1

# Introduction

There have been many authorities who have asserted that the basis of science lies in counting or measuring, i.e. in the use of mathematics. Neither counting nor measuring can however be the most fundamental processes in our study of the material universe – before you can do either to any purpose you must first select what you propose to count or measure, which pre-supposes a classification.

(Crowson 1970, p. 2)

It is perhaps an analytic statement - that is, self-evidently true - that the only way in which members of our species, Homo sapiens, can order their perceptions of the world and the ideas to which they give rise is to produce a classification. There is every reason to believe that this need also applies to other animals and is by no means confined to those animals closely related to man. It is literally vital to any animal that it must have a series of metaphorical compartments in which it places perceived phenomena - food, drink, shelter, danger, own species (sub-divided as parent, sib, rival, sexual partner, etc.). We also know that with greater intelligence, learning ability, and general mental flexibility, the classifications employed by higher animals, although articulated only in our own species, are more complex and subtle than those of lower animals. In humankind our knowledge is ordered in many precise, ranked classifications, although the individual person is unlikely to be aware of the structure of the classification he or she is using.

# 2 Introduction

In the simple example above, two ranks are specified. "Own species" is of a higher rank than those items included within it ("parent, sib, rival, sexual partner ... "). But it is also easy to see that "own species" could be included with other items to compose a taxon of higher rank. That higher taxon could be "all living things", or, more limited but precise, "all living things of about my size/shape/colour/pattern/behaviour". Furthermore, any item in the lower rank could be subdivided into two or more taxa at a lower rank still - for example, "rival whom I can intimidate" versus "rival who can intimidate me". One can now see that it is possible to draw a diagram of the pattern of such a classification, a dendrogram, with a number of ranks each at a different level and with the taxa at each rank grouped together to compose a smaller number of taxa at the next higher rank. The pattern of the dendrogram would look like that of a human family tree of the sort that shows all the descendants of a single patriarch, but whereas the latter is a pedigree representing a sequence of generations in time, the classification is a static inclusive hierarchy.

In my discussion of the classificatory hierarchy. I used the word "taxon" (plural "taxa"). Strictly speaking, this term should be confined to a particular type of classification, the systematic classification of organisms - animals, plants, fungi, and so on. Organisms can of course be classified in a number of ways - by size, by their ecology, or by their use or danger to man: the word "fish", as in food sold by a fishmonger, means something very different in the English language from "fish" (the obsolete class Pisces) as used by a professional taxonomist. The job of the taxonomist is, amongst other things, to produce systematic classifications of groups of organisms, or, in other words, to elucidate their relationships. If that taxonomist believes that he or she is literally producing a pattern of relationships rather than simply a taxonomic hierarchy, then this signals an acceptance of the theory that evolution has occurred. A theory of evolution, in the broadest sense, is a theory that the apparent relationships of classification are real, indicating community of descent.

I have been giving zoology tutorials in the University of Newcastle upon Tyne for nearly thirty-five years. Until recently, I often began one of the first meetings of the year by asking students the question, What is the theory of evolution? Now I tend to *tell* them rather than *ask* them, for in all those years I have never had

# Introduction

an immediate and satisfactory reply. My experience was the same when I was giving undergraduate courses as a visiting professor in both the United States and Australia. In every case I was dealing with intelligent students who had been given a grounding in biology at school and, in most cases, for at least a year at university. It is only fair to add that generally we came up with a satisfactory formulation after a short discussion, but the initial lack of response, apart from being dispiriting, strikes me as remarkable - remarkable because evolution is supposed to be the foundation on which all biological science is built, and most academic biologists at least pay lip-service to its importance. But the truth is that little if anything of evolutionary theory is taught at school, at least in Britain and for different reasons in North America; and in many British universities, particularly those with schools of biology dominated by the biochemical-molecular end of the spectrum of levels of analysis, the treatment is so perfunctory as to be contemptible.

So students cannot characterise the theory of evolution. Admittedly some will give a (usually erroneous) account of the theory of Natural Selection, which they will associate (correctly) with the name of Charles Darwin and, much less often, also with that of Alfred Russel Wallace. Natural selection is one component of evolutionary theory as proposed by Darwin and Wallace, but the other, for which selection is merely a hypothesis of mechanism, is the theory that evolution has occurred. But that theory must have been proposed to explain some body of data and/or lowerlevel theories. What I wanted to hear from my students was what that corpus of knowledge was. The answer, as we have seen, is that the theory of evolution states that the apparent relationships of organisms in a systematic classification are real relationships, because "relationship" in such a classification is not a metaphor but is actually to be ascribed to community of descent. Thus the theory of evolution was proposed by Darwin and Wallace to explain the pattern of relationships in what we may now term "Natural Classification".

