Taxonomy for birders: a beginner's guide to DNA and species problems

Norman Maclean, Martin Collinson & Richard G. Newell

Maclean, N., Collinson, M. & Newell, R.G. 2006. Taxonomy for birders: a beginner's guide to DNA and species problems. *Indian Birds* 2 (3): 50-65. [Reprinted from *British Birds* 98: 512-537 (2005)].

Norman Maclean, School of Biological Sciences, University of Southampton, Southampton, Hampshire SO16 7PX, U.K. Email: *nm4@soton.ac.uk*.

Martin Collinson, School of Medical Sciences, University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, U.K.; Email: *m.collinson@abdn.ac.uk*.

Richard G. Newell, Old Beach Farm, Green End, Landbeach, Cambridge CB48ED, U.K. Email: *dick.newell@zen.co.uk*.

This article was first published in the journal *British Birds* (www.britishbirds.co.uk). It is reproduced here with permission of the authors and BB 2000 Ltd. When citing the paper, please use the following citation: Maclean, N., Collinson, M., & Newell, R. G. 2005. Taxonomy for birders: a beginner's guide to DNA and species problems. *Brit. Birds* 98: 512–537.

ABSTRACT The use of molecular data has revolutionised taxonomy. As a result, there is upheaval within national and regional bird lists, with new species splits and lumps and, perhaps more disruptively, alterations to the sequence in which families and species are listed. These changes, sometimes based on esoteric genetic or mathematical data, affect ornithologists and birders. In this paper, some of the basics of modern taxonomic practice are explained, and we show how these principles may be applied to genetic data to generate molecular phylogenies. Examples are used to illustrate how genetic data may resolve complex taxonomic problems, and also to show some of the reasons why DNA does not offer a simple resolution to the 'species problem'. There are no simple rules to determine species boundaries, and the use of molecular data does not yet change this. There are cases where different DNA sequences tell different stories, which can be different again from phylogenies based on morphological data.

Taxonomy allocates all birds to hierarchical groupings (e.g. orders, families, genera, species and subspecies) on the basis of the perceived evolutionary relationships between them (see reviews in Greenwood 1997 and Newton 2003). Conventionally, taxonomists have to draw conclusions on the basis of what, at some level of analysis, birds or other organisms 'look like', not only with respect to their morphology, but also how they sound and behave. It is tempting to assume that birds which look, behave or sound most similar are closely related. This is usually valid – but one problem is that distantly related animals or plants might look like each other because they live in similar environments. One simple example might be cetaceans, penguins and fish; they are all pointed at the front end and have smooth contours because they face the same environmental problem – moving through water without too much drag. Ignoring the possibility of so-called 'convergent evolution' may give false impressions of close relationships.

Modern taxonomy arranges organisms into groups – **clades** – largely on the basis of the presence of '**shared-derived characters**' (or '**synapomorphies**') (see Ridley 1986 for review). These are characters that several (but not all) members of a species group have inherited from a common ancestor during the evolution of the group. For example, if a group of six closely related species (A–F) is thought to have evolved from a single red-legged ancestral species, and four of these species (A–D) have red legs, whereas two (E and F) have yellow legs, the yellow legs would be a shared-derived character – i.e. a new characteristic that has evolved by mutation from the red-legged ancestors (see fig. 1). In a cladistic classification, and in the absence of contrary information, it would be assumed that yellow legs evolved only once, in a common ancestor of E and F, and hence that both E and F are more closely related to each other than to any of species A–D. Of course, leg colour by itself does not prove anything, but if E and F also share other characters which the others do not (perhaps they both have a wing-bar while the other species do not), then bit by bit the evidence mounts that they really have evolved from a recent common ancestor.

To understand how cladistic taxonomy works, it is also necessary to understand the concept of **monophyly**. Any taxonomic clade (species, genus, family, etc.) should conventionally be **monophyletic**, which means that all members of the group should be descended from a single common ancestor, and should contain all the descendants of that ancestor (see Collinson 2001). For example, the six species in fig. 1 form a monophyletic clade. In contrast, a taxonomic grouping is said to be **paraphyletic** if some of the descendants of the common ancestor are now excluded from the grouping. If species A, C, D, E and F in fig. 1 were in one genus (X), and species B was classified as belonging to another genus (Y), then genus X would be paraphyletic, since as one of the known descendants of the common ancestor of the genus would now be excluded. By convention, paraphyletic groupings are not allowed. Nor, conventionally, are **polyphyletic** groupings, where all the different members of a genus or higher taxonomic rank have evolved from a number of more distantly related ancestors. To a first approximation, polyphyletic groupings are a mishmash of distantly related animals, grouped together to the exclusion of more closely related animals. An example would be if we attempted to group cetaceans, penguins and fish together in an Order of 'swimming things' – in effect an extreme version of paraphyly.

Many superficial characters, such as leg colour or wing length, may change rapidly – indeed too rapidly and too

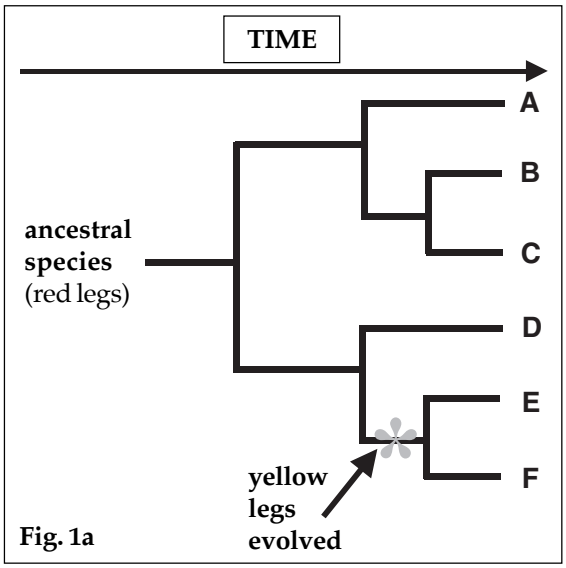

Fig. 1a

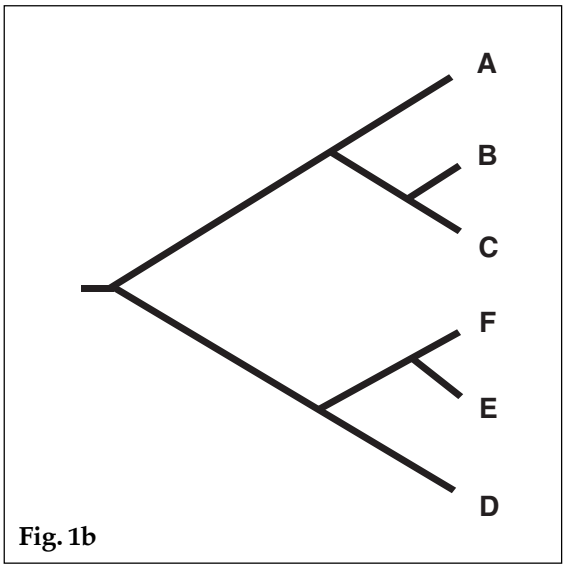

Fig. 1b

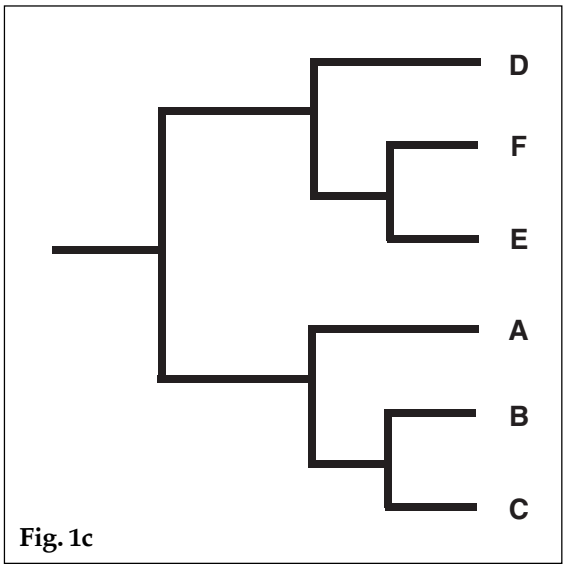

Fig. 1c

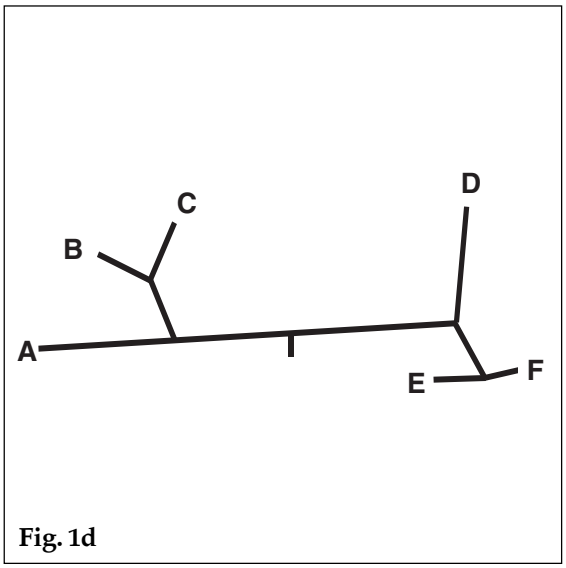

Fig. 1d

Fig. 1a. Phylogenetic figure adapted from Baldauf (2003), showing the relationships among six hypothetical species of bird, A–F (see also text). It postulates that a single ancestral species split into two separate species over evolutionary time. One of these species was the ancestor of A–C, the other was the ancestor of D–F. The A–C group and the D–F group may be regarded as separate clades (see text for definition). In the A–C clade, the ancestor of A–C split into two further species, one of which went on to become present-day species A, while the other subsequently underwent a further split to become species B and C. An identical process of evolution occurred in the D–F clade. Every branching point on the figure represents (in this case) a speciation event, and is called a node. The branches are internodes. Some methods of resolving the phylogeny (but not all) may allow the lengths of the branches to represent the evolutionary distance between the species (i.e. a rough measure of the length of time they have been separated).

The yellow legs of species E and F are a shared-derived character (synapomorphy), as explained in the text. This is evidence that they are sister species and would be used as such in the construction of the tree. If one of the taxa, say B, had green legs, this would be a unique-derived character (**autapomorphy**), but this is phylogenetically rather uninformative as it gives no clue to which of the other taxa are its closest relatives. The red legs of A–D are a retained ancestral character (**symplesiomorphy**) and are also phylogenetically uninformative (within this species group).

Fig. 1b–d. Note that the exact shape of the branches is of no consequence, such that figs. 1b and 1d present exactly the same qualitative phylogenetic information as fig. 1a. It is also possible to rotate the nodes without changing the phylogeny at all; fig. 1c carries identical information to fig. 1a.

All the trees in this figure have a root – the single earliest branch representing the common ancestor of species A–F. Roots can only be indicated if an **outgroup** (a taxon that is only distantly related to species A–F) has been included in the analysis (but not, in the cases above, explicitly shown on the tree). For example, if A–F were species of tit *Parus*, the outgroup taxon might be a species of flycatcher *Ficedula*. If an outgroup has not been included, it is not possible to indicate a root for the tree, but it is perfectly acceptable to construct and display an unrooted tree.

often to tell us anything useful about the evolutionary history of the birds. Experienced 'conventional' phylogeneticists are generally able to recognise which morphological or behavioural characters are likely to be phylogenetically useful. Nonetheless, the potential for physical characters to mislead phylogeneticists never really goes away. The use of molecular evidence (proteins and, especially now, DNA-sequence data) helps to get round this by offering an array of data that are closer to the underlying genetic base of the birds.

