Molecular Palaeontology

Ben Waggoner, *University of Central Arkansas, Comway, Arkansas, USA*

Secondary article

Article Contents

The traces that organisms of the past have left are not restricted to their bones, shells, imprints, and tracks; biogenic macromolecules may also survive in the fossil record. As techniques for studying fossil macromolecules have expanded, information from molecules has complemented and expanded our knowledge of the history of life as derived from other types of fossils.

Introduction

Recovery of molecules from fossils

In the broadest possible sense, molecular palaeontology might encompass any and all chemical traces left by once-living organisms or by their life processes. This would cover fields ranging from isotope geochemistry and biomineralization to molecular biology. However, in a more restricted sense, molecular palaeontology is the study of complex organic molecules made by once-living organisms. In some cases, molecules are the only remaining clues to the existence of certain organisms at certain times. When complex organic molecules are found in close association with fossils of the organisms that produced them, they may provide important information on the organisms' evolutionary relationships, age and/or mode of life.

Four major classes of large, complex biomolecules are usually defined: nucleic acids, proteins, lipids and carbohydrates. Of these, nucleic acids have probably the lowest potential for preservation as fossils. When water is present, nucleic acids usually hydrolyse rapidly (Lindahl, 1993). Most proteins are also poor candidates for preservation, although their survivability depends in part on their structures and on associated minerals. In bones, for instance, collagen disintegrates fairly rapidly, while small compact proteins such as osteocalcin may survive largely intact for much longer periods (Bada, 1991). Certain carbohydrates such as chitin and cellulose may be more resistant, although they have received less attention. Lipids are usually chemically modified after death and fatty acids are likely to degrade, but modified sterol lipids are often preserved well enough that their original chemical form can be deduced. Finally, some important molecular fossils do not fit neatly into any of the four classes. These include porphyrins such as chlorophyll and haem; sporopollenins, extremely resistant polymers that form the walls of pollen and spores; and lignin, a complex polymer of phenolic alcohols found in vascular plant tissues. The most abundant 'molecular fossil' of all, however, is kerogen, a name given to insoluble, high-molecular weight organic matter of uncertain composition.

History 1950s–1996

In the mid-1950s, Abelson found amino acids preserved in fossil shells as old as 360 million years, and suggested that comparing fossil amino acid sequences with those of extant organisms would allow molecular evolution to be studied directly (Abelson, 1954). Since then, fossil peptides have been studied both by amino acid analysis and, beginning in the 1970s, by structural studies on whole peptides and by immunological methods (e.g. de Jong *et al.*, 1974; Westbroek *et al.*, 1979). In the 1960s, oil companies began analysing the composition of petroleum with a view to determining its source and history, opening up another field of molecular palaeontology. Later, in the 1970s, some palaeobotanists began studying molecules isolated from well-preserved fossil plants, such as flavonoids and components of lignin (reviewed in Niklas, 1982). Attempts to extract DNA from preserved and fossil specimens began in the mid-1980s, and became increasingly common with the development of the polymerase chain reaction (PCR) technique.

Methods

Methods for studying molecular fossils are little different from methods used in studying complex molecules from living organisms. PCR is used to amplify fossil DNA, and various types of chromatography, NMR and IR spectroscopy and immunochemical reactions are routinely used to identify molecular fossils. Specialized protocols are sometimes used to detect and avoid contamination, or to compensate for sample degradation, both of which can be serious problems for molecular palaeontological studies. However, the ways in which the results of molecular palaeontological analyses are applied are the discipline's most distinctive features.

Biomarkers

Fairly complex organic molecules, including alcohols, fatty acids and amino acids, are known from certain

