# The Molecular Clock and Evolutionary Rates Across the Tree of Life

Simon Y. W. Ho

## Abstract

The molecular evolutionary clock was proposed in the 1960s and has undergone considerable evolution over the past six decades. After arising from early studies of the amino acid sequences of proteins, the molecular clock became a point of contention between competing theories of molecular evolution. In this chapter, I describe the origins of the molecular clock hypothesis and the mixture of evidence that emerged throughout the 1970s and 1980s, including the discovery of departures from clocklike evolution in proteins and DNA. I review some of the broad patterns of evolutionary rate variation across the tree of life, including rates of spontaneous mutation and long-term evolution in viruses, bacteria, animals, and plants. With the remarkable growth of genomic data over the past two decades, the molecular clock is now primarily seen as a tool for reconstructing evolutionary timescales. In the final parts of this chapter, I summarize the key developments in molecular dating methods and describe how these approaches have been used to infer the timing of major evolutionary events.

## 1.1 Introduction

The molecular evolutionary clock has had a profound influence on molecular evolutionary theory, while also providing an indispensable tool for inferring evolutionary rates and timescales. Starting from the simple premise that evolutionary change at the molecular level proceeds at a relatively constant rate, the molecular clock has undergone considerable evolution over the past six decades (Fig. 1.1). The history of research on the molecular clock has featured an extensive debate over molecular evolutionary theory, persistent challenges to its assumptions and predictions, and applications to questions about the timing of major biological events. Throughout this time, researchers have devoted substantial efforts to understand the causes of evolutionary rate variation across the tree of life, and to apply the principle of the molecular clock in methods for estimating evolutionary timescales. The molecular clock has now confirmed its important role in research in the life sciences, finding applications in such diverse fields as evolutionary biology, molecular ecology, archaeology, and epidemiology.

S. Y. W. Ho (✉)

School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia
e-mail: simon.ho@sydney.edu.au

The idea of a molecular clock emerged from studies of proteins in the mid-twentieth century, a time when new biochemical and genetic data were bringing important insights into evolutionary biology. In particular, efforts to determine the amino acid sequences of proteins were yielding valuable data sets that could inform evolutionary thinking. A series of innovative studies in the early 1960s gave rise to the molecular clock (Zuckerkandl and Pauling 1962, 1965; Margoliash 1963; Doolittle and Blombäck 1964), which soon grew to become an integral part of the neutral theory of molecular evolution (Kimura 1968, 1969). In the ensuing decades, the molecular clock played a central role in the debates between neutralists and selectionists, who supported opposing theories of molecular evolution (Ohta and Gillespie 1996). In the present genomic age, the molecular clock is perhaps most widely recognized as a tool for estimating the timing of evolutionary events (Bromham and Penny 2003).

This book provides an overview of the molecular evolutionary clock, including its theory and practice. It attempts to cover a huge field of research that cannot be satisfactorily summarized in an individual review article; nevertheless, this book can only be considered as an introductory text. Many of the chapters in this book focus on recent developments in this fast-moving field, including the latest endeavours to cope with genome-scale data sets and to combine molecular, phenotypic, and palaeontological data in a biologically meaningful way.

In this opening chapter, I describe the origins of the molecular clock and its evolution over the past six decades. I then provide an overview of the different forms of evolutionary rate variation across the tree of life, ranging from viruses and bacteria to eukaryotes. The chapter concludes with a description of how molecular clocks are used to infer evolutionary timescales, including a summary of some of the major applications of molecular dating. Throughout this chapter, I introduce the contents of the remaining chapters of the book.

## 1.2 The Molecular Clock Hypothesis

### 1.2.1 Origins of the Molecular Clock

The term 'molecular evolutionary clock' was proposed by Emile Zuckerkandl and Linus Pauling in 1965. Zuckerkandl had joined Pauling in the California Institute of Technology in late 1959 and the two worked on the sequencing and analysis of the haemoglobin protein (Morgan 1998). Less than a decade earlier, the first amino acid sequence of a protein, insulin, had been determined. Zuckerkandl and Pauling (1962) noted that the divergence in the amino acid sequence of haemoglobin increased over time with the evolutionary distance between species. They made the inspired assumption that a simple linear relationship existed between the two quantities.

Zuckerkandl and Pauling (1962) raised the possibility of using this clocklike property to develop a tool for estimating the timing of divergence between haemoglobin chains and between vertebrate species. Based on a palaeontological estimate of 100–160 million years (Myr) for the divergence between human and horse, they inferred an evolutionary rate of 1 amino acid substitution per 14.5 Myr (Fig. 1.2a). Their application of this rate to the amino acid sequences yielded estimates of the divergence times between haemoglobin chains, with the $\alpha$ chain splitting from the $\beta$ and $\gamma$ chains about 565–600 Myr ago in the late Precambrian. The divergences between the $\beta$ chain and the $\gamma$ and $\delta$ chains were estimated to have occurred much more recently, at 260 Myr ago in the Permian and 44 Myr ago in the Eocene, respectively.

In their analysis of haemoglobin, Zuckerkandl and Pauling (1962) also obtained an estimate of 11 Myr for the evolutionary split between gorilla and human (Fig. 1.2a). They noted that this estimate was at the lower end of the timing of 11–35 Myr ago suggested by the fossil record. Their estimate, and other molecular estimates of the hominid evolutionary timescale reported in the 1960s (Sarich and Wilson 1967a), were controversial because they were inconsistent with the
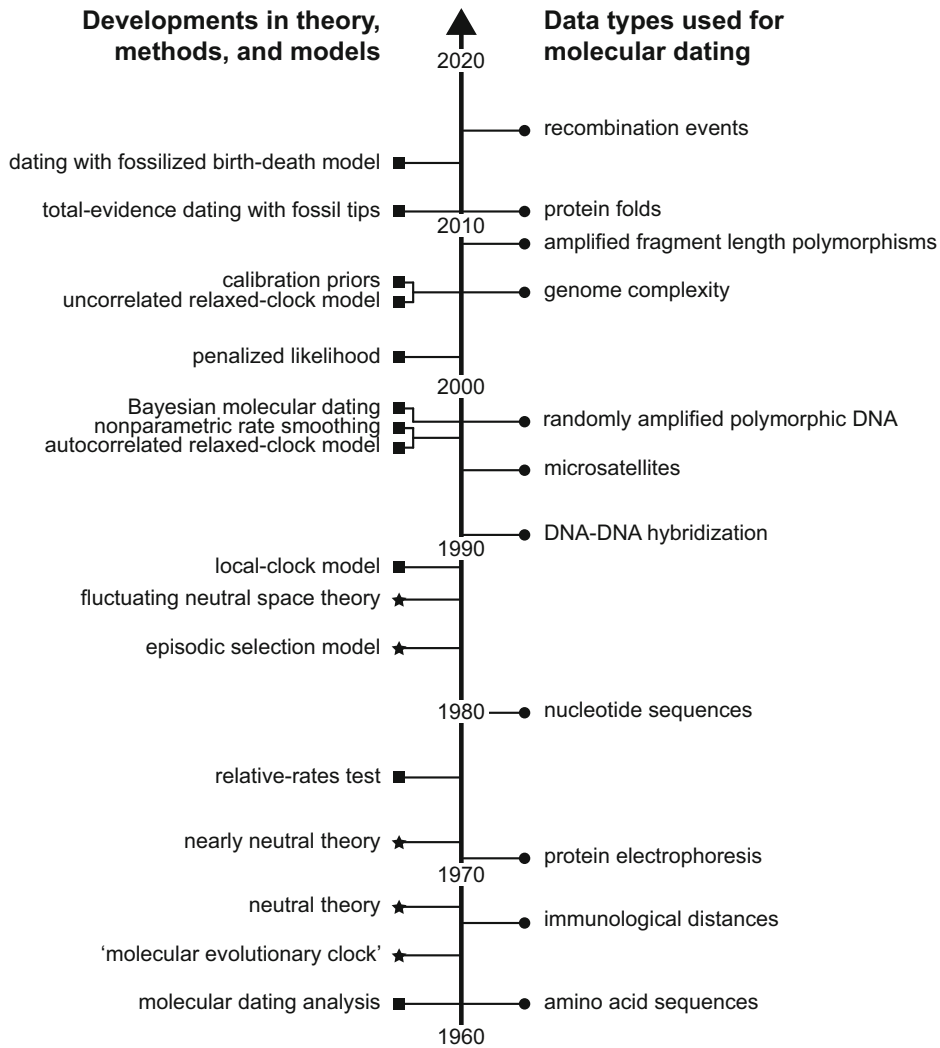
**Fig. 1.1** Timeline of advances throughout the history of the molecular clock, beginning with its application to amino acid sequences (Zuckerkandl and Pauling 1962). The left side of the timeline lists some of the key developments in molecular evolutionary theory (stars) and in molecular dating methods and models of evolutionary rate variation (squares). The term 'molecular evolutionary clock' was introduced in 1965. Most of the developments listed here are referred to explicitly in the main text of this chapter. The right side of the timeline lists the first use of different data types for molecular dating (circles; for references, see Ho et al. 2016). Nucleotide sequences are now the most widely used type of genetic data in molecular dating analyses

prevailing notion of a large evolutionary distance between modern humans and the other great apes (Wilson et al. 1977). However, reports would soon emerge of constant evolutionary rates in the amino acid sequences of cytochrome *c* (Margoliash 1963) and fibrinopeptides (Fig. 1.2b; Doolittle and Blombäck 1964), lending support to the molecular clock hypothesis.

