

Molecular identification: species, individuals, and sex

- Introduction
- The species question
- Dealing with individuals
- Sex
- Bits of individuals: non-invasive sampling
- Molecular identification methods: an appraisal
- Summary
- Review articles
- Useful software
- Questions

Introduction

It may seem surprising that the identification of macroorganisms, mostly animals and plants large enough for visual inspection, is an issue in ecology. After all, this kind of skill is the prerequisite of a good naturalist and most ecologists clearly fall into that category. However, there are two contexts in which significant problems arise with identification, and molecular approaches are relevant to both of them. The first is more profound but, most of the time, is a headache for evolutionary biologists rather than ecologists. It concerns the very fundamentals of taxonomy, deciding what a species really is and where lines are to be drawn between species, subspecies, hybrids, and so on. Although this important theoretical question is usually somewhat removed from the general practice of ecology, there are situations where it has to be considered in the field. It is essentially a problem of group identification. The second problem is conceptually trivial but can be of substantive practical importance, and is therefore quite often of concern to ecologists. This is the issue of identifying the species, sex, or identity of individuals under circumstances where simple morphology

cannot be relied upon. There are many examples of these shortcomings which, in the past, has probably limited the kinds of study that ecologists could carry out. Perhaps the most widespread problem is where organisms readily identifiable as adults have morphologically indistinguishable early life stages, a common situation in invertebrate groups such as beetles and molluscs. It can be useful to know the sex of an individual early in its development, before sexual maturity is attained, even in species that do not undergo dramatic metamorphosis. However, for a high proportion of species this will not be obvious, other than by killing and dissection of the reproductive organs. For some invertebrates even the identification of adults requires this destructive treatment and for many plants morphological identification can only be carried out with certainty at a particular time of year, such as the flowering season. For rare animals of special concern to conservation biologists it may be much easier to find fragments of tissue (fur, feathers) or excrement than to locate individuals directly. In all these situations there are opportunities for molecular methods to provide valuable assistance. As we will see in Chapter 9, both of these problems (fundamental taxonomy and identification) coalesce in microbial ecology to generate particularly daunting difficulties.

● KEY POINT

The two main problems with identification are agreeing about taxonomic units and recognizing specific organisms.

The species question

Defining distinctiveness

Perhaps the most common taxonomic problem faced by ecologists relates to the applied science of conservation biology. The dilemma is essentially one of deciding when a group of animals or plants is sufficiently distinct as to merit conservation in its own right. This can be difficult at a variety of taxonomic levels from species through to subspecies, races, or varieties, designations which in many cases are very poorly defined. It has been realized for sometime that even the long-standing *Biological Species Concept* (BSC) is frequently inadequate as a basis for decisions of this type. The BSC remains the most widespread method for distinguishing whether organisms are of the same or different species, and is based on whether individuals interbreed with one another to produce viable, fertile offspring. However, this apparently simple criterion can be difficult to establish. Most obviously, when populations do not overlap (i.e. when distribution ranges are fragmented) the concept cannot easily be tested. Very often there are good reasons for making conservation priorities below the species level, for example, with morphologically distinct ‘subspecies’. Here the grounds for decision-making become particularly problematic because there is not even a universal principle comparable with the BSC that can be applied. This problem is discussed further in Chapter 8, Conservation genetics.

Genetic analysis can usually reveal the extent of population structuring and thus give an indication of subgroup distinctiveness. However, the outcome of such efforts is highly dependent on the marker used. Allozymes are relatively

● KEY POINT

In conservation biology it is important to decide on the taxonomic unit deserving attention.

insensitive to fine-scale variation in many situations and may suggest uniformity, where more polymorphic loci such as microsatellites reveal significant differentiation into subgroups. Indeed, genotyping across enough polymorphic loci will eventually show that every individual is distinct! Molecular contributions in this area have therefore usually promoted ‘splitting’ rather than ‘lumping’ of taxa, but it is easy to see how this can confuse rather than clarify a conservation issue. It has been suggested, for example, that reciprocal monophyly in mitochondrial DNA (mtDNA) sequences (where each population has its own unique set of haplotypes, see Chapter 7) could be a basis for defining ‘evolutionary significant units’ worthy of independent study and conservation (Ryder 1986; Moritz 1994a, b). However, this definition is problematic because its sensitivity will vary according to the particular region of mtDNA analysed. The control region sequence in mtDNA, for example, is usually much more variable than most other parts of the molecule. The question is, therefore, whether molecular ecology is of any use at all in this aspect of conservation biology. It is surely naive to try and derive a single formula for the universal application of molecular genetic data to conservation issues, but that is not to say the molecular approach is without merit.

One example of this dilemma is the koala bear *Phascolarctus cinereus* of eastern Australia. Koalas are among the most well known of Australia’s wildlife, and although they are still quite widespread with robust populations in some parts of the country, there have been dramatic local declines. Certainly the koala’s biogeographical range (Fig. 3.1) is much more fragmented than in historical

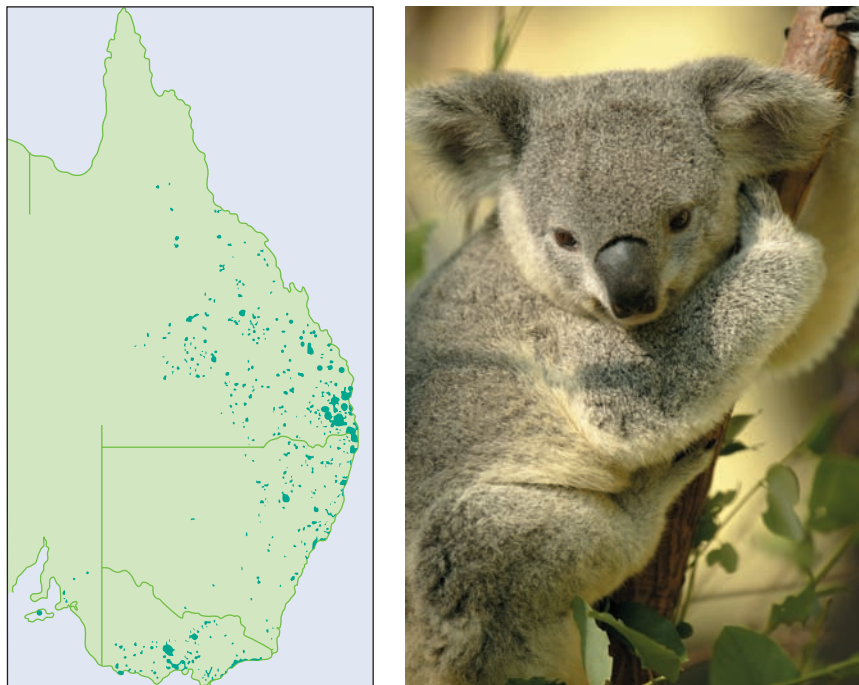


Figure 3.1 Recent (1990s) distribution of koala bears in Australia with picture of koala bear (right). Spots indicate recent distribution records. The total extent of range (north-south) covers approximately 2000 km.

times, mostly due to habitat destruction. It is becoming increasingly important to establish whether koalas are essentially homogeneous as a group, or whether forms exist below the species level that should attract individual conservation priority. Koalas have previously been differentiated on morphological grounds, mostly variations in size and colour, into three subspecies. Should these subspecies be identified as discrete taxonomic units to be maintained individually and in isolation from each other? Molecular studies tell a different story. The first stage in any molecular genetic study involves tissue sampling and extraction of protein or DNA, as outlined in Box 3.1. A battery of molecular genetic markers including multilocus minisatellites, mtDNA, random amplified polymorphic DNAs (RAPDs), and microsatellites failed to give patterns of population subdivision consistent with the nominal subspecies. Rather they implied strong differentiation between northern and southernmost populations that were probably best explained as a continuous cline of progressive, small changes between adjacent populations (Sherwin *et al.* 2000). How, then, can the morphological and molecular observations best be reconciled? Unfortunately such discord is not uncommon and always needs careful interpretation. In some situations, morphological variation arises from environmental rather than genetic causes and is therefore not heritable. Strong selection acting on genes responsible for the morphological differences might also be important, and could give a quite different pattern of genetic variation from that shown by the neutral markers used in this and most other comparable studies. However, it seems likely that in this case the molecular data have highlighted limitations in the original subspecies designations, based as they were on small numbers of

**BOX
3.1****Protein and DNA extraction from tissue samples**

A wide range of tissues are potentially suitable for genetic analyses, including leaves, flowers, roots, shoots, blood, skin, liver, kidney and so on. Wherever possible, of course, non-destructive sampling that leaves the organism alive and healthy should be employed. For protein analyses, tissue consistency (e.g. always using blood) is important because many proteins are expressed in tissue-specific patterns. DNA, on the other hand, will be present in most or all tissues of an organism so sampling regimes can be more flexible. Proteins are normally extracted by homogenizing tissue samples in buffered saline solutions followed by low-speed centrifugation to remove cell debris. The supernatants can then be stored at -80°C as sources of material for allozyme or other analyses (May 1992). DNA from animal tissues can be obtained by several methods. Extraction with organic solvents (phenol–chloroform mixes) to remove

protein followed by precipitation with ethanol is efficient but laborious with multiple samples. Several commercial companies market kits for tissue DNA extractions, mostly based on selective adsorption of DNA onto silica filters. These kits generally yield good-quality DNA, but tend to be expensive. Simpler alternatives include incubating tissues with Chelex resin to liberate DNA and inhibit nucleases, followed by low speed centrifugation and direct use of the DNA-containing supernatant (e.g. Walsh *et al.* 1991). Plant tissues pose extra problems because they often contain substantial quantities of polysaccharides such as cellulose, and polyphenols (Milligan 1992). These interfere with PCR assays and should be removed using specialized purification kits available from several biotechnology companies.