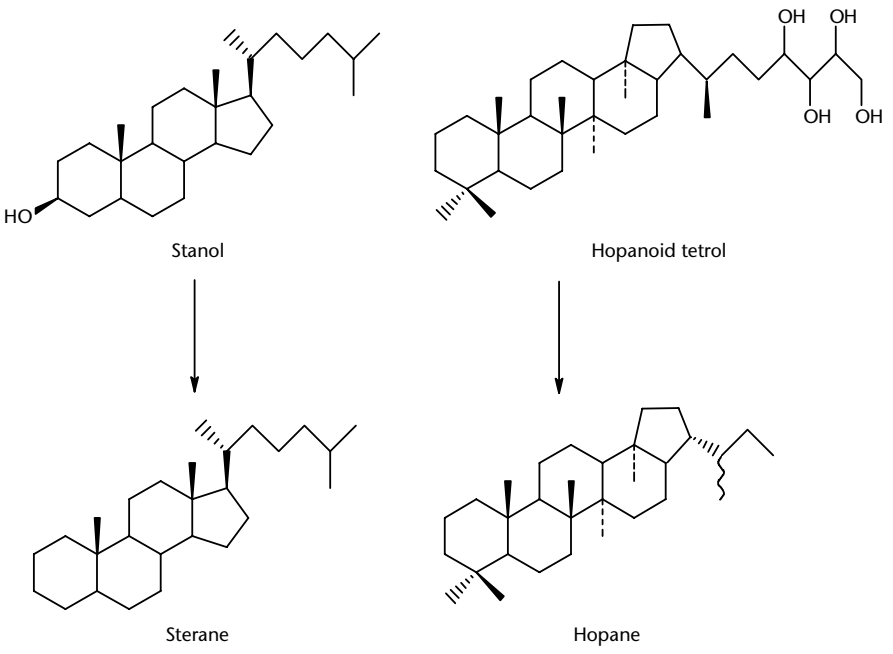

Figure 1 The sedimentary diagenesis of two molecular fossils, a typical bacterial hopane and a typical eukaryotic sterane.

meteorites and from interstellar dust. These are not derived from living organisms, and thus are not fossils in the true sense. They are, however, of interest to researchers on the origin of life, since they demonstrate the ubiquity and apparent ease of synthesis of organic molecules in the cosmos. It is possible that key components of the first proto-life had an extraterrestrial origin (e.g. Deamer *et al*., 1994).

The molecular fossil record proper begins with complex molecules, mostly derived from steroid lipids, known as biomarkers. Petroleum often contains biomarkers that can be identified with reasonable certainty as products of a known taxon. Analysis of biomarkers is useful in the petroleum industry for information about the source and degree of maturation of a petroleum deposit, but it is also useful in documenting biological evolution. The oldest biomarkers currently known are from $\sim$2700 million-year-old rocks of northwestern Australia. Among other molecules, these include both 2-methylhopanes, which are derived from cyanobacteria, and C_{28}–C_{30} sterol derivatives (steranes), which are unique to eukaryotes. These biomarkers confirm the presence of cyanobacteria, previously known from microfossils and stromatolites, and provide strong indirect evidence for the presence of eukaryotes over 500 million years prior to the oldest known body fossils (Brocks *et al*., 1999). Steranes are also known from $\sim$1690 million-year-old rocks of Australia (Brasier and Lindsey, 1998), and a large and growing number of hydrocarbon biomarkers are known from the Proterozoic, including eukaryote, bacterial, and archaean biomarkers as well as a number of 'orphans' (Ourisson, 1994) (**Figure 1**). Biomarkers in younger sediments are useful in evolutionary studies as well. To give just two examples: dinosteranes, biomarkers unique to dinoflagellates, are now known in rocks as old as the Cambrian, nearly 300 million years before the oldest definite dinoflagellate fossils (Moldowan and Talyzina, 1998). Oleananes, a family of pentacyclic triterpenoids, are widely distributed in flowering plants, and their concentration in sediments increases dramatically in the Cretaceous and Tertiary, parallelling the radiation of flowering plants seen in the macrofossil record (Moldowan *et al*., 1994) (**Figure 2**).

Amber, or fossilized plant resin, is composed of 'orphan' biomarkers, labdanes (**Figure 2**), that have polymerized naturally. It is usually not possible, on the basis of macrofossils alone, to identify the botanical source of a given amber sample. However, when amber is analysed by

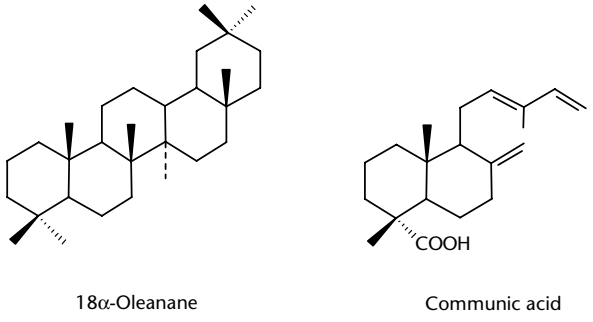

Figure 2 Formulae of oleanane, a biomarker associated with flowering plants, and a labdane (communic acid) found in gymnosperm resins.