In addition to developing a tool for inferring evolutionary timescales, Zuckerkandl and Pauling (1962) foresaw some of the problems that would beset molecular clock analyses in subsequent decades. They referred to the problems posed by repeated substitutions at the same amino acid site (including back-mutations), the potentially confounding impacts of natural selection, and the influence of population size.
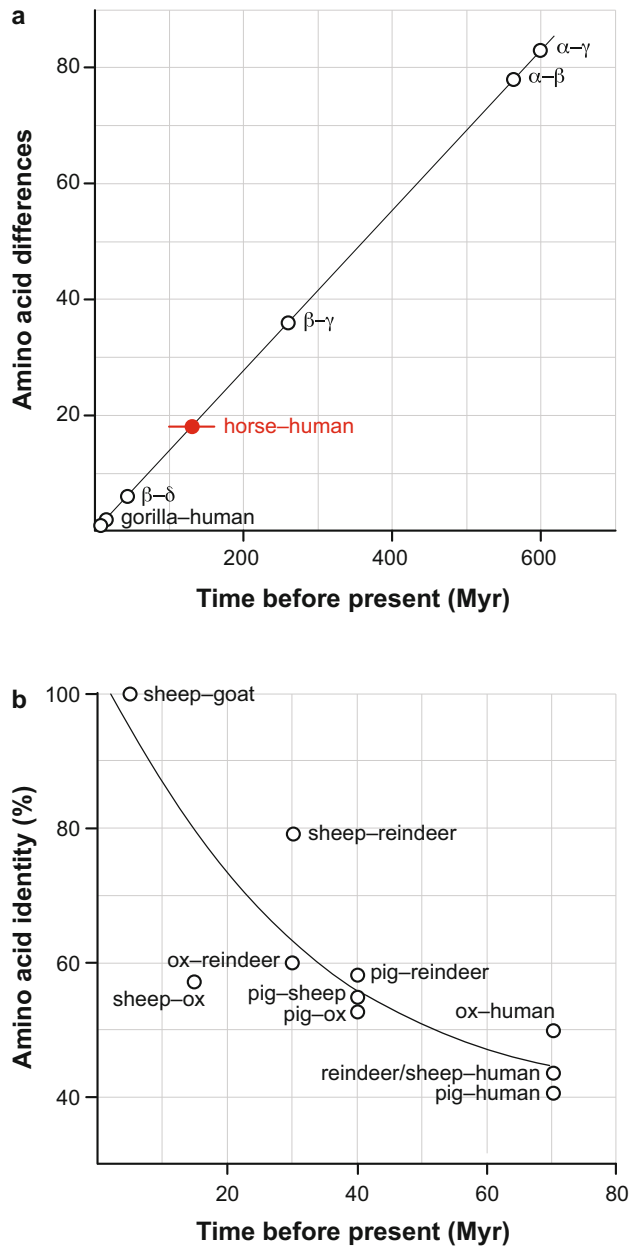
**Fig. 1.2** (**a**) The earliest use of the molecular clock to infer evolutionary divergence times, based on amino acid sequences of haemoglobin (Zuckerkandl and Pauling 1962). The evolutionary rate was calibrated using a palaeontological estimate of the horse–human divergence at 100–160 million years ago. Assuming a constant rate of amino acid replacements, the divergence time of gorilla and human was estimated (represented by two data points corresponding to the divergences between the two $\alpha$ chains and between the two $\beta$ chains), along with the divergence times of various pairs of haemoglobin chains (denoted by Greek letters). Data from Zuckerkandl and Pauling (1962). (**b**) Clocklike evolution in fibrinopeptides, based on pairwise comparisons of amino acid sequences in sheep, goat, reindeer, ox, pig, and human. Pairwise amino acid sequence identity (%) is plotted against the time of divergence estimated from the fossil record. Data from Doolittle and Blombäck (1964)

Their idea of the molecular clock acknowledged an important role for natural selection, although they later surmised that 'the changes that occur at a fairly regular over-all rate would be expected to be those that change the functional properties of the molecule relatively little' (p. 148, Zuckerkandl and Pauling 1965). This statement seemed to anticipate the close association that would soon form between the molecular clock and the neutral theory of molecular evolution (e.g., Kimura 1968; King and Jukes 1969; Wilson and Sarich 1969).

The neutral theory, put forward by Motoo Kimura in 1968, made the bold assertion that the majority of mutations are neutral. This contradicted the dominant view that such mutations are rare or transient (Fisher 1936; Mayr 1963), although the importance of neutral mutations in molecular evolution had been suggested earlier in the same decade (Freese 1962; Sueoka 1962). In Kimura's proposal, the term 'neutral' was not intended to suggest that the corresponding gene lacked function (e.g., Zuckerkandl 1978), but instead meant that the mutation conferred neither an advantage nor disadvantage to the organism and that the fate of the mutation would be governed by genetic drift. Although the molecular clock was influential in the development of the neutral theory (Takahata 2007), Kimura's case for the theory largely rested on estimates of enzyme variability from electrophoretic studies and rates of protein evolution inferred from analyses of amino acid sequences. He argued that these high evolutionary rates greatly exceeded the limits imposed by the 'cost of natural selection' (Haldane 1957), thus suggesting that many of the mutations must be neutral (Kimura 1968).

A significant consequence of the neutral theory is that the rate at which neutral mutations are fixed in the population (known as the 'substitution rate') is approximately equal to the rate at which the mutations are spontaneously generated (Kimura 1968). For this reason, the molecular clock was regarded as an additional source of evidence for the neutral theory (Kimura 1969, 1983). In Chap. 2, Soojin Yi provides an introduction to molecular evolution, including the

neutral theory and its later developments, as well as some of the principles behind the molecular clock. She also explains the relationship between the mutation rate and substitution rate under the neutral theory.

The initial reactions to the proposal of the molecular clock were largely negative (e.g., Stebbins and Lewontin 1972), with criticisms being levelled by a number of eminent evolutionary biologists. For example, Ernst Mayr argued that 'evolution is too complex and too variable a process, connected with too many factors, for the time dependence of the evolutionary process at the molecular level to be a simple function' (p. 137, Zuckerkandl and Pauling 1965). At the time, the evolutionary biologist Morris Goodman was one of the few to recognize the potential applications of the clock (Morgan 1998). With further evidence for the constancy of molecular evolutionary rates, as well as growing appreciation of its great potential for reconstructing the timescale of evolution, the notion of a molecular clock endured. By the late 1970s, Allan Wilson et al. (1977) declared that the 'discovery of the evolutionary clock stands out as the most significant result of research in molecular evolution' (p. 577).

### 1.2.2 Decades of Evolution

The molecular clock was a prominent source of contention in the molecular evolutionary debates throughout the 1970s to 1990s, an era that also saw a shift in focus from protein sequences to DNA sequences (Fig. 1.1; Ohta and Gillespie 1996; Nei et al. 2010). In the early part of this period, there was growing evidence of a discrepancy between the evolutionary dynamics of 'silent' (synonymous or non-coding) and 'replacement' (nonsynonymous) changes in DNA. Replacement substitutions occurred at a constant rate per year, which was cited as support for the neutral theory (Kimura 1969). However, silent substitutions, which are expected to be under much lower selective constraint, appeared to occur at a constant rate per generation (Laird et al. 1969; Kohne 1970). There was evidence of a

slowdown in evolutionary rates of both proteins and DNA in hominoids compared with other primates and mammals, particularly rodents (Goodman 1961; Kikuno et al. 1985; Wu and Li 1985), in accordance with the differences in generation times among these organisms.