● KEY POINTS

Ultimately, molecular markers can differentiate at every level down to individual organism. Molecular information must therefore be interpreted with this in mind.

Molecular measures of population structure do not always coincide with morphological criteria.

● KEY POINT

Careful combination of molecular and other data (morphological or behavioural) is a powerful approach for deciding what are meaningful taxonomic units.

individuals. Conservation strategies should be devised which take account of this new assessment of koala genetic variation, which essentially infers that koalas should be treated as a single taxonomic unit. Thus rather than focusing attention on the maintenance of specific isolated groups it will make more sense to maintain habitat continuity over the whole range, and thus retain a clinal genetic structure. This will have the benefit of reducing long-term risks from genetic drift and inbreeding that can bedevil isolated subpopulations (see Chapters 5 and 8).

Molecular studies also provide interesting surprises of the opposite kind to that revealed in koalas, by confirming the importance of morphological or behavioural differences that were previously disregarded. The herald petrel *Pterodroma heraldica*, a wide-ranging sea bird of the Pacific Ocean, is a case in point. These petrels, in common with many other seabirds, nest in large colonies on islands relatively lacking in predators of eggs or chicks. Although distinct colour varieties of herald petrels, dark and light, have long been known it was thought until recently that this morphological variation had no special significance. However, detailed behavioural studies on one of the Pitcairn Islands in the central Pacific indicated that the two types of birds mated assortatively. Dark birds consorted with dark, and light with light, but rarely if ever were the two forms found together. Did this mean they were distinct taxa? Sequence analysis of part of the mtDNA cytochrome *b* gene from large numbers of birds of both colour morphs revealed five haplotypes unique to the light coloured birds (A, E–H) and a further three confined to the dark variety (B–D), with no haplotypes common to both (Fig. 3.2). Thus the mtDNA study confirmed that in this case we really have two reproductively isolated populations,

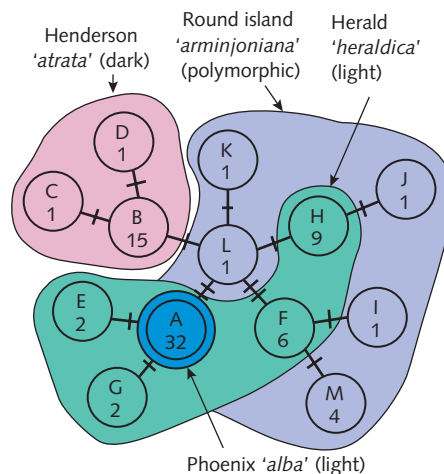


Figure 3.2 Unrooted mtDNA haplotype network of *Pterodroma* species (after Brooke and Rowe). *Pterodroma heraldica* and *P. atrata* are from the Pitcairn Islands, whereas *P. arminjoniana* and *P. alba* are from other islands. Letters refer to specific haplotypes, and numbers to the quantity of individuals with each haplotype. Cross-bars on connecting lines represent the number of mutational changes between the haplotypes.

with no maternal gene flow between them. Indeed, the dark birds have now been ascribed to a new species, the Henderson petrel *Pterodroma atrata* (Brooke and Rowe 1996). This example demonstrates how evidence for taxonomic subdivision based on a combination of molecular and morphological observations can be particularly persuasive. Many computer programs are available for aligning multiple DNA sequences and detecting differences among them (see the Useful software list at the end of this chapter).

Hybrids

Hybrid organisms are those resulting from the interbreeding of two separate species. While taxonomic decisions about group identification can be difficult at the best of times, significant levels of hybridization between group members can confound the issue still further. Organisms differ widely in their propensity to hybridize, though in general it happens more often in plants than in animals. As occasional rarities, hybrids in most situations may not matter much, but there are some interesting exceptions where hybrids are common and contribute significantly to communities. One such situation sometimes arises when two closely related taxa are separated for prolonged periods, as can occur during Ice Ages when they are forced to retreat into separate refugia in warmer latitudes, but then meet again after subsequent range expansions. Hybrid zones may then form at the distribution interface as regions where hybrid individuals are consistently common. Some of these zones are stable over long time periods, and are maintained by natural selection acting on habitat preferences, hybrid inviability, or both. Such regions are of interest to evolutionists and ecologists alike, and have been described as natural experiments because, most unusually, it is possible to measure the ongoing effects of selection in the field. An example of hybridization between two amphibian species is described in Box 3.2.

Ideally a study of hybrid zones would include separate analyses of maternal and paternal gene introgressions. This would provide information about how hybridization was occurring, and in particular whether the sexes differed in their contributions to introgression. Conifer forests are therefore particularly interesting places for molecular studies of hybrid formation, because gymnosperms (pine trees) commonly show paternal rather than maternal inheritance of chloroplast DNA (cpDNA). MtDNA is maternally transmitted in the usual way, which means that where hybrids are suspected it is possible to look separately at introgression of male and female genes, and at the direction of successful crosses. Extensive subalpine coniferous forests occur in Japan, with vertical separation of four major species of tall pines (genus *Abies*) on the mountain slopes of Honshu island. Being closely related, it is interesting to enquire whether hybridization occurs in the overlap zones that are sometimes extensive on the mountain slopes. Using a combination of mtDNA, cpDNA, and RAPD markers Isoda *et al.* (2000) showed that occasional hybrid saplings occurred at one site where two of these trees, *Abies veitchii* and *Abies homolepis*, overlapped at around 1900 metres above sea level. In this study, sex-specific

● KEY POINT

The full geographical extent of stable hybrid zones can be identified using molecular methods.

BOX
3.2

Hybridization in toads

One of the best-studied animal hybrid zones is that between two small toads, *Bombina bombina* (Fig. 3.3) and *Bombina variegata* (Fig. 3.4), in eastern Europe. These amphibians last shared a common ancestor about four million years ago, and probably survived in separate refugia during the

recent Pleistocene glaciation. Although closely related and broadly similar in shape and size, they differ morphologically in a number of distinctive ways including belly colour and spot size, skin thickness, and robustness of skeleton. They also have ecological differences. *Bombina bombina*



Figure 3.3 Toad *Bombina bombina*. Courtesy of Horia Bogden.



Figure 3.4 Toad *Bombina variegata*. Courtesy of Horia Bogden.

(continued overleaf)

is a lowland animal inhabiting shallow but relatively permanent pools, whereas *B. variegata* prefers very shallow upland pools. *Bombina bombina* lays smaller eggs than *B. variegata*, and larval development time is longer in *B. bombina*. Male vocalizations are distinct, but selection has not prevented significant amounts of interspecific reproduction where the two species meet, despite the fact that first generation hybrids show significantly reduced viability relative to the parental types (Nurnberger *et al.* 1995).

