MOLECULAR PALEONTOLOGY

The morphology (form and structure) of organisms provides a wealth of information about evolutionary history. However, additional information contained in molecules also can contribute to our understanding of past life. Molecular data are used to build phylogenetic trees (family trees based upon genetic information) and to estimate times of divergence. These data are especially useful for groups with relatively few morphological characters and poor fossil records (e.g., fungi, bacteria), although our knowledge of some well-known groups (e.g., vertebrates) also has increased greatly as a result of molecular studies. Some molecular characters (nucleotides and amino acids) have been conserved throughout the history of life and can be used to study ancient relationships. In some cases, molecular data have been obtained directly from fossil organisms.

Molecular Phylogenetics

A major advance in the field was the development of the polymerase chain reaction (PCR) in the 1980s (Mullis 1990). In this method, a region of interest in the genome, typically a portion of a gene that is 200 to 800 base pairs long, is amplified to a million or more copies in order to facilitate additional molecular analysis, such as DNA sequencing. The process is called a chain reaction because it involves repeated cycles of heating and cooling, during which the DNA fragments are split, short pieces of DNA (primers) are joined, and then new strands are created. Before PCR, DNA fragments were amplified by bacterial cloning, a much longer process.

The technique of PCR, combined with automated methods for sequencing DNA (determining the order of nucleotides and genes), has revolutionized evolutionary biology. A virtually unlimited number of characters can be obtained from a large number of species to address interesting systematic questions. The most popular molecule for study in animals has been the mitochondrial chromosome. This circular structure contains only a small proportion of genes (in animals, it is about 16 kilobases in length). Also, the greater technical ease and rapid rate of change of mitochondrial DNA, compared with typical nuclear genes, are responsible for its popularity in recent years. In plants, DNA contained in another organelle, the chloroplast, has been the primary source of sequence data for evolutionary studies.

A seemingly limitless number of nuclear genes also are available for tree-building. Unlike mitochondrial and chloroplast genes, which are tightly linked, nuclear genes used in estimating phylogeny usually are unlinked. One feature of nuclear genes in many eucaryotes (organisms with membrane-enclosed nuclei in their cells) is the presence of large amounts of noncoding DNA (introns) interspersed within regions that control the production of substances (coding regions). For example, a typical gene with only 900 base pairs of coding region (for a protein of 300 amino acids) may consist of 10,000 base pairs because of the presence of introns. While fast-evolving introns provide information for relationships of individuals or populations, they may not be useful in applications dealing with larger categories such as species phylogenies. In those cases, it is more informative to study only the coding region found in the messenger RNA (usually without introns) of that gene.

There are different methods of building evolutionary trees from molecular data (usually DNA or amino acid sequences). A common feature of all methods is that they minimize the amount of inferred change (substitutions of nucleotides or replacements of amino acids). Maximum likelihood methods seek the tree that matches the data set with highest probability, minimum evolution (neighbor-joining) searches for the tree with shortest overall length, and the maximum parsimony method tries to find the tree requiring fewest character-state changes. The groups of spe-

cies defined within a tree can be evaluated by statistical tests to determine if they are significant. The most commonly used test is the bootstrap method, which involves sampling the sites randomly with replacement many times, constructing new trees at each cycle, and then determining the frequency that a particular group appears among the many bootstrap samples. For example, if 97 out of 100 bootstrap trees join human with chimpanzee, then we infer that the chimp-human relationship is supported with 97% confidence.

Molecular Clocks

For most groups in the fossil record, the time of divergence between two lineages is not known with accuracy. The minimum time of divergence is the date of the oldest fossil assigned to one of the two lineages. The actual divergence is assumed to have occurred even earlier. In those rare cases where the fossil record is excellent, the minimum age may be close to the divergence time. However, for most groups of organisms, an accurate time frame has yet to be established using data from fossils. For this reason, evolutionary biologists frequently turn to molecules to determine dates of divergence.