Kimura (1983) later recognized that the neutral theory should predict a constant substitution rate per generation rather than per year, while admitting that evidence of the constancy of evolutionary change per unit time presented a 'difficult problem' (p. 246) for the theory. The different dynamics observed for silent and replacement substitutions were partly reconciled in the nearly neutral theory, developed by Tomoko Ohta (1972, 1973). The nearly neutral theory proposed that many mutations have a small impact on fitness and are mildly deleterious or mildly advantageous (see Chap. 2), and predicts a constant evolutionary rate per unit of time. However, this prediction relies on a negative correlation between population size and generation time, which was assumed but not explicitly demonstrated by Ohta (1972, 1973). In any case, as described by Gillespie (1991), Kimura 'quickly retreated from the [per-year constancy of mutation rates] when he adopted Ohta's mildly deleterious theory' (p. 274). Nevertheless, upon considering the evidence of a generation-time effect, Kimura (1987) noted that the departures from rate constancy across lineages were not as great as would be expected on the basis of differences in generation time.

A somewhat different challenge to the hypothesis of a molecular clock was that the occurrences of substitutions were often found to be more erratic than expected. Zuckerkandl and Pauling (1965) had suggested that amino acid substitutions occur stochastically, following a Poisson point process. Under this stochastic process, the variance in the number of substitutions per unit time is equal to the expected number of substitutions per unit time. The ratio of these quantities, known as the index of dispersion, provides a measure of the departure from a Poisson process; values exceeding 1 indicate overdispersion. Studies of proteins found that overdispersion was widespread among proteins

(Ohta and Kimura 1971; Langley and Fitch 1974; Gillespie 1984, 1989), contradicting the expectations under the molecular clock. One attempt to explain this overdispersion within the framework of the neutral theory was based on a model of fluctuating neutral space (Takahata 1987), in which each neutral mutation changes the rate of neutral mutations. However, most explanations appealed to the effects of natural selection, with overdispersion being a potential outcome under some conditions of episodic, fluctuating, or negative selection (Gillespie 1984, 1993; Cutler 2000). There is now a body of evidence showing that some features of molecular and genomic evolution cannot be adequately explained by the neutral theory (e.g., Kreitman and Akashi 1995; Kern and Hahn 2018).

The molecular clock gradually moved away from its conspicuous role in the selectionist–neutralist debate and became increasingly appreciated for its practical applications in evolutionary biology. Although there is continued interest in the causes of evolutionary rate variation, the molecular clock is now most widely known as a tool for inferring evolutionary timescales. However, the utility of the molecular clock as a dating tool is potentially diminished by the presence of evolutionary rate variation. There have been considerable efforts to rescue the molecular clock from this quagmire, leading to major advances in molecular dating methods over the past two decades.

## 1.3 Evolutionary Rate Variation

### 1.3.1 Partitioning Variation in Rates

Evolutionary rate variation occurs in different modes and across a range of temporal, molecular, and biological scales. Early studies considered differences in rates across nucleotide or amino acid sites (site effects), across genes or loci (gene or locus effects; Fig. 1.3a), and across lineages (lineage effects; Fig. 1.3b). For a given gene, any overdispersion that remained after accounting for lineage effects was ascribed to residual effects (e.g., Langley and Fitch 1974;
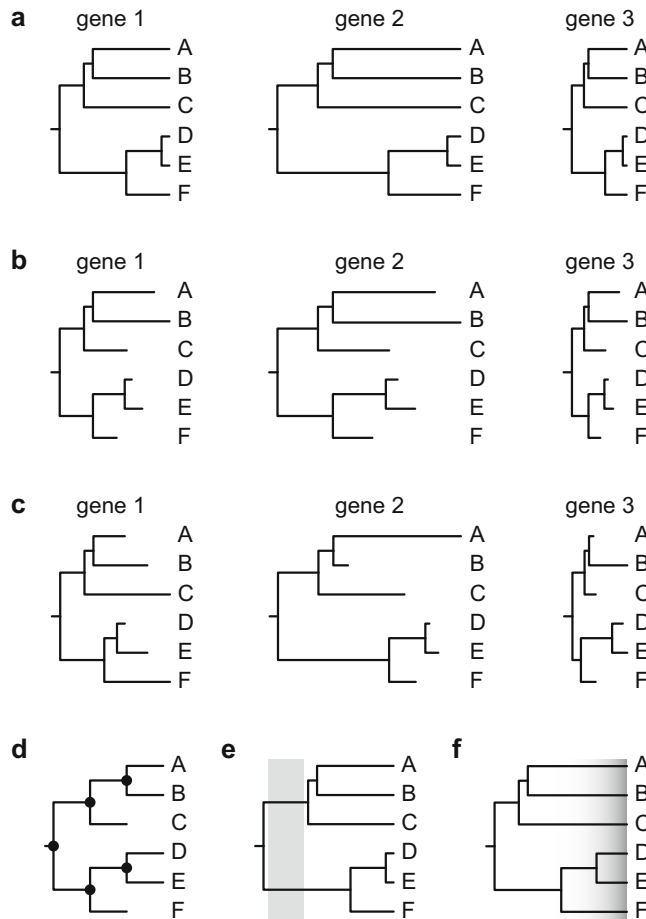
**Fig. 1.3** Evolutionary rate variation depicted in phylogenetic trees with branch lengths proportional to the amount of genetic change. Each tree consists of six taxa, labelled A to F. (**a**) Gene effects lead to rate variation across genes, so that there are different total amounts of genetic change in the phylogenetic trees from genes 1, 2, and 3. (**b**) Lineage effects lead to rate variation across the branches of the tree, but this effect is shared by all of the genes. Accordingly, the branch lengths of the three phylogenetic trees share the same proportions. (**c**) Gene-by-lineage interactions lead to different patterns of among-lineage rate variation across the phylogenetic trees from the three genes. (**d**) Punctuated evolution proposes that bursts of genetic change occur at speciation events, denoted by circles. This leads to a pattern in which the length of the path between the root and any tip of the tree is approximately proportional to the number of speciation events along that path. (**e**) Epoch effects occur when there is a different evolutionary rate during a particular period of time, as indicated by the grey shaded area. (**f**) Time-dependent rates lead to a bias whereby evolutionary rates are higher when estimated over recent, short-term timescales than over longer periods of time

Gillespie 1991). In their comprehensive review of evolutionary rate variation in plants, Gaut et al. (2011) used an approach inspired by an analysis of variance that had been conducted 8 years earlier (Smith and Eyre-Walker 2003). Specifically, in addition to site effects, gene effects, and lineage effects, they considered the two- and three-way interactions among these three components: site-by-gene effects, site-by-lineage effects, gene-by-lineage effects, and site-by-gene-by-lineage effects. For molecular clocks, the most important of these effects are caused by gene-by-lineage interactions (Fig. 1.3c); these are analogous to residual effects (Gillespie 1991).

### 1.3.2 Site Effects and Gene Effects

Site effects can be caused by differences in selective constraints on individual nucleotides or amino acids and by heterogeneity in mutation rates (Hodgkinson and Eyre-Walker 2011). Functionally or structurally important sites tend to evolve more slowly than other sites, or might even be invariant to change, and such amino acid sites in cytochrome *c* and haemoglobin were discussed at length by Zuckerkandl and Pauling (1965). Differences in the proportions of such constrained sites were argued to be the main cause of rate variation across proteins under the neutral theory (King and Jukes 1969). In the nucleotide sequences of protein-coding genes, nonsynonymous mutations are more likely to be selected against than are synonymous mutations. The distinction between 'silent' and 'replacement' dynamics in DNA sequences was already well appreciated by the 1970s (e.g., King and Jukes 1969; Jukes and Kimura 1984), and its varying effects on rates at the three codon positions in protein-coding genes are now routinely taken into account in analyses of nucleotide sequences (e.g., Shapiro et al. 2006).

Mutation rates can also vary among nucleotides and according to the local context, with mutations at cytosine-guanine dinucleotides ('CpG') occurring at higher rates than at other dinucleotides partly because of the vulnerability of the cytosine to deamination (see Chap. 2). Studies of genomic data have revealed other forms of site effects, such as higher mutation rates in parts of the genome linked to insertions and deletions (Tian et al. 2008). In analyses of molecular sequence data, site effects are typically accommodated by modelling the site rates using a gamma distribution (Yang 1996). As with many models in biology, this approach aims to capture an important feature of sequence evolution without attempting to resolve the underlying mechanisms.