Species concepts and phylogenetic trees

Species concepts, as they apply to birds, have been reviewed extensively (Cracraft 1997; Collinson 2001; Helbig *et al*. 2002; Newton 2003; Parkin 2003) and we do not intend to repeat the arguments here. In broad terms, these can be divided into: **'biological' species concepts** in which species are defined primarily on the basis of present-day reproductive behaviour (a tendency not to hybridise); and **'phylogenetic'** or **'evolutionary' species concepts** in which species are defined according to their evolutionary history.

Phylogenetic trees, such as those presented in fig. 1, are reconstructions of the perceived evolutionary history of the taxa involved, based on evidence which can normally be obtained only from the present day. Whatever the species concept used, any speciation event where one species evolves into two species which no longer breed with each other can be represented by a node ('**branching point**') in a phylogenetic tree. Speciation may be due to range fragmentation (see Tokeshi 1999 for review). For example, many remote islands have endemic bird species, probably resulting from the chance arrival of vagrants and the subsequent evolution and divergence of these 'founders' away from their ancestral stock, with which they no longer breed. The separation of two species into different evolutionary lineages in a phylogenetic tree does not necessarily preclude the possibility of ongoing hybridisation. Over 10% of bird species have been known to hybridise with at least one other species in the wild (Grant & Grant 1992). Sometimes, especially when related species are fortuitously (or unwittingly) brought together, this hybridisation may be quite extensive. For example, Ruddy Ducks *Oxyura jamaicensis* breed freely with White-headed Ducks *O. leucocephala*. Hybridisation between Hawaiian Ducks *Anas wyvilliana* and Mallards *Anas platyrhynchos*, following the introduction of the latter, is apparently leading to the virtual extinction of the original form of Hawaiian Duck (US Fish & Wildlife Service 1999). These are extreme cases, where the species pairs did not meet in the past and have done so now only through human intervention. Nonetheless, the possibility for gene flow (hybridisation) between different species is ever-present, and may confuse genetic analysis of species boundaries. The rest of this paper deals with the analysis of the genetic structure within and between species. First, it is necessary to introduce the genetic material – DNA.

An introduction to DNA

Deoxyribonucleic acid (DNA) is the heritable genetic material. Every cell in our bodies carries copies of the DNA we inherited from our parents. It is the 'instruction manual' for the cells, known as the **genome**, the so-called 'book of life', which is written with a code using only four letters – the chemical bases adenine, thymine, cytosine and guanine (A, T, C and G). Hanging off a repetitive sugar-phosphate 'washing line' is the 'laundry' of As, Ts, Cs and Gs; the order in which these letters occur along the line underlies the genetic code and makes us what we are. The sequence of letters – genes – can be read by molecular machinery in the cell that tracks along them and puts proteins together on the basis of the instructions it sees in the DNA (see the extensive review text in Lewin 2002). The familiar double helix of DNA consists of two complementary 'washing lines' of bases, bonded to one another across the strands, such that T always bonds with A and G always bonds with C. Although DNA is generally extremely stable, sometimes, these A, T, C and G bases may fall off, or be chemically changed, or new bases may be added, or big chunks of DNA may break off, or turn round – these are the processes of mutation (reviewed in Majerus *et al*. 1996). Mutation usually arises as a result of accidental damage or mistakes in replication of DNA, sometimes due to radiation or exposure to environmental chemicals. Mutations occur randomly in all our cells all the time, but as described below, may cause disadvantageous changes. Usually our cellular repair proteins spot mutations and correct them. In the time it takes you to read this paper, over 20 potentially lethal mutations will occur somewhere in the cells of your body, and you have to hope that they are repaired.

Proteins carry out all the important jobs required to keep body cells working, and are clearly critical to survival. If the coding regions of DNA, which provide the template for these proteins, are defective, as a result of mutation, then the proteins may not work. If that happens, there is a good chance that the mutant animal will not survive so, over evolutionary time, these DNA sequences do not change much. There are also regulatory regions of DNA that do not themselves make proteins, but control where and when different proteins are made in which different cells of the body, for example building bone cells or brain cells in the right places. Regulatory regions are critical too. Nonetheless, mutations in regulatory or coding regions of DNA can still occur because not all changes are necessarily bad. So, over evolutionary time, mutations may build up in and around genes that enable us to distinguish different individuals and species.

The DNA in a typical vertebrate cell includes about six billion base pairs, enough to provide for about one million genes, but most vertebrates have only about 20,000–30,000 genes (Genome Sequencing Consortium 2001). Therefore, the bulk of DNA in higher organisms appears not to do anything. Many of these bits of DNA may have had functions in the past – the minority are recognisable as genes that are no longer functional. Some are, or were, viral sequences that were integrated into the DNA and never left, and are now inactive. Other, very similar bits of DNA may occasionally replicate and move about within the DNA (**'jumping' elements**), with or without causing any harm. Some are strings of repeated sequences that are the product of mistakes during DNA replication. For example, molecular studies of bird taxa are frequently carried out by exploiting

microsatellite DNA sequences. These are non-coding and consist of a long series of short (usually two to six bases) motifs (Majerus *et al*. 1996). Many different microsatellites are spread throughout the genome of most organisms and may be used as indicators of variation between individuals or populations. Being non-coding, they also evolve quickly. Whatever their origin, these areas of DNA comprise vast strings of 'nonsense'. These apparently non-functional 'neutral' sequences of DNA can mutate freely, because mutation of nonsense, to make more nonsense, has no adverse effect (Kimura 1983).

For genes that encode proteins which affect key metabolic processes like respiration, mutations that change the proteins are likely to be lethal and are soon eliminated from the population – these areas of DNA change relatively slowly. They are said to be highly conserved. Bits of DNA that have no apparent function may be free to acquire mutations very rapidly. This is an important consideration; if a phylogeneticist is studying closely related taxa that diverged from each other only recently, there is little point in looking at slowly evolving, highly conserved parts of DNA – they will be too similar in the different taxa to tell us anything useful about their evolution. Conversely, when studying old evolutionary events, it is better to use these highly conserved stretches of DNA, for which informative mutations are likely to have built up over millions of years. The non-conserved, variable bits of DNA will have mutated so much that it is impossible to compare them among the relevant taxa. It soon becomes clear that DNA-based approaches to taxonomy are potentially extremely complex.

The mechanics of molecular evolution

Most mutations are repaired. Some, however, are not repaired, and persist. Mutations in sperm or egg cells may be passed on to the next generation, meaning that different individuals in a single species are genetically different. For example, humans may have red, blond or black hair, and blue, brown or green eyes – these are genetic differences within a single species. Although we all carry the same genes, there are many slightly different versions (**alleles**) of these genes in the population as a whole, thanks to mutations. All animals carry pigment genes, but because individuals may have different alleles of these genes, they are differently pigmented, e.g. dark- and light-phase Arctic Skuas *Stercorarius parasiticus*. Mendel's famous peas (Mendel 1866) were selected on the basis of having different alleles of the same gene giving, say, wrinkled or smooth coats on the seeds. When scientists talk about the genetic variation within a population or a species, they are referring to the number and proportion of different alleles they can find for any single gene or other stretch of DNA.

There are several forces that can change the genetic structure of a species. Mutations are ultimately the source of all genetic variation, and the occurrence of different mutations in different populations of a single species is the raw material of speciation. Mutations persist or are eliminated partly as a result of **natural selection**, which changes the genetic structure of a population either by allowing the survival of a newly arisen mutation that confers a survival advantage to the bearer, or by eliminating individuals that are disadvantaged. Because the environment plays a major role in natural selection, populations of one species which lives in different environments may, through natural selection, evolve to look or behave differently. Random processes can also change the genetic structure of a population; for example, when a cow treads on a lark's nest, the nestlings will perish irrespective of genetic traits (whether 'fit' or 'unfit'). Such random mortality is inevitable, and may lead ultimately to genetic differences between different populations of a single species, or it may lead to the accidental loss of useful mutations purely by chance (**genetic drift**).

Random dispersal of individuals between populations may slow down divergence; the permanent movement of individuals from one population to another is called '**migration**' by population biologists (distinct from the seasonal movement of individuals between breeding and wintering areas more familiar to birders). Dispersive 'migration' may often work against speciation (but not always; see the study of Great Tits *Parus major* by Garant *et al*. 2005). Perhaps counterintuitively, sex may also work against speciation. Sexual reproduction provides a mechanism for new gene alleles or gene combinations which arose in one population of a species to spread through other populations of the same species, and is a great homogenising factor. Sex also allows for hybridisation, by which gene sequences may be shared between different species.

At the level of a single DNA base pair it can be seen that there are several forces at work, either increasing or reducing divergence between individuals within a species. When the necessary mathematics are performed, the homogenising forces (e.g. sex, dispersal) tend to win (Dover 1982). Thus, unless there are special natural or sexual selective circumstances that maintain unusually high diversity, individuals of a species in an interbreeding population tend to have very similar (but not usually identical) DNA.

So, if there are large fixed (consistent) DNA differences between two populations or subspecies of a single species, it indicates that there has been little or no interbreeding. When two or more taxa no longer interbreed, gene flow between them can no longer occur. So the DNA sequences of reproductively isolated populations are free to diverge, by mutation, with respect to each other. Taxa that stopped interbreeding only recently will have very similar DNA sequences; taxa that have been reproductively isolated for many millions of years will have very different DNA sequences. The ability of scientists to read the same bits of DNA sequence in individuals of different taxa, to analyse how similar they are and decide, on the basis of sequence comparisons, which taxa are most closely related to which others, is the basis of DNA-based molecular phylogeny. While the mathematics are complicated, the principles are not dissimilar to classical morphology-based attempts to reconstruct phylogeny – those taxa that are closely related will tend to look most similar, and can be arranged into clades on the basis of shared-derived mutations in their DNA.

Nuclear and mitochondrial DNA

The vast majority of the DNA that all animals, plants and fungi carry in their cells occurs in the cell nucleus. At fertilisation, the nucleus of the sperm, carrying one complete 'set' of DNA (**chromosomes**), fuses with the nucleus of the egg, carrying another complete set of DNA, to form an embryo that is **diploid**, i.e. it carries two complete sets of chromosomal DNA – one from each parent. This means that every cell in our bodies has two copies (alleles) of every single gene and every bit of non-coding or 'junk' DNA. Normally, both alleles of any particular gene will be functional. However, a benefit of being diploid is that if one allele of the gene, say the one inherited from the father, is mutated and not working, the other copy, from the mother, is likely to be functional – there is back-up. Note also that, while an individual may possess two variant alleles of any one nuclear gene, within a population, there may be a large number of different alleles of the same gene. This is what is meant by the term '**gene pool**'.

Some DNA resides outside the nucleus in structures such as chloroplasts (responsible for photosynthetic processes in plants) or mitochondria (the 'batteries' of the cell, responsible for producing energy). Mitochondrial DNA (mtDNA) is a relatively small circular molecule (the ends are connected), and carries a small number of genes, not directly related to external morphology, but to energy production (see Lewin 2002).

Although both sperm and eggs carry mitochondria, the sperm mitochondria do not enter the egg at fertilisation, and are quickly broken down. This means that for birds (and all other animals), mitochondria are inherited only down the female line. That is, although all the cells in our bodies contain nuclear DNA from both our parents, we have only our mother's mitochondrial DNA, whether we are male or female.