The two forms currently hybridize in extensive contact zones in eastern Europe. Given that hybridization is occurring all the time, and presumably has done so for millennia, a question arises about how much genetic exchange occurs between populations of the parent species beyond the hybrid zone. Hybrids in the middle of a zone may be quite easy to identify, but backcrossing of hybrids with the parental species is certain to make this job progressively

more difficult towards or beyond the apparent zone margins. In the case of the *Bombina* toads, allozyme studies showed that there was extensive linkage disequilibrium in the centre of a continuous (clinal) but narrow hybridization zone about 20 km across, and which spanned an altitudinal gradient where *B. bombina* occurred in the lowlands and *B. variegata* in the highlands (Szymora and Barton 1986). However, there was also evidence of gene transmission on both sides of the area well beyond where hybrids could be readily identified (Fig. 3.5). Five allozyme loci were found in which different alleles were always fixed in each of the parental populations distant from the hybrid zone, but which could be detected across a much wider area (a so-called 'introgression zone') at progressively decreasing frequency well into the range of the complementary species. The width of this introgression zone was some 200 km, much greater than the 20 km hybrid zone identified by

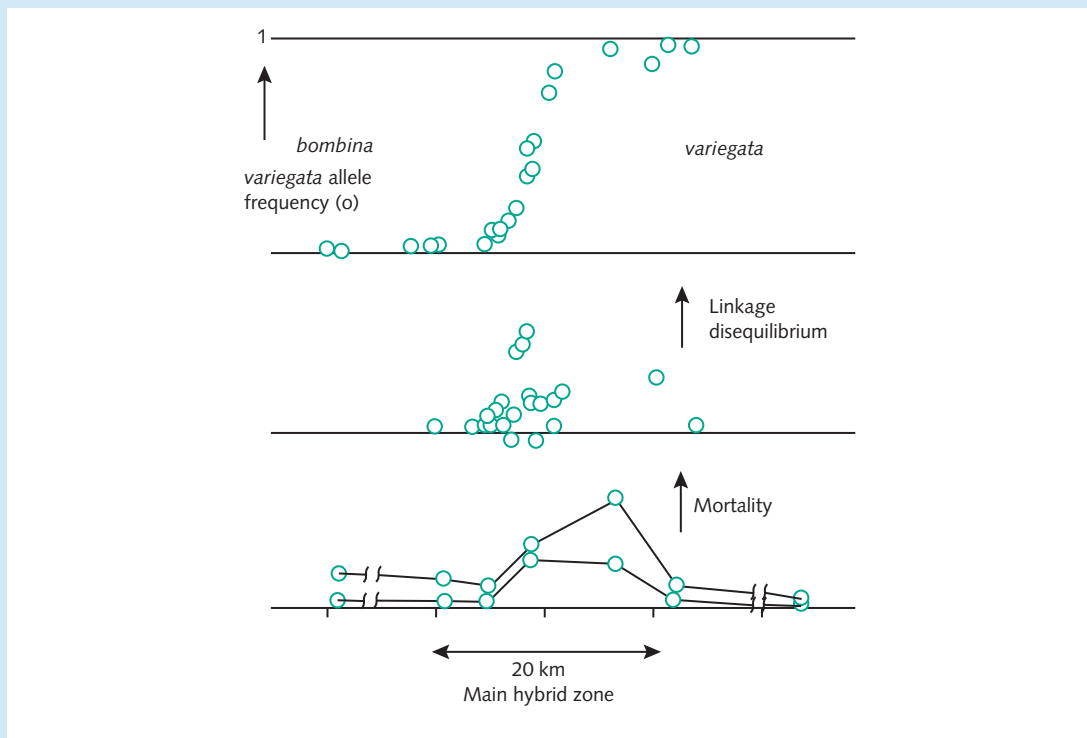


Figure 3.5 Changes in gene frequency, linkage disequilibrium, and embryonic mortality across a toad hybrid zone in Poland (after Szymora and Barton 1986). Introgression extends about 100 km on each side of the main hybrid zone.

(continued overleaf)

morphology. Work with mtDNA, using restriction fragment length polymorphism (RFLP) analysis of the whole molecule, showed that although most natural hybrids were formed as a result of male *B. variegata* x female *B. bombina* pairs, there was no evidence of *B. bombina* mtDNA beyond the hybrid zone in *B. variegata* populations. Males are the heterogametic sex in *Bombina*, so it is unlikely in this case that the consequences of Haldane's rule (the heterogametic sex being most prone to infertility or inviability in hybrids) explain the lack of mtDNA introgression. Strongly co-adapted gene complexes in the *Bombinas* seemed to have provided a firm but slightly leaky (at least with respect to nuclear genes) barrier to extensive introgression beyond the main hybrid zone.

In a different and much broader region where the two species meet there was no altitudinal cline and no simple, linear cline of hybrid frequency across the district (Vines *et al.* 2003). In this case there was however a clear association of habitat type and microsatellite allele frequencies diagnostic for *variegata* or *bombina* within the zone. The genetic data implied substantial migration of individuals between habitats, especially from deeper ponds with pure *B. bombina* into shallower pools occupied by *B. variegata*. This pattern of gene flow predicts an eventual breakdown of linkages between neutral loci and loci under selection, and a loss of differentiation as measured by neutral markers such as microsatellites. Another important question, however, relates to how exactly the toads interact with each other. Animals in the hybrid zone often show heterozygote deficiency relative to Hardy–Weinberg expectations. Is this due to selective mating among the different genotypes, hybrid inviability, or both? Genotypic data from a suite of six neutral markers (one microsatellite, one simple length polymorphism and four single-stranded conformational polymorphisms) were used to investigate mating patterns among the *Bombinas* (Nurnberger *et al.* 2005). In one analysis, the probabilities of the observed offspring genotypes arising from the available adult (putative parental)

genotypes were estimated on the basis of explicit models of mate choice. In a second approach, associations among parents were estimated directly from a set of full sibship genotypes. Both methods gave concordant results, and suggested that mating was completely random. Heterozygote deficiency in hybrid zone animals is probably due entirely to the reduced viability of many F1 clutches. Molecular methods have therefore offered a powerful approach to defining the extent of hybrid zones as well as understanding the genetic processes occurring within them.

Essay topic

Describe the morphological and ecological differences between *Bombina bombina* and *B. variegata*. Discuss how the use of molecular markers has demonstrated the genetic consequences of hybridization between these species in areas where they are sympatric.

Lead references

- Szymora, J. M. and Barton, N. H. (1986) Genetic analysis of a hybrid zone between fire-bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in Southern Poland. *Evolution*, **40**, 1141–1159.
- Szymura, J. M., Uzzell, T. and Spolsky, C. (2000) Mitochondrial DNA variation in the hybridizing fire-bellied toads, *Bombina bombina* and *B. variegata*. *Molecular Ecology*, **9**, 891–899.
- Vines, T. H., Kohler, S. C., Thiel, M., Ghira, I., Sands, T. R., MacCallum, C. J., Barton, N. H. and Nurnberger, B. (2003) The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution*, **57**, 1876–1888.
- Nurnberger, B., Barton, N. H., Kruuk, L. E. B. and Vines, T. H. (2005) Mating patterns in a hybrid zone of fire-bellied toads (*Bombina*): inferences from adult and full-sib genotypes. *Heredity*, **94**, 247–257.

investigation was based on mitochondrial *nad1* gene intron 2, and chloroplast intergenic (*trnL* and *trnF* gene) spacer sequences. Multiple samples of these DNA sequences, obtained by polymerase chain reaction (PCR) amplifications, were compared using single-strand conformation polymorphism (SSCP) analysis. Discord within an individual between mtDNA and cpDNA haplotypes as

expected from the parental species was used to indicate a hybrid. Interestingly, the hybrids that were found could resemble either parent morphologically but were invariably derived from male *A. homolepis* pollen fertilizing *A. veitchii* mother trees. It remains to be seen whether extensive hybrid zones occur in these forests, but the lack of adult hybrids may suggest a low viability of this particular cross. Of course, in some organisms (such as mammals) there is the potential to use Y-chromosome markers to track male-specific gene flow and contrast it with maternal mtDNA.

Unstable hybrid zones, which vary in their geographical position and do not persist for long time periods, also exist and are interesting for different reasons. In many cases they have arisen recently due to human movements of species around the world, permitting crosses between taxa that would not otherwise have ever met. These not infrequently develop into substantive problems for biodiversity conservation, sometimes to the point where the newcomer colonizes and hybridizes so successfully that the native form becomes threatened with extinction. Invasion of the American west coast by smooth cord grass *Spartina alterniflora*, a saltmarsh plant native only to the east coast, is a worrying example (Ayres *et al.* 1999). *Spartina alterniflora* was introduced into San Francisco Bay during the 1970s. In the intervening years it has spread extensively at the expense of the western native *Spartina foliosa*. Recent RAPD analysis of plants in San Francisco Bay has confirmed very extensive hybridization, to the extent that some areas of the bay now contain only *S. alterniflora* and hybrids (Fig. 3.6). Ninety-six 10-mer primers were screened in this exercise, yielding 10 useful ones which in turn generated 10 species-specific bands (five for each parental type). Pure *S. foliosa* was thus scored as 0 per cent *S. alterniflora*, and nine hybrid categories (from 10 to 90 per cent depending

KEY POINT

Sexual bias in hybrid formation can be investigated in gymnosperms because these trees harbour male- and female-specific genetic markers.

KEY POINT

Unstable hybrid zones can occur following introduction of alien species, and their spread can be monitored by molecular markers.