Gene effects were widely recognized during the development of the molecular clock (Fig. 1.3a), with evidence of evolutionary rate variation across haemoglobin, cytochrome *c*, fibrinopeptides, and other gene products (e.g., Zuckerkandl and Pauling 1965; Dickerson 1971). In her extensive survey of protein sequences, the pioneering biochemist Margaret Dayhoff (1978) found nearly 400-fold variation in evolutionary rates across proteins. Many of the causes of site effects also lead to rate variation across genes, so the two forms of variation are closely linked. However, the evolutionary rates of genes and proteins are most strongly correlated with their levels of expression (e.g., Rocha and Danchin 2004; Park et al. 2012) and not their functional importance (Wang and Zhang 2009). A negative relationship between the expression level of a protein and its evolutionary rate has been found across a wide range of organisms, including bacteria and eukaryotes, but the specific causes of this relationship remain unclear (Zhang and Yang 2015). In contrast with the rate variation across proteins, rates of synonymous substitutions show little variation across protein-coding genes in mammalian genomes (Kumar and Subramanian 2002).

On a broader scale, evolutionary rates can show substantial disparities between nuclear and organellar genomes. A widely recognized pattern in metazoans is that mutation rates are much higher in the mitochondrial genome than in the nuclear genome (Brown et al. 1979; Miyata et al. 1982). However, the ratio of mitochondrial to nuclear evolutionary rates has been found to be considerably greater in birds, reptiles, and other vertebrates than in insects and arachnids (Allio et al. 2017). The comparatively high information content of mitochondrial DNA ensured that it held a long reign as the preferred marker in studies of population genetics, molecular systematics, and phylogenetics in humans and other animals (Avise et al. 1987). The popularity of mitochondrial DNA declined with the advent of high-throughput sequencing technologies, which enabled nuclear genome data to be obtained efficiently and on large scales, and with growing concerns about excessive reliance on a single genetic marker.

In contrast with the trends observed in animals, elevated mutation rates are not seen in the mitochondrial genomes of other eukaryotes

(Baer et al. 2007). In plants, nuclear genomes evolve more rapidly than chloroplast genomes, which evolve more rapidly than mitochondrial genomes (Wolfe et al. 1987). This pattern is particularly pronounced in angiosperms, but less so in gymnosperms (Drouin et al. 2008). The reasons for the low evolutionary rates in the chloroplast and mitochondrial genomes of plants are not entirely clear, but might be related to DNA repair mechanisms (Christensen 2013). In plastid-bearing eukaryotes other than land plants, mitochondrial genomes have a higher evolutionary rate than plastid genomes (Smith 2015).

### 1.3.3   Lineage Effects and Gene-by-Lineage Interactions

Evidence of lineage effects emerged soon after the proposal of the molecular clock and continued to grow in the ensuing decades (Fig. 1.3b). The generation-time effect, as described in Sect. 1.2.2, appeared to be the most prominent form of evolutionary rate variation across lineages. The hominoid slowdown in evolutionary rates, first quantified by Goodman (1961), has been confirmed in genome-scale analyses of primates (Kim et al. 2006; Chintalapati and Moorjani 2020). A generation-time effect has now been found in a variety of organisms, including bacteria (Weller and Wu 2015), birds (Mooers and Harvey 1994), and invertebrates (Thomas et al. 2010), and broadly across animals (Allio et al. 2017). However, evolutionary rates appear to show a more complex relationship with generation time in plants, in which the germline is segregated at a late stage of their growth (Lanfear et al. 2013).

Lineage effects can be detected using a variety of methods. Sarich and Wilson (1967b, 1973) described a framework for comparing the relative rates between a pair of taxa, which was later developed into a statistical test (Fitch 1976; Wu and Li 1985). The relative-rates test has largely been superseded by methods that can test for among-lineage rate heterogeneity across an entire phylogenetic tree. These include the likelihood-ratio test, which can be used to compare a model in which the phylogeny is constrained to be ultrametric (all tips being equally distant from the root of the tree) against a model in which the branch lengths are unconstrained (Felsenstein 1981). The rapid increase in genetic data throughout the 1980s and 1990s led to an accumulation of evidence of evolutionary rate variation (Britten 1986; Drake et al. 1998). Some of the major patterns of rate variation across the tree of life are described in Sect. 1.4.

Gene-by-lineage interactions (Fig. 1.3c), which comprise the variation in evolutionary rates that are not accounted for by gene effects or lineage effects, represent an additional layer of complexity in patterns of rate variation (e.g., Gillespie 1989; Ayala 1997). These interactions have been found to be more prominent in nonsynonymous than synonymous rates in plant chloroplast genomes (Muse and Gaut 1997). Gene-by-lineage interactions appear to account for a small proportion of evolutionary rate heterogeneity in mitochondrial and nuclear genes from eutherian mammals (Smith and Eyre-Walker 2003), but are potentially important when large sets of genes are being analysed for the purposes of inferring evolutionary timescales. Variation across genes and across lineages are the dominant forms of genome-scale rate heterogeneity (Snir et al. 2012), although gene-by-lineage interactions have been detected in genomic data from eutherian mammals (Duchêne and Ho 2015) and flowering plants (Duchêne et al. 2016a). Further genomic analyses will allow the different forms of evolutionary rate variation to be characterized for other groups of organisms.

### 1.3.4   Other Forms of Evolutionary Rate Variation

The framework used in the previous section provides a helpful means of partitioning rate variation into its major components, allowing consideration of the biological and evolutionary drivers of rates of mutation and substitution (Fig. 1.3). Nevertheless, there are several important features of evolutionary rate variation that do not fit neatly

into this classification. Here I describe three of these phenomena: punctuated evolution, epoch effects, and time-dependent rates. These forms of rate variation can pose substantial challenges for using molecular clocks to infer evolutionary timescales.

The punctuated equilibrium theory was put forward in an attempt to explain patterns in the fossil record, which appears to feature long periods of stasis punctuated by rapid bursts of morphological change (Eldredge and Gould 1972). Inspired by this theory, molecular evolutionary biologists have sought evidence of bursts of genetic change caused by founder effects at speciation events (Fig. 1.3d; Webster et al. 2003; Pagel et al. 2006). These can potentially be detected using a phylogenetic approach to analyse molecular sequence data, because the theory predicts that a measurable proportion of genetic change is correlated with the number of speciation events along any lineage in the evolutionary tree. However, tests of punctuated evolution have been seriously hindered by a problem known as the node-density effect, which produces patterns similar to those expected under punctuated molecular evolution (Fitch and Beintema 1990). Newly developed phylogenetic models of evolutionary rates might be able to shed further light on the occurrence of punctuated molecular evolution (Manceau et al. 2020).

Rates of molecular evolution can vary across time periods, leading to epoch effects (Fig. 1.3e; Lee and Ho 2016). For example, some external factors, such as environmental conditions, might raise evolutionary rates across an entire population or even an entire assemblage of organisms. One potential example is a several-fold increase in phenotypic and genomic evolutionary rates during the rapid diversification of metazoan phyla in the Cambrian, an event that is often referred to as the 'Cambrian explosion' (Lee et al. 2013). Epoch effects are particularly difficult to identify unless the period of evolutionary rate elevation can be bracketed by reliable age constraints from the fossil record. For example, epoch effects cannot be detected by a likelihood-ratio test for clocklike evolution, in which the null hypothesis is that all of the tips are the same distance from the root of the tree (Yang 2014).

The study of evolutionary rates has been hindered by a time-dependent bias, which causes rate estimates to scale negatively with the timeframe of their measurement (Fig. 1.3f). This pattern can be caused by various factors, including the effects of purifying selection and substitution saturation (Ho et al. 2011). On short timeframes, estimates of evolutionary rates can be inflated by the inclusion of deleterious mutations, which tend to be removed from the population by purifying selection over longer periods of time. Substitution saturation can cause underestimation of the amount of genetic change across longer evolutionary timescales, and this bias is exacerbated by model misspecification (Soubrier et al. 2012). The most striking disparities are seen when the short-term rate estimates from pedigrees and mutation-accumulation lines are compared with those inferred using phylogenetic analysis (e.g., Howell et al. 2003). There is evidence of a time-dependent pattern in evolutionary rate estimates from viruses (Duchêne et al., 2014; Aiewsakun and Katzourakis 2016), bacteria (Duchêne et al. 2016b; but see Gibson and Eyre-Walker 2019), and metazoan mitochondrial genomes (Molak and Ho 2015). The evidence for time-dependent biases in metazoan nuclear genomes has so far been limited, although spontaneous mutation rates appear to be greater than long-term evolutionary rates estimated using phylogenetic methods (with modern humans being at least one exception to this pattern; Scally 2016; Chintalapati and Moorjani 2020).