Mitochondrial DNA is subject to random mutation just like nuclear DNA but the mutation rate of the mtDNA is substantially higher, not least because mutations are not efficiently repaired. Mutations may be passed from mother to offspring and hence, over time, different taxa that no longer interbreed develop sequence differences in their mtDNA. The use of mtDNA for building evolutionary trees is so widespread, at least in part, because:

1) It is an abundant, fairly small molecule that is easy to isolate. Every cell has many mitochondria, each with several or many copies of (usually) identical mtDNA.
2) It carries mutations only through the female line with virtually none of the mixing (recombination) of maternal and paternal chromosomes that complicates analysis of nuclear DNA.
3) It has genes for respiratory enzymes (e.g. *cytochrome b*) that are highly conserved and which mutate fairly slowly. It also has regulatory regions, e.g. certain parts of the control region, which mutate very quickly.

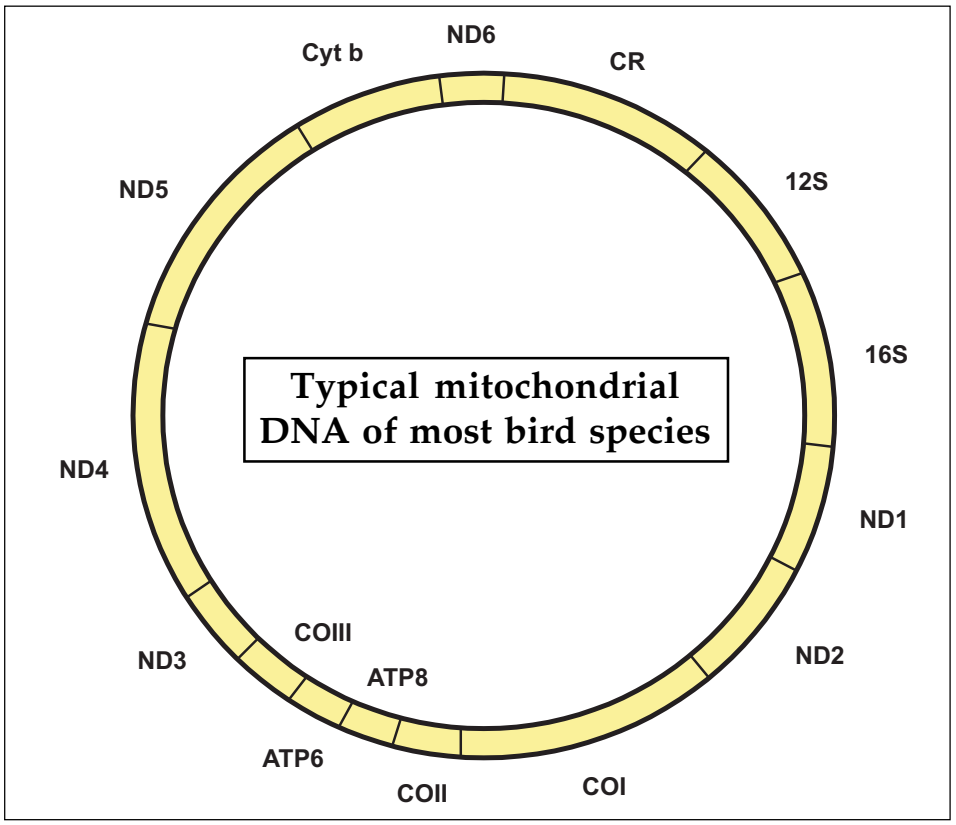

Fig. 2. Diagram of the mitochondrial DNA of a bird. This is the most common form, found in the domestic chicken *Gallus gallus* and most other birds. A slightly different form is found in some bird groups such as falcons (Falconidae) and woodpeckers (Picidae); see Mindell *et al.* (1998). The mitochondrial-DNA molecule consists of a closed circle, consisting of about 17,000 base pairs (compared with about 3,000 million base pairs in the nuclear DNA).

In the diagram, the genetically significant parts are labelled. 12S and 16S are genes coding for ribosomal RNA; ND1–ND6 are genes coding for subunits of a large protein called NADH dehydrogenase; *cytochrome b* is a gene coding for the enzyme *cytochrome b*; and CO I, II, and III are genes coding for cytochrome oxidases. ATP6 and AT8 are genes coding for subunits of the protein adenosine triphosphate. CR is the control region containing the sequences where DNA replication is initiated – the origin of replication. Any of these regions of DNA may be used to extract phylogenetic information – some mutate slowly and are of use for determining ancient phylogenetic events, whereas others, such as the control region, have very quickly mutating (hypervariable) sequences that can be used to resolve recent evolutionary events, such as the radiation of the large white-headed gulls *Larus argentatus–fuscus–michahellis–cachinnans*.

[NB 'RNA' is ribonucleic acid – a nucleic acid found in all living cells that is essential for the manufacture of proteins according to the instructions carried by genes; a 'ribosome' is a subcellular particle consisting of RNA and associated proteins concerned with the manufacture of proteins.]

Consequently, different bits of the mitochondrial genome can be analysed to determine evolutionary patterns between recently diverged (closely related) or distantly related species (fig. 2).

What rates of divergence in mitochondrial DNA should we expect? Calibration of the amount of divergence expected over, say one million years, can be achieved by looking at the percentage DNA differences between taxa whose divergence can be dated to known geological events. For example, Hawaiian honeycreepers (Drepanididae) are believed to have speciated by occupying and diversifying on new islands as these appeared following undersea volcanic activity (Fleischer & McIntosh 2001). The times of appearance of the different islands in the chain is known, thus allowing determination of when new honeycreeper species evolved. The estimates are that in one million years of evolution, the mitochondrial *cytochrome b* sequence has diverged by about 1.6% on average, and the whole mitochondrial sequence has diverged by about 2%. Most of the latter variation is, of course, concentrated in the faster evolving bits.

As discussed by Newton (2003), different subspecies of a single species do not usually show more than about 2.5% divergence in the sequences of their mitochondrial *cytochrome b* gene, whereas different sister species may show up to 10% divergence or more in this sequence. Similar calculations would apply to some other mitochondrial gene sequences (examples in Helbig *et al*. 1995). But, as emphasised by Newton (2003), not all species fit in with these expectations. Some populations of birds that display good species separation on other grounds may show <1% divergence in sequence (e.g. Blue-winged Teal *Anas discors* and Cinnamon Teal *A. cyanoptera*), and some subspecies may have more than 2.5% divergence (e.g. Mountain Chickadee subspecies *Parus gambeli gambeli* and *P. g. baileyae*). So, it is of little value to say, in isolation, that 'DNA studies indicate a 3% difference' either in support of or against a proposed species split, and such arguments based on percentage divergence are dangerously circular anyway.

The mechanics of molecular analysis

DNA can be extracted from tiny amounts of fresh tissue (e.g. blood, muscle, feathers, faeces) and is even recoverable from museum specimens. To perform a molecular analysis, DNA is isolated from individuals of all the taxa under study, and either directly sequenced or subjected to some other chemical procedure which cuts up (**restricts**) or makes copies of (**amplifies**) sections of it using the **polymerase chain reaction** (PCR). Genes such as *cytochrome b* are best analysed by direct sequencing of the series of As, Ts, Cs and Gs. Various pieces of non-coding repetitive sequence, such as microsatellites, are usually analysed indirectly by restriction or amplification, producing a series of DNA bands of different sizes that are characteristic of the taxon.

Once the DNA has been sequenced, cut or amplified by PCR into an analysable form, the actual processes of analysis are not in principle much different from those employed for conventional morphological or behavioural analyses. Taxonomists look for informative mutations – shared-derived characters (synapomorphies) that can be used to link taxa together. For example, if there are six species A–F, and at position 123 of their *cytochrome b* gene, all of them have a thymine base ('T'), this does not help to resolve their phylogeny. If however, A–D have a 'T', but E and F have a guanine base 'G', this may be taken as evidence that E and F are closely related. By itself, this is not strong evidence, but if several different parts of the *cytochrome b* gene tell the same story, then the evidence gets stronger. Mathematical algorithms are used to determine, on the basis of DNA sequences, which species are most similar, and to build evolutionary trees. Different methodologies determine those trees requiring the minimum amount of mutation (**maximum parsimony**), or those trees that are most likely on the basis of prior knowledge about how DNA mutates (**maximum likelihood**) or those that progressively clump together the most similar species (**neighbour-joining**) (full review in Nei 1987). These techniques are described more fully in Appendix 1.

Sibley and Ahlquist exploited DNA–DNA hybridisation to classify the bird families of the world (Sibley *et al*. 1988; Sibley & Ahlquist 1990). Their method did not analyse the genome sequences in detail. Instead, it was assumed that if the DNAs of two different species of bird are very similar, they will bind very tightly together because of complementary pairing of the As with Ts, and Cs with Gs, as described earlier. Conversely, if two species are only distantly related, their DNAs will be more divergent, so when mixed together they will not bind very tightly (rather like a zip-fastener with a lot of missing teeth). Sibley and Ahlquist's technique measured the melting temperature (the temperature at which the two complementary strands fall apart) of the hybridised DNA. The more similar the sequences in the compared samples, the more tightly they will combine and the higher the temperature needed to induce separation. The technique showed, for example, that New World barbets (Capitonidae) have closer molecular affinities to toucans (Ramphastidae) than to the barbets of the Old World (Megalaimidae and Lybiidae). The technique is at best only suited to broad-based taxonomic studies.

Problems with DNA

If DNA methods are so powerful, we should ask why answers to all of the outstanding taxonomic questions are not quickly forthcoming. DNA analysis is in practice probably no more objective than other analyses. Determining a phylogenetic tree from a DNA sequence is not simple. The same mutation may occur independently in distantly related species, giving a false impression of a close relationship. Similarly, over time, a mutated DNA base may randomly mutate back to its original state, giving a false impression of no mutation. The assembly and alignment of DNA sequences from several species of bird may require some heavy-duty computer work. Mistakes here seriously bias the data. Subsequent treatment of the data, whether by different analyses (neighbour-joining, maximum parsimony, etc.) compounded by uncertain assumptions about frequencies of different sorts of mutations, may result in several plausible-looking phylogenies from one dataset. On the other hand, if different

Box 1. ***Can a DNA sequence serve as a bar-code for species identity?***

The idea of using a short DNA sequence as a 'bar-code' to allow molecular categorisation of every species was originally proposed by Hebert *et al*. (2003), using a 650-base-long sequence of the mitochondrial CO1 gene. This idea has now been applied to a study of 260 species of North American birds by these authors (Hebert *et al*. 2004). All species had a different bar-code (some had more than one), and differences between related species were about 18 times higher than those within species. Some interesting anomalies were revealed, all of which serve to support the idea of using this sequence as a species bar-code. In four exceptional cases (Solitary Sandpiper *Tringa solitaria*, Eastern Meadowlark *Sturnella magna*, Marsh Wren *Cistothorus palustris* and Warbling Vireo *Vireo gilvus*), there were deep sequence divergences within a species. Interestingly, all bar the sandpiper have been proposed for possible splitting by some taxonomists, so these species may well include other cryptic molecular species. Thirteen species within four genera revealed rather low levels of molecular differentiation. These included seven species of large white-headed gulls *Larus*, the two North American oystercatchers *Haematopus bachmani* and *H. palliatus*, Mallard *Anas platyrhynchos* and Black Duck *A. rubripes*, and American *Corvus brachyrhynchos* and Northwestern Crow *C. caurinus*, all of these being species with known levels of hybridisation or recent allopatric divergence.

treatments of the DNA data produce the same, or very similar, phylogenetic trees, we can have more confidence that the correct picture is emerging.

Analysis of any gene or other sequence from mitochondrial or nuclear DNA of a group of species produces only a phylogeny *of the gene* (or whatever part of the DNA was analysed), for which each node in the tree is a point at which a mutation occurred. It is usual to assume that the phylogeny of the gene in these different species accurately reflects the evolution of the species that carry the gene. This is probably normally a reasonable assumption, but it need not necessarily be the case, especially for groups of species (or subspecies) that have separated only recently. More than one mtDNA lineage may exist within one species population. Trees inferred from various parts of the nuclear genome will not necessarily be congruent to trees from the mtDNA. Using any of these trees, based on a single gene or other DNA sequence as a dependable guide for the species tree, can be dangerous.

