Molecular taxonomy—concept of DNA taxonomy; construction of phylogenetic trees using mitochondrial DNA/ or other markers

I INTRODUCTION

Molecular Taxonomy, modern branch of taxonomy (the classification of living and extinct organisms into evolutionary groups) that makes use of genetic data from proteins and DNA to estimate phylogeny (the history of branches in the tree of life), and thereby assists in classifying species according to their genealogical connections through shared ancestors.

II HISTORY

Traditionally, taxonomists classified organisms by comparing their morphology (body form) and behaviour. For example, when European naturalists first arrived in Australia, they noticed some small birds that in terms of general appearance and behaviour resembled Old World nuthatches. Accordingly, they classified Australian “sittellas” in Sittidae, the taxonomic family that includes European nuthatches. Much later, in the 1980s, scientists re-examined these birds using a molecular genetic technique (DNA-DNA hybridization) that in effect compares sequences of DNA bases at large numbers of points in the genome (see Molecular Biology). It turned out that the sittellas of Australia had not been classified properly. They proved to be evolutionarily distant from northern hemisphere nuthatches, and phylogenetically closer to certain other avian groups including whistlers and jays. Thus, the behavioural and morphological features shared by sittellas and nuthatches did not arise from their sharing a recent ancestor, but instead must have evolved independently (in convergent fashion) from different ancestors (see Convergent Evolution). Accordingly, the molecular systematists who made these discoveries (Charles Sibley, Jon Ahlquist, and Burt Monroe) reclassified the sittellas into Corvidae, the taxonomic family of jays and their allies.

III TECHNIQUES

DNA-DNA hybridization was one of the earliest laboratory techniques employed by molecular taxonomists. Other molecular methods first adopted during the 1960s and 1970s included protein electrophoresis, microcomplement fixation (an immunological approach for quantifying the differences in homologous proteins), and a variety of DNA-level assays. Over the years, many of the DNA methods in particular have been improved and streamlined considerably. Each laboratory technique has particular strengths (and weaknesses), and provides good phylogenetic resolution at a restricted set of evolutionary timescales. A key consideration is the sequence divergence rate of particular genes—that is, the rate at which they evolve (see Molecular Clocks below). Genes that encode ribosomal RNA, for example, tend to evolve very slowly, so they are excellent sources of information for characterizing ancient (deep) branches that separated tens or hundreds of millions of years ago in a phylogenetic tree. By contrast, some genes in the animal mitochondrial genome evolve so rapidly as to reveal historical relationships of conspecific (same-species) populations that may have separated just centuries or millennia ago. Another class of molecular assays known as DNA fingerprinting is ideal for revealing the shallowest possible genetic connections in a genealogical tree—genetic paternity and maternity across a single generation.

IV MOLECULAR CLOCKS

One great boon to molecular taxonomy was the discovery that particular genes, or nucleotide positions therein, tend to evolve at fairly constant rates. Thus, the magnitude of genetic divergence between various species is at least roughly correlated with the absolute evolutionary time (as gauged by evidence from fossils or the geological record) since they last shared common ancestors. This led to the notion of molecular...
clocks, of which there are many, depending partly on functional constraints on the particular gene examined—that is, whether or how severely, most new mutations disrupt the DNA's viable function in organisms.

Ribosomal genes evolve slowly because they are relatively conserved in function and structure, whereas, for example, many non-coding DNA sequences (such as those characterizing introns and pseudogenes) evolve much faster (see Ribosomes). Molecular clocks are far from being like metronomes, and there is considerable debate regarding their exact calibration and phylogenetic usefulness in particular instances. But they do provide a general temporal framework, often otherwise lacking, for phylogenetic analyses. Furthermore, it is by no means critical in all phylogenetic applications that molecular evolution should have a strict time-dependent nature. Whether or not DNA evolves in clock-like fashion, clades (monophyletic groups of organisms) can be recognized by virtue of sharing uniquely derived molecular characteristics.