## 1.4 Evolutionary Rates Across the Tree of Life

### 1.4.1 Estimating Rates of Mutation and Evolution

Across the tree of life, evolutionary rates show striking variation and span multiple orders of magnitude. This variation can be considered at a range of biological scales: within individuals, between generations, between populations, among species, and across clades. Lying at one end of this spectrum are rates of spontaneous mutation, which have commonly been estimated

by studying laboratory populations but are increasingly based on genome sequencing of closely related individuals or even of different tissues within the same individual. These rates have typically been difficult to estimate directly, because of the small numbers of mutations between generations and because studies often compare the genomes of somatic rather than germline cells. However, improvements in the efficiency and cost of genome sequencing have led to a stunning increase in studies of spontaneous mutation rates, even in multicellular eukaryotes that experience very few mutations per generation. In Chap. 3, Susanne Pfeifer presents an overview of the major approaches that have been used to estimate spontaneous mutation rates, along with a summary of the estimates that have been published so far. These studies have revealed considerable variation in mutation rates across species (Drake et al. 1998; Baer et al. 2007).

Given that most mutations have negative impacts on fitness, the question arises as to why mutation rates are nonzero (Sturtevant 1937). This can be understood in terms of the fitness costs of reducing mutation rates, because cellular and energetic resources are needed for proofreading and error correction (Kimura 1967). A nonzero mutation rate also provides genetic variation, allowing populations of organisms to adapt to changes in environmental conditions. These factors have led to the idea that mutation rates themselves are evolvable; the optimal mutation rate is expected to vary along the genome and across species (Baer et al. 2007). However, some have argued that mutation rates represent a balance between genetic drift and selection for reduced copying errors (Lynch 2010; Lynch et al. 2016). In Chap. 4, Lindell Bromham describes the current state of knowledge of the causes of rate variation across the tree of life, including the factors that affect rates of spontaneous mutation and the rates of fixation of these mutations (i.e., substitution rates).

In many phylogenetic studies using molecular clock models, evolutionary rates and timescales are jointly estimated. These analyses have produced a comprehensive picture of evolutionary rate variation across the diversity of life. In these cases, evolutionary rates are averaged along branches of the phylogeny, meaning that these estimates represent long-term quantities and are partly dependent on taxon sampling (Lanfear et al. 2010). Furthermore, they are somewhat removed from the underlying rates of spontaneous mutation because they have also been shaped by the effects of selection and drift. Some researchers have attempted to use rates estimated from noncoding or synonymous sites as an approximation of mutation rates. In any case, a more complete understanding of rate variation can be achieved by considering both spontaneous mutation rates and phylogenetic estimates of evolutionary rates.

### 1.4.2 Viruses and Bacteria

The genomes of viruses and bacteria show a remarkable range of mutation rates and evolutionary rates. Among viruses, rates broadly vary with the structure and composition of the genome. Viruses with single-stranded genomes evolve more rapidly than those with double-stranded genomes (Duffy et al. 2008; Sanjuán et al. 2010), although the reasons for this pattern remain unclear (Peck and Lauring 2018). RNA viruses copy their genomes using RNA-dependent RNA polymerases, which lack proofreading ability, so most of these viruses are unable to correct any copying errors that occur during genome replication. As a consequence, they generally have higher mutation rates than DNA viruses, especially double-stranded DNA viruses. There is also a negative correlation between genome size and evolutionary rate (Sanjuán et al. 2010), which is particularly noticeable in viruses but is also seen across a broad range of taxa (Drake 1991; Drake et al. 1998).

The most rapidly evolving viruses tend to be those with single-stranded RNA genomes, such as influenza virus, dengue virus, and coronaviruses. These viruses experience substitution rates as high as $10^{-3}$ substitutions per site per year (Duffy et al. 2008). At the other end of the spectrum, double-stranded DNA viruses, such as

variola virus (which causes smallpox), can evolve at rates below $10^{-5}$ substitutions per site per year (Firth et al. 2010). For rapidly evolving viruses, evolutionary rates can be estimated using time-structured data sets in which genomes have been sampled at different points in time (Rambaut 2000; Drummond et al. 2001). In contrast, slowly evolving viruses, such as hepatitis B virus, might not undergo a sufficient amount of genetic change over such timeframes to permit any reliable inference of their substitution rate. In some of these cases, evolutionary rates can be estimated by assuming that viruses have codiverged with their hosts (e.g., Bernard 1994; Paraskevis et al. 2013). Virus-host codivergence appears to be more common in double-stranded DNA viruses than in RNA viruses (Geoghegan et al. 2017).

Bacteria have larger genomes than viruses and tend to evolve more slowly. Analyses of genomic data sets have revealed a wide variation in evolutionary rates among bacterial taxa (Duchêne et al. 2016b). The most rapidly evolving bacterial species, such as *Neisseria gonorrhoeae*, *Helicobacter pylori*, and *Enterococcus faecium*, experience nucleotide substitution rates of about $10^{-5}$ substitutions per site per year. In contrast, rates below $10^{-7}$ substitutions per site per year are seen in *Mycobacterium tuberculosis* and the plague bacterium *Yersinia pestis* (Duchêne et al. 2016b). The variation in evolutionary rates in bacteria has been ascribed to differences in generation time (Gibson and Eyre-Walker 2019), but attempts to resolve these patterns have been hindered by strong time-dependent biases in rate estimation (Rocha et al. 2006; Duchêne et al. 2016b). Nevertheless, a generation-time effect can be seen in the lower evolutionary rates of spore-forming bacteria compared with bacteria that do not form spores (Weller and Wu 2015).

### 1.4.3 Eukaryotes

Rates of molecular evolution in eukaryotes, particularly multicellular eukaryotes with long generation times, are generally lower than those of viruses and bacteria. Estimates of mutation rates in unicellular eukaryotes include $1.9 \times 10^{-11}$

substitutions per site per generation for the protist *Paramecium tetraurelia* (Sung et al. 2012) and about $2 \times 10^{-10}$ substitutions per site per generation for the yeasts *Saccharomyces cerevisiae* (Zhu et al. 2014) and *Schizosaccharomyces pombe* (Farlow et al. 2015). There have been relatively few estimates of spontaneous mutation rates in the nuclear genomes of animals, but these are growing rapidly with the application of high-throughput sequencing to pedigrees and parent-offspring trios (see Chap. 3).

Animal nuclear genomes evolve slowly, so per-generation mutation rates are difficult to estimate because of the confounding impacts of sequencing error. Inference of mutation rates is also complicated by rate differences between sexes and between the soma and germline. Analyses of genomes from pedigrees and parent-offspring trios have produced a range of estimates of the spontaneous mutation rate in modern humans, centred on a value of $5 \times 10^{-10}$ mutations per site per year (Scally 2016). Spontaneous mutation rates have also been estimated for the nuclear genomes of the nematode worm *Caenorhabditis elegans*, the common fruit fly *Drosophila melanogaster*, Western honey bee *Apis mellifera*, collared flycatcher *Ficedula albicollis*, house mouse *Mus musculus*, and common chimpanzee *Pan troglodytes*, among other animal species (see Chap. 3; Smeds et al. 2016).

An alternative approach to estimating mutation rates has involved analyses of rates of synonymous substitutions and changes at third codon positions, which are under weaker selective constraints and so are believed to provide an approximation of mutation rates. These analyses have revealed that mitochondrial mutation rates vary considerably across birds and mammals (Nabholz et al. 2008, 2009) and invertebrates (Thomas et al. 2010). In contrast, studies of mitochondrial substitution rates in birds and mammals have identified a relative degree of constancy across lineages, with a mean rate of about 0.01 substitutions per site per Myr (Weir and Schluter 2008; but see Pereira and Baker 2006; Nguyen and Ho 2016). This has led to the notion of a 1% mitochondrial clock in birds and mammals. A

similar 'universal' mitochondrial clock has been widely used in studies of invertebrates (Brower 1994; but see Papadopoulou et al. 2010).

