeastern Australia between Cooktown and Townsville: Windsor Tableland (WT), Carbine Tableland (CT), Atherton Tableland (AT), and Cardwell Ranges (CR). Note the deep phylogenetic break (long branches) separating the WT/CT populations in the north from the CR/AT populations in the south for all three species (prickly skink, chowchilla, and gray-headed robin). These results suggest long-term isolation for numerous species between the northern and southern rain forests. These regions merit conservation as separate systems. From Moritz and Faith (1998).

A promising phylogeographic approach is nested clade analysis (NCA) (Templeton 1998). NCA involves three steps: (1) building a parsimony network of alleles such that hierarchically nested clades are identified (i.e., recent derived clades versus ancestral clades); (2) testing for statistically significant geographic structuring of alleles within clades; and (3) interpreting the biological cause of structuring (e.g., isolation by distance, recent fragmentation, or range expansion). Step 3 uses an inference key that lists expectations of each cause of a given structuring pattern. For example, NCA predicts that under isolation by distance (i.e., restricted gene flow), the derived alleles will be localized geographically whereas ancestral alleles will be less localized. This is because under restricted gene flow new alleles will not have had time to spread geographically. This pattern is not expected under range expansion (Templeton 1998).

There has been substantial debate over the usefulness of NCA (e.g., Knowles and Maddison 2002). The main shortfall of NCA is that it does not incorporate error or

uncertainty. This is the same problem with most phylogenetic approaches. For example, NCA does not consider interlocus variation (as do coalescent-based population genetic models; e.g., Appendix A9 and Figure A12). Thus NCA might provide the correct inference about phylogeographic history, but we cannot quantify the probability of it being correct. Another occasionally cited shortfall is that NCA is somewhat ad hoc in using an inference key in order to distinguish among many different historical processes.

Fortunately, the emerging field of "statistical phylogeography" promises to combine the strengths of NCA with formal modeling and statistical tests to allow for more objective testing of alternative hypotheses that could explain phylogeographic patterns (Knowles and Maddison 2002). Until formal modeling and validated statistical phylogeography approaches are available, it seems prudent to use NCA in combination with other approaches such as AMOVA (analysis of molecular variation) that use genealogical information in ways similar to or complementary to NCA (e.g., see Turner et al. 2000).

16.5 Units of conservation

It is critical to identify species and units within species to help guide management, monitoring, and other conservation efforts, and to facilitate application of laws to conserve taxa and their habitats. In this section we consider the issues of identifying species and intraspecific conservation units.

16.5.1 Species

Identification of species is often problematic, even for some well-known taxa. One problem is that biologists cannot even agree on the appropriate criteria to define a species. In fact, more than two dozen species concepts have been proposed over the last decades. Darwin (1859) wrote that species are simply highly differentiated varieties. He observed that there is often a continuum in the degree of divergence from between populations, to between varieties, species, and higher taxonomic classifications. In this view, the magnitude of differentiation that is required to merit species status can be somewhat arbitrary.

The biological species concept (BSC) of Mayr (1942, 1963) is the most widely used species definition, at least for animals. This concept emphasizes reproductive isolation and isolating mechanisms (e.g., pre- and postzygotic). Criticisms of this concept are that: (1) it can be difficult to apply to allopatric organisms (because we cannot observe or test for natural reproductive barriers in nonoverlapping populations); (2) it cannot easily accommodate asexual species (that may not interbreed only because they are selfing); and (3) it has difficulties dealing with introgression between highly distinct forms. Further, an emphasis on "isolating mechanisms" implies that selection counteracts gene flow. However, the BSC generally does not allow for interspecific gene flow, even at a few segments of the genome (i.e., limited introgression) (Wu 2001).

The phylogenetic species concept (PSC; Cracraft 1989) relies largely on monophyly, such that all members of a species must share a single common ancestor. This concept has fewer problems dealing with asexual organisms (e.g., many plants, fish, etc.) and with allopatric forms. However, it does not work well under hybridization and it can lead to oversplitting, for example as more and more characters are used (e.g., using powerful DNA sequencing techniques) more "taxa" might be identified.

A problem using the PSC can arise if biologists interpret fixed DNA differences (monophyly) between populations as evidence for species status. For example, around the world many species are becoming fragmented, and population fragments are becoming fixed (monophyletic) for different DNA polymorphisms. Under the PSC, this could cause the proliferation of "new" species if biologists strictly apply the PSC criterion of monophlyly for species identification. This could result in oversplitting and the waste of limited conservation resources. This potential problem of fragment-ation-induced oversplitting is described in a paper titled "Cladists in wonderland" (Avise 2000a). To avoid such oversplitting, multiple independent DNA sequences (e.g., not mtDNA alone) should be used, along with many nongenetic characters when possible.

Other species definitions include the ecological species concept based on a distinct ecological niche (Van Valen 1976), and the evolutionary species concept often used by paleontologists to identify species based on change within lineages through time but without splitting (anagenesis) (Simpson 1961). The different concepts overlap, but emphasize different types of information. Generally, it is important to consider many kinds of information or criteria when identifying and naming species. If most criteria (or species concepts) give the same conclusion (e.g., species status is warranted), than we can be more confident in the conclusion.

African cichlid fishes illustrate some of the difficulties with the different species concepts. Approximately 1,500 species of cichlids have recently evolved a diverse array of morphological differences (e.g., mouth structure, body color) and ecological differences (e.g., feeding and behaviors such as courtship). Morphological differences are relatively pronounced among cichlids. However, the degree of genetic differentiation among cichlids is relatively low compared to other species, due to the recent radiation of African cichlid species (less than 1–2 million years ago!). Further complicating species identification using molecular markers, is that reproductive isolation can be transient. For example, some cichlid species are reproductively isolated due to mate choice based on fixed color differences between species. However, this isolation breaks down during years when murky water prevents visual color recognition and leads to temporary interspecific gene flow (Seehausen et al. 1997)!