In groups for which molecular data can be obtained, such as all living and some fossil organisms, it is possible to estimate divergence times using a molecular clock. To do this, molecular divergence must be measured between the two taxa being compared and at least one additional taxon with a known or presumed time of divergence (for calibration). One requirement is that the rate of change along all lineages should not be significantly different. Although many of the molecular changes used for timing divergences probably do not result in a functional change in the protein or organism (i.e., are neutral), this is not a requirement of a molecular clock.

Sources of error in molecular clocks include the estimates of molecular divergence and the calibration time used. Calibrations are taken from the fossil record or from well-dated geologic (or climatic) events (if they are assumed to be the cause of the phylogenetic divergence). There is no "universal molecular clock" because different genes evolve at different rates. For instance, genes involved in the immune reaction must change rapidly to keep pace with a diversity of antigens (e.g., viruses), whereas genes involved in the most basic of cell functions rarely change because of their universal importance. However, a rapid rate of change in synonymous substitutions (those that do not cause changes in amino acids) is common to virtually all protein coding genes.

Genes that show strong positive selection in particular lineages are unlikely to behave as good molecular clocks. On the other hand, some genes, such as serum albumin in vertebrates, appear to evolve in a relatively clocklike fashion and have proven useful in dating evolutionary divergences. Ideally, divergence times should be estimated from a large number of genes in order to reduce the error of the time estimate (Hedges et al. 1996). The strong need to know times of divergence and the rapidly expanding databases of sequences have maintained a prominent role for molecular clocks in evolutionary biology.

Fossil Biomolecules

The possibility that proteins in the fossil record may have been preserved was suggested as early as the 1950s, when P.H. Abelson used techniques such as thin layer chromatography to identify organic components in fossils. The hypothesis also was supported by structures seen in electron microscopy. Researchers identified collagen-like fibers in fossil specimens from as long ago as the Early Paleozoic (545 million years ago), and these structures also were present in dinosaur bones of the Mesozoic era (251 to 65 million years ago).

Genetic information about the immune properties of proteins can be used to provide information on phylogenetic relationships. A small region of the whole protein molecule is all that is needed for a functional immune system to recognize an invader and form antibodies against it. A sequence of amino acids determines the shape of these three-dimensional regions, called epitopes. They are created by the complex folding of the protein molecule into its functional form. As few as five amino acids in the proper "shape" are enough for antibody recognition. Only small fragments of molecules are needed, and phylogenetic relationships can be inferred from the degree or strength of the bonds that bind the antibody and epitope together. Species that are more closely related usually have a greater number of antibody-epitope bonds.

Amino acid identification also was used very early to suggest the preservation of fossil proteins and continues to yield valuable information today. Individual proteins have been identified from fossil specimens, including the bone proteins (e.g., collagen, osteocalcin, and osteonectin). Proteins abundant in blood, such as albumin and hemoglobin, have been identified in several ancient specimens from as far back as the Mesozoic (Schweitzer et al. 1997). Using these preserved proteins as phylogenetic tools was suggested by J.M. Lowenstein and G. Scheuenstuhl (1991) through the application of immunological techniques, including Western blots and ELISA assays (tests). The degree of binding of antibodies to ancient protein fragments has been used to investigate the relationships of fossil taxa.

Variation in protein composition and structure also is useful in phylogenetic analysis. Researchers can study chemical content, hydrophobic characteristics, and amino acid composition, as well as the unique ways that proteins curve and fold into three-dimensional shapes, because these characteristics differ greatly between proteins and among the same proteins in different taxa. Additionally, while the phylogenetic information found in DNA is contained in the sequence of bases, phylogenetic information from proteins may be obtained indirectly from their three-dimensional structure as well as directly from their amino acid sequence. Finally, different classes of proteins can be determined by their function as well as by their constituent amino acids and the various ways that members of these classes are preserved.