Discrepancies between morphological and genetic divergence

New species may evolve rapidly over comparatively short time spans (Fryer & Iles 1972). The best-known avian examples are Darwin's finches (Emberizidae) on the Galapagos and Hawaiian honeycreepers, which must have evolved over thousands rather than millions of years because geological evidence shows that the volcanic archipelagos on which they live were created only recently – see Sato *et al.* (1999). Conversely, some organisms, e.g. ginkgo trees (Ginkgoaceae) and coelacanths (Latimeriidae), have remained morphologically almost unchanged for about 300 million years. Ospreys *Pandion haliaetus* look morphologically remarkably uniform across a virtual world range, while Common Chaffinches *Fringilla coelebs* and Wrens *Troglodytes troglodytes* have many described subspecies, with local song variations even within the UK, as well as clinal plumage variations across continents and discrete isolated or island forms (Vaurie 1959; Catchpole & Rowell 1993).

The morphology of a species is determined largely by its genes. Morphology is also influenced by the environment, however, and when the environment changes rapidly, morphology may change rapidly. On the other hand, animals or plants that live in constant, unchanging environments, such as the deep ocean, may show very little morphological change over geological timescales, even though mutations in their DNA are inevitably piling up. This has implications for genetic analysis. Groups of species which have evolved rapidly and recently may all be genetically uniform; their evolution may be difficult to resolve using DNA. Yet single species with a large and long-standing world range may show high intraspecific genetic variation; for example, all Barn Owls *Tyto alba* have a similar appearance but they exhibit a fairly complex genetic structure, so they must have spread across the planet some considerable time ago (König *et al*. 1999).

The strongest reason for the frequent lack of concordance between molecular and morphological approaches to taxonomy is that the molecular data are almost never drawn from the genes controlling aspects of morphology that are conventionally used to delimit species. The normal analyses of fast-evolving and selectively neutral non-coding sequences, or of slowly mutating metabolic or respiratory enzymes like *cytochrome b*, are fine for measuring the time elapsed since the divergence of two taxa, but give no indication of the often rapid changes in morphology and reproductive behaviour which may drive speciation. Thus, while the Barnacle Goose *Branta leucopsis* is, on the basis of the molecular sequences analysed, closely related to the Lesser Canada Goose *B. hutchinsii*, its morphology and plumage have changed quite dramatically (Paxinos *et al*. 2002; see discussion below). Although this may appear to be a bad thing, it is in fact the strength of genetic analysis – DNA mutation allows a direct analysis of evolution without the complications that result from the influence of the physical and biological environment on external morphology. Molecular analyses could be based on genes which *are* involved in morphological evolution important to speciation, but relatively few of these sequences have been characterised and none have been studied sufficiently to confirm they give reliable phylogenetic data.

Some examples of the use of DNA

While advising caution for birders in the interpretation and application of DNA data, we do not want to paint an unnecessarily bleak picture. The rest of this paper demonstrates how DNA has contributed to our understanding of some interesting species issues, describing cases where it has provided

a clear phylogenetic signal, and cases where it has not.

Phylogenies based solely on DNA sequences may reveal species boundaries that were not obvious solely on the basis of morphology. For example, it is now largely accepted that Western *Phylloscopus bonelli* and Eastern Bonelli's Warblers *Ph. orientalis* are separate species, based largely on differences in their calls, and also on the DNA evidence, which showed that they may have diverged several million years ago. The species have remained separate in spite of continued occasional opportunities to hybridise (Helbig *et al*. 1995). More recently, DNA evidence has suggested other splits, based both on the absolute genetic distance between taxa that were previously assumed to be closely related, and on the relationship of those taxa with other species. Bearing in mind that any taxonomic grouping (species, genus, family, etc.) should be monophyletic, one of the arguments for splitting Eurasian *Anas crecca* and Green-winged Teals *A. carolinensis* was that, on the basis of mtDNA, *carolinensis* appeared to be genetically closer to a South American species, Specked Teal *A. flavirostris*, than to *crecca* (Johnson & Sorenson 1999). There were two possible taxonomic arrangements that could satisfy the requirement for monophyly in the taxonomic arrangement of these teals, both different from the traditional arrangement: either lump all three taxa into one species (which would be bizarre, given the highly divergent plumage of *flavirostris*) or split them into three species (Sangster *et al*. 2001). It was suggested that, at some stage, there was a single Holarctic ancestral teal with separate populations which diverged into Palearctic and Nearctic species. This may have been followed by range expansion of the Nearctic species into South America. The North American *carolinensis* maintained an appearance close to the Palearctic *crecca* but, in isolation and under its own selective pressures, the South American *flavirostris* evolved into something that looked superficially quite different. On the other hand, is it a mere coincidence that the Yellow-billed Pintail *Anas georgica*, sympatric with Speckled Teal, shares traits with *flavirostris* not present in the other teals? Has hybridisation caused *flavirostris* to look quite different from *crecca* and *carolinensis*?

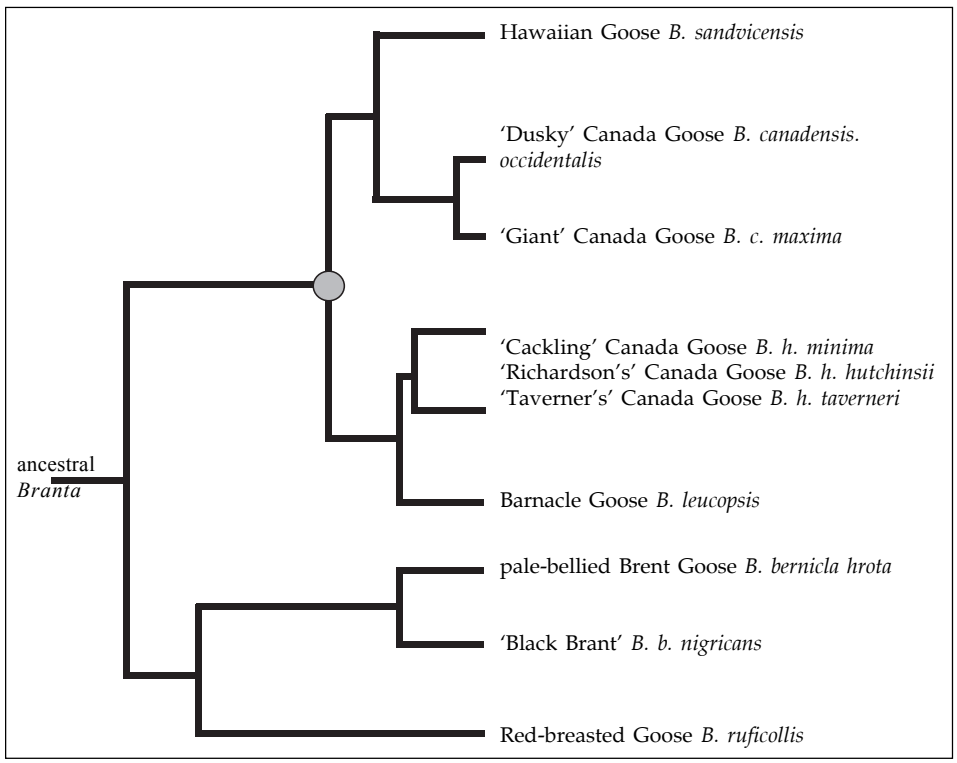

Fig. 3. Relationships among the black geese *Branta*, adapted from Paxinos *et al*. (2002), based on mtDNA sequences. The tree suggests that Hawaiian Goose is the sister taxon of the large Canada Geese ('Dusky' and 'Giant') and that Barnacle Goose is the sister species of small Canada Geese ('Cackling', 'Richardson's' and 'Taverner's'). The last common ancestor of the 'traditional' Canada Goose (marked as a red node here) also gave rise to Hawaiian Goose (and other extinct Hawaiian taxa not shown here) and Barnacle Goose, which are both different species. The traditional Canada Goose assemblage is, therefore, paraphyletic. Splitting large and small Canada Geese recreates 'acceptable' monophyly, wherein each species is descended from one common ancestral taxon, and all the descendants of that taxon are included in the species. Note that this figure is diagrammatic only, and that branch lengths have been altered slightly.

Canada Geese

A similar situation may exist with respect to the Canada Geese *Branta canadensis*. Morphologically, these may be divided into 'small' and 'large' subspecies. Paxinos *et al*. (2002) published a phylogeny based on mtDNA sequences of two large subspecies ('Dusky' *B. c. occidentalis* and 'Giant' Canada Goose *B. c. maxima*) and three small subspecies ('Cackling' *B. c. minima*, 'Richardson's' *B. c. hutchinsii* and 'Taverner's' Canada Goose *B. c. taverneri*), and also included other *Branta* species in the analysis (see fig. 3). They showed that the traditional 'Canada Goose' species was paraphyletic – Barnacle Goose *B. leucopsis* appeared to be closely related to the small Canada Geese while the Hawaiian Goose *B. sandvicensis* (with other extinct Hawaiian geese) was suggested to be closely related to the large Canada Geese. In contrast, small and large Canada Geese were more distantly related to each other. Barnacle and Hawaiian Geese are clearly different species from Canada Goose, so the requirement to maintain monophyly demands that Canada Goose should be split into two species, Greater *B. canadensis* and Lesser Canada Goose *B. hutchinsii* (Banks *et al*. 2004). It is possible that, assuming that the mtDNA data represent the true phylogeny of these goose taxa, their last common ancestor had plumage resembling modern-day Canada Geese, but that two isolated ancestral populations at the extreme edges of the range, Hawaiian and Barnacle Goose, evolved rapidly into different species with quite different

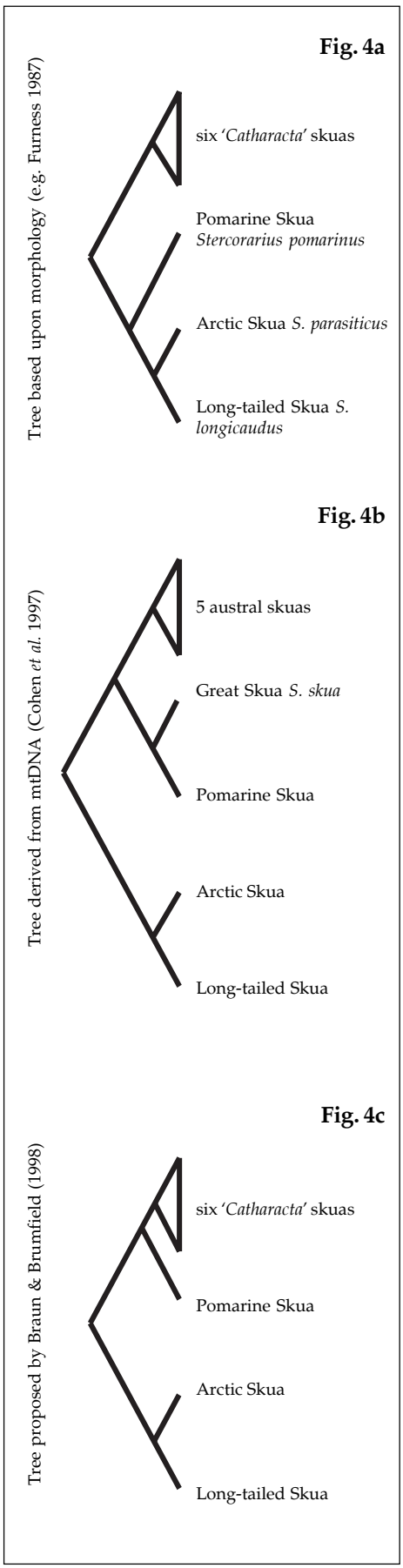

Fig. 4. Different interpretations of skua *Stercorarius* taxonomy – see text for details and discussion.

plumage patterns. Situations where molecular data do not accord with phylogenies based on morphology or plumage are, however, hard to interpret. Ideally, each line of enquiry – nuclear and mtDNA and morphology – should give the same result.