infrared or NMR spectroscopy, the spectra can be compared with those of modern tree resins, and the botanical source of the amber can be identified (Lambert *et al*., 1990). This is not only useful in palaeontology but has applications in archaeology as well: amber artefacts found at a site can be traced by their spectroscopic 'fingerprints' to their source deposits, and thus ancient trade routes can be reconstructed.

Dating

There have been several attempts to use slow chemical reactions that occur in fossil biomolecules (other than radioactive decay) as 'clocks' for dating the samples. The most promising has been amino acid racemization (AAR). Amino acids are chiral molecules and, with very few exceptions, all amino acids in living organisms are found in the L-configuration. Over time, L-amino acids racemize; that is, they are converted to mixtures of L- and D-forms. Since the proportion of D-amino acids in a sample is proportional to its age, in principle, amino acid racemization can be used to date fossils. In practice, applying this method is difficult: contamination and leakage can pose problems, racemization rates are different for each amino acid, and rates also depend on temperature and on the associated mineral matrix. However, in certain cases, such as tooth enamel from Olduvai Gorge, Tasmania, racemization can be calibrated against radiometric dates, and has proved useful (Bada, 1991). Amino acid proportions can sometimes provide age estimates for fossils that are too old for ^{14}C dating (e.g. Martin *et al*., 1996).

Another molecular dating method has been applied to amber. Particular peaks in the NMR spectra of amber samples (exomethylene resonances) are weak or absent in older samples (Lambert *et al*., 1990). Thus, amber can be dated directly using the size of the exomethylene peaks. Since burial conditions can affect the size of the peaks, this is not a very accurate dating method, but it is at least broadly consistent with ages obtained in other ways.

Systematics

The most famous molecular fossils are the segments of DNA isolated and sequenced from several different fossil contexts. Fossil DNA, which will be treated separately in the next section, has received a great deal of scientific and popular attention, but it is certainly not the only molecule that can provide information on the systematics of extinct organisms. Fossil proteins can be compared with modern ones by various immunological techniques. Typically, antisera are prepared to protein extracts from both fossils and living animals using standard immunodiffusion, enzyme-linked immunosorbent assay (ELISA), complement fixation or radioimmunoassay protocols. This method has been used to compare fossil and living cephalopods, proboscideans and rodents, as well as recently extinct organisms such as the quagga, Tasmanian wolf, and Steller's sea cow (Westbroek *et al*., 1979; Lowenstein and Scheuenstuhl, 1991).

The biomarker oleanane is widely distributed in living angiosperms but is rare in other plant taxa. Oleanane is most common in Cretaceous and Cenozoic rocks and petroleum, a time interval that corresponds to the radiation of terrestrial angiosperms in the macrofossil record. It is uncommon in pre-Cretaceous rocks, but has been found in Cretaceous fossil bennettitaleans (an extinct group formerly included in the 'cycadeoids'), which are believed on morphological grounds to be close relatives of flowering plants. Oleanane is also found in Permian gigantopterids, a group of 'seed ferns' that may be more closely related to flowering plants than was previously thought. Oleanane has not yet been found in other fossil seed-bearing plants, but may well be present in other fossil taxa that are related to the angiosperms. This shows that nonprotein fossil molecules can confirm hypotheses of evolutionary relationship, and can suggest new hypotheses (Moldowan *et al*., 1994; Taylor *et al*., 1998; see Niklas, 1982, for many other examples).

Palaeoecology

Modern invertebrates contain lipid sterols associated with the mineral components of their skeletons. These lipids are derived largely from the animals' diets, and may include distinctive lipids that are derived from particular food items (e.g. dinosterane from dinoflagellates and diatoms, 22-dehydrocholesterol from rhodophytes). Since lipids are relatively stable molecular fossils and can be extracted from invertebrate fossils, lipid analysis should be useful in inferring the diets and feeding modes of fossil invertebrates (CoBabe and Ptak, 1999). Chemical analyses of desiccated Pleistocene coprolites (fossil faeces) have been used to identify food items of animals and humans (Poinar *et al*., 1998). Human diets and hunting habits have been studied using another technique: protein residues on stone points have been identified as coming from various animals that were killed and butchered (Lowenstein and Scheuenstuhl, 1991).