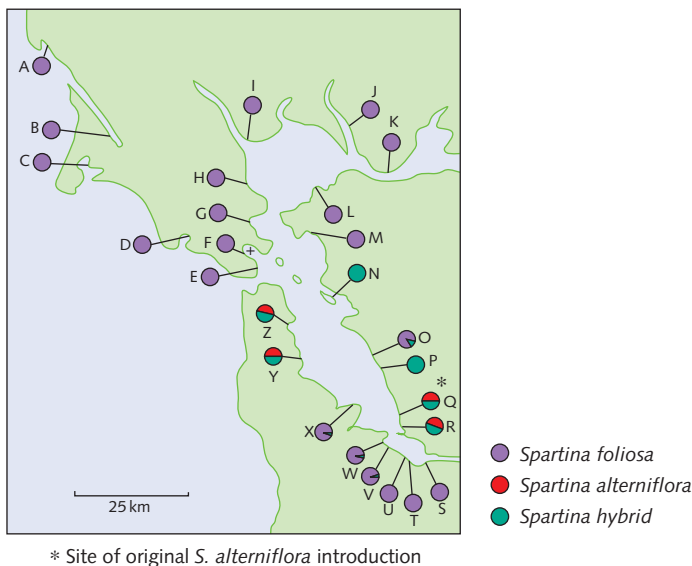


Figure 3.6 Distribution of *Spartina* (cord grass) species and hybrids around San Francisco Bay in 1997–8 (after Ayres *et al.* 1999). Letters refer to sampling sites.

on the number of *S. alterniflora* bands present) could be identified. Many of these hybrids were morphologically indistinguishable from one or other of the parental types, so the RAPD analysis provides a much more sensitive assessment of the extent of the problem than would otherwise be possible. *Spartina alterniflora* is competitively dominant over *S. foliosa* and also colonizes a much wider area of saltmarsh habitat. It therefore poses a threat not only to *S. foliosa*, but also to the structure of the entire ecosystem. In this case molecular monitoring should be useful in the early detection of hybrids elsewhere on the Californian coast, should this pernicious weed begin to spread further afield. This in turn should improve the chances of local eradication before the alien or hybrids become established at new sites.

In some situations hybrids can be investigated at the cytological level, by examining chromosome patterns (karyotypes) where these differ in the parental forms. Common shrews *Sorex araneus* in Europe occur as several chromosomal races with distinctly different karyotypes. In Poland, two such races meet and hybridize. Surprisingly, however, the distinctive karyotypic races were not strongly differentiated from each other at the level of specific genetic loci (seven nuclear microsatellites), suggesting that quite dramatic differences in chromosome structure patterns do not always preclude successful hybridization (Jadwiszczak *et al.* 2006). This kind of result, which has also been found in some other studies, cautions against the simple use of different karyotypes as indicators of reproductive barriers.

Dealing with individuals

Basic identification

It is impossible to know how many good research ideas in ecology have been shelved for the apparently trivial reason that identification of cryptic taxa posed an insuperable practical problem, but the issue is real enough. Fortunately the advent of molecular markers has already made an impact in this area, and undoubtedly will continue to do so as the methodology becomes progressively cheaper and less daunting to perform. There are innumerable published examples of molecular identifications, across a huge range of taxonomic groups, often using relatively simple methods such as RAPD analysis to generate species-specific banding profiles.

Competition between toads and the problem of indistinguishable larvae

Competition processes are of widespread interest in ecology, but competition is often strongest between closely related taxa, which in turn can lead to identification problems. Just such a difficulty arose with studies of competition

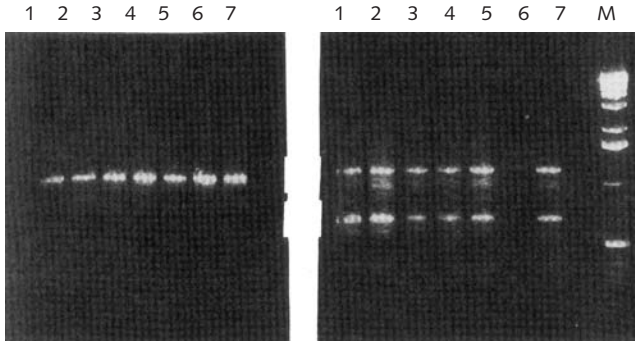


Figure 3.7 RAPD identification of larvae of two toad species. *Bufo calamita* (left) samples 1–7, and *Bufo bufo* (right), samples 1–7 with No. 6 missing. M, size markers. After Bardsley *et al.* (1998).

between common and natterjack toads (*Bufo bufo* and *Bufo calamita*) in Britain. Natterjacks are competitively inferior to common toads, meaning that the latter species thrives better than the former in habitats where the two coexist. Such habitats are relatively rare in pristine environments, but have become more common in some places following land use changes. This in turn has intensified competition to the extent that some populations of natterjack toads, previously rare in any case, have been exterminated as a result. Adult toads are easy to identify, but larvae are virtually indistinguishable, being uniformly black, and larvae commonly occur as mixed communities in ponds. It is at the larval stage that competition is manifest so an ability to identify larvae, and thus assess competition strength, is clearly essential to carry out research on this issue. Fortunately either RAPD analysis or protein profiling, using small tissue samples from the tadpole tailfins, provides a highly reliable identification guide (Fig. 3.7). A single 10 base pair (bp) oligonucleotide primer used in a RAPD PCR identifies the two toads unambiguously. This has made possible a quantitative analysis of the extent of competition, and the threat it poses, in a series of natural ponds in northern England (Bardsley and Beebee 1998). At several sites few or no natterjack tadpoles survived to metamorphosis in the presence of common toad larvae. Habitat management designed to reduce interspecific competition between these amphibians can now be attempted, secure in the knowledge that its effects can be assessed quickly and accurately in the tadpole communities.

Speeding up identification: the use of SNP-based probes

The evident power of molecular markers to identify cryptic species is partly compromised by the time involved. This issue was addressed by Itoi *et al.* (2005) in their development of a novel procedure for distinguishing two species of eels, *Anguilla japonica* and *A. anguilla*. In the first instance, a short (79 bp) section of the mtDNA 16S rRNA gene was identified that contained a single and completely reproducible nucleotide difference (i.e. a SNP) between the two species. The 16S rDNA segment from samples of the two species was then

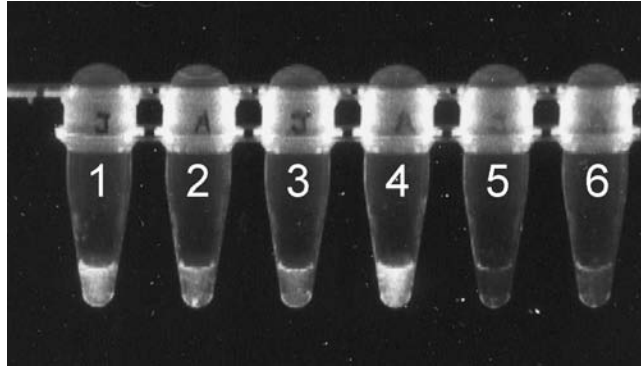


Figure 3.8 Identification of PCR products from two eel species by visual inspection. Only tubes 1–4 fluoresced and therefore derived from the species under test.

● **KEY POINT**

Molecular identification of individuals can make possible new kinds of study in ecology.

amplified in PCRs with conserved primers, but also with short (21 bp) probes including the SNP region and with the fluorophore FAM and quencher MGB at their 5' and 3' ends respectively. In PCRs with the fully complementary probe, the 5'-exonuclease action of *Taq* DNA polymerase degraded the transiently hybridized probe, generating a fluorescence signal (because the fluorophore was liberated from close proximity to the quencher), essentially the same procedure used in RT-PCR analysis. When the probe with the mismatch was present, none of this happened (because it did not form a stable, transient hybrid) and there was no consequent fluorescence. However, in this analysis it was not necessary to use a specialized RT-PCR machine because the relative levels of fluorescence in pairs of assay tubes with the two probes could be visualized directly over a suitable light source (Fig. 3.8). The entire procedure took less than an hour for 192 samples (two PCRs each) in a 384-well plate. Use of different dyes on the probes could allow single PCRs for each sample. Developments of this kind offer the prospects of rapid and reliable molecular identifications of many different species in future, though of course there will usually be a preliminary requirement for identifying diagnostic SNPs, with all the time and effort that entails.