Molecular genetic data can help identify species, especially cryptic species that have similar phenotypes (see also Section 20.1). For example, the neotropical skipper butterfly was recently identified as a complex of at least 10 species, in part by the sequencing of a standard gene region (**DNA "barcoding"**). The 10 species have only subtle differences in adults and are largely sympatric (Hebert et al. 2004). However, they have distinctive caterpillars, different caterpillar food plants, and a relatively high genetic divergence (3%) in the mitochondrial gene cytochrome *c* oxidase I (COI) gene.

Molecular data can also help identify taxa that are relatively well studied. For example, a recent study of African elephants used molecular genetic data to detect previously unrecognized species. Elephants from tropical forests are morphologically distinct from savannah elephants. Roca et al. (2001) biopsy-dart sampled 195 free-ranging elephants from 21 populations. Three populations were forest elephants in central Africa, 15 were savannah elephants (located north, east, and south of the forest populations), and three were unstudied and thus unclassified populations. DNA sequencing of 1,732 base pairs from four nuclear genes revealed 52 nucleotide sites that were phylogenetically informative (i.e., at least two individuals shared a variant nucleotide).

All savannah elephant populations were closer genetically to every other savannah population than to any of the forest populations, even in cases where the forest population was geographically closer (Roca et al. 2001). Phylogenetic analyses revealed five fixed site differences between the forest and savannah elephants (Figure 16.14). By comparison, nine fixed differences exist between Asian and African elephants. Hybridization was considered

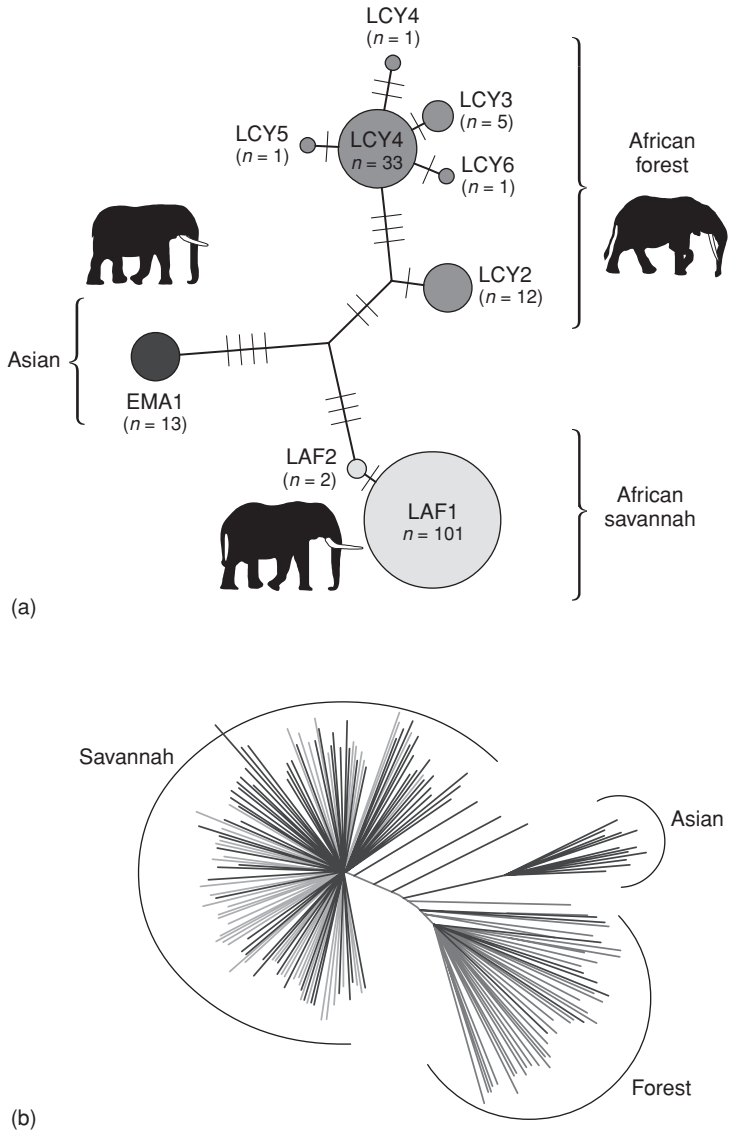


Figure 16.14 (a) Minimum spanning network showing relationships among nine haplotypes observed for the X-linked *BNG* gene for African forest and savannah elephants and the Asian elephant. Each slash mark along branches separating each haplotypes represents one nucleotide difference. From Roca et al. (2001). (b) Neighbor-joining cluster analysis of 189 African elephants and 14 Asian elephants based on proportion of shared alleles (Dps) at 16 microsatellite loci. From Comstock et al. (2002).

to be "extremely limited", although the number of individuals sampled was only moderate, and one savannah individual apparently contained one nucleotide diagnostic for the forest elephants. The genetic data (see also Comstock et al. 2002; Figure 16.14b), combined with the morphological and habitat differences, suggests that species-level status is warranted. This study represents a nice example of combining genetic and nongenetic data. The results could influence conservation strategies, making it more urgent to protect and manage these increasingly endangered taxa separately.

Sometimes genetic data may show that recognized species are not supported by reproductive relationships. Some authors have recognized the black sea turtle (*Chelonia* spp.) as a distinct species on the basis of skull shape, body size, and color (Pritchard 1999). However, molecular analyses of mtDNA and three independent nuclear DNA fragments suggest that reproductive isolation does not exist between the black and green forms (Karl and Bowen 1999). Over the years, taxonomists have proposed more than a dozen species for different *Chelonia* populations, with oversplitting occurring in many other taxa as well. Nonetheless, for conservation purposes, it is clear that black turtles are distinct and could merit recognition as an intraspecific conservation unit (see Section 16.5.2) that posses potential local adaptations. Unfortunately, populations are declining, and additional data on potentially adaptive differences are needed (e.g., food sources and feeding behavior, etc.).

16.5.2 Evolutionary significant units

An **evolutionary significant unit** (**ESU**) can be defined broadly as a population or group of populations that merit separate management or priority for conservation because of high distinctiveness (both genetic and ecological). The first use of the term ESU was by Ryder (1986). He used the example that five (extant) subspecies of tigers exist, but there is not space in zoos or captive breeding programs to maintain viable populations of all five. Thus sometimes we must choose which subspecies to prioritize for conservation action, and perhaps maintain only one or two global breeding populations (each perhaps consisting of more than one named subspecies). Since Ryder, the term ESU has been used in a variety of frameworks for identifying conservation units (Example 16.6).

