CellPress

entirely clear, and it may also involve telomerase-independent mechanisms.

Recently, somatic mutations of POT1 were reported in 3.5% of cases of chronic lymphocytic leukemia. Most of these mutations clustered within the POT1-OB (oligonucleotide/oligosaccharide-binding) DNA-binding folds, and hence might compromise binding of POT1 to the single-stranded 3' overhang. Furthermore, rare germline variants of POT1 have been identified in familial cases of glioma and melanoma.

Is shelterin conserved in other species? All eukaryotes protect their chromosome ends with a telomere-binding protein complex. However, a shelterin-like complex is not always present. As in mammals, fission yeast telomeres are bound by a shelterin-like complex, consisting of a TPP1/POT1-like dimer, Tpz1–Pot1, and a TRF-like protein, Taz1. The Tpz1–Pot1 complex is connected to Taz1 via Rap1 and Poz1, establishing a link between the double-stranded and single-stranded telomeric DNA-binding factors. In contrast, the architecture of the telomere-binding complex and the proteins involved are quite distinct in budding yeast. Telomeres in budding yeast are bound by Rap1, which is the only structurally conserved shelterin component, although the mammalian and fission yeast Rap1 do not bind DNA. The single-stranded telomeric DNA in yeast is protected by the yeast CST complex.

Where can I find out more?

Arnoult, N., and Karlseder, J. (2015). Complex interactions between the DNA-damage response and mammalian telomeres. Nat. Struct. Mol. Biol. *22*, 859–866.

de Lange, T. (2010). How shelterin solves the telomere end-protection problem. Cold Spring Harb. Symp. Quant. Biol. *75*, 167–177.

Holohan, B., Wright, W.E., and Shay, J.W. (2014). Cell biology of disease: Telomeropathies: an emerging spectrum disorder. J. Cell Biol. *205*, 289–299.

Palm, W., and de Lange, T. (2008). How shelterin protects mammalian telomeres. Annu. Rev. Genet. *42*, 301–334.

Schmidt, J.C., and Cech, T.R. (2015). Human telomerase: biogenesis, trafficking, recruitment, and activation. Genes Dev. *29*, 1095–1105.

Laboratory for Cell Biology and Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10065, USA.
*E-mail: delange@rockefeller.edu

Primer

Molecular clocks

Michael S.Y. Lee[1] and Simon Y.W. Ho[2]

In the 1960s, several groups of scientists, including Emile Zuckerkandl and Linus Pauling, had noted that proteins experience amino acid replacements at a surprisingly consistent rate across very different species. This presumed single, uniform rate of genetic evolution was subsequently described using the term 'molecular clock'. Biologists quickly realised that such a universal pacemaker could be used as a yardstick for measuring the timescale of evolutionary divergences: estimating the rate of amino acid exchanges per unit of time and applying it to protein differences across a range of organisms would allow deduction of the divergence times of their respective lineages (Figure 1).

In the 50 years since, leaps in genomic sequencing technology and new computational tools have revealed a more complex and interesting reality: the rates of genetic change vary greatly across the tree of life. The term 'molecular clock' is now used more broadly to refer to a suite of methods and models that assess how rates of genetic evolution vary across the tree of life, and use this information to put an absolute timescale on this tree. Modern molecular clocks are thus critical to inferring evolutionary timescales and understanding the process of genetic change. Analyses of genomic data using clock models that accommodate variation in evolutionary rates have shed new light on the tree of life, as well as the organismal and environmental factors driving genetic change along its branches. However, some major theoretical, empirical and computational challenges remain.

Evolutionary rate variation

Modern molecular clocks can handle various forms of evolutionary rate heterogeneity. Rates can vary across different parts of the genome (site effects), across taxa (lineage effects), and across time (here termed 'epoch effects'). Site effects occur when different parts of the genome evolve at distinct rates (Figure 2A). A widely recognized example involves protein-coding genes, which have a higher rate of evolution at the third position of codons than at the first and second. This is because changes at first and second codon sites are more likely to change the encoded amino acid, with potential consequences for protein function. In animals, mitochondrial DNA evolves faster than nuclear DNA, for reasons that are still debated. These site effects were the first major sources of rate heterogeneity to be characterized and accounted for during genetic analysis.

Lineage effects occur when different taxa exhibit distinct rates of molecular evolution (Figure 2B). For example, rodents have higher rates of genetic change than do other mammals, partly due to their short generation times. Likewise, parasitic plants evolve more rapidly than their free-living relatives. The importance of this form of rate variation took longer to be appreciated, but was confirmed in the 1970s when formal statistical tests of among-lineage rate variation were developed. This led to the introduction of 'relaxed-clock' approaches, which attempt to statistically model rate variation across branches of the evolutionary tree. These methods allow evolutionary timescales to be estimated using molecular clock approaches even when rates vary across lineages.

Epoch effects occur when rates of evolution differ across different time slices (Figure 2C). For instance, evolutionary rates in influenza were found to have undergone a sharp increase around 1990. Such temporal heterogeneity is harder to detect and model than either site effects or lineage effects. This is partly because it generates patterns of genetic divergence among living taxa that are very similar to those expected when rates have remained constant through time.