Skuas

The skuas *Stercorarius* provide a good example of the difficulties that taxonomists face in trying to unravel evolutionary history. Morphological analysis produces an unambiguous phylogenetic tree with two well-separated clades, consistent with the separation of skuas into two genera, *Catharacta* and *Stercorarius* (fig. 4a; Furness 1987), representing the 'large' and 'small' skuas respectively. On the other hand, Cohen *et al.* (1997) reported the results of several independent studies which analysed nuclear DNA, mtDNA and ectoparasites, with supporting evidence from behaviour and structure. These data implied that Pomarine *S. pomarinus* and Great Skuas *S. skua* are more closely related to each other than either is to its congeners. A phylogenetic tree, based solely on the analysis of mtDNA, grouped Pomarine and Great Skuas together, forming a sister clade to the 'austral' skuas: South Polar *S. maccormicki*, Chilean *S. chilensis* and Brown Skuas *S. antarcticus* (fig. 4b). This tree is inconsistent with the traditional grouping of two genera for the large and small skuas, because both genera would be paraphyletic.

But there is a difficulty with the Cohen *et al.* (1997) tree because, depending upon the assumptions made about the morphology of the original ancestral skua, either the 'large skua' ('*Catharacta*') morphology (of Great, South Polar, Brown and Chilean Skuas) or the 'small skua' morphology (of Pomarine, Arctic *S. parasiticus* and Long-tailed Skuas *S. longicaudus*) must have evolved independently twice, something which is believed to be highly unlikely. They explored several hypotheses to explain this, though popular attention settled on the possibility that Pomarine Skua is a result of past hybridisation between a large '*Catharacta*' skua species and a small *Stercorarius* skua population.

Braun & Brumfield (1998) reanalysed the data of Cohen *et al.* (1997). They looked at the results from nuclear DNA, which suggested that Great Skua is closer to South Polar Skua than it is to Pomarine Skua, and proposed an alternative tree, with Pomarine basal to the '*Catharacta*' clade. This tree has a much more plausible explanation, which is that the ancestral skua species split into two lineages, one of which evolved into Arctic and Long-tailed Skuas while the other evolved into Pomarine Skua and all the large '*Catharacta*' skuas (fig. 4c). The Braun & Brumfield tree implies that the ancestral skua had small-skua-type morphology; it is attractive because it does not rely on 'far-fetched' explanations such as inter-generic hybridisation to explain skua evolution. However, it does not really explain why Pomarine and Great Skuas have such similar mtDNA and ectoparasites. Andersson (1999) calculated a phylogenetic tree based upon the ectoparasite species found on skuas, used as a taxonomic indicator of their hosts, and this tree was similar to that generated from mtDNA. It is highly unlikely that this concordance between the two trees could occur fortuitously, or by convergent evolution, so Andersson concluded the most likely explanation was that hybridisation *had* occurred. The evidence indicates that Great Skua perhaps evolved from a (presumably very small) population of migrant male austral skuas that lingered in the north and could not find mates of the same species and so, in desperation, took the best mates that they could find – female Pomarine Skuas! Given the uncertainty about skua evolution, the best solution at present seems to be to place all skuas, large and small, in a single genus, *Stercorarius*. In any case, there is no protocol to deal with the constraint of monophyly when a taxon derived by hybridisation is interposed between two otherwise perfectly good monophyletic groupings.

Greenish Warblers

Greenish Warbler *Phylloscopus trochiloides* shows a clinal gradation from northwest Europe through Central Asia, forming (notwithstanding a 1,000-km gap in China) a 'ring' around the Himalayas (Irwin 2000, 2001, 2005; reviewed in Collinson 2001, Collinson *et al*. 2003). In the area of overlap they behave as two species: 'Two-barred

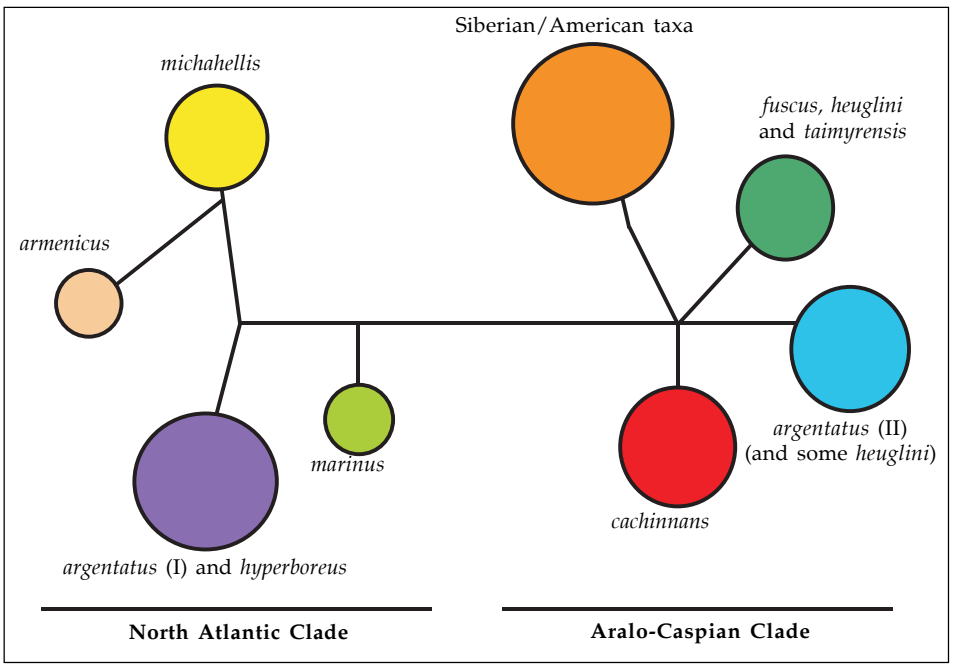

Fig. 5. Pictorial representation of the genetic relationships between 'large white-headed gulls'. The lines represent genetic distance, the lengths of the lines being roughly proportional to the number of DNA mutations differentiating among the taxa. The filled circles represent a cluster of individual gulls with very similar or identical mtDNA sequences. The orange circle representing 'Siberian/American' taxa contains individuals of Iceland Gull *Larus glaucoides*, Glaucous-winged Gull *L. glaucescens*, American individuals of Glaucous Gull *L. hyperboreus*, Slaty-backed Gull *L. schistisagus*, as well as individuals of *smithsonianus*, *taimyrensis*, *mongolicus*, *vegae* and *heuglini*. The figure is based on Liebers *et al*. (2004), but very much simplified and with several anomalies removed. The occurrence of European and British Herring Gulls (*argentatus/argenteus*) with mtDNA sequences that fall within either clade may be due either to past hybridisation or to retention of ancient DNA sequences that were present in the ancestors of all the gulls.

Greenish Warbler' *P. t. plumbeitarsus* does not interbreed with *P. t. viridanus*. On vocal and morphological grounds, there is no point in the ring at which one can say where a Greenish Warbler turns into a Two-barred Greenish Warbler, so they are classified as one species (see further detail in Newton 2003). DNA analysis demonstrates that there are eastern and western genetic clades, but that these genetic clades do not correspond with any morphological or other biological distinction between any of the taxa. This is perhaps the best avian example of a ring species, where classifying *P. t. viridanus* and *P. t. plumbeitarsus* as either subspecies on one hand, or separate species on the other, does not in either case fully describe the subtlety of the evolving biological scenario – a case of a rigid nomenclature being unable to describe what is going on in the real world. It would appear that the ancestral 'greenish warbler' evolved along the southern rim of the Himalayas, and moved north during interglacial periods, undergoing selection pressure for more complex songs as it did so. On top of this, it is possible that periodic range fragmentation allowed for development of clearly separate eastern and western genetic clades.

Large white-headed gulls

On the other hand, the best available DNA evidence has probably torn apart another erstwhile example of an avian ring species, that involving the large white-headed gulls. The circumpolar distribution of northern Herring *Larus argentatus* and Lesser Black-backed Gull *L. fuscus* taxa was interpreted as an example of a ring species, based on an apparent cline in mantle and wing coloration of adults; from the darkest birds (Lesser Black-backed), east through Heuglin's *L. ?f. heuglini*, Taimyr *L. ?f. taimyrensis*, Vega *L. ?a. vegae/birulai* and American Herring Gull *L. ?a. smithsonianus* to the palest birds, European Herring Gulls *L. a. argentatus/argenteus* (Mayr 1963; Grant 1986). Early attempts to determine the evolutionary relationships between these gulls, and the more southern 'yellow-legged' taxa *atlantis*, *michahellis*, *cachinnans*, *barabensis* and *mongolicus*, were difficult to interpret because the DNA sequences analysed did not mutate fast enough to resolve their apparently rapid evolution. Crochet *et al*. (2002, 2003) and Liebers *et al*. (2001, 2004) analysed DNA sequences from both the *cytochrome b* gene and the fast-mutating hypervariable region of mtDNA from 'Herring Gull' and other white-winged and Siberian/Pacific taxa. It was found that the deepest genetic split divided gulls broadly into a 'North Atlantic' clade (containing well-differentiated groups of DNA sequences from individuals of *michahellis*, *armenicus*, *argentatus* but also containing a group of DNA sequences from Great Black-backed Gulls *L. marinus* and European individuals of Glaucous Gull *L. hyperboreus*) and an 'Aralo-Caspian' clade, containing everything else (see fig. 5). If the mtDNA evidence truly reflects the evolution of these gulls, then the Herring Gull is not monophyletic, but has both Aralo-Caspian and North Atlantic origins.

The large white-headed gulls probably evolved against a background of periodic glaciations which restricted northern birds to isolated refugia, splitting once-widespread species into smaller subpopulations where they then evolved in isolation. During warmer interglacial periods, their ranges would expand again (see Collinson 2001). Birds from the North Atlantic refugium, the possible ancestors of the subspecies *atlantis*, expanded into the Mediterranean and along both the European and the American Atlantic coasts, and differentiated to form Armenian Gull *L. armenicus*, Yellow-legged Gull *L. m. michahellis/atlantis*, European Herring Gull and Great Black-backed Gull. Birds from the Aralo-Caspian refugium migrated north and west, where they evolved into *heuglini* and the Lesser Black-backed taxa; and east, where they evolved into other Siberian and Pacific taxa. Several interesting points emerged from the

analysis. In particular that the American Herring Gull *smithsonianus* is not closely related to European Herring Gulls *argentatus/argenteus*, but is a member of the Siberian group of gulls descended from the Aralo-Caspian refugium. This suggests that the original interpretation of a ring species was wrong, because if 'Herring Gull' was a ring species, *argentatus/argenteus* and *smithsonianus* would have to be closely related. Similarly Caspian Gull *L. a. cachinnans* and Yellow-legged Gull are apparently not closely related – they have also evolved from different refugial populations.