Evolutionary rates show considerable heterogeneity across plant lineages, but a few general trends can be observed. The nuclear genomes of gymnosperms evolve at rates that are several times lower, on average, than those of angiosperms (De La Torre et al. 2017). This pattern can potentially be explained by the longer generations and large genomes of gymnosperms. Within flowering plants, there is evidence of a substantial increase in evolutionary rates in the early evolution of the grasses (Christin et al. 2014), whereas palms have evolved much more slowly (Gaut et al. 1992). Evolutionary rates are higher in annual plants than in perennial plants, a pattern that has been found in sequence analyses of the internal transcribed spacer of nuclear ribosomal DNA (Kay et al. 2006) and in larger sets of chloroplast and nuclear genes (Yue et al. 2010). Similarly, herbaceous flowering plants have higher rates of molecular evolution than woody plants with shrub or tree habits (Smith and Donoghue 2008). These patterns in rate variation between annual and perennial plants, and between herbaceous and woody plants, are believed to reflect broad differences in generation time.

Mutation rates in nuclear genomes have been estimated for a number of plant species, including thale cress *Arabidopsis thaliana* (Ossowski et al. 2010), common oak *Quercus robur* (Schmid-Siegert et al. 2017), Sitka spruce *Picea sitchensis* (Hanlon et al. 2019), and yellow box eucalypt *Eucalyptus melliodora* (Orr et al. 2020). Some of these studies were able to trace somatic mutations across the plant, such as along tree branches. For example, 300 mutations were identified along 90.1 metres of branch length in an individual tree of *Eucalyptus melliodora*, allowing the somatic mutation rate to be calculated at $2.75 \times 10^{-9}$ mutations per nucleotide for each metre of tree branch (Orr et al. 2020). A detailed genomic analysis of eight plant species revealed evidence of higher per-year mutation rates in roots than in shoots in perennial plants, but such a pattern was not seen in annual plants

(Wang et al. 2019). In addition, mutation rates were found to be higher in petals than in leaves. These studies have revealed the complexities of mutation rate variation in plants, while highlighting the difficulty in understanding the relationships of these rates to the long-term evolutionary rates in these taxa.

## 1.5 The Molecular Clock as a Tool for Inferring Timescales

### 1.5.1 Molecular Dating

In modern genetics and genomics, the molecular clock has its most prominent role as a tool for inferring evolutionary timescales. This application of the molecular clock is sometimes referred to as molecular clock dating, divergence-time estimation, or simply molecular dating. There is a rich history of development of molecular dating methods (Fig. 1.1), with much of the progress in this field being tied to advances in phylogenetic methods and computational power (Bromham and Penny 2003; Kumar 2005). In Chap. 5, Susana Magallón describes the principles behind molecular dating methods and the steps involved in using these methods to infer evolutionary timescales from molecular sequence data.

Research on molecular dating has led to the development of a range of phylogenetic dating methods and statistical models of evolutionary rates (Heath and Moore 2014; Ho and Duchêne 2014; Yang 2014; Kumar and Hedges 2016). These have included methods to cope with among-lineage rate variation, such as nonparametric rate smoothing (Sanderson 1997) and penalized likelihood (Sanderson 2002), as well as models of evolutionary rate variation across branches (Hasegawa et al. 1989; Thorne et al. 1998). Notably, much of the recent progress in molecular clocks has focused on phenomenological rather than mechanistic models, leaving these developments somewhat decoupled from the earlier theoretical context of the molecular clock.

Molecular dating was first performed using amino acid sequences (Zuckerkandl and Pauling 1962) and immunological comparisons by

microcomplement fixation (Sarich and Wilson 1967a), but is now overwhelmingly based on the analysis of nucleotide sequences. The most important developments have been in the use of genome-scale data sets for inferring evolutionary timescales. Alongside these efforts, there have been various attempts to use other forms of genetic, genomic, and protein data for molecular dating (Fig. 1.1; Ho et al. 2016). For example, the timing of intraspecific events has been estimated using molecular clocks based on microsatellites (Goldstein et al. 1995), whereas deeper events have been dated using protein folds (Wang et al. 2011).

The application of Bayesian approaches to phylogenetic analysis has led to major developments in molecular dating (dos Reis et al. 2016; Bromham et al. 2018). In Chap. 6, Tianqi Zhu provides an introduction to the Bayesian framework for molecular dating, which permits the application of complex, parameter-rich models that would not be tractable using other methods. These include sophisticated models of evolutionary rate heterogeneity (clock models), models of lineage diversification (in the form of the tree prior), and various means of incorporating data from the fossil record (dos Reis et al. 2016; Bromham et al. 2018).

In Bayesian molecular dating, models of among-lineage rate variation have seen particularly active development. The most widely used are the relaxed-clock models, which allow a distinct rate of evolution along each branch of the phylogenetic tree. The earliest relaxed-clock models were inspired by the work of Gillespie (1991), who suggested that the substitution rate might evolve along lineages. Relaxed-clock models that allow such autocorrelation in the evolutionary rate were implemented in Bayesian dating methods in the late 1990s and subsequently expanded (e.g., Thorne et al. 1998; Kishino et al. 2001; Aris-Brosou and Yang 2002). Later work saw the appearance of relaxed-clock models that allow independent or uncorrelated rates across branches (e.g., Drummond et al. 2006; Rannala and Yang 2007).

The methods developed for molecular dating have also been applied, with some modifications, to analyses of morphological data. In Chap. 7,

Michael Lee describes the use of phenotypic traits for estimating evolutionary timescales, focusing on the analysis of discrete morphological characters. The use of morphological clocks has produced useful insights into the evolution of birds and other groups of organisms (e.g., Polly 2001; Lee et al. 2014), although there continue to be various shortcomings that need to be addressed (Puttick et al. 2016). For example, questions persist about the strength of the association between molecular and morphological rates of evolution (Davies and Savolainen 2006; Seligmann 2010). Nevertheless, with continued advances in models of phenotypic evolution (e.g., Álvarez-Carretero et al. 2019), phylogenetic dating analyses of morphological characters present a promising avenue for further research.

Unless there is a priori information about the evolutionary rate, molecular dating methods need to calibrate the clock so that it gives date estimates measured in absolute time. The most widely used types of calibrating information are those based on palaeontological, geological, and biogeographic evidence. In Chap. 8, Jacqueline Nguyen and I describe the use of fossil evidence for calibration, which has a rich history of development and has fostered productive collaborations between geneticists and palaeontologists. In Chap. 9, Michael Landis explains how information from biogeography and palaeogeography can be used to calibrate the molecular clock, based on the timing of geological events such as the separation of landmasses.

Some phylogenetic methods have been extended to account for the inclusion of genomes and morphological data that have been sampled at distinct points in time. In Chap. 10, Sebastián Duchêne and David Duchêne describe the use of sampling times for calibration in analyses of rapidly evolving viruses and bacteria, and when analysing data sets containing ancient DNA sequences. Distinct sampling times are also a feature of morphological data sets that include fossil taxa. In Chap. 11, Alexandra Gavryushkina and Chi Zhang describe the analysis of combined morphological and molecular data, including the development of diversification models that explicitly include extinct species and fossil

sampling (e.g., Ronquist et al. 2012; Heath et al. 2014).

The past two decades have seen remarkable growth in genomic data, which has been made possible by the development of high-throughput sequencing methods. This has provided a vast wealth of molecular sequence data for understanding molecular evolution at the genomic scale, but has also brought substantial challenges to molecular dating (Ho 2014; Tong et al. 2016). In Chap. 12, Qiqing Tao, Koichiro Tamura, and Sudhir Kumar review a range of methods that are designed to perform rapid molecular dating, allowing the analysis of data sets containing large numbers of sequences. In Chap. 13, Sandra Álvarez-Carretero and Mario dos Reis describe the application of Bayesian phylogenetic dating to genome-scale data sets, including some of the techniques that have been used to improve computational feasibility. These two closing chapters present a promising picture of how the molecular clock will retain its relevance and utility in the coming years.

### 1.5.2 Evolutionary Timescales

The molecular clock has been used extensively to reconstruct evolutionary timescales across the tree of life. Early studies focused on the divergence times of humans and related primates (Zuckerkandl and Pauling 1962; Sarich and Wilson 1967a), but often included other mammals (Margoliash 1963; Doolittle and Blombäck 1964). There continued to be a focus on the evolutionary rates and timescales of mammals, particularly eutherian mammals, primarily because of the availability of molecular data for this group of organisms. Developments in automated DNA sequencing in the late 1980s and early 1990s led to rapid growth in molecular sequence data, allowing a considerable expansion of the scope of molecular dating studies.