There is considerable confusion and controversy in the literature associated with the term ESU. For example, the US Endangered Species Act (ESA) lacks any definition of a distinct population segment (DPS; see Example 16.1). Waples (1991) suggested that a population or group of populations (of salmon) would be a DPS if it is an ESU. This has lead to some confusion because some biologists equate a DPS and an ESU. We will use the term DPS when referring to officially recognized "species" under the ESA, and the term ESU in the more generally accepted sense.

It can be difficult to provide a single concise, detailed definition of the term ESU because of the controversy and different uses and definitions of the term in the literature. This ESU controversy is analogous to that surrounding the different species concepts mentioned above. The controversy is not surprising considering the problems surrounding the definition of species, and the fact that identifying intraspecific units is generally more difficult than identifying species (Waples 1991). It is also not surprising considering the different rates of evolution that often occur for different molecular markers and phenotypic traits used in ESU identification. Different evolutionary rates lead to problems analogous to that in the classification of birds (Aves) as a separate taxonomic class (due to their rapid evolution), when in fact birds are monophyletic within the class Reptilia (see Figure 16.3).

Example 16.6 Proposed definitions of evolutionary significant units

1. **Ryder (1986)**: populations that actually represent significant adaptive variation based on concordance between sets of data derived by different techniques. Ryder (1986) clearly argues that this subspecies problem is "considerably more than taxonomic esoterica". (Main focus: zoos for potential *ex situ* conservation of gene pools of threatened species.)
2. **Waples (1991)**: populations that are reproductively separate from other populations (e.g., as inferred from molecular markers) and that have distinct or different adaptations and that represent an important evolutionary legacy of a species. (Main focus: integrating different data types, and providing guidelines for identifying "distinct population segments" or DPSs (of salmon) which are given "species" status for protection under the United States Endangered Species Act.)
3. **Dizon et al. (1992)**: populations that are distinctive based on morphology, geographic distribution, population parameters, and genetic data. (Main focus: concordance across some different data types, but always requiring some degree of genetic differentiation.)
4. **Moritz (1994)**: populations that are reciprocally monophyletic (see Figure 16.15) for mtDNA alleles and that show significant divergence of allele frequencies at nuclear loci. (Main focus: defining practical criteria for recognizing ESUs based on population genetics theory, while considering that variants providing adaptation to recent or past environments may not be adaptive (or might even retard the response to natural selection) in future environments.)
5. **USFWS and NOAA (1996b)** (US policy for all vertebrates): (1) discreteness of the population segment (DPS) in relation to the remainder of the species to which it belongs; and (2) the significance of the population segment to the species to which it belongs. This DPS policy does not use the term ESU, but has a framework similar to that of Waples' (1991) salmon ESU policy.
6. **Crandall et al. (2000)**: populations that lack: (1) "ecological exchangeability" (i.e., they have different adaptations or selection pressures – e.g., life histories, morphology, quantitative trait locus variation, habitat, predators, etc. – and different ecological roles within a community); and (2) "genetic exchangeability" (e.g., they have had no recent gene flow, and show concordance between phylogenetic and geographic discontinuities). (Main focus: emphasizing adaptive variation and combining molecular and ecological criteria in a historical timeframe. Suggests returning to the more holistic or balanced and two-part approach of Waples.)

In practice, an understanding of the underlying principles and the criteria used in the different ESU frameworks will help when identifying ESUs. The main criteria in several different ESU concepts are listed in Example 16.6, and synthesized at the end of this section (see also Fraser and Bernatchez 2001). Here we discuss some details about three widely

used ESU frameworks, each with somewhat different criteria as follows: (1) reproductive isolation and adaptation (Waples 1991); (2) **reciprocal monophyly** (Moritz 1994); and (3) "exchangeability" of populations (Crandall et al. 2000). This will provide a background on principles and concepts, as well as a historical perspective of the controversy surrounding the different frameworks for identifying units of conservation.

Isolation and adaptation

Waples (1991) was the first to provide a detailed framework for ESU identification. His framework included the following two main requirements for an ESU: (1) long-term reproductive isolation (generally hundreds of generations) so that an ESU represents a product of unique past evolutionary events that is unlikely to re-evolve (at least on an ecological time scale); and (2) ecological or adaptive uniqueness such that the unit represents a reservoir of genetic and phenotypic variation likely important for future evolutionary potential. This second part requiring ecological and adaptive uniqueness was termed the "evolutionary legacy" of a species by Waples (1991, 1998). This framework has become the official policy of the United States Fish and Wildlife Service and the National Marine Fisheries Service (USFWS and NOAA 1996b).

Waples (2005) recently argued that ESU identification is often most helpful if an intermediate number of ESUs are recognized within each species, e.g., when the goal is to preserve a number of genetically distinct populations within a species. Waples (2005) reviewed many of the published ESU concepts and criteria (e.g., see Example 16.6) and concluded that they could often identify only a single ESU or a large number (hundreds) of ESUs in Pacific salmon species. This is a tentative conclusion based on the published criteria for other ESU concepts, many of which are subjective or qualitative. There is a need for more empirical examples in which multiple ESU concepts are applied to a common problem (as in Waples 2005).

Reciprocal monophyly

Moritz (1994) offered simple and thus readily applicable molecular criteria for recognizing an ESU: "ESUs should be reciprocally monophyletic for mtDNA (in animals) and show significant divergence of allele frequencies at nuclear loci". Mitochondrial DNA is widely used in animals because it has a rapid rate of evolution and lacks recombination, thus facilitating phylogeny reconstruction. Cytoplasmic markers are often used in plants as they also lack recombination. "Reciprocally monophytic" means that all DNA lineages within an ESU must share a more recent common ancestor with each other than with lineages from other ESUs (Figure 16.15). These molecular criteria are relatively quick and easy to apply in most taxa because the necessary molecular markers (e.g., "universal" PCR primers) and data analysis software have become widely available. Further, speed is often important in conservation where management decisions may have to be made quickly, and before thorough ecological studies of a species can be conducted.