The study of ancient DNA is a younger field, dating only to the early 1980s. Since then, several studies have successfully resolved relationships of extinct organisms, including the Tasmanian wolf (Krajewski et al. 1997). Others have reported finding ancient DNA in fossils, including those of dinosaurs and amber-encased insects, from as early as the Mid-Mesozoic. However, many of these studies have been received with caution, and no results have been replicated independently for fossils older than

100,000 years (Austin et al. 1997). Absolute time limits on molecular preservation have not yet been demonstrated under naturally occurring conditions, and, therefore, recovery of very ancient DNA remains a possibility. Authenticity and independent replication of results continue to be major concerns.

Problems with Contamination

Fossil bone and tissue have been exposed to many contaminants during decay, burial, and transformation into rock. In addition, human and laboratory contaminants can be introduced into samples despite the most careful controls. Also, the sensitivity of PCR greatly increases the potential for amplifying contaminant DNA molecules. Moreover, contaminant molecules may be more abundant and less damaged than ancient molecules within the same fossils (Austin et al. 1997). Thus, contamination is a major concern among researchers in this field.

Some contamination problems can be reduced through careful design of PCR primers. Primers can be designed to rule out amplification of the more common contaminant sources, such as microbial and human DNA. The chances of amplification of DNA from ancient sources successfully are increased by consideration of phylogenetic relationships among modern taxa. For example, in designing primers to amplify DNA from dinosaur bones, one would want to select DNA sequences unique to modern birds and modern crocodiles, the two living groups most closely related to the extinct dinosaurs.

Besides contamination, one may encounter other difficulties in working with ancient biomolecules. In the case of DNA, factors such as acids from organic humus may inhibit the action of the polymerase enzymes used in amplification attempts, or the DNA may be damaged enough to introduce misleading artifacts (artificial substances produced, inadvertently, by the process). Also, the ancient DNA may be degraded to strands that are too short for binding with PCR primers. Protein sequences obtained from ancient sources have not been reported in the literature and may be difficult to obtain, either because the minimal amounts of protein yield concentrations too low for sequencing or because modifications of binding sites make enzymatic degradation ineffective.

Molecular approaches have provided paleontology with new tools to answer old questions, and the result has been a revolution in our understanding of evolutionary history. In some instances, where molecular phylogenies have contrasted with long-standing views based on morphology, it might appear that information from fossils and morphology is no longer needed. On the contrary, molecular information is unlikely to replace the history of adaptations reflected in morphology and the fossil record. Molecular approaches will continue to complement classical approaches to paleontology in the foreseeable future.

S. Blair Hedges and Mary H. Schweitzer

Works Cited

Abelson, P.H. 1956. Paleobiochemistry. *Scientific American* 195:83–92.

Austin, J.J., A.J. Ross, A.B. Smith, R.A. Fortney, and R.H. Thomas. 1997. Problems of reproducibility: Does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society of London*, ser. B, 264:467–74.

Hedges, S.B., P.H. Parker, C.G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* 381:226–29.

Krajewski, C., L. Buckley, and M. Westerman. 1997. DNA phylogeny of the marsupial wolf resolved. *Proceedings of the Royal Society of London*, ser. B, 264:911–17.

Lowenstein, J.M., and G. Scheuenstuhl. 1991. Immunological methods in molecular palaeontology. *Philosophical Transactions of the Royal Society of London*, ser. B, 333:375–80.

Mullis, K.B. 1990. The unusual origin of the polymerase chain reaction. *Scientific American* (April):56–65.

Schweitzer, M.H., M. Marshall, K. Carron, D. Bohle, S. Busse, E. Arnold, D. Barnard, J. Horner, and J. Starkey. 1997. Heme compounds in dinosaur trabecular bone. *Proceedings of the National Academy of Sciences, USA* 94:6291–96.

Further Reading

Ambler, R.P., and M. Daniel. 1991. Proteins and molecular palaeontology. *Philosophical Transactions of the Royal Society, London*, ser. B, 333:381–89.

Herrmann, B., and S. Hummel (eds.). 1993. *Ancient DNA*. London: Springer-Verlag; New York: Springer-Verlag, 1994.

Li, W.H. 1997. *Molecular Evolution*. Sunderland, Massachusetts: Sinauer.

Pääbo, S. 1993. Ancient DNA. *Scientific American* (November):86–92.