Identification of prey in predator guts

Molecular identification can also assist with one of the most intractable practical problems in ecology, notably the study of predator–prey interactions. Particularly with small organisms, identifying prey and especially prey range can be a daunting task. Mostly this has to be done by the manual examination of gut contents. Unfortunately digestive processes rapidly destroy many morphological clues to prey identity, but molecules or fragments of them may persist for longer. In studies on invertebrates, monoclonal antibodies have been used successfully to identify specific items in gut contents, but monoclonal development is expensive and takes a long time. While practicable for detecting

specific prey species, it is much less so for screening a full range of possible victims. DNA-based methods offer a useful alternative, and have recently been employed to detect multiple potential prey items simultaneously. Harper *et al.* (2005) developed a system in which more than ten invertebrates (potential prey of the ground beetle *Pterostichus melanarius*) could be identified using molecular markers. Taxon-specific primers were used to amplify short (< 250 bp) sections of mtDNA 12S ribosomal RNA genes from earthworms and molluscs, and cytochrome oxidase (*COI*) genes from aphids and beetles, in multiplex PCRs. MtDNA has the advantage in this kind of study of high copy number, thus maximizing the chances of detection by PCR before its complete destruction by digestive enzymes. Mostly the reactions generated species-specific fragments that differed in size and could be identified, when fluorescently labelled, using an automated DNA sequencer. For earthworms, however, it only proved possible to create group-specific primers because of large (overlapping) intra-specific variation of amplicon sizes in these animals. Nevertheless, feeding trials in the laboratory demonstrated that all these prey items could be identified reliably in the guts of the predatory beetles with half-lives relative to the time after consumption of between 9.7 and 88.5 hours, depending on species. Beetles caught in the field were also analysed successfully, with prey taxa identified in 80 per cent of them. Of course many challenges remain, for example with respect to converting presence/absence data into reliable estimates of how many of each prey species are consumed. The different half-lives of prey DNAs contribute to this problem, but so also does the issue of secondary predation. Some species detected by sensitive PCR amplifications may be scavenged, or the prey of prey consumed by the predator under investigation, rather than being directly consumed by the predator itself. Sheppard *et al.* (2005) explored an extreme case in which the danger of confusing primary and secondary predation was particularly acute. The ground beetle *P. melanarius* was again used, this time in comparison with another predator, the spider *Tenuiphantes tenuis*, in trials where both were fed with the aphid *Sitobion avenae* and where, in some experiments, the beetle was fed with spiders that had previously eaten aphids. Specific PCR amplification of 110 and 245 bp fragments of aphid *COI* DNA showed that secondary predation was readily detectable for up to eight hours after beetles fed on spiders that had previously consumed aphids. This is a particularly problematic case because spiders can reduce their metabolic rates between feeds, and prey DNA remained detectable in them much longer than it did in beetle guts (Fig. 3.9). Many other predator–prey systems may be less susceptible to the detection of secondary predation, and thus to errors in food chain determination, but clearly there is a need for caution and for further control studies in work of this kind. Despite these caveats, issues of uncertainty inevitably arise during the development of new methods and PCR-based prey detection will undoubtedly become an important tool in future investigations of food web structures.

● KEY POINT

Even gut contents can sometimes be identified using DNA-based methods.

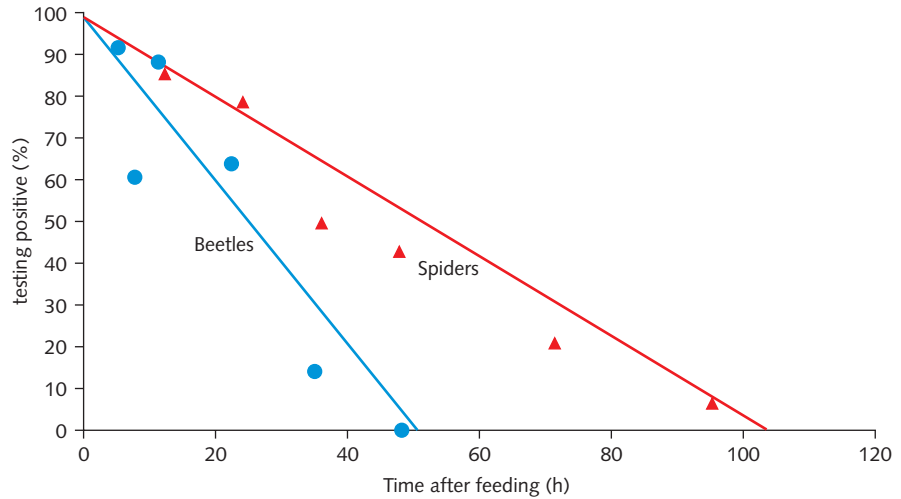


Figure 3.9 Different survival times of aphid mtDNA in spider and beetle guts.

● KEY POINT

Molecular identifications have many forensic applications.

Forensic investigations

In a quite different context, RAPD analysis has been successfully employed to identify fly maggots found in and under a human corpse (Benecke 1998). Such information can be very valuable both in establishing the time elapsed since death occurred (because different insects use corpses at different stages of decay), and in showing whether bodies have been moved prior to discovery. The present work was of particular interest because it rigorously investigated the reliability of the RAPD procedure, a method notorious for difficulties in reproducibility between laboratories. This issue is of course of paramount importance in forensics. In this case it was possible to show that high levels of reproducibility were achievable with RAPDs provided that template DNA concentration was maintained between certain (20–60 ng) limits. PCR ‘beads’, dehydrated reaction components including buffers, *Taq* polymerase and deoxyribonucleotides supplied ready for use in PCR tubes, were also found to contribute to reproducibility. Variation in the thermocycler machine used did not affect the results. RAPD phenotypes (banding patterns) were quantified in this case on an automated DNA sequencer to permit very detailed comparisons between samples. Maggots on the body were in this instance identified as green bottle blow flies (*Lucilia* species), whereas pupae found under the corpse were of a different unidentified species. Blue bottle blow flies (*Calliphora erythrocephala*) from another case also yielded a distinct and different RAPD profile.

There have also been forensic applications in plant biology. In one litigation, unauthorized commercial use of a patented variety of strawberry (Marmolada®) was challenged partly on the basis of molecular data. This particular strain has many advantages, including high productivity, and cold and mould resistance. Farmers suspected of growing Marmolada® unlawfully were challenged and it was shown by a combination of morphological and RAPD data that 13 out of 31 samples were indeed Marmolada®. Intriguingly, before carrying out the

test the 13 plants taken from the illicit growing areas had been mixed, unknown to the testers, with 18 plants of different varieties. The RAPD analysis, using six primers, therefore unambiguously identified the commercial strain with 100 per cent efficiency. The court was convinced, and the farmers lost the case (Congiu *et al.* 2000).

DNA barcoding: towards an inventory of life

Molecular markers offer a wide range of options for identifying species, but the question arises as to whether a general, universally applicable method might be found. If this proved possible it would make a substantial contribution to the Global Taxonomic Initiative of the Convention on Biological Diversity (<http://www.biodiv.org>) and to the Global Biodiversity Information Facility (<http://www.gbif.org>). So-called DNA-barcoding could provide just such an approach. The idea here is to select one or a few genes that are shared by most, if not all organisms on earth and which show large interspecific but small intraspecific levels of variation. The sequences of such genes could then become the equivalent of species-specific barcodes, and in the not too distant future provide a data set representative of earth's biodiversity. Optimists suggest that hand-held, portable DNA sequencers will be available within a few years and allow identification of species in the field even by people with little or no training in taxonomy (Savolainen *et al.* 2005). Indeed, not only would such an approach permit identification of already known species but it should also lead to the discovery of new, previously unrecognized ones as barcodes not already in the database come to light.

This all sounds too good to be true, and it probably is. Early work focused on the mitochondrial cytochrome oxidase subunit 1 (*cox1*, usually referred to as COI in barcoding studies). This gene has many desirable properties. It is required by all aerobic organisms, and being mitochondrial it is usually present in high copy number per cell. A partial COI sequence of about 650 bp has turned out to have, in a wide variety of organisms from insects to birds, high interspecific but low intraspecific variation. However, there are some significant difficulties with this locus. Anaerobic organisms are excluded, and interspecific variation of mtDNA (including COI) in plants other than algae is often too low to be useful. Prokaryotes, which include most of the earth's biodiversity (see Chapter 9) are essentially excluded. There is also a risk of errors with any single locus from lineage sorting in recently diverged species where reciprocal monophyly has not yet been achieved. It looks as if at the very least barcoding will have to include sequences from several different genes to be universally applicable. In the case of plants, plastid loci such as ribulose biphosphate carboxylase (*rbcL*) and nuclear ribosomal DNA intragenic spacer (ITS) regions have been proposed.