An occasionally cited advantage of the Moritz (1994) monophyly criterion is that it can employ population genetics theory to infer the time since population divergence. For example, it takes a mean of $4N_e$ generations for a newly isolated population to coalesce to a single gene copy and therefore become reciprocally monophyletic (through drift and

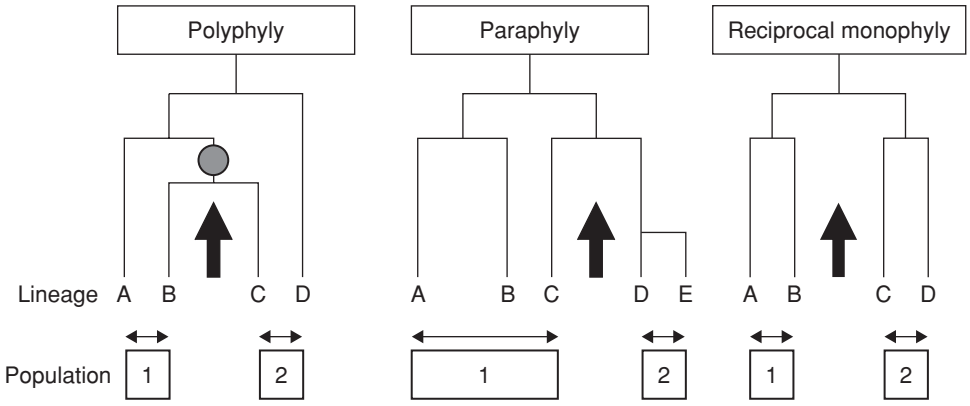


Figure 16.15 Development of phylogenetic structure of alleles between populations. After a population splits into two, the phylogenetic relationship of the alleles in the two daughter populations usually proceeds from polyphyly through paraphyletic conditions to reciprocal monophyly. The alleles (or lineages) are labeled A, B, C, D, and E. When two populations (1 and 2) become isolated, **both** will initially have some alleles that are more closely related to alleles in the other population (**poly**phyly). The filled circle at the root of the B and C branches represents the most recent common ancestor between B and C. After many generations of isolation, one population might become monophyletic, e.g., for alleles for D and E in the paraphyly example (see also the black duck, Figure 16.7). But the other population might maintain an allele that is more related to an allele in the other population (e.g., the mallard duck, Figure 16.7). After approximately four N_e generations, both daughter populations will usually be monophyletic with respect to each other (reciprocal monophyly). Modified from Moritz (1994).

mutation) at a nuclear locus (Neigel and Avise 1986). This means that if a population splits into two daughter populations of size $N_e = 1{,}000$, it would take an expected 1,000 generations to become reciprocally monophyletic for mtDNA. For mtDNA to become monophyletic it requires fewer generations because the effective population size is approximately four times smaller for mtDNA than for nuclear DNA; thus lineage sorting is faster (see Section 9.5). Here it is important to recall that adaptive differentiation can occur in a much shorter time period than does monophyly (see, for example, Guest Box 8).

A disadvantage of the Moritz ESU concept is it generally ignores adaptive variation, unlike the two-step approach that also incorporates the "evolutionary legacy" of a species (Waples 1991). The framework of Moritz is based on a cladistic phylogenetic approach (Section 16.2) using neutral loci. Thus, unfortunately, the Moritz approach makes it more likely that smaller populations (e.g., bottlenecked populations) will be identified as ESUs. Small populations quickly become monophyletic due to drift or lineage sorting. Worse, natural selection is most efficient in large populations. Consequently, the strict Moritz framework is unlikely to identify ESUs with substantial adaptive differences.

One limitation of using only molecular information is that a phylogenetic tree might not equal the true population tree. This is analogous to the "gene tree versus species tree" problem discussed in Section 16.3.2 (Example 16.7). This problem of population trees not equaling gene trees is worse at the intraspecific level because there is generally less time

Example 16.7 Lack of concordance between mtDNA and nuclear genes in white-eyes

Degnan (1993) compared dendrograms from mtDNA and nuclear DNA for two species of white-eye from Australia. The mtDNA data yielded a single gene tree that does not reflect the organismal tree based on phenotypic characters. In contrast, the two nuclear DNA loci revealed phylogeographic patterns consistent with the traditional classification of the two species (Figure 16.16). The author concluded that the discordance between the mtDNA and nuclear DNA (and phenotype) likely results from past hybridization between the two species of white-eye and mtDNA introgression. Evidence for hybridization might have been lost in nuclear genes through recombination.

This study provides a clear empirical demonstration that single gene genealogies cannot be assumed to accurately represent the true organismal phylogeny. Further, it emphasizes the need for analyses of multiple, independent DNA sequences when inferring phylogeny and identifying conservation units. This is especially true for populations within species where relatively few generations have passed. For example, if we were trying to identify ESUs (or species) in this study by using mtDNA alone, we might identify three ESUs (corresponding to the three mtDNA haplogroups in Figure 16.16); However these three are not concordant with the two groups identified by phenotype, nuclear DNA, and geography.