Another interesting point was that the 'herring gull' phylogeny contained other non-controversial 'good' species embedded within it, such as Great Black-backed Gull, but also (within the 'Siberian' group), individuals of Glaucous-winged Gull *L. glaucescens*, Iceland Gull *L. glaucoides* and Slaty-backed Gull *L. schistisagus*. This suggests that the rapid evolution of morphology and reproductive isolation in gull taxa is not well correlated with genetic distance between the taxa – new species may be able to evolve very rapidly in the absence of much genetic change (even in the highly variable bits of their mtDNA). This was rather counterintuitive, since one might expect that if these gulls were well-formed species, then the clades of similar DNA would correspond to morphologically recognisable species or subspecies; but this was not always the case. Among the Siberian/American group, for example, individuals from nine different taxa representing five uncontroversial species had virtually identical mtDNA.

The concept of a ring species of northern 'herring gulls' is not entirely dead, but it now excludes European Herring Gull *L. argentatus*! The pattern of DNA divergence of the 'Aralo-Caspian'-derived clade is consistent with a 'broken ring', from Lesser Black-backed Gulls in northwest Europe, with no undisputed species boundaries along the (stepped) cline of *fuscus–heuglini–taimyrensis–vegae–smithsonianus*. If Lesser Black-backed Gulls *L. f. graellsii* continue their colonisation of northeast North America and breed alongside *smithsonianus*, then some sort of ring-like species scenario may once again be completed.

Lots of other interesting and as yet unresolved hypotheses about gull taxa can be postulated on the basis of recent genetic data. These gulls demonstrate that any simplistic approach to analysing their DNA is not going to reveal their true phylogeny. Much more work needs to be done, perhaps with nuclear DNA, before there is any chance of uncovering the whole story.

Wagtails

It is perhaps not a coincidence that another group of closely related taxa, the wagtails *Motacilla*, which share some aspects of their glacial evolutionary history with the gulls (i.e. recent interglacial expansion from southern refugia), also share a confusing mtDNA phylogeny. Furthermore, their mtDNA, nuclear DNA and morphology all tell different stories, though results from morphological studies are closer to those from nuclear DNA (Voelker 2002; Pavlova 2003; Odeen & Bjorklund 2003; reviewed in Alström *et al*. 2003). Analysis of both nuclear and mtDNA suggested that there is a genetic divide between the eastern and western forms of Yellow Wagtail *M. flava*, which mostly correlates with marked differences in their vocalisations. In terms of their genetics, 'eastern' yellow wagtails include *M. f. tschutschensis*, *taivana* and *macronyx*, whereas 'western' subspecies are *flava*, *flavissima*, *iberiae*, *cinereocapilla*, *pygmaea*, *feldegg*, *lutea*, *leucocephala* and probably *thunbergi* (although eastern populations of 'Grey-headed Wagtail' *M. f. thunbergi* may fall within the eastern clade). Furthermore, the phylogeny derived from mtDNA shows that the conventional 'Yellow Wagtail' species turns out to be paraphyletic, because subspecies of Citrine Wagtail *M. citreola*, Grey Wagtail *M. cinerea* and the White Wagtails *M. alba*, and possibly other taxa, are nested within and between the eastern and western clades of *M. flava*. Interestingly, there is little if any genetic evidence for splitting 'Black-headed Wagtail' *M. f. feldegg* from other 'western' yellow wagtails. A case could be made for splitting at least the eastern and western *M. flava* clades, even though each group has forms with morphology similar to the other. This has been done by the American Ornithologists' Union (Banks *et al*. 2004), but Alström *et al.* (2003) considered that more research is needed before taking such a step. It is suggested that the Yellow Wagtails evolved their bewildering variety of male plumages, in the absence of much genetic change, as a result of unpredictable forces of sexual selection as pioneer birds advanced northwards during interglacials (Voelker 2002; Odeen & Bjorklund 2003). Under these conditions of low population density and perhaps reduced natural selective forces, male plumage evolution was driven rapidly by the sexual 'whims' of the females.

DNA sequence from the control region, 12S ribosomal and the ND2 gene suggested that the *citreola* and *calcarata* subspecies of Citrine Wagtail were not closely related (Odeen & Bjorklund 2003), and if we acted on this evidence, they would have to be split. However, nuclear DNA (from the CHD1Z gene) produced a more 'sensible' phylogenetic tree, which placed these two taxa next to each other, as close relatives. This illustrates how cautious we have to be when interpreting genetic studies; major taxonomic changes can be adopted safely only when supported by multiple lines of evidence. Birds may carry the 'wrong' mitochondrial DNA as a result of previous hybridisation with other taxa.

If wagtails and gulls demonstrate how plumage variation and reproductive isolation may evolve rapidly without leaving a clear, or even particularly meaningful, genetic paper-trail, then perhaps even more extreme examples may ultimately rewrite the speciation textbooks.

Galapagos finches and Crossbills

The species cluster of 'Darwin's finches' *Geospiza*, *Camarhynchus*, *Certhidea* and *Pinaroloxias* has been regarded for many years as a classic example of island species evolution (see Grant 1986 for a general account). The cluster consists of 14 well-described species, of which 13 live on the Galapagos Island archipelago and one on Cocos Island, almost 800 km to the northeast. They are all apparent descendants of a single ancestral mainland species, presumed to be a relative of the Dull-coloured Grassquit *Tiaris obscurus*.

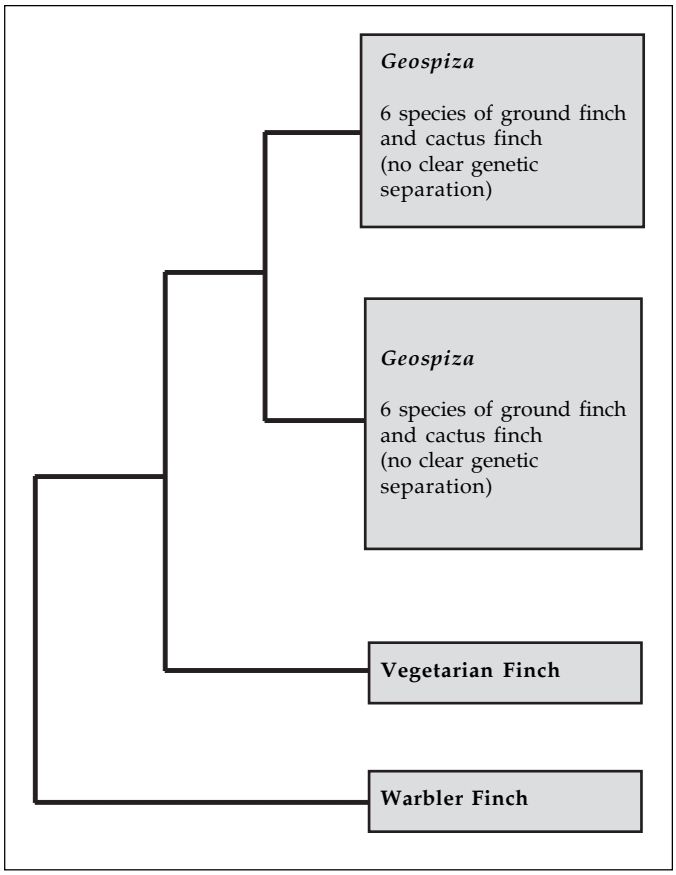

Fig. 6. Molecular phylogeny of Galapagos and Cocos Island Finches, simplified from Sato *et al.* (1999).

Sato *et al*. (1999) published the results of a substantial study of all of these birds using two distinct mitochondrial sequences for comparison, the *cytochrome b* region and the control region. They confirmed that the Cocos Island Finch *Pinaroloxias inornata* belongs with the Galapagos finch grouping, supporting a line of descent from a common ancestor with the Dull-coloured Grassquit; and they identified the Warbler-Finch *Certhidea olivacea* and Vegetarian Finch *Camarhynchus crassirostris* as being slightly distant from the main group of tree finches. Indeed, the Warbler Finch seems on molecular grounds to be basal to the Darwin's finch assemblage (fig. 6).

In contrast, the six species of ground finch *Geospiza* continue to hybridise frequently, and are not well-separated genetically. This was something of a surprise, although an earlier study based on protein **allozyme** (allozymes are differing forms of an enzyme encoded by different alleles at the same locus) variation had also suggested this (Yang & Patton 1981).

The different species of ground finch differ in body size and bill depth, and are presumably adapted for different food items. However, long-term research has suggested that in most years there may be such an abundance of easily accessible food that selection in this respect may be weak, and that virtually all birds may survive, whatever their bill size or shape. In these conditions, there may be little or no selection against hybrids, and even relatively rare hybridisation would allow for gene flow between species, obscuring their molecular phylogenies. In years of drought, however, when food is scarce, strong selection pressure may act to eliminate birds with suboptimal bill sizes.

The Medium Ground Finch *Geospiza fortis* recently gave a dramatic demonstration of just how fast a population can evolve. When, in 1977, a serious drought caused plants yielding smaller seeds (the food supply of smaller individuals) to fail, on Daphne Island the mean body size and bill size of the population increased significantly within one year (Boag & Grant 1981). Most of the smaller birds died as they could not cope with larger and tougher seeds. In subsequent seasons, however, with the return of the small-seed food supply, mean body size slowly decreased again. Had the drought lasted sufficiently long, then alleles for small size could have been eliminated completely and the mean size would not have decreased again so quickly, if at all.

The situation of this finch, where the processes driving speciation occur rapidly, all the time, may give some insights into the evolutionary relationships among the crossbills *Loxia*. Based on nuclear microsatellite DNA and mitochondrial control regions, there are no clear differences among the DNA of any of the three currently recognised west European species, Parrot *L. pytyopsittacus*, Scottish *L. scotica* and Common Crossbill *L. curvirostra* (Piertney *et al.* 2001), somewhat similar to the lack of genetic differentiation among *Geospiza* finches. These three crossbills are diagnosable only by mean body size, bill shape and call. As in the case of the bills of Galapagos ground finches, the three mean bill sizes of these crossbill taxa may be optimal for different sizes of cone, but this does not restrict their feeding most of the time, when cones are abundant. It is possible that bill size is selected strongly only when cones are scarce or if the environment is particularly harsh. Their preferences for different-sized cones are sufficient to restrict, but not eliminate, the gene flow between them so that bill size never becomes distinctly different in each form.

A comparable lack of genetic differentiation was reported among redpoll *Carduelis* taxa (Ottvall *et al*. 2002). For redpolls, in common with crossbills, it may be that the different species have evolved recently. Furthermore, their nomadic lifestyles (which allow for gene flow across large geographic distances), combined with generally large populations (which make it difficult for new mutations to spread throughout the populations), and occasional hybridisation, prevent very much genetic differentiation becoming established, even between good morphological species.

Cuckoos

There are other avian examples where genetically distinguishable forms are not necessarily different species. One fascinating example of this is the female Common Cuckoo *Cuculus canorus*. Female Common Cuckoos are genetically distinguishable according to their host species (Gibbs *et al*. 2000); however, males are not similarly distinguishable, and they will mate with a female of any kind. The different host-specific females are known as **gentes** (singular gens). Present evidence suggests that the gene which determines egg colour is in the female-specific W-chromosome. The ability to lay eggs in nests of the right

host is learned behaviour (imprinting). Apart from the W-chromosome, a Meadow Pipit *Anthus pratensis* Cuckoo female can pass its nuclear genes to a Reed Warbler *Acrocephalus scirpaceus* Cuckoo granddaughter via a male son. For a full account of the cuckoos, see Davies (2000).

Where does the future lie?