A study of North American mayflies shows the potential value of DNA barcoding for a large group of insects in which identification of adults by morphology requires substantial training, and for which larval identification can be highly

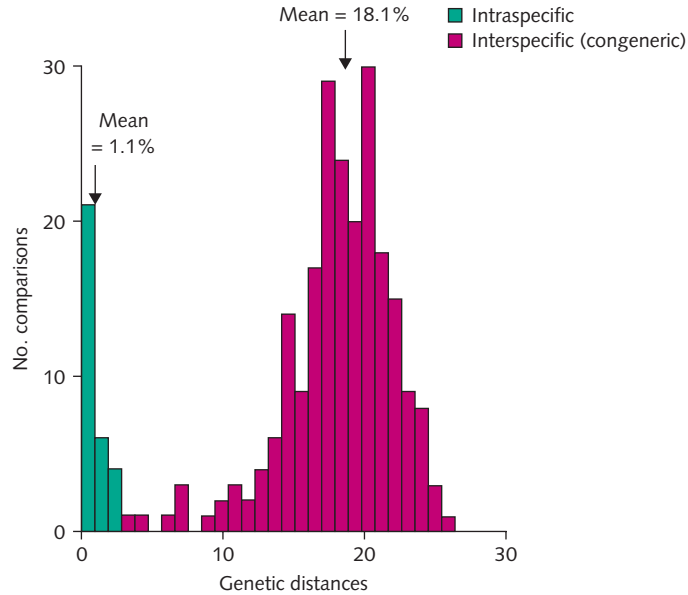


Figure 3.10 Inter- and intraspecific differences in COI sequences among mayflies. Genetic distances are based on the extent of nucleotide sequence divergence.

● KEY POINT

DNA barcoding may provide a way of documenting all the world's biodiversity.

problematic (Ball and Hebert 2005). These flies are widely used in the biomonitoring of aquatic habitats, where larvae are often more easily obtained than adults. A reference data set of 630 bp sequences of COI from single specimens of 80 species has been generated. Seventy new individuals from 32 of these species were then barcoded, and correctly identified in 69 cases. The single misidentified specimen might, it was later realized, be an unrecognized cryptic species. The average intraspecific sequence divergence was around 1 per cent, whereas mean differences among congeneric species was about 18 per cent (Fig. 3.10). Only in one case was there an overlap between intra and interspecific divergence levels. Certainly for this group of insects COI barcoding offers the prospect of much more reliable identification than morphology, especially for early instar larvae, damaged specimens or fragments of specimens. DNA microarrays using COI sequences may soon offer the prospect of multiple, simultaneous identifications for entire samples from ponds, lakes and rivers. Even so, only time will tell whether DNA barcoding, now widely proven as a useful method within several particular taxonomic groups, will justify the rather high ambitions of its proponents in a universal context.

Sex

Situations arise in which it is important to know the sex of an individual but where this cannot be determined just by looking. This is most obviously an issue with immature animals in which secondary sexual characters have not developed. Yet there are many reasons for wanting to find out about sex,

whether the study be of population structure, behaviour, or overall distribution. The discovery of the testis-determining *SRY* gene on the mammalian Y chromosome proved a major breakthrough in this regard. *SRY* genes are male-specific in mammals and can be detected by nucleic acid hybridization or, more often these days, by PCR amplification using specific primers. Because failure to amplify for technical reasons could generate ‘false females’ it is always essential to have an appropriate control, usually another nuclear gene such as actin.

Cetaceans are of particular interest on account of their complex social behaviour and, in some cases, because they are critically endangered. However, in encounters where they can only be viewed from a distance out at sea, it can be difficult to identify their sex even as adults. PCR amplification of a 147 bp fragment of the *SRY* gene from tissue samples taken by harpoon biopsy from live northern bottlenose whales (*Hyperoodon ampullatus*, Fig. 3.11), using primers developed for the sperm whale *SRY* gene, has provided an accurate sexing procedure for these animals. The method was verified by comparison with photographs taken at the same time, and with material taken from dead whales of known sex (Gowans *et al.* 2000). This approach is particularly valuable for immature individuals that cannot be sexed in any other way.

KEY POINT

Sex-specific genes permit sex identification in situations where morphology cannot be used.



Figure 3.11 Male (above) and female (below) bottle-nosed whales. Photos courtesy of Shannon Gowans, the Whitehead Laboratory.

Unfortunately sex-specific genes are not widely conserved across the broad taxonomic spectrum of animals and plants, though this is not particularly surprising because sex determination is also not conserved. In lizards *Calotes versicolor*, for example, males have an *SRY* gene but so do 50 per cent of females! In birds, as in some other groups such as moths, it is the females that are the heterogametic sex. In the case of birds other than some flightless species (ratites) suitable markers for sex determination now exist, notably the *chromobox-helicase-DNA-binding* gene localized on the W (female-specific) chromosome (*CHD1W*), albeit with a homologue (*CHD1Z*) on the Z chromosome. W and Z chromosomes in species where females are the heterogametic sex are the equivalents of Y and X in others, such as mammals, where males are heterogametic. Fortunately PCR primers can be designed to amplify the intron of the *CHD1* gene and generate fragments which consistently differ in size between the W and Z versions (Griffiths *et al.* 1998). The method has been shown to work in 50 species from 11 different orders, providing a very valuable tool (Fridolfsson and Ellegren 1999). In plants, too, molecular sex determination is now possible. Only the unfertilized flowers of female hops (*Humulus lupulus*) are required for beer production, generating a commercial incentive for the development of early sex identification so that only female plants need be grown in quantity. In this case a RAPD ‘shotgun’ approach, screening 900 primers, ultimately yielded a single sequence from which primers reliably generate a Y- (male-) specific PCR product (Polley *et al.* 1997). In principle this approach to obtaining sex-specific markers seems likely to work with almost any organism, given the availability of sufficient time and money.

● KEY POINT

Sex-specific molecular markers are not universal and need to be developed independently for different classes of organisms.

Bits of individuals: non-invasive sampling

Quite a number of organisms leave signs of their passing, even when rare or for other reasons difficult to find. Skin, fur, and feathers, for example, are commonly available in the field from reptiles, mammals, and birds respectively. Excrement, too, is sometimes easy to locate especially from species that are inclined to mark their territories with it. All of these materials are potential sources of DNA, and thus of individual identification, given the exquisite sensitivity of the PCR and thus the ability to make do with very tiny amounts of tissue. This leads, however, to a very important methodological point (Box 3.3). Starting with small quantities makes PCR susceptible to contamination, particularly if (as is usually the case) the laboratory is routinely working with the species in question and there are likely to be other tissue or DNA samples close by. It is therefore essential to take special precautions when the starting DNA concentration is very low (sub-ng). Despite the technical difficulties, there have been some fascinating and very valuable studies based almost entirely on sampling that makes no contact with the animals under investigation.

● KEY POINT

Skin, fur, feathers, and faeces are all useful sources of DNA.

**BOX
3.3****Special circumstances: museum material and non-invasive sampling**

One of the great benefits of PCR-based methods is the ability to obtain information from very old or very tiny tissue samples that may, in extreme cases, contain just a few DNA molecules. It is fortunate that many museum specimens are stored in ethanol, an ideal preservative for DNA, and tissue samples from such specimens can be used to provide historical genetic data for comparison with extant populations. Also, non-invasive sampling methods can extract usable DNA from such unlikely sources as hair, feathers, scales, shed skin, faeces, and urine. These techniques are of immense value when studying species that are difficult to locate in the field, but which leave traces that can be found relatively easily. There are potential risks, however, which arise because the samples contain so little DNA (Taberlet and Luikart 1999). The slightest contamination of PCR assays with DNA from other sources can be enough to generate products from the contaminant rather than the intended sample. Primers for PCR assays may in some cases not be species-specific, in which case human or other DNA can be amplified (e.g. from tiny skin flakes) in the reaction rather than the intended target. Even when primers are species-specific, there is a risk of contamination if there has been extensive previous PCR of DNA from the same species in a laboratory. There have been well-publicized examples of mistakes arising from PCR

contamination, especially with the use of apparently 'ancient' DNA that has turned out to be embarrassingly recent. A second problem with trace amounts of DNA is 'allelic dropout', in which only one allele of a heterozygote becomes amplified and thus appears to be a homozygote. Even 'false alleles' can sometimes be generated by the PCR when starting from tiny traces of DNA. To address the contamination issue it is necessary to take elaborate precautions. Taberlet and Luikart (1999) provide valuable guidelines for non-invasive genetic sampling, including ways of minimizing the pitfalls listed above. Sample extractions and PCRs are carried out in dedicated laboratories separate from where routine work is done, using aseptic techniques typical of good microbiological practice. All materials (tubes, pipette tips, etc.) are sterilized before use and special pipette tips with piston plungers, preventing contamination of pipette barrels, are employed. Negative controls (i.e. with no sample DNA added) are essential, and in the absence of contamination should of course generate no PCR products. The problem of allelic dropout can be countered by a multiple tubes approach, in which each sample is replicated in several separate PCRs. Only if every product is identical can a homozygote be assumed. This is the best solution but in practice it is sometimes limited by the amount of sample available.