The purpose of this book is, firstly, to deal with the complex relationship between the concepts of natural classification and evolution, or "transmutation" as Darwin and Wallace and their forerunners would originally have called it. In a broader sense I shall be dealing with what scientists and philosophers saw (and see) as

#### 4 Introduction

the natural order of living things and the way in which they explain that order. Darwin and Wallace saw the order as an irregular hierarchical pattern, but they were not the first evolutionists. Other patterns were proposed and sometimes explained as caused by transmutation. Nor was it always taken *a priori* that the correct pattern of classification should be the same as the natural order of organisms. We must therefore explore both.

But if the theory that evolution has occurred was proposed to explain the phenomenon of natural classification, the latter, in logic, cannot be "evidence for evolution"; so we must ask whether there is any evidence that is not merely composed of taxonomic data.

The enormous diversity of organisms on Earth is unique beyond all reasonable doubt. Furthermore, the differences among individual organisms, among species of organisms, and among the taxa at every rank into which those species can be grouped must, if we accept that evolution has occurred, be based on contingent properties of the organisms that form the basis of the hierarchy. One must therefore ask whether there are any general statements that can be made about living things. One view of the proper business of scientists is that it should be to produce statements of ever increasing generality until a series of statements that are asserted to be universally true is produced. This view of science is claimed to be correct by many physical scientists, by some biologists who suffer from "physics envy", and by those philosophers of science for whom physics is the paradigm science. In the latter part of this book, I shall be asking if there are, or possibly can be, any universal statements in biology.

Tackling such a fundamental question probably seems like *hubris* of a high order. Because of this I have felt constrained to present the evidence on which my opinions are based in some detail, but I hope this detail may be of use to the reader in other ways. Thus I give an account of methods of classification and phylogeny reconstruction which I hope will be accessible to the layperson but still sufficiently up-to-date, detailed, and rigorous to be useful to students of biology and professional colleagues. I have also tried to give each subject I discuss a historical basis, and, while I would not claim that any part of the book is a work of historical scholarship, I have attempted to play fair by making it clear, but without

#### Introduction

overt statement, when I have used primary and when secondary historical sources.

But mostly the book is meant to be one long logical argument. Thus in the spirit of those scholars, particularly professional philosophers, who summarise their work in their introduction in terms of "I will first show...I will then go on to demonstrate ...", I now present an abstract of that argument, itemised as the content of individual chapters.

Thus in Chapter 2: Until early in the nineteenth century, the dominant idea of the natural order was that of a *scala naturae*, an unbroken sequence from the most primitive to the most advanced organisms. Various other patterns were suggested, but the natural order was not accepted as an irregular hierarchy until the time of Darwin and Wallace, despite the fact that the pattern of classification was accepted as hierarchical long before. Subsequent classifications usually claimed to be phylogenetic and emphasised either grade, as in the *scala naturae*, or the pattern of branching. Two recent techniques, phenetics and cladistics, yield dichotomous dendrograms, but the latter is based on a hierarchy of characters, the former on aggregate similarity.

In Chapter 3: Theories of evolution were proposed to explain the authors' perceptions of the natural order. At the beginning of the nineteenth century, Lamarck's original theory, subsequently modified, was formulated to explain the *scala naturae*. Darwin and Wallace emphasised the contingent nature of the hierarchy, but later authors revived the *scala* by accepting an evolutionary pattern of classification similar to the ancient "Tree of Porphyry" pattern of classifications. The *scala* was also present in the search of evolutionists for ancestor-descendant sequences, but cladistics took the hierarchical pattern *a priori*, originally as representing phylogeny but latterly as a pattern to be explained by phylogeny.

In Chapter 4: If the pattern of classification is logically prior to phylogeny, the characters on which it is based should have logical priority over the pattern. There should therefore be a Natural Hierarchy of characters, whose similarity in all the members of a taxon is recognised as *homology*. Diagnostic homologies are *taxic homologies*, whereas characters incongruent with the pattern are *homoplastic*. These concepts pre-date proposals of phylogeny and are based on comparative anatomy, as is that of transformational

#### 6 Introduction

homology. Homology is not evidence for evolution, but the existence of vestigial organs is.