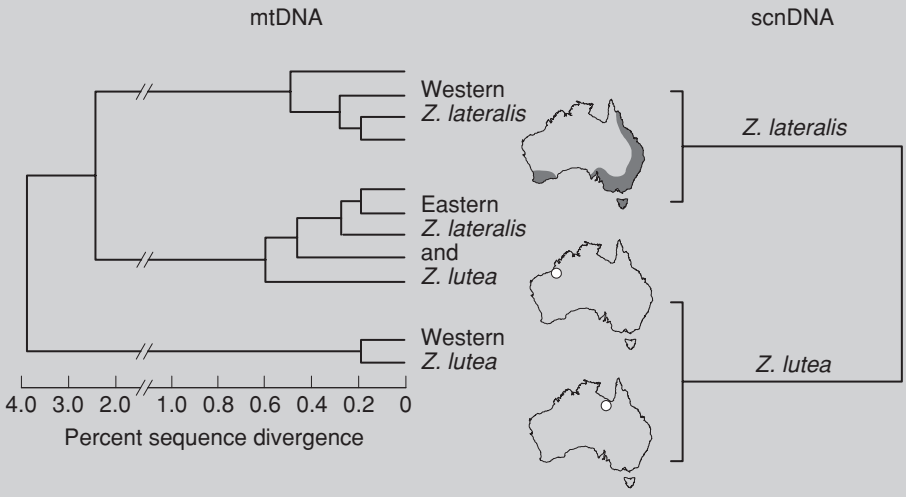


Figure 16.16 UPGMA dendrograms for mtDNA haplotypes (left) and scnDNA (single copy nuclear DNA) genotypes (right) in silver-eyes (*Zosterops lateralis*) and yellow white-eyes (*Z. lutea*). The distribution of silver-eye haplotypes is shown in solid black (top map), and for yellow white-eyes in white circles (bottom two maps). Note the middle map shows that the yellow white-eye samples from northwestern Australia group in the mtDNA tree with silver-eye samples from eastern Australia, but group with the yellow white-eyes in the scnDNA tree. Modified from Degnan (1993).

since reproductive isolation at the intraspecific level, and thus more problems with lineage sorting and paraphyly. Consequently, problems of gene trees not matching population trees (and different genes giving different trees) will be relatively common at the intraspecific level. Unfortunately, in the conservation literature, mtDNA data are often used alone to attempt to identify ESUs. This should occur less often as nuclear DNA markers become more readily available.

Exchangeability

Crandall et al. (2000) suggested that ESU identification be based on the concepts of ecological and genetic "exchangeability". The idea of exchangeability is that individuals can be moved between populations and can occupy the same niche, and can perform the same ecological role as resident individuals, without any fitness reduction due to genetic mechanisms (e.g., outbreeding depression). If we can reject the hypothesis of exchangeability between populations, then those populations represent ESUs. Ideally, exchangeability assessment would be based on heritable adaptive quantitative traits. Strengths of this approach are that it integrates genetic and ecological (adaptive) information and that it is hypothesis based.

Exchangeability can be tested using common-garden experiments and reciprocal transplant experiments. For example, if two plant populations from different locations have no reduced fitness when transplanted between locations, they might be exchangeable and would not warrant separate ESU status (see also Section 2.4, Figure 2.9, and Figure 8.1).

The main problem with this approach is it is not generally practical – i.e., it is difficult to test the hypothesis of exchangeability in many species. For example, it is difficult to move a rhinoceros (or most any endangered species) from one population to another and then to measure its fitness and the fitness of its offspring. Such studies are especially problematic in endangered species where experiments are often not feasible. Although difficult to test, exchangeability is a still worthy concept to consider when identifying ESUs. Even when we cannot directly test for exchangeability, we might consider surrogate measures of exchangeability, such as life history differences, the degree of environmental differentiation, or the number of functional genes showing signatures of adaptive differentiation (see Section 16.6). Surrogates are often used when applying Waples' ESU definition (see Guest Box 16).

Synthesis

Substantial overlap in criteria exists among different ESU concepts. Several concepts promote a two-step approach involving isolation and adaptive divergence. The main principles and criteria are the following: reproductive isolation (no gene flow), adaptive differentiation, and concordance across multiple data types (e.g., genetic, morphological, behavioral, life history, and geographic). The **longer the isolation** and the **more different the environment** (i.e., selection pressures), the more likely are populations to represent distinct units worthy of preservation and separate management. We should not rely on any single criterion, such as reciprocal monophyly of mtDNA. In fact, the greater the number of different data types showing concordant differentiation between populations, the stronger the evidence for ESU status.

16.5.3 Management units

Management units (**MUs**) usually are defined as populations that are demographically independent. That is, their population dynamics (growth rate) depend mostly on local birth and death rates rather than on immigration. The identification of these units, similar to "stocks" recognized in fisheries biology, would be useful for short-term management – such as delineating hunting or fishing areas, setting local harvest quotas, and monitoring habitat and population status.

MUs, unlike ESUs, generally do not show long-term independent evolution or strong adaptive differentiation. MUs should represent populations that are important for the long-term persistence of an entire ESU (and/or species). The conservation of multiple populations, not just one or two, is critical for insuring the long-term persistence of species (Hughes et al. 1997; Hobbs and Mooney 1998).

MUs are generally smaller than ESUs, such that an ESU might contain several MUs. MUs often are divergent subpopulations within a major metapopulation that represents an ESU. For example, fish populations are often structured on hierarchical levels such as small streams (as MUs) that are nested within a major river drainage (an ESU, e.g., Guest Box 16). Moritz (1994) defined the term "management unit" as a population that has substantially divergent allele frequencies at many loci.

One potential limitation of using allele frequency differentiation (e.g., F_{ST}) to identify MUs is that F_{ST} cannot directly be interpreted as evidence for demographic independence. For example, large populations experience little drift (and little allele frequency differentiation) and thus can be demographically independent even if allele frequencies are similar. The same Nm (and hence F_{ST}) can result in different migration rates (m) for different population sizes (N). As N goes up, m goes down for the same F_{ST} (Table 16.4). So in a large population, the number of migrants can be very small, and the population could be demographically independent yet have a relatively low F_{ST}.

A related difficulty is determining if migration rates would be sufficient for recolonization on an ecological time scale, e.g., if a MU became extinct or overharvested. Allele frequency data can be used to estimate migration rates (Nm), but at moderate to high rates of migration ($Nm > 5$–10) genetic estimators are notoriously imprecise, such that confidence intervals on the Nm estimate might include infinity (Waples 1998). Unfortunately, the

Table 16.4 Inferring demographic independence of populations by using genetic differentiation data (F_{ST}) requires knowledge of the effective population size (N_e). Here, the island model of migration was assumed to compute $N_e m$ and m (proportion of migrants) from the F_{ST} (as in Figure 9.9). Recall that the effective population size is generally far less than the census size in natural populations (see Section 7.10).