Clarification of Actual Preservation of DNA

Techniques to isolate, sequence and analyse protein and DNA sequences in living organisms have revolutionized biological systematics over the past three decades. These techniques have been applied to the analysis of molecular fossils as well. By far the most celebrated development has been the sequencing of ancient DNA (aDNA), its fame being largely due to the overwhelming popular success of

the book and movie *Jurassic Park*, which deal with fictional attempts to resurrect dinosaurs from DNA sequences preserved in Mesozoic amber. The premise of the book is sheer fantasy: among other difficulties, under most preservational conditions, fossil DNA is so fragmented and degraded that recovery of anything like a complete fossil genome would be virtually impossible. Hydrated DNA at moderate temperatures degrades to short fragments on a time scale of thousands of years. However, high ionic strength, partial dehydration, and anoxia can slow the degeneration process (Lindahl, 1993), making it possible in theory for DNA to survive for much longer periods if preserved under certain conditions.

Several reports of aDNA millions of years old have generated a great deal of excitement. The earliest such report was of DNA from Miocene leaves from the 'Clarkia beds' of Idaho, USA (Golenberg *et al*., 1990). There have been several reports of both animal and plant DNA sequences from Cretaceous and Cenozoic amber; the oldest is from a Lower Cretaceous weevil from Lebanese amber (reviewed in Poinar, 1999). Woodward and coworkers (1994) isolated putative DNA, and sequenced a portion of the mitochondrial cytochrome *b* gene, from well-preserved dinosaur bones from a Late Cretaceous coal seam in Utah, USA. In all of these cases, preservation is so good that cellular and even ultracellular structures can be observed. Specimens yield molecular fossils of other kinds, including chlorophylls from the 'Clarkia' fossils and collagen from the Utah dinosaur bones.

However, rigorous attempts to test these results have not been encouraging. High-molecular weight DNA from the 'Clarkia beds' has been interpreted as contamination, probably from bacterial sources (Sidow *et al*., 1991; Lindahl, 1993). Cladistic analyses of the Utah 'dinosaur DNA' have shown that most of the sequences cluster with human or other mammalian sequences. While a few of the sequences could conceivably be of dinosaur origin, most and probably all are either contaminants or too degraded to be identified positively (Young *et al*., 1995). Finally, repeated attempts to reproduce aDNA extractions from amber have failed, suggesting that DNA does not normally survive over millions of years in amber (Austin *et al*., 1997). In short, the existence of identifiable DNA millions of years old is uncertain at best.

Analyses of associated molecules other than nucleic acids have been used to estimate the probability of aDNA survival. Poinar and colleagues (1996) found that the extent of aspartic acid racemization in fossil material correlates with the degree of DNA preservation. Dinosaur bone and leaves from the Clarkia site (but not insects in amber) show extensive racemization and also contamination, and thus are unlikely to have yielded authentic aDNA. Leaves from the Clarkia locality retain original lignin but have lost all proteins and polysaccharides (Logan *et al*., 1993); this further suggests that nucleic acids are likely to have been lost in this material. An additional complication is that aDNA can bind to surrounding molecules, such as silica and hydroxyapatite. This can prolong its survival in some cases (Lindahl, 1993), but it can also make the DNA impossible to extract and sequence without special techniques (Poinar *et al*., 1998).