Northern hairy-nosed wombats *Lasiorhinus krefftii* have been investigated on the basis of hair samples. Wombats resemble oversized guinea pigs and are native to parts of eastern Australia. They are nocturnal marsupials, inhabiting burrows in daytime and emerging only after dark to graze on nearby vegetation. Habitat destruction, especially land use changes that favour sheep and cattle grazing, have caused severe declines in wombat populations. Most precarious is the position of the northern hairy-nosed wombat, once found throughout a region more than 3000 km long but now reduced to less than 100 individuals in a small part of central Queensland. The dilemma in this case is how to monitor animals that are very sensitive to disturbance, and thus gain information that could be useful for their conservation, without actually making things worse by increasing their stress levels. Sloane *et al.* (2000) came up with the ingenious idea of simply suspending double-sided sticky tape above burrow runways. Hairs that stuck to the tape as the animals passed underneath were removed the following morning and DNA extracted from the associated root cells. The

● KEY POINT

DNA from hair root cells can be used to identify individual animals.

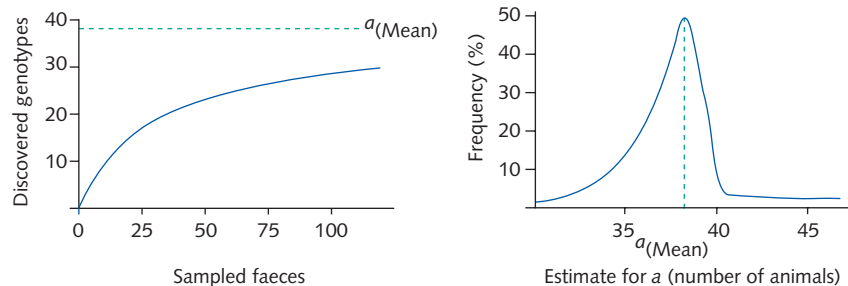
samples were then genotyped at 12 microsatellite loci and tested for sex using a 115 bp fragment of the Y-linked *Ube1Y* gene together with a 175 bp fragment of the X-linked *G6PD* locus, following PCR amplifications. By this method it proved possible to identify individual wombats, including some individuals not previously known, thus opening the way for very detailed population monitoring without any significant stress risks. Similar success has been achieved with other very small populations of endangered and secretive species, such as brown bears in the French Pyrenees. In situations like this it is feasible, and usually desirable, to identify down to the individual rather than just the species level. Long-term studies can then, in a completely non-invasive way, reveal many interesting facts about individual home-ranges, longevities, and breeding success.

A study of North American coyotes (*Canis latrans*) gives some idea of just how powerful non-invasive sampling can be in ecology. Kohn *et al.* (1999) exploited the fact that coyotes in mountainous habitat near Los Angeles mostly defecate along trails and territorial boundaries. Six hundred and fifty-one samples of carnivore faeces were collected, of which 238 were randomly chosen for analysis. DNA was extracted and genotyped after PCR amplification using mtDNA to confirm species, the *SRY* gene to confirm sex, and three microsatellite loci to identify individuals. Canid-specific primers were used to amplify part of the mtDNA control region, followed by *MvaI* restriction digestion which generated species-specific fragment patterns. This stage of the analysis demonstrated that 188 of the faecal samples (79 per cent) were definitely from coyotes. Extensive controls were carried out to establish the accuracy and reproducibility of the marker systems, all of which demonstrated impressively low error rates. An important step in this type of analysis is to determine the power of the markers to identify individuals unequivocally, a task that usually requires high levels of polymorphism (Box 3.4). Thirty coyotes were detected by this method with very high statistical power (the probability of random match between genotypes was 0.0065), using just the three microsatellite loci. There were approximately equal numbers of the two sexes, and from the rate of decline in detecting new genotypes during the faecal sampling it was possible to estimate the local population size (Fig. 3.12). This estimate, of 38 individuals, corresponded remarkably well with an independent assessment of 41 carried out by a mark–recapture protocol. The study area was not discrete, and coyotes were free to wander in and out of it. This research also showed that the number of

● KEY POINT

Systematic collection of DNA by non-invasive methods can provide extensive information about population sizes, sex ratios, individual movement, and breeding success.

Figure 3.12 Estimation of coyote population size from faecal sampling. In the left hand figure, a (mean) is the asymptote position predicted from the data curve. The right hand figure shows a frequency distribution of a as deduced from computer simulations. After Kohn *et al.* (1999).



BOX
3.4

Probability of identity and population size estimation

To establish whether samples come from two separate individuals, we need to know what the chances are that any two individuals within the population will share the same genotype. This in turn requires information on allele frequencies in the population, because common alleles are much more likely to be shared than rare ones. In practice it is desirable to aim for a probability of shared genotype identity between any two individuals (P_{ID}) of < 0.05 . This can be estimated for a population with random mating according to the formula derived by Paetkau and Strobeck (1994):

$$P_{ID} = \sum_i p_i^4 + \sum_{i \neq j} (2p_i p_j)^2$$

where, p_i and p_j are the frequencies of the i th and j th alleles at each locus in the population. Overall P_{ID} across multiple loci is the product of individual P_{ID} s for each locus. Modifications to this equation are required where population sizes are known to be small, or where high numbers of sibs are likely to be present in the sample.

As genotype data accumulate, for example from faecal samples as in the coyote study, estimates can be made

of total population size. There are several ways of doing this, one of which involves rarefaction analysis. In this approach, the asymptote of the curve showing the relationship between the cumulative number of unique genotypes found and the number of samples taken (see Fig. 3.12) is estimated. Again there are several equations that can be used, the value of each depending upon factors such as sampling effort. Kohn *et al.* (1999) used:

$$y = \frac{ax}{b+x}$$

to fit the rarefaction curve, where y = cumulative number of unique genotypes, a = estimate of population size, x = number of samples taken, and b = rate of decline in the value of the slope. Estimates of a and b , and their confidence limits, can be obtained through iterative non-linear regression using statistics programs such as SAS. Computer programs to estimate P_{ID} and population size from genotypic data are widely available: see Useful software at the end of this chapter.

faeces deposited by individuals matched their relative use of the study area. Finally, relatedness studies based on the faecal genotypes indicated that each coyote had an average of 6–7 close (parent or sib) relatives, a number consistent with field observations. In one case a suspected paternity was also confirmed. The amount of ecological information about these wild canids obtained without ever meeting them was quite remarkable, but there is no reason to suppose that the limits of such indirect approaches have yet been reached. Usable DNA has even been obtained from urine deposited in snow by wolves in the French Alps. It seems that virtually any and every biological material now has potential for molecular identification and thus yielding the valuable information that stems from it.

Non-invasive methods are easiest to apply when the species under study are relatively large, but there is increasing interest in their application to small organisms including invertebrates. Some of these methods are quite ingenious. Kawai *et al.* (2004) simply allowed marine snails *Nucella freycineti* to crawl along microscope slides for a few minutes. Mucus deposited on the slide contained epithelial and blood cells sufficient to yield high quality DNA, following standard extraction techniques, that functioned as a template in PCR amplifications of three microsatellite loci. Other methods of tissue removal involved injury to the snail with subsequent effects on behaviour and reproductive success.



Figure 3.13 Dragonfly exuvia.

Watts *et al.* (2005) collected exuviae, the cast-off skins of dragonfly nymphs that are left on emergent vegetation around ponds and streams after the adults metamorphose and fly away (Fig. 3.13). They used three different methods for extracting DNA from the exuviae, testing the products for their ability to support PCR amplification of five polymorphic microsatellite loci. All the methods were successful to varying degrees, in the order tissue kits > proteinase K/ethanol precipitation > chelex extraction. The commercial tissue kits generated DNA with very low failure rates (< 2 per cent of PCRs). Clearly there is much scope for imaginative non-invasive sampling even of very small organisms, and this will be increasingly important in studies of endangered species.

Molecular identification methods: an appraisal

Essentially just two types of molecules are used for molecular identification in ecology: protein and DNA. Applications involving protein are far fewer than those with DNA, for a number of reasons. Perhaps most importantly, protein

polymorphism is usually much less than that detectable in DNA, thus greatly limiting the resolving power of protein methods. Protein analysis generally requires relatively large amounts of tissue, and for small organisms can therefore be dangerously destructive. Proteins are also often differentially expressed both in space (tissue specificity) and time (developmental regulation), limiting their availability. Finally, protein is also more difficult than DNA to maintain in a non-denatured state under field conditions, which may or may not matter according to the type of analysis involved.

Protein analysis

Despite the limitations recounted above, protein analysis can still be useful in certain circumstances. Amphibian eggs, for example, are surrounded by proteinaceous jelly completely lacking in DNA. It is easy to remove jelly without damaging eggs and identify the species by 'protein profiling' on SDS-polyacrylamide gels. This is a simple procedure and protein can be collected into ethanol, just like DNA, because denaturation is inherent in the protocol. Allozyme analysis is rarely used in identification work because it has relatively low resolving power. For this kind of study, protein must be maintained (usually by deep freezing) in a non-denatured state. Immunological methods are, of course, also protein based. Monoclonal antibodies have been used to probe gut contents in predation studies, but their application in identification work is likely to remain very limited. They are costly and time-consuming to produce, and have few attributes not superseded in the molecular ecological context by DNA methods.