Most of the easy taxonomic decisions in ornithology have probably already been made. It is a fact that more significance is given to the rank of species than to any other rank, whether by birdwatchers, professional ornithologists, or government environmental ministries. This is a pity, as the importance of each classifiable form should be assessed independently of whether current taxonomy considers them species or not (Collinson 1999). Bird taxonomy is in perpetual flux. This is normal and perhaps desirable, if it is a symptom of discovery of new relevant facts by continuing research. Ornithologists in the Western Palearctic have perhaps forgotten this, having been spoilt by the stability engendered by the Voous List of Holarctic birds. Stability becomes stagnation when it stifles the application of relevant research. While there are perhaps some arguments for managing the pace of taxonomic change for practical reasons, stability *per se* is not a good thing. We would argue that, while unnecessary change is to be avoided, unnecessary stability may be equally harmful to ornithology. Nor is it always 'safer' to lump than to split, for the same reasons. Furthermore, taxonomic decisions are *hypotheses* – it is not necessary to wait until the evidence becomes overwhelming before proposing a taxonomic change; it is sufficient that the proposal is justified after taking into account all the relevant known facts or, alternatively, to await more facts when the conclusion looks improbable.

What we hope we have shown in this paper is that although the advent of DNA studies and other molecular techniques has enabled enormous strides in determining the inter-relationships between species and higher-level categories, DNA is not a panacea. There are some sets of species that have indistinguishable DNA, at least on the basis of the sequences so far analysed, for example the crossbills, as well as single species which have highly diverse DNA in their gene pools. Thus it is necessary to apply the same standards of critical evaluation to molecular evidence as for evidence based on aspects of morphology, vocalisations, behaviour, ecology and physiology.

Taxonomy is a science; but it is performed by humans and its end-product, classification, is, at least in part, for humans. The challenge for taxonomists is to get their science and their judgements right, in a climate of political pressure. When Fea's *Pterodroma feae* and Zino's Petrels *P. madeira* were split (Bourne 1983), the scientific underpinnings of the decision were perhaps of less immediate significance than the subsequent attention directed to the plight of these two relict populations. More recently, proposed splits of albatross species (e.g. Shirihai 2002) may well attract attention for the same reason. And what would be the consequences of lumping Scottish Crossbill (Britain's only endemic bird species) with Common Crossbill (or even with both the other west European species) on the conservation status of the forests in which they live? We are not advocating political taxonomy, but the political consequences of taxonomic decisions are sometimes far-reaching.

Acknowledgments

We thank Richard Porter for encouraging us to start this review, and both Prof. Ian Newton and Dr Stephen Votier for helpful comments and suggestions. We would like to thank the many people whose ideas have influenced the text. We especially thank a forthright colleague who wished to retain anonymity.

References

Alström, P., Mild, K., & Zetterström, D. 2003. *Pipits and Wagtails of Europe, Asia and North America.* Christopher Helm, London.

Andersson, M. 1999. Hybridisation and skua phylogeny. *Proc. Roy. Soc. Lond.* B 266: 1579–1585.

Avise, J. C. 2004. *Molecular Markers, Natural History, and Evolution*. 2nd edn. Sinauer Associates, Sunderland, MA.

Baldauf, S. L. 2003. Phylogeny for the faint of heart: a tutorial. *Trends Genet*. 19: 345–351.

Banks, R. C., Cicero, C., Dunn, J. L., Kratter, A. W., Rasmussen, P. C., Remsen, J. V., Rising, J. D., & Stotz, D. F. 2004. Forty-fifth supplement to the American Ornithologists' Union Check-list of North American Birds. *Auk* 121: 985–995.

Boag, B. T., & Grant, P. 1981. Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galapagos. *Science* 214: 82–85.

Bourne, W. R. P. 1983. The Soft-plumaged Petrel, the Gon-gon and the Freira, *Pterodroma mollis*, *P. feae* and *P. madeira*. *Bull. BOC* 103: 52–58.

Braun, M. J., & Brumfield, R. T. 1998. Enigmatic phylogeny of the skuas: an alternative hypothesis. *Proc. Roy. Soc. Lond.* B 265: 995–999.

Catchpole, C. K., & Rowell, A. 1993. Song sharing and local dialects in a population of European Wren *Troglodytes troglodytes*. *Behaviour* 125: 67–78.

Cohen, B. L., Baker A. J., Bleschschmidt K., Dittmann, D. L., Furness, R. W., Gerwin, J. A., Helbig, A. J., De Korte, J., Marshall, H. D., Palma R. L., Peter, H. U., Ramli, R., Siebold, I., Willcox, M. S., Wilson R. H. & Zink, R. M. 1997. Enigmatic phylogeny of skuas (Aves: Stercorariidae). *Proc. Roy. Soc. Lond.* B 264: 181–190.

Collinson, M. 1999. Subspecies – more than meets the eye. *Brit. Birds* 92: 118–119.

— 2001. Shifting Sands: taxonomic changes in the world of the field ornithologist. *Brit Birds* 94: 1–24.

—, Knox, A. G., Parkin, D. T., & Sangster, G. 2003. Specific status of taxa within the Greenish Warbler complex. *Brit. Birds* 96: 327–331.

Cracraft, J. 1997. Species concepts in systematics and conservation – an ornithological viewpoint. In: Claridge, M. F., Dawah, H. A., & Wilson, M. R. (eds.), *Species: the units of biodiversity*: 325–339. Chapman & Hall, London.

Crochet, P., Lebreton, J., & Bonhomme, F. 2002. Systematics of large white-headed gulls: patterns of mitochondrial DNA variation in Western European taxa. *Auk* 119: 603–620.

—, Chen, J. J. Z., Pons, J. M., Lebreton, J. D., Hebert, P. D. N., & Bonhomme, F. 2003. Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution* 57: 2865–2878.

Davies, N. B. 2000. *Cuckoos, Cowbirds and Other Cheats*. Poyser, London.

Dover, G. A. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111–117.

Fain, M. G., & Houde, P. 2004. Parallel radiations in primary clades of birds. *Evolution* 58: 2558–2573.

Fleischer, R. C., & McIntosh, C. E. 2001. Molecular Systematics and Biogeography of the Hawaiian avifauna. *Stud. Avian Biol*. 22: 51–60.

Fryer, G., & Iles, T. D. 1972. The Cichlid Fishes

of the Great Lakes of Africa: their biology and evolution. Oliver & Boyd, London.

Furness, R. W. 1987. *The Skuas.* Poyser, Calton.

Garant, D., Kruuk, L. E., Wilkin, T. A., McCleery, R. H., & Sheldon, B. C. 2005. Evolution driven by differential dispersal within a wild bird population. *Nature* 433: 60–65.

Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

Gibbs, H. L., Sorenson, M., Marchetti, K., Brooke, M. de L., Davies, N. B., & Nakamura, H. 2000. Genetic evidence for female host-specific races of the Common Cuckoo. *Nature* 407: 183–186.

Grant, P. J. 1986. *Gulls: a guide to identification.* 2nd edn. Poyser, Calton.

Grant, P. R. 1986. *Ecology and Evolution of Darwin's Finches.* Princeton University Press, Princeton.

— & Grant, B. R. 1992. Hybridisation of bird species. *Science* 256: 193–197.

— & — 1997. Genetics and the origin of bird species. *Proc. Natl. Acad. Sci. USA* 94: 7768–7775.

Greenwood, J. J. D. 1997. The diversity of taxonomies. *Bull. BOC* 117: 85–96.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & de Waard, J. R. 2003. Biological identifications through DNA barcodes. *Proc. Roy. Soc. Lond.* B 270: 313–322.

—, Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. 2004. Identification of birds through DNA Barcodes. *PloS Biology* 2: 1657–1663.

Helbig, A. J., Seibold, I., Martens, J., & Wink, M. 1995. Genetic differentiation and phylogenetic relationships of Bonelli's Warbler *Phylloscopus bonelli* and Green Warbler *Ph. nitidus*. *J. Avian Biol.* 26: 139–153.

—, Knox, A. G., Parkin, D. T., Sangster, G., & Collinson, M. 2002. Guidelines for assigning species rank. *Ibis* 144: 518–525.

—, Martens, J., Seibold, I., Henning, F., Schotter, B., & Wink, M. 1996. Phylogeny and species limits in the Palearctic Chiffchaff *Phylloscopus collybita* complex: mitochondrial genetic differentiation and bioacoustic evidence. *Ibis* 138: 650–666.

Irwin, D. E. 2000. Song variation in an avian ring species. *Evolution* 54: 998–1010.

—, Bensch, S., & Price, T. D. 2001. Speciation in a ring. *Nature* 409: 333–337.

—, —, Irwin, J. H., & Price, T. D. 2005. Speciation by distance in a ring species. *Science* 307: 414–416.

James, H. F., Erikson, P. G. B., Slikas, B., Lei, F., Gill, F. B., & Olson, S. L. 2003. *Pseudopoces humilis*, a misclassified terrestrial tit of the Tibetan plateau. *Ibis* 145: 185–202.

Johnson, K. P., & Sorenson, M. D. 1999. Phylogeny and biogeography of dabbling ducks (genus *Anas*): a comparison of molecular and morphological evidence. *Auk* 116: 792–805.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

König, C., Weick, F., & Becking, J-H. 1999. *Owls – a guide to the owls of the world*. Pica Press, Robertsbridge.

Lewin, B. 2002. *Genes VII*. Oxford University Press/Prentice Hall, Oxford.

Liebers, D., Helbig, A. J., & de Kniff, P. 2001. Genetic differentiation and phylogeography of gulls in the *Larus cachinnans–fuscus* group (Aves: Charadriiformes). *Mol. Ecol.* 10: 2447–2462.

—, de Knijff, P., & Helbig, A. J. 2004. The Herring Gull complex is not a ring species. *Proc. Roy. Soc. Lond.* B 277: 893–901.

Majerus, M., Amos, W., & Hurst, G. 1996. *Evolution: the four billion year war*. Longman: Harlow, UK.

Mayden, R. L. 1997. A hierarchy of species concepts: the denouncement in the saga of the species problem. In: Claridge, M. F., Dawah, H. A., & Wilson, M. R. (eds.), *Species: the units of biodiversity*: 381–424. Chapman & Hall, London.

Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, Mass.

— 1987. The ontological status of species: scientific progress and philosophical terminology. *Biol. Philos.* 2: 145–166.

Mendel, G. 1866. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brunn* 4: 3–47.

Mindell. D. P. (ed.) 1997. *Avian Molecular Evolution and Systematics*. Academic Press, London.

—, Sorenson. M. D., & Dimcheff, D. E. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA* 95: 10693–10697.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press.

Newton, I. 1972. *Finches*. Collins, London.

— 2003. *The Speciation and Biogeography of Birds*. Academic Press, London.

Odeen, A., & Bjorklund, M. 2003. Dynamics in the evolution of sexual traits: losses and gains, radiation and convergence in Yellow Wagtails (*Motacilla flava*). *Mol. Ecol.* 12: 2113–2213.

Ottvall, R., Bensch, S., Walinder, G., & Lifjeld, J. T. 2002. No evidence of genetic differentiation between Lesser Redpolls *Carduelis flammea cabaret* and Common Redpolls *Carduelis f. flammea*. *Avian Science* 2: 237–244.

Parkin, D. T. 2003. Birding and DNA: species for the new millennium. *Bird Study* 50: 223–242.

Pavlova. A., Zink, R. M., Drovetski, S. V., Red'kin, Y., & Rohwer, S. 2003. Phylogeographic patterns in *Motacilla flava* and *Motacilla citreola*: species limits and population history. *Auk* 120: 744–758.

Paxinos, E. E., James, H. F., Olson, S. L., Sorensen, M. D., Jackson, J., & Fleischer, R. C. 2002. MtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada Goose (*Branta canadensis*). *Proc. Natl. Acad. Sci. USA* 99: 1399–1404.

Piertney, S. B., Summers, R., & Marquiss, M. 2001. Microsatellite and mitochondrial DNA homogeneity among phenotypically diverse crossbill taxa in the UK. *Proc. Roy. Soc. Lond.* B 268: 1511–1517.