Perhaps more useful, if less glamorous, has been the analysis of aDNA from much younger fossil and subfossil remains, ranging from 1 to 100 000 years old. Mitochondrial aDNA is usually the target, since it exists in thousands of copies per cell. aDNA fragments from late Pleistocene vertebrate remains have been successfully used in phylogenetic studies. For instance, Yang and colleagues (1996) sequenced cytochrome *b* genes from fossils of *Mammut americanum* (American mastodon) and *Mammuthus primigenius* (woolly mammoth); comparison of these sequences with those of living elephants suggested that the mammoth is closer to the Indian elephant than to the African elephant, with the mastodon as the outgroup, confirming the results of both morphological studies and immunological studies of fossil proteins. Höss and coworkers (1996) sequenced mitochondrial aDNA of the extinct North American ground sloth *Mylodon darwinii*. Phylogenetic analysis suggested that *Mylodon* was closer to living two-toed sloths than to living three-toed sloths, implying that the arboreal lifestyle evolved at least twice in sloths. Their results also implied that the mammalian class Edentata may have arisen about 80 million years ago, in the Cretaceous period. Poinar and colleagues (1998) were able to isolate aDNA from a coprolite of a different ground sloth, *Nothrotheriops shastensis*. Some of the aDNA sequences turned out to be from the sloth itself, while others were derived from plants in the sloth's diet; these workers were able to identify sequences from grasses, yuccas and agaves, mustards, borages and other plants, some of which match the macroscopic plant remains identified in the coprolite.

A large number of studies have been carried out on aDNA from prehistoric and historic human remains, including skeletons, mummies, frozen remains, and bodies preserved in peat. Especially noteworthy is the work of Krings and colleagues (1997), who successfully isolated and sequenced mitochondrial aDNA from the original specimens of Neanderthal man. The Neanderthal sequence falls well outside the range of sequence variation seen in modern humans, suggesting that Neanderthals did not contribute to the gene pool of modern humans. Nonhuman aDNA from archaeological contexts has also proved useful: for instance, DNA from ancient plant seeds may shed light on the origin of modern crops. aDNA from the pathogen *Mycobacterium tuberculosis*, isolated from the lung of a 1,000-year-old-Peruvian mummy, resolved a long-standing controversy by showing that tuberculosis did exist in the Americas before European contact (Salo *et al*., 1994). Finally, DNA is now routinely extracted from museum and herbarium specimens. These include preserved specimens of organisms that have become extinct in

historic times, such as the quagga and the Tasmanian wolf (Higuchi *et al*., 1984; Thomas *et al*., 1989).

References

Abelson PH (1954) Organic constituents of fossils. *Carnegie Institute of Washington Yearbook* **53**: 97–101.

Austin JJ, Ross AJ, Smith AB, Fortey RA and Thomas RH (1997) Problems of reproducibility – does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society of London Series B*: **264**: 467–474.

Bada JL (1991) Amino acid cosmogeochemistry. *Philosophical Transactions of the Royal Society of London Series B* **333**: 349–358.

Brasier M and Lindsay JF (1998) A billion years of environmental stability and the emergence of eukaryotes: new data from northern Australia. *Geology* **26**: 555–558.

Brocks JJ, Logan GA, Buick R and Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* **285**: 1033–1036.

CoBabe EA and Ptak AJ (1999) Comparison of in situ mineral-associated lipid compositions in modern invertebrate skeletons: preliminary evidence of dietary and environmental influence. *Paleobiology* **25**: 202–211.

de Jong EW, Westbroek P, Westbroek JF and Bruining JW (1974) Preservation of antigenic properties of macromolecules over 70 Myr. *Nature* **252**: 63–64.

Deamer DW, Mahon EH and Bosco G (1994) Self-assembly and function of primitive membrane structures. In: Bengtson S (ed.) *Early Life on Earth*, pp. 107–123. New York: Columbia University Press.

Golenberg EM, Giannasi DE, Clegg MT *et al*. (1990) Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature* **344**: 656–658.

Higuchi R, Bowman B, Freiberger O, Ryder A and Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**: 282–284.

Höss M, Dilling A, Current A and Pääbo S (1996) Molecular phylogeny of the extinct ground sloth *Myledon darwinii*. *Proceedings of the National Academy of Sciences of the USA* **93**: 181–185.

Krings M, Stone A, Schmitz RW *et al*. (1997) Neanderthal DNA sequences and the origin of modern humans. *Cell* **90**: 19–30.

Lambert JB, Frye JS and Poinar GO (1990) Analysis of North American amber by carbon-13 NMR spectroscopy. *Geoarcheology* **5**: 43–52.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**: 709–715.

Logan GA, Boon JJ and Eglinton G (1993) Structural biopolymer preservation in Miocene leaf fossils from the Clarkia site, northern Idaho. *Proceedings of the National Academy of Sciences of the USA* **90**: 2246–2250.