An extra layer of interest and complexity emerges when two or more sources of rate heterogeneity interact. Site and lineage effects interact when different genes have different patterns of rate variability across taxa (Figure 2D). Mitochondrial DNA has greatly accelerated rates of evolution in snakes and dragon lizards

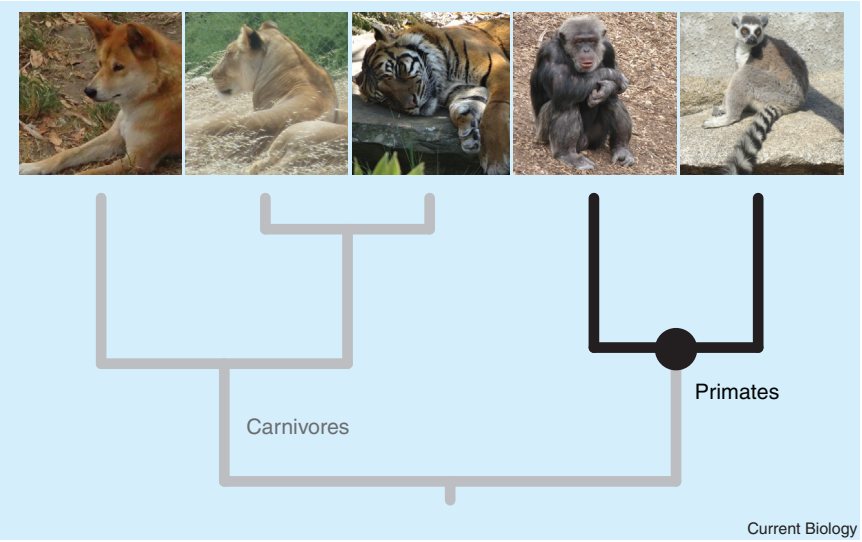

Figure 1. The simplest molecular clock approach for inferring evolutionary timescales.
The rate of genetic change is first ascertained for one part of the tree of life (e.g. primates), often by calibrating the amount of genetic divergence to the absolute age of divergence as suggested by the fossil record. This rate is then extrapolated across the rest of the tree, allowing relative genetic divergences between all other taxa (e.g. carnivores) to be translated into absolute time, even without recourse to fossil evidence.

compared with typical lizards, but nuclear DNA shows no such trend. Genomic analyses suggest that such interactions are widespread. Selection might be relaxed on particular genes in particular taxa and thus lead to rapid molecular evolution. For example, the genes coding for tooth enamel are no longer under stabilizing selection in toothless mammals such as anteaters and sloths. Thus, those genes evolve much more rapidly in these lineages, but this pattern is not seen for most other genes. Such complex patterns of rate variation can be accommodated using partitioned clock models, where different portions of the genome are recognized as evolving according to separate clocks or 'pacemakers'.

Calibrating the molecular clock

Genetic divergences alone, even when analysed using the most sophisticated molecular clock models, are only able to provide a relative timescale. For example, DNA evidence suggests that the major lineages (orders) of living birds diverged from each other during the first quarter of the evolutionary history of modern birds, but does not tell us the absolute timeframe of this diversification. The molecular clock needs to be calibrated in order to translate these relative dates into absolute ones. We would then be able to make statements such as "the major lineages of birds diverged in an interval of 20 million years spanning the end of the Cretaceous period".

Calibrations are typically derived from the fossil record: for instance, when dating a molecular tree of vertebrates, the clade 'modern birds' must be at least as old as the most ancient fossil that can be robustly assigned to that group, currently the 67-million-year-old *Vegavis*. Well-dated geological events — such as island formation, continental rifting or river capture — can also be used to constrain the ages of evolutionary divergences between taxa presumed to have been affected by these events. For very shallow trees, spanning short time periods, such as those of virus epidemics, the ages of 'fossilized' genomes sampled across real time can be used to calibrate the molecular clock. With caution, previous estimates of evolutionary rates and divergence times can also be used for calibration.

Most attention has focused on statistical approaches for capturing information from the fossil record for calibrating molecular trees, but similarly rigorous approaches are now being developed for biogeographic information. Different types and numbers of calibrations can be used simultaneously, but this can change the balance between the signals from the calibrations and from the genetic data. Incorporating as much calibrating information as possible can severely constrain the possible range of inferred timescales. On the other hand, using a smaller subset of temporal information allows the molecules more latitude to speak for themselves.

Evolutionary timescales

Molecular clocks are vital to reconstructing the detailed timescale and branching pattern of the tree of life, especially in soft-bodied groups that have left few or no fossils. In turn, this can shed light on how major evolutionary events have been influenced by Earth history. However, the use of inappropriate clock models or erroneous calibrations can produce highly misleading estimates of evolutionary timescales. These issues have led to vigorous debates about the timing and drivers of major evolutionary events, including the origins of animal phyla, the ordinal divergences of birds and mammals or the radiation of flowering plants.

Some of the earliest molecular clock analyses of divergences between animal phyla concluded that metazoans diverged about a billion years ago — nearly twice the age of the explosion of animal fossils in Cambrian rocks. These results were at least partly driven by failure to account for lineage effects: genetic change generally occurs more slowly in vertebrates than in invertebrates, but early molecular analyses extrapolated the slow vertebrate evolutionary rate across the entire animal tree. This caused the estimates of animal divergence times to be stretched deep into the Precambrian. Subsequent analyses with better models of rate variation and more carefully chosen calibrations moved the initial radiation of animals to a later time — into the early Ediacaran period, when the world was gripped by several massive glaciation events ('snowball earth'). Nevertheless, this still precedes the first definitive metazoan fossils by tens of millions of years.

In other groups of organisms, improved molecular clock analyses have also often increased the congruence between timescales