Molecular dating gained widespread attention in the 1990s when researchers began analysing large data sets to reconstruct the timescales of major evolutionary events. These studies often involved spectacular claims about the antiquity of major branches of the tree of life. These questions have held perennial interest, including the timing of the divergences among the kingdoms of life (e.g., Doolittle et al. 1996), the divergences among metazoan phyla (the 'Cambrian explosion'; e.g., Wray et al. 1996; dos Reis et al. 2015), the diversification of angiosperms (e.g., Martin et al. 1989; Magallón et al. 2015), and the radiations of eutherian mammals and modern birds (e.g., Hedges et al. 1996; Easteal 1999; Springer et al. 2003; dos Reis et al. 2013). The molecular date estimates for these events have often been at odds with the timescales supported by a literal reading of the palaeontological evidence, leading to deliberation about the relative merits of the fossil record and molecular clocks (Smith and Peterson 2002; Benton and Ayala 2003; Brochu et al. 2004). For example, many molecular estimates for the age of crown angiosperms have been greater than 200 Myr, whereas the oldest fossil evidence dates to about 136 Myr in the Early Cretaceous (Magallón et al. 2015). The debates over the discrepancies between molecular and fossil evidence identified some important shortcomings in molecular dating methods, which provided a strong impetus for methodological innovation. Improved modelling of evolutionary rate variation and use of fossil evidence has narrowed some of the gaps between molecular and palaeontological date estimates.

Molecular dating has been particularly valuable for understanding the evolutionary history and epidemiological dynamics of pathogens (Pybus and Rambaut 2009). Fine-scale sampling of pathogens, for example during contemporary virus outbreaks, can allow a detailed reconstruction of evolutionary rates, transmission dynamics, and phylogeographic spread (Pybus and Rambaut 2009). Over longer evolutionary timescales, molecular clocks can be used to determine when pathogens crossed species barriers and infected new hosts, and whether these pathogens continued to codiverge with the host populations.

One of the more surprising applications of molecular dating has been to estimate the ages of the biological samples from which genomic data have been obtained (Shapiro et al. 2011;

Moorjani et al. 2016). This approach can be used to estimate or validate the ages of any samples that have uncertain or contentious dates, such as those that are beyond the 50,000-year reach of radiocarbon dating or where the cost of direct radiometric dating is prohibitive. For example, a Bayesian dating analysis was used to estimate the age of a 400,000-year-old hominin sample from Sima de los Huesos in Spain (Meyer et al. 2014). Ancient hominin genomes have also been dated using a molecular clock based on the accumulation of recombination events over time (Moorjani et al. 2016).

Continued development of molecular clocks will allow evolutionary and demographic timescales to be resolved with increasing confidence. Some of the most promising areas of research include better techniques for incorporating fossil data, mechanistic models of evolutionary rate variation among lineages, and molecular dating methods that are able to process genome-scale data sets from large numbers of taxa. At the same time, these efforts will be substantially aided by advances in understanding of genomic evolution and other biological processes.

## 1.6 Concluding Remarks

This book is intended to provide an overview of the state of the art of molecular clocks, although the continual and rapid expansion of the field prevents a comprehensive treatment from being achievable. Nevertheless, I hope that this book provides a useful starting point for researchers and students interested in molecular evolutionary clocks. The field is likely to carry on developing at a great pace in response to the growth of genomic data. With international efforts to sequence the genomes of all vertebrates, invertebrates, and other eukaryotes, we will continue to make great strides towards placing a timescale on the tree of life.

## References

Aiewsakun P, Katzourakis A (2016) Time-dependent rate phenomenon in viruses. J Virol 90:7184–7195

Allio R, Donega S, Galtier N, Nabholz B (2017) Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. Mol Biol Evol 34:2762–2772

Álvarez-Carretero S, Goswami A, Yang Z, dos Reis M (2019) Bayesian estimation of species divergence times using correlated quantitative characters. Syst Biol 68:967–986

Aris-Brosou S, Yang Z (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst Biol 51:703–714

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489–522

Ayala FJ (1997) Vagaries of the molecular clock. Proc Natl Acad Sci USA 94:7776–7783

Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet 8:619–631

Benton MJ, Ayala FJ (2003) Dating the tree of life. Science 300:1698–1700

Bernard H-U (1994) Coevolution of papillomaviruses with human populations. Trends Microbiol 2:140–143

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Brochu CA, Sumrall CD, Theodor JM (2004) When clocks (and communities) collide: Estimating divergence times from molecules and the fossil record. J Paleontol 78:1–6

Bromham L, Penny D (2003) The modern molecular clock. Nat Rev Genet 4:216–224

Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW (2018) Bayesian molecular dating: opening up the black box. Biol Rev 93:1165–1191

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc Natl Acad Sci USA 91:6491–6495

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967–1971

Chintalapati M, Moorjani P (2020) Evolution of the mutation rate across primates. Curr Opin Genet Dev 62:58–64

Christensen AC (2013) Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol 5:1079–1086

Christin P-A, Spriggs E, Osborne CP, Strömberg CAE, Salamin N, Edwards EJ (2014) Molecular dating, evolutionary rates, and the age of the grasses. Syst Biol 63:153–165

Cutler D (2000) Understanding the overdispersed molecular clock. Genetics 154:1403–1417

Davies TJ, Savolainen V (2006) Neutral theory, phylogenies, and the relationship between phenotypic change and evolutionary rates. Evolution 60:476–483

Dayhoff MO (1978) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington, DC

De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK (2017) Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol Biol Evol 34:1363–1377

Dickerson RE (1971) The structure of cytochrome *c* and the rates of molecular evolution. J Mol Evol 1:26–45

Doolittle RF, Blombäck B (1964) Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. Nature 202:147–152

Doolittle RF, Feng D-F, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

dos Reis M, Donoghue PCJ, Yang Z (2013) Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. Biol Lett 10:20131003

dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PCJ, Yang Z (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr Biol 25:2939–2950

dos Reis M, Donoghue PCJ, Yang Z (2016) Bayesian molecular clock dating of species divergences in the genomics era. Nat Rev Genet 17:71–80

Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 88:7160–7164

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686

Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol 49:827–831

Drummond AJ, Forsberg R, Rodrigo AG (2001) The inference of stepwise changes in substitution rates using serial sequence samples. Mol Biol Evol 18:1365–1371

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLOS Biol 4:e88

Duchêne S, Ho SYW (2015) Mammalian genome evolution is governed by multiple pacemakers. Bioinformatics 31:2061–2065

Duchêne S, Holmes EC, Ho SYW (2014) Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc R Soc B 281:20140732

Duchêne S, Foster CSP, Ho SYW (2016a) Estimating the number and assignment of clock models in analyses of multigene datasets. Bioinformatics 32:1281–1285

Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC (2016b) Genome-scale rates of evolutionary change in bacteria. Microb Genom 2:e000094

Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9:267–276

Easteal S (1999) Molecular evidence for the early divergence of placental mammals. BioEssays 21:1052–1058

Eldredge N, Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM (ed) Models in paleobiology. Freeman, San Francisco, CA, pp 82–115

Farlow A, Long H, Arnoux S, Sung W, Doak TG, Nordborg M, Lynch M (2015) The spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*. Genetics 201:737–744

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A (2010) Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol 27:2038–2051

Fisher RA (1936) The measurement of selective intensity. Proc R Soc B 121:58–62

Fitch WM (1976) Molecular evolutionary clocks. In: Ayala FJ (ed) Molecular evolution. Sinaeuer Associates, Sunderland, MA, pp 160–178

Fitch WM, Beintema JJ (1990) Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense branching as exemplified by ribonuclease. Mol Biol Evol 7:438–443

Freese E (1962) On the evolution of base composition at DNA. J Theor Biol 3:82–101

Gaut B, Muse SV, Clark WD, Clegg MT (1992) Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J Mol Evol 35:292–303

Gaut B, Yang L, Takuno S, Eguiarte LE (2011) The patterns and causes of variation in plant nucleotide substitution rates. Annu Rev Ecol Evol Syst 42:245–266

Geoghegan JL, Duchêne S, Holmes EC (2017) Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLOS Pathog 13:e1006215

Gibson B, Eyre-Walker A (2019) Investigating evolutionary rate variation in bacteria. J Mol Evol 87:317–326

Gillespie JH (1984) The molecular clock may be an episodic clock. Proc Natl Acad Sci USA 81:8009–8013

Gillespie JH (1989) Lineage effects and the index of dispersion of molecular evolution. Mol Biol Evol 6:636–647

Gillespie JH (1991) The causes of molecular evolution. Oxford University Press, Oxford, UK

Gillespie JH (1993) Substitution processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. Genetics 134:971–981

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based

on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Goodman M (1961) The role of immunologic differences in the phyletic development of human behavior. Hum Biol 33:131–162