Lowenstein JM and Scheuenstuhl G (1991) Immunological methods in molecular paleontology. *Philosophical Transactions of the Royal Society of London Series B*: **333**: 375–380.

Martin RE, Wehmiller JF, Harris MS and Liddell WD (1996) Comparative taphonomy of bivalves and foraminifera from Holocene tidal flat sediments, Bahia la Choya, Sonora, Mexico (northern Gulf of California): taphonomic grades and temporal resolution. *Paleobiology* **22**: 80–90.

Moldowan JM and Talyzina NM (1998) Biochemical evidence for dinoflagellate ancestors in the Early Cambrian. *Science* **281**: 1168–1170.

Moldowan JM, Huizinga BJ, Dahl J *et al*. (1994) The molecular fossil record of oleanane and its relationship to angiosperms. *Science* **265**: 768–771.

Niklas KJ (1982) Chemical diversification and evolution of plants as inferred from palaeobiochemical studies. In: Nitecki MH (ed.) *Biochemical Aspects of Evolutionary Biology*, pp. 29–91. Chicago: University of Chicago Press.

Ourisson G (1994) Biomarkers in the Proterozoic record. In: Bengtson S (ed.) *Early Life on Earth*, pp. 259–269. New York: Columbia University Press.

Poinar GO Jr (1999) Ancient DNA. *American Scientist* **87**: 446–457.

Poinar HN, Höss M, Bada JL and Pääbo S (1996) Amino acid racemization and the preservation of ancient DNA. *Science* **272**: 864–866.

Poinar HN, Hofreiter M, Spaulding WG *et al*. (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**: 402–406.

Salo WL, Aufderheide AC, Buikstra J and Holcomb TA (1994) Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proceedings of the National Academy of Sciences of the USA* **91**: 2091–2094.

Sidow A, Wilson AC and Pääbo S (1991) Bacterial DNA in Clarkia fossils. *Philosophical Transactions of the Royal Society of London Series B* **333**: 429–433.

Taylor DW, Li H, Dahl J *et al*. (1998) The molecular fossil oleanane and preliminary data on its occurrence in gigantopterids, anthophytes and other seed plants. *American Journal of Botany* **85**(6): 81–82.

Thomas RH, Schaffner W, Wilson AC and Pääbo S (1989) DNA phylogeny of the extinct marsupial wolf. *Nature* **340**: 465–467.

Westbroek P, van der Meide PH, van der Wey-Kloppers JS *et al*. (1979) Fossil macromolecules from cephalopod shells: characterization, immunological response and diagenesis. *Paleobiology* **5**: 151–167.

Woodward SR, Weyand NJ and Bunnell M (1994) DNA sequence from Cretaceous period bone fragments. *Science* **266**: 1229–1232.

Yang H, Golenberg EM and Shoshani J (1996) Phylogenetic resolution within the Elephantidae using fossil DNA sequence from the American mastodon (*Mammut americanum*) as an outgroup. *Proceedings of the National Academy of Sciences of the USA* **93**: 1190–1194.

Young DL, Huyen Y and Allard MW (1995) Testing the validity of the cytochrome *b* sequence from Cretaceous period bone fragments as dinosaur DNA. *Cladistics* **11**: 199–209.

Further Reading

Eglinton G and Logan GA (1991) Molecular preservation. *Philosophical Transactions of the Royal Society of London Series B* **333**: 315–328.

Herrmann B and Hummel S (eds) (1994) *Ancient DNA: Recovery and Analysis of Genetic Material from Paleontological, Archaeological, Museum, Medical, and Forensic Specimens*. New York: Springer-Verlag.

Logan GA, Collins MJ and Eglinton G (1991) Preservation of organic biomolecules. In: Allison PA and Briggs DEG (eds) *Taphonomy: Releasing the Data Locked in the Fossil Record*, pp. 1–24. New York: Plenum Press.

Lowenstein JM and Scheuenstuhl G (1991) Immunological methods in molecular paleontology. *Philosophical Transactions of the Royal Society of London Series B*: **333**: 375–380.

Poinar GO Jr (1999) Ancient DNA. *American Scientist* **87**: 446–457.