In general, molecular markers offer several advantages over traditional taxonomic data in the estimation of phylogeny. First and foremost, they are genetic (fully heritable), a situation that does not always hold for morphological or behavioural traits that may be influenced by environmental conditions. Second, the precise nature and amount of genetic information is immediately evident in molecular data sets, and this permits researchers to place confidence statements on particular phylogenetic inferences. This contrasts with the insecure knowledge on the number or make-up of genes affecting most of the (non-molecular) phenotypes employed in traditional taxonomy. Third, molecular methods provide direct access to an enormous volume of historical genetic information. For example, the human genome contains about 3 billion nucleotide pairs, an exceptional tally for multicellular organisms (see Human Genome Project). Molecular assays involve sampling, often more or less at random, from such vast genealogical archives. Fourth, molecular data provide common yardsticks for systematics (the study of the phylogenetic relationships of organisms). Creatures from microbes to whales all share many assayable genes and enzymes, such as those involved in central metabolic pathways for processing food and energy. This contrasts with the paucity or absence of morphological traits common to such disparate creatures as birds and fishes, much less microbes and whales. Finally, molecular methods have opened the entire biological world to genetic (and hence phylogenetic) scrutiny. Prior to the molecular revolution in the latter half of the 20th century, genetic analyses of any sort had been confined to a small handful of species (such as pea plants, fruit flies, and mice) that could be reared and crossed easily.

V CURRENT EFFORTS

As palaeontologist George Gaylord Simpson noted in 1945: "The stream of heredity makes phylogeny; in a sense, it is phylogeny. Complete genetic analysis would provide the most priceless data for the mapping of this stream." Since the 1960s, improvements in molecular techniques have given scientists greater and more direct access to this hereditary chronicle. Today it is routine to analyse homologous DNA sequences totalling thousands of nucleotides from dozens of species, and thereby reveal the phylogenetic history of the group. To pick one recent example, detailed analyses of nucleotide sequences from multiple genes led scientists to a remarkable conclusion: several seemingly unrelated mammalian taxa apparently shared a distant ancestor. This discovery prompted the compelling "Afrotheria" notion, which posits that about one third of all placental mammals in the world, including creatures ranging from elephants to aardvarks to hyraxes, originated from a common ancestor that inhabited the African continent about 75 million years ago. The Afrotheria hypothesis parallels in some respects the aforementioned hypothesis, from molecular data, that a major adaptive radiation of birds (as well as marsupial mammals) occurred in the Australian region.

These examples also illustrate how the fields of molecular phylogenetics and historical geography can inform each other.
Molecular genetic methods provide taxonomy with a powerful tool to describe species as they apply to all organisms irrespective of developmental stage, sex, or body part and offer universal, quantifiable characters. With DNA barcoding, the foundation is laid for automated and accelerated taxon identification. Molecular taxonomy is particularly effective in combination with other methods, usually with morphology.

A phylogenetic tree is a graphical representation of the evolutionary relationships among entities that share a common ancestor. Those entities can be species, genes, genomes, or any other operational taxonomic unit (OTU). More specifically, a phylogenetic tree, with its pattern of branching, represents the descent from a common ancestor into distinct lineages. It is critical to understand that the branching patterns and branch lengths that make up a phylogenetic tree can rarely be observed directly, but rather they must be inferred from other information.

The principle underlying phylogenetic inference is quite simple: Analysis of the similarities and differences among biological entities can be used to infer the evolutionary history of those entities. However, in practice, taking the end points of evolution and inferring their history is not straightforward.

**Steps in phylogenetic tree reconstruction:**
1. Select a sequence of interest. This could correspond to a whole gene, a region of a gene (coding or noncoding regions can be used), a regulatory region for a gene, a transposable element, or even a whole genome.

1. Identify homologs. Acquire sequence data for objects that are homologous to the sequence of interest.

1. Align sequences. Align the sequence of interest and the homologous regions to generate a sequence data matrix.