Haldane JBS (1957) The cost of natural selection. J Genet 55:511–524

Hanlon VCT, Otto SP, Aitken SN (2019) Somatic mutations substantially increase the per-generation mutation rate in the conifer *Picea sitchensis*. Evol Lett 3:348–358

Hasegawa M, Kishino H, Yano T (1989) Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea. J Hum Evol 18:461–476

Heath TA, Moore BR (2014) Bayesian inference of species divergence times. In: Chen M-H, Kuo L, Lewis PO (eds) Bayesian phylogenetics: methods, algorithms, and applications. CRC Press, Boca Raton, FL, pp 277–318

Heath TA, Huelsenbeck JP, Stadler T (2014) The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc Natl Acad Sci USA 111:E2957–E2966

Hedges SB, Parker PH, Sibley CG, Kumar S (1996) Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226–229

Ho SYW (2014) The changing face of the molecular evolutionary clock. Trends Ecol Evol 29:496–503

Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol 23:5947–5965

Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A (2011) Time-dependent rates of molecular evolution. Mol Ecol 20:3087–3101

Ho SYW, Chen AZY, Lins LSF, Duchêne DA, Lo N (2016) The genome as an evolutionary timepiece. Genome Biol Evol 8:3006–3010

Hodgkinson A, Eyre-Walker A (2011) Variation in the mutation rate across mammalian genomes. Nat Rev Genet 12:756–766

Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet 72:659–670

Jukes TH, Kimura M (1984) Evolutionary constraints and the neutral theory. J Mol Evol 21:90–92

Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol Biol 6:36

Kern AD, Hahn MW (2018) The neutral theory in light of natural selection. Mol Biol Evol 35:1366–1371

Kikuno R, Hayashida H, Miyata T (1985) Rapid rate of rodent evolution. Proc Japan Acad 61:153–156

Kim S-H, Elango N, Warden C, Vigoda E, Yi SV (2006) Heterogeneous genomic molecular clocks in primates. PLOS Genet 2:e163

Kimura M (1967) On the evolutionary adjustment of spontaneous mutation rates. Genet Res 9:23–34

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

Kimura M (1969) The rate of molecular evolution considered from the standpoint of population genetics. Proc Natl Acad Sci USA 63:1181–1188

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kimura M (1987) Molecular evolutionary clock and the neutral theory. J Mol Evol 26:24–33

King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164:788–798

Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol 18:352–361

Kohne DE (1970) Evolution of higher-organism DNA. Q Rev Biophys 3:327–375

Kreitman M, Akashi H (1995) Molecular evidence for natural selection. Annu Rev Ecol Syst 26:403–422

Kumar S (2005) Molecular clocks: four decades of evolution. Nat Rev Genet 6:654–662

Kumar S, Hedges SB (2016) Advances in time estimation methods for molecular data. Mol Biol Evol 33:863–869

Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. Proc Natl Acad Sci USA 99:803–808

Laird CD, McConaughy BL, McCarthy BJ (1969) Rate of fixation of nucleotide substitutions. Nature 224:149–154

Lanfear R, Welch JJ, Bromham L (2010) Watching the clock: studying variation in rates of molecular evolution between species. Trends Ecol Evol 25:495–503

Lanfear R, Ho SYW, Davies TJ, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP (2013) Taller plants have lower rates of molecular evolution. Nat Commun 4:1879

Langley CH, Fitch WM (1974) An examination of the constancy of the rate of molecular evolution. J Mol Evol 3:161–177

Lee MSY, Ho SYW (2016) Molecular clocks. Curr Biol 26:R387–R407

Lee MSY, Soubrier J, Edgecombe GD (2013) Rates of phenotypic and genomic evolution during the Cambrian explosion. Curr Biol 23:1–7

Lee MSY, Cau A, Naish D, Dyke GJ (2014) Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. Syst Biol 63:442–449

Lynch M (2010) Evolution of the mutation rate. Trends Genet 26:345–352

Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL (2016) Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet 17:704–714

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol 207:437–453

Manceau M, Marin J, Morlon H, Lambert A (2020) Model-based inference of punctuated molecular evolution. Mol Biol Evol 37:3308–3323

Margoliash E (1963) Primary structure and evolution of cytochrome *c*. Proc Natl Acad Sci USA 50:672–679

Martin W, Gierl A, Saedler H (1989) Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339:46–48

Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge, MA

Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga J-L, Martínez I, Gracia A, Bermúdez de Castro JM, Carbonell E, Pääbo S (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature 505:403–406

Miyata T, Hayashida H, Kikuno R, Hasegawa M, Kobayashi M, Koike K (1982) Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J Mol Evol 19:28–35

Molak M, Ho SYW (2015) Prolonged decay of molecular rate estimates for metazoan mitochondrial DNA. PeerJ 3:e821

Mooers AØ, Harvey PH (1994) Metabolic rate, generation time, and the rate of molecular evolution in birds. Mol Phylogenet Evol 3:344–350

Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, Reich D (2016) A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc Natl Acad Sci USA 113:5652–5657

Morgan GJ (1998) Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. J Hist Biol 31:155–178

Muse SV, Gaut BS (1997) Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. Genetics 146:393–399

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. Mol Biol Evol 25:120–130

Nabholz B, Glémin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evol Biol 9:54

Nei M, Suzuki Y, Nozawa M (2010) The neutral theory of molecular evolution in the genomic era. Annu Rev Genomics Hum Genet 11:265–289

Nguyen JMT, Ho SYW (2016) Mitochondrial rate variation among lineages of passerine birds. J Avian Biol 47:690–696

Ohta T (1972) Evolutionary rate of cistrons and DNA divergence. J Mol Evol 1:150–157

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98

Ohta T, Gillespie JH (1996) Development of neutral and nearly neutral theories. Theor Pop Biol 49:128–142

Ohta T, Kimura M (1971) On the constancy of the evolutionary rate of cistrons. J Mol Evol 1:18–25

Orr AJ, Padovan A, Kainer D, Külheim C, Bromham L, Bustos-Segura C, Foley W, Haff T, Hsieh J-F, Morales-Suarez A, Cartwright RA, Lanfear R (2020) A phylogenomic approach reveals a low somatic mutation rate in a long-lived plant. Proc R Soc B 287:20192364

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327:92–94

Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119–121

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Mol Biol Evol 27:1659–1672

Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SYW, Belshaw R, Allain J-P, Hatzakis A (2013) Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57:908–916

Park C, Qian W, Zhang J (2012) Genomic evidence for elevated mutation rates in highly expressed genes. EMBO Rep 13:1123–1129

Peck KM, Lauring AM (2018) Complexities of viral mutation rates. J Virol 92:e01031–e01017

Pereira SL, Baker AJ (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol 23:1731–1740

Polly PD (2001) On morphological clocks and paleophylogeography: towards a timescale for *Sorex* hybrid zones. Genetica 112–113:339–357

Puttick MN, Thomas GH, Benton MJ (2016) Dating placentalia: morphological clocks fail to close the molecular fossil gap. Evolution 70:873–886

Pybus OG, Rambaut A (2009) Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet 10:540–550

Rambaut A (2000) Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics 16:395–399

Rannala B, Yang Z (2007) Inferring speciation times under an episodic molecular clock. Syst Biol 56:453–466

Rocha EPC, Danchin A (2004) An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol Biol Evol 21:108–116

Rocha EPC, Maynard Smith J, Hurst LD, Holden MTG, Cooper JE, Smith NH, Feil EJ (2006) Comparisons of d*N*/d*S* are time dependent for closely related bacterial genomes. J Theor Biol 239:226–235

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP (2012) A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst Biol 61:973–999

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14:1218–1231

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101–109

Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. J Virol 84:9733–9748

Sarich VM, Wilson AC (1967a) Immunological time scale for hominid evolution. Science 158:1200–1203

Sarich VM, Wilson AC (1967b) Rates of albumin evolution in primates. Proc Natl Acad Sci USA 58:142–148

Sarich VM, Wilson AC (1973) Generation time and genomic evolution in primates. Science 179:1144–1147

Scally A (2016) The mutation rate in human evolution and demographic inference. Curr Opin Genet Dev 41:36–43

Schmid-Siegert E, Sarkar N, Iseli C, Calderon S, Gouhier-Darimont C, Chrast J, Cattaneo P, Schütz F, Farinelli L, Pagni M, Schneider M, Voumard J, Jaboyedoff M, Fankhauser C, Hardtke CS, Keller L, Pannell JR, Reymond A, Robinson-Rechavi M, