Molecular clocks

Simon Y. W. Ho
School of Biological Sciences, University of Sydney, Sydney, NSW, Australia

Definition

Molecular clock. A hypothesis that predicts a constant rate of molecular evolution among species. It is also a method of genetic analysis that can be used to estimate evolutionary rates and timescales using data from DNA or proteins.

Introduction

The tempo and mode of evolution are central themes of biological research. This places importance on the estimation of evolutionary timescales, which provide the backdrop for our interpretations of evolutionary patterns and processes. Traditionally, such inferences were made from the fossil record, coupled with radiometric dating. Fossils can provide an estimate of when different lineages first appeared and when species diverged from each other. In many cases, however, such data are unavailable, forcing us to look elsewhere for a source of temporal information. The “molecular clock,” proposed in the 1960s (Zuckerkandl and Pauling, 1962, 1965), allows evolutionary timescales to be estimated using genetic data. Molecular clocks have provided insights into significant evolutionary events such as human evolution and the radiations of birds, mammals, and flowering plants. Molecular-clock methods continue to undergo improvement, which is crucial if we are to take advantage of the growing wealth of genomic data (Kumar, 2005).

The molecular-clock hypothesis posits that rates of molecular evolution, as reflected in changes in DNA or protein sequences through time, are constant among lineages (but not across different regions of the genome). Nucleotide mutations cause DNA sequences to change
over time, whereas mutations of amino acids lead to evolution in proteins. These mutations occur with a constant probability rather than at a constant frequency – that is, the molecular clock is stochastic rather than metronomic (Zuckerkandl and Pauling, 1965). The occurrence of mutations in a DNA sequence is often modelled using a Poisson process.

A corollary of the molecular-clock hypothesis is that the genetic difference between any two species is proportional to the time since they last shared a common ancestor. This is a useful relationship because it allows evolutionary timescales to be estimated from genetic data, provided that the rate of molecular evolution is known. The molecular clock represents the only method for studying the evolutionary timescales of organisms that have failed to leave any trace in the fossil record, including many groups of invertebrates, bacteria, and viruses. It also allows us to estimate the timing of events that are too recent to be resolved by fossil evidence, such as divergences among conspecific populations (Arbogast et al., 2002).

History of the molecular clock

The molecular-clock hypothesis was put forward by Emile Zuckerkandl and Linus Pauling (1962), who assumed a constant evolutionary rate in their analysis of globin proteins from vertebrates. They observed about 18 differences in amino acids between horse and human and estimated the mutation rate by assuming that the divergence between these two species occurred 100–160 Ma ago. Upon extrapolating this rate, Zuckerkandl and Pauling (1962) estimated that humans diverged from gorillas about 11 Ma ago. They also estimated that different copies of globin genes first diverged from each other in the late Precambrian. When reporting these estimates, Zuckerkandl and Pauling (1962) warned about the possible confounding effects of a number of factors, including natural selection, variations in population size, and mutational saturation.

In the following years, further studies produced evidence of clocklike evolution in other proteins. Doolittle and Blomback (1964) found a simple relationship between sequence identity and time since divergence in mammalian fibrinopeptides (Figure 1). A year later, Zuckerkandl and Pauling (1965) coined the term “molecular evolutionary clock.” It promised to be a useful tool in biological research, as demonstrated shortly afterward by Sarich and Wilson (1967a, b) in their pioneering studies of the evolutionary timescale of hominids and other primates.

At the time, the idea of a constant substitution rate was controversial (Morgan, 1998). The molecular clock was received unenthusiastically by George Gaylord Simpson (1964), among others. Criticisms of the molecular clock were partly motivated by the apparent lack of uniformity in the pace of morphological evolution, which was presumably linked to molecular evolution. There was no evidence to suggest that adaptive change occurred at a uniform rate and it seemed implausible that the complex process of molecular evolution could be described by such a simple statistical model.

Meanwhile, there was growing evidence of high evolutionary rates in various proteins, suggesting that a large proportion of the changes in amino acids must have a negligible impact on evolutionary fitness. As a response, Motoo Kimura (1968) proposed the neutral theory of molecular evolution, which states that many mutations have such a small effect on the fitness of an organism that they can be considered as “neutral.” This can be explained by the fact that many amino acids in a protein can be exchanged for other amino acids with similar biochemical properties, with negligible impact on the overall function or structure of the protein. At the level of DNA, many mutations in protein-coding genes are “synonymous” because they do not result in any changes in the amino acid sequence of the protein. The fate of these “neutral” mutations is determined stochastically by a process known as genetic drift. The neutral theory was proposed independently by Jack King and Thomas Jukes (1969), who cited a range of evidence to support their hypothesis.