F_{ST}	N_e	m	$N_e m$	Demographic independence
0.10	50	0.040	2	Unlikely
0.10	100	0.020	2	Likely
0.10	1,000	0.002	2	Yes

range of *Nm* we are most interested in for MU identification is often moderate to high (5–50). Additional problems with *Nm* estimation exist (e.g., Whitlock and McCauley 1999; and Section 9.8.1).

The identification of conservation units can be difficult when population differences are subtle or if hierarchical structure is complex. For example, in green turtles (mentioned in Section 16.5.1), two ESUs have been proposed: the Atlantic Ocean and the Indo-Pacific region (e.g., Karl and Bowen 1999). Within each ESU, more than 10 MUs have been recognized; however population differences and demographic independence is difficult to delineate. In the humpback whale, extensive molecular and demographic studies have suggested the presence of one ESU containing numerous MUs, most of which correspond to major stocks identified by migration routes (Baker et al. 1998). A similar scenario was proposed for koalas in Australia, which was suggested to contain only one ESU but many MUs, based on mtDNA, microsatellite DNA variation, and biogeography (e.g., Houlden et al. 1999).

Two general errors can occur in MU diagnosis (as with ESU identification): identifying too few or too many units. Recognizing too few units could lead to underprotection and then to the reduction or loss of local populations. This problem could arise, for example, if statistical power is too low to detect genetic differentiation when differentiation is biologically significant.

For example, too few MUs (and underprotection) may be established if only one MU is identified when the species is actually divided into five demographically independent units. Consider that the sustainable harvest rate is 2% per year on the basis of total population, but that all the harvest comes from only one of the five MUs. Then the actual harvest rate for the single harvested MU is 10% (assuming equal size of the five MUs). This high harvest rate could result in overexploitation and perhaps extinction of the one harvested MU population. For example, if the harvested population's growth rate is only 4% per year and the harvest rate is 10%, overexploitation would be a problem (Taylor and Dizon 1999).

Here, undersplitting could result from either a lack of statistical power (e.g., due to too few data) or to the misidentification of population boundaries (e.g., due to cryptic population substructure). To help avoid misplacement of boundaries, researchers should sample many individuals that are widely distributed spatially, and use recently developed, individual-based statistical methods (see Section 16.4.2 and caveats therein; see also Manel et al. 2003).

Diagnosing too many MUs (oversplitting) could lead to unnecessary waste of conservation management resources. This error could occur if, for example, populations are designated as MUs because they have statistically significant differences in allele frequencies, but this differentiation is not associated with important biological differences. This becomes a potential problem as more and more molecular markers are used that are highly polymorphic (and thus statistically powerful).

For example, if many highly polymorphic microsatellites are genotyped, and different populations have "significantly" different allele frequencies ($P < 0.01$), they might not necessarily warrant recognition as different MUs. This is because the magnitude of differentiation could be low (e.g., $F_{ST} << 0.01$) even though this relatively small difference is significantly different ($P < 0.01$). Note that an F_{ST} of 0.01 suggests that populations probably exchange numerous migrants, on average, per generation (recall that if $F_{ST} = 0.1$, then approximately two migrants per generation would occur, assuming the island model,

expression 9.3). In this case, the populations might not be demographically independent and not merit separate MU status. Researchers and managers must be careful to understand the difference between the biological significance and the statistical significance of genetic differentiation measures (e.g., see Waples 1998; Hedrick 1999). Also recall that the island model has numerous assumptions unlikely to be met in natural populations (see Section 9.8.1; Whitlock and McCauley 1999).

16.6 Integrating genetic, phenotypic, and environmental information

Many kinds of information should be integrated, including life history traits, environmental characteristics, phenotypic divergence, and patterns of gene flow (isolation/phylogeography), for the identification of conservation units (Figure 16.17; Guest Box 16). For example, if two geographically distant populations (or sets of populations) show large molecular differences that are concordant with life history (e.g., flowering time) and morphological (e.g., flower shape) differences, we would be relatively confident in designating them as two geographic units important for conservation (perhaps ESUs).

Researchers should always consider if the environment or habitat type of different populations has been different for many generations, because this could lead to adaptations (even in the face of high gene flow) that are important for the long-term persistence of species. The more kinds of independent information that are concordant, the more sure one can be that a population merits recognition as a conservation unit. The principle of considering multiple data types and testing for concordance is critical for identifying conservation units.

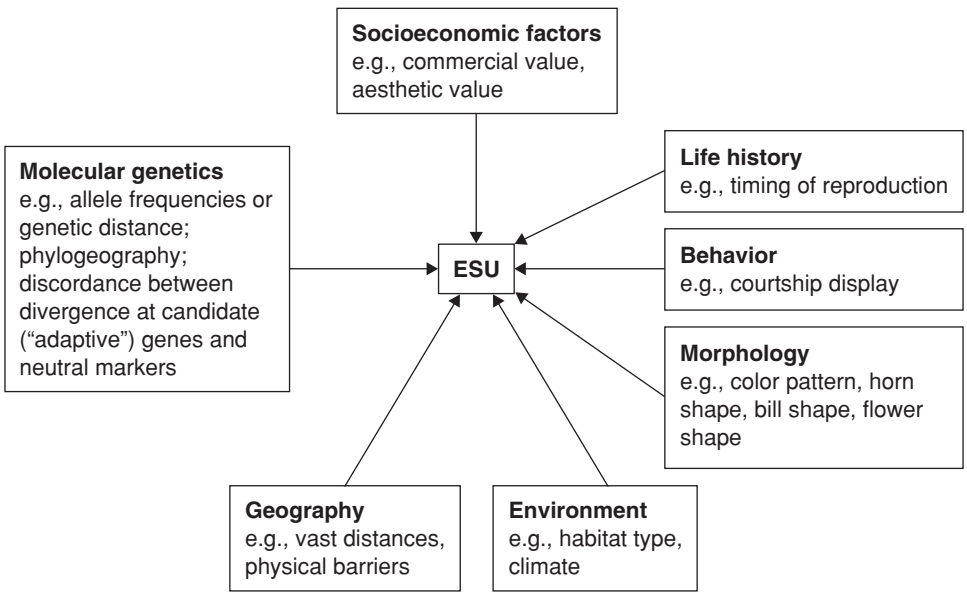


Figure 16.17 Sources of information that can help diagnose an evolutionary significant unit. Modified from Moritz et al. (1995).