In Chapter 5: Both geology and biogeography yield evidence for evolution that is independent of the pattern of classification. Stratigraphic palaeontology corroborates a picture of "progression" through geological time which cannot be explained by successive catastrophic extinction and creation: evolution is the only rational answer. This is reinforced by chronoclines of transformational homology and of species within major taxa. The patterns of geographical distribution of animals cannot be explained by adaptation to environmental conditions alone; the biogeographical history of organisms must be invoked together with evolution. The cladistic technique of vicariance biogeography reconstructs patterns and sequences of geological events, not of phylogeny. Thus it is not logically prior to phylogeny, but is powerful evidence for evolution.

In Chapter 6: Techniques of classification date back to the logical division of Plato and Aristotle, later to appear as the tree of Porphyry. The same methods were used by Linnaeus but were regarded as a means of summarising knowledge rather than representing the natural order. The two aims were reconciled by the time of Darwin's Origin of Species, but post-Darwinian taxonomy remained largely *ad hoc*, with contradictions between the pattern of classification and that of phylogeny arising from the different emphases given to phyletic evolution (and thus grade), to cladogenesis (the pattern of splitting in evolution), and also to pre-Darwinian tradition.

In Chapter 7: Phenetics and cladistics were rival techniques developed after the middle of this century; their methods are described in some detail. Phenetics claims objectivity and freedom from any theoretical presumptions. Aggregate differences among taxa are represented as distances in character hyperspace, and a hierarchy is constructed from the clusters so formed. There is a null hyothesis of no hierarchical structure in the data. Cladistics (originally "Phylogenetic Systematics") was introduced by Hennig as a technique based on the pattern of speciation, division in evolution of one species into two or more, but with the *a priori* assumption of a hierarchical pattern of taxic homologies. Like phenetics it aims to produce dichotomous dendrograms, but these are based on unique characters at every rank ("shared derived

#### Introduction

characters"). Thus characters have to be "polarised" as unique to a given taxon or characterising a taxon of higher rank – originally by "out-group comparison". Cladograms are "Steiner trees" with real "terminal taxa" (those to be classified) only at the tips of the branches; as phylogeny the internal nodes can only represent hypothetical ancestors. Thus real fossils cannot be recognised as ancestors by cladists.

In Chapter 8: "Transformed cladistics" was developed from phylogenetics by dropping the assumption that speciation yields the branch points in the cladogram, thus restoring the priority of classification to evolution: the pattern of natural classification is the *explanandum* of which evolution is the *explanans*. But this leaves the *a priori* assumption of a hierarchy of characters with no justification. Transformed cladists polarise characters using patterns of ontogeny (individual development) where these are available; this is compared with out-group comparison. Otherwise their methods are similar to those of phylogenetics, but with no assumption of a phylogeny, the distinction between homology and homoplasy is vital to the *a priori* hierarchy. The distinction is made by the use of parsimony. Of three taxa, those two with a majority of shared unique characters are "sister-groups"; characters suggesting any other pairing are "mistakes".

In Chapter 9: "Numerical cladistics" and techniques of phylogeny reconstruction based on biochemical and molecular data developed concurrently with transformed cladistics but emphasised rather than rejected the presumption of phylogeny. Numerical cladistics, which diverged from phenetics and converged on cladistics, is described in detail. Like cladistics it polarises characters and uses parsimony, but to minimise branch lengths in the dendrogram representing total character difference among taxa. But early molecular techniques, notably immunology and DNA hybridisation, produced only distance data comparable to the calculated distances in phenetics. Comparing sequences of amino acids between homologous proteins of different organisms or bases in DNA does, however, yield unit character differences, but, given the limited possible number of acids or bases, they are not unique characters. Techniques of aggregate similarity (as in phenetics), parsimony, and "likelihood" (based on models of random change) are rivals for resolving them.