One of the predictions of the neutral theory is that rates of molecular evolution are constant among lineages. However, this prediction refers specifically to the rate of genetic change per generation. As a consequence, we expect to see a generation-time effect, whereby species with shorter generations tend to evolve more quickly per unit of time. For example, a higher rate would be observed...
in rodents than in whales. This is based on the assumption that most mutations occur during the replication of germline DNA. Such a generation-time effect was observed in the analyses of non-coding DNA (Laird et al., 1969; Kohne, 1970). This was in contrast with the rate of protein evolution, which appeared to be independent of generation time.

In response to the shortcomings of the neutral theory, Tomoko Ohta (1972, 1973) proposed the nearly neutral theory of molecular evolution. In this framework, there is a large class of “nearly neutral” mutations that have small effects on an organism’s fitness. In contrast with the results of the neutral theory, the nearly neutral theory states that the population sizes of species have a significant influence on the molecular evolutionary process. Many mutations are slightly harmful and are gradually removed from the population. In large populations, natural selection is effective at removing mutations that are detrimental to the organism’s fitness. In small populations, natural selection tends to be ineffective and most genetic change is driven by genetic drift. Generally, drift acts more quickly in small populations than does selection in large populations. However, a greater number of mutations appear each generation in large populations, simply because there are more individuals that can experience mutations. These two effects offset each other, leading to a relatively constant evolutionary rate among species per unit of time.

Through the ensuing decades, the molecular clock was used to study the evolutionary timescales of a range of organisms. Some of the most prominent studies involved analyses of the metazoan evolutionary timescale, with a focus on the divergences among animal phyla. In considerable conflict with the “Cambrian explosion” scenario supported by the fossil record, estimates from the molecular clock suggested a protracted Precambrian timescale for basal metazoan divergences (e.g., Dickerson, 1971; Runnegar, 1982; Doolittle et al., 1996). Similarly, molecular evidence pointed to much deeper diversifications of avian and mammalian orders than those suggested by paleontological evidence (e.g., Cooper and Penny, 1997; Springer, 1997). These discrepancies remain some of the key sources of debate about the accuracy of the molecular clock (Bromham and Penny, 2003).

Meanwhile, there was a growing body of evidence that pointed to rate variation among lineages. This came partly from various statistical tests that had been developed to test for clocklike evolution in genetic data (Wu and Li, 1985; Tajima, 1993). Departures from the molecular clock provided a challenge for studies of evolutionary timescales. It was not until the late 1990s, however, when molecular-clock methods were developed that were able to account for variation in evolutionary rates (Sanderson, 1997; Thorne et al., 1998). Known as “relaxed” molecular clocks because they relax the assumption of rate constancy among lineages, these have become the standard techniques in molecular evolutionary analysis.

### Rates of molecular evolution

Absolute rates of molecular evolution vary among regions of the genome because of differing constraints. Functionally important genes often evolve at an extremely slow pace, because many new mutations are detrimental to the organism and are rapidly removed by natural selection. As a consequence, such genes are often conserved even among distantly related organisms, making them useful genetic markers for studying deep evolutionary relationships. For example, histones, which are proteins that play an important role in binding and packaging DNA, have a very low substitution rate. In contrast, non-coding DNA has fewer functional constraints and can evolve at a higher rate than protein-coding DNA.

The mitochondrial genome evolves rapidly in animals and experiences about $10^{-8}$ substitutions per nucleotide per year, which is an order of magnitude higher than the average rate in the nuclear genome. This is in contrast with the pattern of molecular evolution in plants, where nuclear genomes evolve more quickly than either chloroplast or mitochondrial genomes. Rates of change in the genomes of viruses are higher by several orders of magnitude. In rapidly evolving RNA viruses, such as influenza virus and human immunodeficiency virus (HIV), mutation rates can exceed $10^{-3}$ mutations per nucleotide per year, and measurable genetic change can occur over a matter of weeks.

There is substantial variation in rates of molecular evolution across the tree of life. This variation is possibly driven by biological factors such as generation time, population size, longevity, and body temperature, as well as abiotic factors such as ultraviolet radiation. The relative importance of each of these factors is still poorly understood (Bromham, 2009). Studies have found evidence that molecular evolutionary rates are associated with longevity in mammals (Welch et al., 2008), with generation time in invertebrates (Thomas et al., 2010), and with height in plants (Lanfear et al., 2013). Research into the factors governing rate variation among organisms is ongoing.