When concordance is lacking among data types, difficulties arise. For example, imagine that two populations show morphological differences in size or color, but show evidence of extensive recent gene flow. This scenario has arisen occasionally in studies that measure phenotypic traits from only small samples or nonrepresentative samples of individuals from each population (e.g., only 5–10 individuals of different sexes or ages). In this example taxonomic "oversplitting" results from biased or limited sampling, and status is not warranted. This hypothetical example relates to the green/black turtle "species" dilemma described above, where more extensive sampling and study (including life history and adaptive trait information) is needed.

It is prudent to not use only molecular or only morphological information to identify ESUs, because adaptive differences can exist between populations even when little molecular or morphological differentiation is detectable (e.g., see Merilä and Crnokrak 2001). It is especially important to not base gene flow estimates on only one type of molecular marker (e.g., mtDNA). Rather, researchers should combine many loci along with ecological information. When ecological information is scarce (for example life history information can be difficult to collect), researchers could at least consider climate, habitat type, adaptive gene markers, etc., when identifying an ESU.

One example of how to integrate adaptive and "neutral" molecular variation is to consider them on two separate axes in order to identify populations with high distinctiveness for both adaptive and neutral diversity (Figure 16.18). For many species (e.g., mammals, salmon, and agricultural plants and their relatives), it is becoming feasible to detect adaptive molecular variation by genotyping numerous mapped markers including candidate genes (Luikart et al. 2003; Morin et al. 2004).

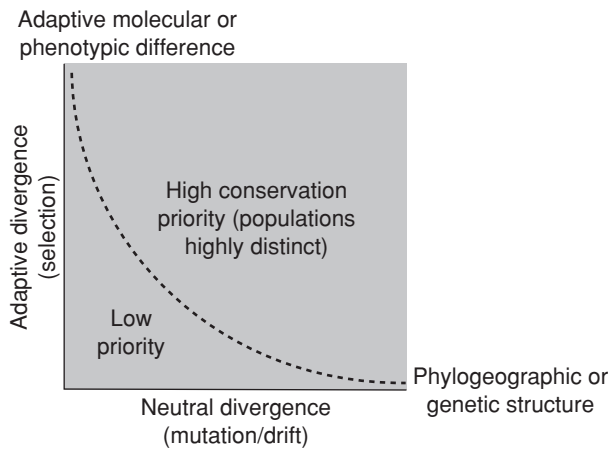


Figure 16.18 Adaptive information, including adaptive genes under selection, could be integrated with information from neutral markers and information on long-term isolation. Such an approach could help identify the most appropriate source population (i.e., non-adaptively differentiated population) from which to translocate individuals into small, declining populations that require supplementation. This approach could also help rank or prioritize populations for conservation management. From Luikart et al. (2003); modified from Moritz (2002).

Guest Box 16 Identifying conservation units in Pacific salmon
Robin S. Waples

Pacific salmon populations considered for listing as threatened or endangered "species" under the US Endangered Species Act (ESA) have been evaluated by the National Marine Fisheries Service (NMFS) using a concept of evolutionary significant units (ESUs) developed by Waples (1991, 1995). Under this framework, a population or (more often) group of populations is considered an ESU if it is substantially isolated reproductively and contributes substantially to ecological and genetic diversity of the species as a whole. Molecular genetic data are particularly informative for the first criterion. The second criterion emphasizes adaptive differences, but direct information about adaptations is generally lacking, so life history and ecological data are typically used as proxies.

In one example application, the NMFS evaluated a petition to list steelhead in the Illinois River in southern Oregon under the ESA. The Biological Review Team (BRT; Busby et al. 1993) found some support for the petitioners' claims of local differences in phenotypic and life history traits between Illinois River steelhead and nearby Rogue River steelhead, but in a broader geographic survey the traits of Illinois River fish were found to be shared by many other populations in southern Oregon and northern California. Furthermore, three of four genetic samples collected from within the Illinois River drainage were more similar to a population from outside the basin than to any of the other Illinois River samples. As a consequence of these findings, which illustrate the importance of an appropriate geographic context for evaluating distinctiveness, the BRT concluded that Illinois River steelhead do not, by themselves, constitute an ESU.

The BRT again expanded the geographic scope of its evaluations to determine the boundaries of the ESU to which Illinois River steelhead belong. Several lines of evidence suggest that Cape Blanco, which forms the northern boundary for the Klamath Mountains Geological Province (KMP), is also the northern boundary for this ESU. The KMP is distinctive geologically and ecologically (interior valleys receive less precipitation than any other location in the Pacific Northwest west of the Cascade Range) and supports a large number of endemic species. In the marine environment, the strength and consistency of coastal upwelling south of Cape Blanco yields high productivity in nearshore waters utilized by salmon. Tagging studies suggest that coho salmon and steelhead from south of Cape Blanco may not be strongly migratory, remaining instead in these productive oceanic waters.

Identifying the southern extent of this ESU was more problematical. The KMP and the distinctive Klamath–Rogue freshwater zoogeographic zone include the Klamath River basin but not areas further south. However, Cape Mendocino (well to the south of the Klamath River) is a natural landmark associated with changes in ocean currents and represents the approximate southern limit of two important life history traits for steelhead: adult fish that return to fresh water in the summer, and subadults that spend only a few months at sea before returning to fresh water on a false spawning run at a size that inspired their name, "half-pounders". Finally, the area of increased upwelling extends well into central coastal California.