In Chapter 10: The merits of these latter techniques and that of

## 8 Introduction

compatibility analysis, described previously, are discussed with respect to both molecular taxonomy and taxonomy in general. It is concluded that *if* the hierarchical pattern of the natural order of organisms, and the hierarchy of characters on which it is based, are accepted *a priori*, *then* cladistics is the only valid technique for discovering that order. This re-establishes the logical priority of classification to phylogeny. But there is no extrinsic method of establishing that the natural order of organisms is a divergent hierarchy, and empirical evidence from hybrid species and the difficulty of classifying animal taxa of high rank suggest rejection of the universal hierarchy. Without it evolutionary theory loses its *explanandum*. There is no solution to this paradox. Furthermore, if all the features of organisms result from their individual history, one can ask whether there are any other universal statements in biology.

In Chapter 11: Perhaps a true hypothesis of the mechanism of evolutionary change might be such a statement. A history of such theories of mechanism is presented. They comprise internal factors, producing bodily change in organisms, and external factors to which organisms respond so as to produce adaptation to the environment. Lamarck's theory, and that of many post-Darwinian naturalists, proposed directional (orthogenetic) internal factors and the inheritance of responses to the environment. Darwin and Wallace proposed "random" individual changes (of unknown causality) and the differential inheritance of those better adapted (*Natural Selection*). It is shown that the theory of natural selection is not "tautologous" but has empirical content. The development of genetics in the twentieth century added the "missing ingredient" to Darwin's theory but at first suggested a rival theory of evolution by "saltation".

In Chapter 12: The reconciliation came with the development of population genetics, but by concentrating on changes of gene frequency in populations, the "Synthetic Theory" is both tautologous and can make no predictions about emergent properties in evolution beyond the species level. Other criticisms are that adaptation by selection is taken as *a priori* for all characters of an organism, and that insufficient attention is paid to systematic constraints in development of the characters on which selection is supposed to act.

In Chapter 13: Two problems have dogged epistemology (the-

## Introduction

ory of knowledge) since the time of the ancient Greeks; they are causality and induction. The problem of induction is that, despite the claims of Logical Positivism, there is no logical way of proceeding from individual scientific data to universal laws. But the alternative of testing such laws, proposed *ad hoc*, by falsification is also flawed. Laws (or theories) may succeed one another sequentially, or replace one another for sociological reasons. It is even doubtful whether universal laws play much part in the progress of science.

In Chapter 14: A distinction can be made in the philosophy of science between Natural History and Natural Philosophy, in which the former uses general principles to explain particular phenomena. Laws, theories, and other empirical generalisations occur in all sciences, but comparison of the theories of plate tectonics and evolution, both theories in Natural History, suggest a special status for the taxonomic hierarchy, the explanandum for evolutionary theory. The special status of taxonomic statements arises from this but poses questions about the philosophical status of the entities classified: species properly defined are individuals, as are higher taxa to the phylogenetic cladist; thus their taxonomic hierarchy must be a unique and contingent entity. But all taxa in transformed cladistics are classes and logically prior to phylogeny. Cladistics should use *methodological* essentialism to construct cladograms, which may subsequently be interpreted as phylogeny and tested as classifications. Laws in biology concern classes of entities such as taxonomic categories, but predictive generalisations about individuals are taxonomic statements.

2

# Patterns of classification

It is taken for granted today, at least by zoologists, that systematic classifications of organisms can be represented by branching diagrams (dendrograms: Mayr, Linsley, and Usinger 1953) that represent hierarchical arrangements - Darwin's (1859) "groups within groups". The nested groups are taxa, each of which belongs to a category that represents its level in the hierarchy (Simpson 1961a). In the tenth edition of Linnaeus's Systema Naturae (1758), he proposed the following categories: Regnum (Kingdom), Classis, Ordo, Genus, Species, to which the categories Phylum and Family were added later. All the taxa at the same level in the hierarchy occupy the same rank and are given the same category. Thus "the rank of a taxon is that of the category of which it is a member" (Simpson 1961a). Modern biological classification is therefore a process of "ordinally stratified hierarchical clustering" (Jardine and Sibson 1971, p. 127), and the result is an aggregational hierarchy (Mayr 1982, pp. 64-6) in which the units, usually species, which constitute its lowest rank, are aggregated in successively higher ranks. The hierarchy is also an inclusive one (Mayr 1982, pp. 205-8) as opposed to an *exclusive* one:

> Military ranks from private, corporal, sergeant, lieutenant, captain up to general are a typical example of an exclusive hierarchy. A lower rank is not a subdivision of a higher rank; thus lieutenants are not a subdivision of captains. The *scala naturae*... is another good example of an exclusive hierarchy. Each level of perfection was considered an advance (or degradation) from the