### Molecular clocks and phylogenetic analysis

Molecular clocks are typically used in phylogenetic analyses, which aim to reconstruct evolutionary trees that show the relationships among species of interest (Figure 2). Internal nodes in the tree represent evolutionary divergence events. The timing of these events can be estimated using molecular clocks. A number of statistical methods are available for testing the molecular-clock hypothesis for a given set of DNA or protein sequences. When the molecular clock is rejected for a data set, one can use a statistical model to account for rate variation when estimating evolutionary timescales (Welch and Bromham, 2005).

When estimating evolutionary timescales in a phylogenetic analysis, the molecular clock needs to be “calibrated.” This can be done by assigning an absolute
Molecular Clocks, Figure 2 Phylogenetic tree showing the relationships of species from three genera, A, B, and C. Genera A and B are found on the mainland, whereas species of Genus C are found on an offshore island that was formed by volcanic activity. The tree is drawn to an unspecified timescale, so that the branch lengths are proportional to time. The tree illustrates the use of two molecular-clock calibrations. First, a fossil taxon has been assigned to the lineage leading to Genus A. This places a minimum age constraint on the divergence of A from B and C. Second, a geological calibration is placed on the divergence between B and C. The split between these genera is assumed to coincide with the formation of the offshore island on which Genus C is found. The age of the island is estimated using radiometric dating. The uncertainty in this date estimate is incorporated into the calibration, as indicated by the black bar. Gray error bars represent the uncertainty in the molecular-clock estimates of lineage divergence times.

age to one or more nodes in the phylogeny, which can then act as a reference point for estimating the ages of the remaining nodes. Calibrations are often based on the fossil record, which can provide date estimates for the divergence events in the phylogeny (Figure 2). If a fossil taxon can be reliably assigned to one of the lineages in the tree, then the divergence of that lineage from its sister lineage must be older than the age of the fossil. Calibrations can also be based on geological events, including the separation of continents or the emergence of islands, if they are linked to evolutionary divergences. An example of this might involve two sister genera, one of which is endemic to the mainland and the other endemic to an offshore island that formed as a result of volcanic activity (Figure 2). The timing of the divergence between the two genera can be estimated by the age of the island which, in turn, can be estimated using radiometric dating. In some cases, previous genetic estimates of dates are employed as calibrations. These are known as “secondary” calibrations and are generally used when other calibrating information is unavailable.

Once the phylogenetic tree is calibrated by fixing or constraining the ages of one or more nodes in the phylogeny, the ages of the remaining nodes can be estimated using a molecular-clock analysis of the genetic data. This process is referred to as “molecular dating” or “divergence-time estimation.” There is a large range of methods and models that have been developed for this purpose, implemented in various statistical frameworks. These methods are available in a range of computer programs, including most standard phylogenetic software.

Universal molecular clocks

There is abundant evidence of evolutionary rate variation among species, dispelling any hope of a universal molecular clock across the tree of life. However, some researchers entertain the idea of homogeneous rates of mitochondrial evolution within certain groups of organisms, such as birds, mammals, and arthropods. This is a convenient assumption because it allows evolutionary rates to be applied in molecular-clock analyses when fossil or geological calibrations are otherwise unavailable.

Many studies of birds assume that the mitochondrial genome evolves at about 1% per million years, a value that was initially based on a study of five species of geese (Shields and Wilson, 1987). Various analyses of avian mitochondrial DNA have supported this estimate, including an extensive study involving 74 calibrations and genetic data from 12 orders of birds (Weir and Schluter, 2008). The avian mitochondrial clock has been used to investigate key questions in the evolution of birds, including the impact of Pleistocene glaciation on the diversification of passerines. Some have questioned the reliability of this clock, citing evidence of significant rate variation among lineages and across timescales (García-Moreno, 2004; Ho, 2007). Despite this criticism, the avian mitochondrial clock is still widely used.

In a similar fashion, a universal mitochondrial clock of 1% per million years is often assumed for mammals. This rate was initially based on a mitochondrial study of primates and rodents (Brown et al., 1979), but it soon found support from estimates derived from other organisms (Wilson et al., 1985). Comprehensive analyses of mammalian DNA have demonstrated that there is substantial variation in rates among species, partly driven by differences in longevity (Nabholz et al., 2008; Welch et al., 2008).

Studies of evolutionary timescales in arthropods have often employed a rate of 1.15% per million years, which is based on an analysis of insects and crustaceans (Brower, 1994). This arthropod mitochondrial clock is widely used, partly because of the relative paucity of reliable fossil calibrations for many invertebrate taxa. However, this clock was based on a very small number of data points, leading to its validity being questioned (Papadopoulou et al., 2010; Ho and Lo, 2013). Moreover, there is strong evidence of mitochondrial and nuclear rate variation among invertebrate species (Thomas et al., 2010).