Ridley, M. 1986. *Evolution and Classification: the reformation of cladism*. Longman, London.

Ryder, O. A. 1986. Species conservation and systematics: the dilemma of subspecies. *Trends Ecol. Evol.* 1: 9–10.

Sangster, G., Collinson, M., Helbig, A. J., Knox, A. G., Parkin, D. T., & Prater, T. 2001. The taxonomic status of Green-winged Teal. *Brit. Birds* 94: 218–226.

Sato, A., O'Huigin, C., Figueroa, F., Grant, P. R., Grant, B. R., Tichy, H., & Klein, J. 1999. Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proc. Natl. Acad. Sci. USA* 96: 5101–5106.

Shirihai, H. 2002. *Complete Guide to Antarctic Wildlife*. Alula Press, Finland.

Sibley, C. G., & Ahlquist, J. E. 1990. *Phylogeny and Classification of Birds*. Yale University Press, New Haven.

—, —, & Monroe, B. L. 1988. A classification of the living birds of the world based on DNA–DNA hybridization studies. *Auk* 105: 409–423.

Tokeshi, M. 1999. *Species Coexistence: ecological and evolutionary perspectives*. Blackwell, Oxford.

US Fish & Wildlife Service. 1999. *Revised Recovery Plan for Hawaiian Waterbirds, Second Revision.* Portland, Oregon (http://ecos.fws.gov/docs/recovery_plans/1999/990709.pdf).

Vaurie, C. 1959. *The Birds of the Palearctic Fauna*. Witherby, London.

Voelker, G. 2002. Systematics and historical biogeography of wagtails: dispersal versus vicariance revisited. *Condor* 104: 725–739.

Yang, S. Y., & Patton, J. L. 1981. Genic variability and differentiation in the Galapagos finches. *Auk* 98: 230–242.

Yésou, P. 2002. Systematics of *Larus argentatus-cachinnans-fuscus* complex revisited. *Dutch Birding* 24: 271–298.

Appendix 1
A 'cheat-sheet' for those trying to understand the primary literature

Constructing phylogenetic trees from DNA sequence data
This appendix describes the basics of phylogenetic analysis, based on direct sequencing of DNA, to reveal the sequence of As, Ts, Cs and Gs, as described in the text. Many of the principles described here are, however, also applicable to other methods that analyse the lengths of DNA fragments without identifying the exact sequences.

(i) Sequence the same region of DNA in individuals of all the taxa you wish to include in the analysis
There are, at the first level of analysis, two problems here. The first is to ensure that you are really sequencing the same (orthologous) bit of DNA in all taxa. During evolution, DNA duplication events have led to extra copies of genes being created. Sometimes, these extra genes remain functional (gene families), with different members accumulating different mutations. Sometimes the extra genes may become non-functional and accumulate new mutations rapidly – they look somewhat like functional genes, but have no role ('**pseudogenes**'). Some mitochondrial genes may have lookalike copies in the cell nucleus. At the molecular level, it is possible to get all these variants confused, and if the same single gene is sequenced in, say, five taxa, but by accident a mutant pseudogene is sequenced in a sixth closely related taxon, the molecular analysis is going to suggest, erroneously, that this sixth species is only distantly related to the other five. For this reason, phylogenetic papers will often describe the precautions taken to ensure that orthologous sequences were compared, usually by looking hard for non-orthologous copies and making sure that, if they were present, they were found and identified.

The second problem is that, as explained in the text, different individuals from the same taxon will often have slightly different DNA sequences (different alleles). The scientist needs ideally to sample several individuals from the same species or subspecies to quantify the intra-taxon variation. It may be necessary to produce a 'consensus' (average) sequence for each taxon. Ideally, the intra-taxon variation should be small compared with the variation among the taxa to be analysed.

(ii) Line up of DNA sequences ready for analysis
If one has a DNA sequence from one taxon (for example ten bases ATCGATCGAG), then the power of phylogenetic analysis is maximised only when it is possible to be confident that each base can be compared with the equivalent base in other taxa. This would be easy if no bases were ever gained or lost. If there were only ever going to be ten bases in the orthologous sequence in other taxa, then if a sequence from another species or subspecies of ATCGATGGAG was derived, we would know that there had been one mutation in the second taxon (a C to G substitution at position 7). As described in the main text, however, it is also possible, within the normal range of mutations, to lose bases (**deletions**) or gain them (**insertions**). If, in a third taxon, the C at position 7 has been deleted, we get a sequence of ATCGATGAG. Lining that up simply against the first sequence gives the impression that in fact four mutations (40% divergence!) have occurred – three base substitutions and a deletion, i.e.

ATCGATCGAG
ATCGATGAG–

The phylogeneticist, or his computer, has to spot that, in fact, it is more parsimonious to assume that only one mutation has occurred, the deletion at position 7 (a 10% divergence), and line up the sequences accordingly.

ATCGATCGAG
ATCGAT–GAG

This is the process of sequence alignment, and is critical to the process of obtaining reliable phylogenetic information. While this may seem obvious, it has to be mentioned here because sequence alignment can be extremely tricky when more than about five sequences are compared. Furthermore, the last step in this largely computerised process remains to check them and adjust by eye. If the sequence alignment produces an unexpected phylogenetic tree, it may be necessary to go back and realign the sequences. This is an obvious source of subjectivity and potential error.

Insertions or deletions, leading to gaps in sequences, are a biological big deal. They often lead to a gene becoming completely inactive, so conceptually the creation of a new gap is rarer than a simple swapping of one base for another in the DNA. So, although sequence comparisons may ignore sites with gaps (see below), the presence or absence of different gaps in different taxa is a valuable source of phylogenetic information; if two taxa in a family share a derived gap in sequence that is not shown by their other relatives, it is extremely strong evidence that they are sister taxa. Indeed, it is possible to create large phylogenies based on the presence or absence of gaps (Fain & Houde 2004). For the purpose of sequence analysis, however, gaps and the bases immediately adjacent to them are normally ignored.

(iii) Analysis of the sequence homology
If the sequence alignment has been performed correctly, the phylogeneticist should ideally be left with complete orthologous sequences of the same length from all the taxa to be examined. These sequences can now be fed into a computer to determine which are the most similar, and/or what phylogenetic tree linking the taxa can best explain the observed pattern of mutations in the DNA. The mathematics are complicated, and the details of no interest to the average birder, but the methods can be classified broadly into two groups (Nei 1983).

Analyses that link the most closely related sequences together and piece by piece build up a tree of 'sister groups' are called ***distance-matrix methods*** – they appear in papers as 'Unweighted Paired-group Means Analysis' (UPGMA), 'Fitch-Margoliash method', or 'neighbour-joining'. Put simply, they measure the genetic divergence between different

sequences and put the most similar ones together.

The second set of analyses, ***discrete data methods***, regard each position in a set of sequences from the different taxa as a small piece of information about the evolutionary history of the taxa, and build a tree on the basis of a consensus of sum total of all these little pieces of information from all the orthologous positions in the sequence. They make certain assumptions about the nature of mutation – which mutations are rarer than others, for example – and reconstruct a tree based on the minimum amount of evolution required to explain the observed sequences ('maximum parsimony'), or which phylogenetic tree is most likely, given what we think we know about the processes of molecular change ('maximum likelihood').

If analyses of the same data by two or more methods, from both classes of analysis described above, give the same results, this increases our confidence in the conclusions.

(iv) Testing the phylogenetic tree

Once the computers have done their work, the resulting tree is meant to reflect the relationships between the taxa included in the DNA analysis. But how do we know whether the tree is reliable? Were the molecular data clear-cut, such that the computer could really reconstruct only one plausible tree; or were the DNA sequences so confusing that the computer could not really sort them out, and has just made the best of a bad job? Each phylogenetic tree needs testing for reliability, and the most universal way of doing this is by a technique known as '**bootstrapping**'; so called because it can obtain reliable statistical information out of almost any dataset, hence 'pulling it up by its bootstraps'.

If 100 bases of DNA sequence have been compared among, say, four taxa, then in fact there are 100 independent pieces of information, each suggesting a phylogeny. The consensus phylogenetic tree is the combined 'voice' of these 100 parts. For example, the preferred phylogenetic tree may suggest that two species, A and B, are sister taxa, more closely related to each other than to any other taxa. If this result is clear-cut and obvious from the 100 base-pair sequence comparison, then it is argued that it should also be quite obvious from a subset of the data – so if, instead of taking 100 pieces of information, we take 50, then A and B should still emerge as sister taxa. This is essentially what bootstrapping does. Thus for a 100 bp sequence, the computer would pick 100 random numbers between 1 and 100 (note that, like rolling a dice six times, some numbers might come up more than once, and some not at all) so that a subset of unique sites in the sequence is derived. The phylogeny would then be reconstructed using *only* the DNA bases at those randomly selected positions. The process might be repeated 100 or 1,000 times, and the percentage of times that A and B emerge as sister taxa on the basis of these random subsets of the data is calculated. This percentage can be presented unmodified as 'bootstrap support' and shown on the consensus phylogenetic tree at the node (branch point) between A and B. If the relationship between A and B is robust, we would expect bootstrap support of 90–100%, and certainly more than 70%. Anything less than 70% has to be looked at critically; it suggests there are other arrangements of taxa in the phylogenetic tree that might equally be plausible.

Ideally, in published molecular phylogenetic trees using methods that are open to bootstrapping, the bootstrap support for each node will be shown. The closer the number is to 100, the more you can believe it.

This appendix is based loosely on Baldauf (2003).

Waterbird surveys along coastal Myanmar

Thet Zaw Naing

Naing, Thet Zaw. 2006. Waterbirds survey in mouth of Yangon River and Ayeyarwaddy (Irrawaddy) delta. *Indian Birds* 2 (3): 65-71.
Naing, Thet Zaw, Secretary, Myanmar bird and Nature Society. Email: *sst@mptmail.net.mm.*

ABSTRACT In December 2005 and March 2006, waterbirds surveys were conducted at various points in the mouth of Yangon River and in the Ayeyarwaddy (Irrawaddy) Delta. One new species and one new subspecies for Myanmar were recorded namely: Great Frigatebird *Fregata minor* and Long-tailed Shrike *Lanius s. schach* respectively. In addition, three globally threatened species—Spotted Greenshank *Tringa guttifer,* Lesser Adjutant-Stork *Leptoptilos javanicus,* and Greater Spotted Eagle *Aquila clanga*—and four Near-threatened species—Darter *Anhinga melanogaster*, Painted Stork *Mycteria leucocephala,* Oriental White Ibis *Threskiornis melanocephalus* and Brown-winged Kingfisher *Halcyon amauroptera*—were also recorded.

Introduction

The wetlands of Myanmar include rivers and streams, shallow fresh water lakes and marshes, water storage reservoirs, fish ponds, seasonal flooded cultivated plains, and estuaries with extensive mangrove swamps. With a 2,278 km long coast, most wetlands in Myanmar are directly or indirectly associated with river systems, several very large estuarine and delta systems and numerous offshore islands. There are an estimated 517,000 ha of mangrove forest in Myanmar, mostly located in Ayeyarwaddy (Irrawaddy) Delta, on the Tanintharyi and Rakhine coasts and offshore islands.

The Ayeyarwaddy Delta covers an estimated 1,100,000 ha and is located in Ayeyarwaddy and Yangon (Rangoon) divisions. The mouth of the Yangon River is situated in Yangon division in the gulf of Mottama. These sites fall within the southern part of the ornithological regions in Myanmar (Robson 2000). These areas have been very poorly covered by ornithological surveys (Armstrong 1876; Inskipp et al. 2001, 2003; Salter 1982; Naing & Aung 2002; van der Ven & Naing 2005).