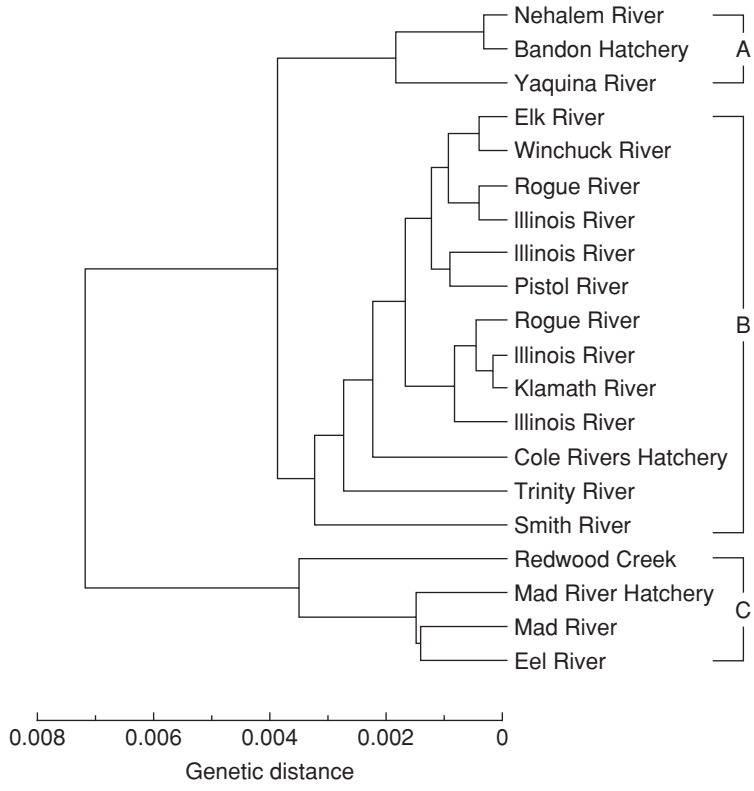


Figure 16.19 Dendogram constructed using UPGMA and pairwise genetic distances between populations (Nei's 1978 unbiased distance) computed from allele frequencies at 39 polymorphic allozyme loci. The population groups A, B, and C are in different ESUs. From Busby et al. (1994).

This issue was resolved by additional genetic sampling from the northern California coast, which showed a sharp genetic transition south of the Klamath River. At several genetic loci, alleles that were rare or absent north of the Klamath River suddenly appeared at appreciable frequencies (Figure 16.19). These results suggest considerable reproductive isolation between steelhead from the Klamath River and populations to the south (those in cluster C), and as a result the BRT concluded that the Klamath River forms the southern boundary for this ESU (Busby et al. 1994).

This example illustrates how combining different kinds of information can help identify intraspecific units for conservation. Here, the kinds of information included life history (migratory behavior), geology (river drainage system), ecology and environment (precipitation, ocean currents, and productivity), and demography (tagging and movement studies), as well as genetics (allele frequency differences).

Problem 16.1

What are the three hierarchical levels of biodiversity recognized by the IUCN and some other organizations? Name two additional organizational levels. Is any one level most important to focus on for conservation efforts? Why or why not?

Problem 16.2

Name three temporal aspects of biodiversity conservation. Which is most important and should be prioritized?

Problem 16.3

Describe several scenarios, mechanisms, and evolutionary processes that can lead to isolated populations failing to show reciprocal monophyly of DNA lineages.

Problem 16.4

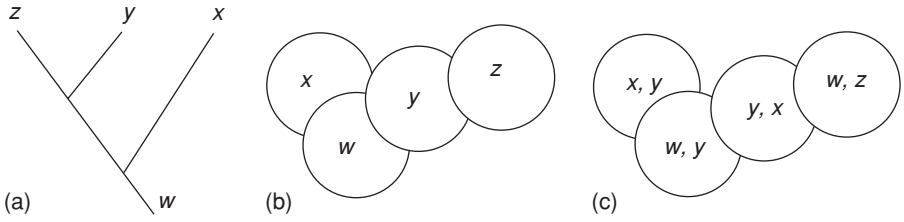
What are the three main schools of taxonomic classification? Which school is a combination of the other two? Which is most widely used today? Which is most appropriate for studies of evolutionary history?

Problem 16.5

Define paraphyly and polyphyly. Does our currently accepted classification of birds (relative to reptiles) represent an example of paraphyly or polyphyly? Are reptiles monophyletic in our currently accepted classification? (Consider Figure 16.3.)

Problem 16.6

The figure below (a) shows a hypothetical phylogenetic tree with three derived alleles *x*, *y*, and *z*, that arose from the ancestral allele *w*. Figure (b) shows circles representing geographic areas in which each allele from (a) is distributed (modified from Moritz and Faith 1998). Conduct a phylogeographic analysis and overlay the phylogeny of alleles onto the geographic distribution (b) of alleles. Is there evidence for phylogeographic structuring? Why or why not? Now, imagine another geographic distribution of alleles (c); does figure (c) reveal phylogeographic structuring?



Problem 16.7

What is an ESU? What are the main principles and criteria used for ESU identification?

Problem 16.8

How does a management unit differ from an ESU? How might molecular markers be used to help identify management units?

Problem 16.9

What kinds of information are most useful for identifying units for conservation?

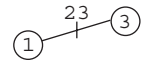
Problem 16.10

Below are four allelic DNA sequences from a species of domestic ungulate (family Bovidae). Nucleotide positions are numbered 1 to 40 (numbers are written vertically) above the sequences. For example, nucleotide position 7 has a substitution (**G**) in sequence 2 and 4, which is different from the "t" in reference sequence 1.

(a) Build a parsimony network (as in Figures 16.12 and 16.14a) by hand by connecting the sequences based on their similarity at polymorphic nucleotide sites (in bold and capitals). The two circles (connected by a line) below the sequences are to get you started drawing the network. Circles represent sequences 1 and 3 and the line connecting the circles show the one substitution at base pair position 23 that exists between the two sequences (haplotypes) 1 and 3.

```
                Base pair position (1-40)
          1234567891 1111111112 2222222223 3333333334
                   0 1234567890 1234567890 1234567890

1) gagtattata agggcgagtg tcatttcttc aacgggaccg
2) gagtatGCta agAgcgagtg tcatttcttc aaccggacgg
3) gagtattata agggcgagtg tcTtttcttc aacgggaccg
4) gagtatGCta agAgcgagtg tcatttGttc aacgggacgg
```



(b) Conduct a BLAST search (at http://www.ncbi.nlm.nih.gov/BLAST/) with sequence number 1, and identify the gene and species of origin of the sequence.