DNA analysis

Most, indeed almost all DNA-based identification procedures are now PCR-based and the critical considerations are (i) protocols for obtaining clean DNA that can be used for PCR, and (ii) choice of suitable primer sets to amplify appropriate diagnostic sequences. In some cases, as discussed in the section Bits of individuals, particularly thorough precautions must also be taken to minimize any risks of contamination.

There are now many relatively quick, simple, and reliable ways of obtaining DNA from multiple small tissue samples, including those based on chelex resins and kits from commercial suppliers that exploit selective adsorption of DNA onto silica. Specialized kits are available that remove PCR-inhibiting contaminants from plant tissues (polyphenols, polysaccharides, and so on) and from faeces. Serious difficulties at this stage are now gratifyingly rare.

Choice of primers is a potentially more complex problem. MtDNA genes such as cytochrome *b* have been sequenced from a wide range of species and it is now usually straightforward (using the information in GenBank) to develop new primer variants, or sometimes just to use existing ones. For quick identification it is often possible to restrict the PCR products, thus generating

species-specific patterns, and then observe the resulting fragments after gel electrophoresis. Increasingly, however, it should be possible (as with the eel rDNA) to identify species directly on the basis of fluorescence after PCR amplification with universal primers and species-specific probes. RAPDs are potentially powerful at many levels of identification from genus down to individual, but may require lengthy screening procedures to find primers of use in a particular study. Reproducibility of RAPDs is also notoriously sensitive to PCR conditions, especially DNA concentration, which can be tedious with large sample sizes. Similarly, sex identification will for many species require the development of new primers depending on which sex-linked genes have been identified in the organism in question. Much the same goes for microsatellites and the identification of individuals. Isolation and characterization of microsatellite loci is a complex, expensive, and moderately long procedure, albeit usually with great rewards at the end. Of course the best plan is first to screen existing databases to discover whether suitable loci have already been identified, either in the species of interest or in closely related ones. This can save a lot of time, though unfortunately there is no guarantee that markers will be directly transferable even between congeners. As time goes by the increasing range of markers available for multiple different taxa is progressively reducing the need to develop new primers.

Finally, where identification down to individual level is required, some kind of pilot study on the feasibility of the work is highly desirable. In the case of non-invasive techniques, this should include an assessment of error rates from allelic dropout and other technical limitations. For all individual identification studies, however, it is important to be sure that the statistical power of the marker loci is up to the job, for example, by carefully measuring the extent of marker polymorphism in the subject population. This essentially means estimating the probability of unambiguous genotype assignment given the size and allele frequency distribution of the population under study.

● SUMMARY

- This chapter has outlined the uses of molecular methods as tools to facilitate ecological identifications.
- The two main problems that can be addressed by molecular methods are the definition of taxonomic units and identification of individuals to or within species.
- Molecular assessments of taxonomic groupings are sometimes at variance with morphological ones, and can therefore complicate rather than simplify a situation. Nevertheless, addressing taxonomic problems with a combination of molecular and other forms of data normally generates a more robust outcome than any single approach used in isolation.
- The significance of hybrids and the extent of hybrid zones can be more fully explored using sensitive molecular methods than by relying solely on morphological criteria.

- A wide range of individual identification problems from cryptic life stages to detecting prey in the guts of predators and forensic analysis can be greatly helped by molecular markers. RAPD methods remain useful in this context despite concern about their unreliability with intraspecific (population) studies. Fluorescent, species-specific probes in PCRs can make identification very quick. In the longer term, DNA barcoding may offer the prospect of a systematic inventory for species identification.
- It is possible to identify the sex of individuals of many species using molecular markers, such as those that are Y-chromosome-specific in mammals. This can be particularly useful for immature animals and plants.
- Animals that leave evidence of their passing in the environment, such as fragments of skin, hair, feathers, or excrement can be identified to species or to individual level by PCR amplification of suitable markers from trace amounts of DNA. This non-invasive approach is particularly valuable for rare or easily stressed species and can produce surprisingly extensive ecological information.
- Protein and DNA polymorphisms can be used in molecular identifications, but DNA methods based on the PCR are by far the more broadly applicable. Isolation of DNA from small tissue samples is usually straightforward, but appropriate primers for the subsequent analysis will usually require careful development.

● REVIEW ARTICLES

Hegarty, M. J. and Hiscock, S. J. (2005) Hybrid speciation in plants: new insights from molecular studies. *New Phytologist*, **165**, 411–423.

Piggott, M. P. and Taylor, A. C. (2003) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildlife Research*, **30**, 1–13.

Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K. and Lane, R. (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B*, **360**, 1805–1811.

Sheppard, S. K. and Harwood, J. D. (2005) Advances in molecular ecology: tracking trophic links through predator–prey food webs. *Functional Ecology*, **19**, 751–762.

● USEFUL SOFTWARE

T-COFFEE. A program that will align multiple protein or DNA sequences, and thus permit investigation of haplotype differences among individuals. Available at <http://www.ebi.ac.uk/t-coffee/> as freeware.

GIMLET v.1.3.2. (Valière 2002) Used to estimate genotyping errors, kinship between individuals, probability of identity (P_{ID}), and population size from genotyped samples. Available at <http://pbil.univ-lyon1.fr/software/Gimlet/gimlet%20frame1.html> as freeware.

● QUESTIONS

1. In a study of vole populations on an offshore island you notice that individuals differ subtly in a range of morphological and behavioural characters (coat colour, tail length, male–male aggression in the breeding season) from conspecifics on the nearby mainland. List the tests you would apply to address the question of whether the voles were two distinct species.
2. You wish to study the comparative ecologies of two species of water beetles that have a high spatial niche overlap. Adults are easy to identify on morphological grounds, but larvae are impossible to determine in this way. Ponds teem with larvae, and it will be important to identify them to species if your study is to succeed. Describe how you might develop a molecular marker system to accomplish this task.
3. You are interested in using mtDNA sequence data to develop a DNA barcoding system for a group of closely related spiders. Short (43 bp) amplified sequences from 20 individuals (five each from four species) are shown below. Discuss the potential use of this DNA fragment in barcoding the spiders, and how you would further test its suitability.

Species	Sequence
1	AGTTCGGTCCATGCAATTGACTTGGGCAAGCCGTAACCTTAGG
1	AGTTCGGTCCATGCATTTGACTTGGGCAAGCCGTAACCTTAGG
1	AGTTCGGTCCATGCAATTGACTTGGGCAAGCCCTAACCTTAGG
1	AGTTCGGTCCATGCATTTGACTTGGGCAAGCCGTAACCTTAGG
1	AGTTCGGTCCATGCATTTGACTTGGGCAAGCCGTAACCTTAGG
2	AGTACGGTCCATGCAATTGAGTTGGGCATGGCGTAACCTTAGG
2	AGTCCGGTCCATGCATTTGACTTGGGCATGGCGTAACCTTAGG
2	AGTACGGTCCATGCATTTGACTTGGGCATGGCGTAACCTTAGG
2	AGTACGGTCCATGCAATTGACTTGGGCATGGCGTAACCTTAGG
2	AGTACGGTCCATGCATTTGACTTGGGCATGGCGTAACCTTAGG
3	AGTTGCGACCATGCATTTGACTTGGGCATGCCGTAAGCTTAGG
3	AGTGCGACCATGCATTTGACTTGGGCATGCCGTAAGCTTAGG
3	AGTGCGACCATGCAATTGACTTGGGCATGCCGTAAGCTTAGG
3	AGTGCGGTCATGCAATTGACTTGGGCATGCCGTAAGCTTAGG
3	AGTGCGGACCATGCAATTGACTTGGGCATGCCGTAAGCTTAGG
4	AGTCCGGTCCATGCATTTGACTTGGGCATGCCGTAACCTTTTG
4	AGTCCGGTCCATGCATTTGACTTGGGCATGCCGTTACCTTTTG
4	AGTCCGGTCCATGCATTTGACTTGGGCATGCCGTAACCTTTTG
4	AGTCCGGTCCATGCATTTGACTTGGGCATGCCGTAACCTTTTG
4	AGTCCGGTCCATGCATTTGACTTGGGCATGCCGTAACCTTTTG

4. Two microsatellite loci each have multiple alleles with the frequencies shown below in a population of badgers. How useful would each of them be by themselves, and how useful in combination, for the reliable identification of individuals from genotype data?

Allele	Locus 1	Locus 2
1	0.1	0.4
2	0.15	0.3
3	0.25	0.3
4	0.1	–
5	0.4	–