The employment of “standard” mitochondrial clocks, such as those described above, has often been criticized.
This is because the mitochondrial genome is subject to differing degrees of natural selection among species, which can lead to heterogeneity in evolutionary rates. Similar reasoning applies to evolutionary rates in the nuclear genome. Given that rates can show substantial variation even among closely related species, the use of standard clocks has the potential to yield date estimates that are highly misleading.

Controversies

The molecular clock has had a long history of controversy, beginning with the criticisms of the assumption of rate constancy and continuing with the debate over the relative merits of the neutral theory compared with natural selection. Current debates include the use of calibrations in genetic dating analyses and the constancy of evolutionary rates across timescales (Pulquério and Nichols, 2007).

The choice of calibrations has a substantial influence on the outcome of a molecular-clock analysis. Inadequate modelling of calibrations can lead to highly misleading estimates of evolutionary timescales. Accordingly, recent work has focused on accounting for the uncertainty in fossil and geological calibrations (Ho and Phillips, 2009). There is also ongoing development of methods for modelling and quantifying the uncertainty in fossil calibrations (e.g., Marshall, 2008; Wilkinson et al., 2011). Some aspects of calibration methodology are poorly understood, such as the impact of using different methods for modelling the uncertainty in calibrations. One of the key points of agreement is that best practice involves the use of multiple calibrations throughout the phylogenetic tree.

The use of secondary calibrations in molecular dating has been the target of strong criticism (Graur and Martin, 2004). Although the use of secondary calibrations is sometimes unavoidable, good practice involves taking into account the uncertainty in previous molecular-clock estimates. Ignoring this uncertainty can lead to artificially precise estimates of evolutionary timescales. Instead, the uncertainty should be incorporated into the dating analysis so that it can be included in the resulting estimates. This can readily be done in a Bayesian framework, in which the user can choose prior distributions that reflect the degree of uncertainty in the parameters and node times.

There is growing evidence that estimates of rates depend on the timescale of observation. Low rates of evolution are seen across long evolutionary time frames, such as those analyzed in phylogenetic analyses of distant species. In contrast, high evolutionary rates have been estimated in studies of populations and pedigrees (Howell et al., 2003). The exact causes of this disparity remain unclear, although they are likely to include the effects of natural selection and inaccurate modelling of the molecular evolutionary process (Ho et al., 2011). If the time dependence of molecular rates is not taken into account, evolutionary timescales can be under- or overestimated by an order of magnitude.

Conclusions

The molecular clock has undergone considerable evolution during its long history. It is useful as a hypothesis in molecular evolution and as a tool for estimating evolutionary rates and timescales. New molecular-clock methods are being developed in order to take advantage of the large amounts of genomic data that are being generated. With further refinement and development, the molecular clock will continue to play an important role in understanding the evolution of life on Earth.

Bibliography


Molecular clocks, human evolution


Cross-references
Gene Sequencing
Molecular Clock Calibration
Molecular Clocks, Human Evolution
Molecular Clocks, Relaxed Variant
Molecular Dating of Evolutionary Events
Molecular Rate Variation (Molecular Clocks)
Polymerase Chain Reaction DNA Amplification

MOLECULAR CLOCKS, HUMAN EVOLUTION

Simon Y. W. Ho¹ and Phillip Endicott²
¹School of Biological Sciences, University of Sydney, Sydney, NSW, Australia
²Département Hommes, Natures, Sociétés, Musée de l’Homme, Paris, France

Definition
Human evolution (molecular clocks). The timescale of human evolution and migration can be estimated from genetic data that have been sampled from living, historical, and ancient humans. This can be done using statistical methods based on the molecular clock hypothesis.

Introduction
The timescale of human evolutionary origins and migration across the globe has been a long-standing focus of scientific research. Genetic studies of the human evolutionary timescale were first performed in the 1960s, with the divergence time between humans and chimpanzees being estimated using data from the albumin protein (Sarich and Wilson, 1967). The complete sequence of the human mitochondrial genome, consisting of about 16.5 thousand nucleotides, was published in 1981 (Anderson et al., 1981). Two decades later, the International Human Genome Consortium (2001) released
Encyclopedia of Scientific Dating Methods
Rink, W.J.; Thompson, J.W. (Eds.)
2015, XXIX, 978 p. 375 illus., 260 illus. in color. In 2 volumes, not available separately., Hardcover
ISBN: 978-94-007-6303-6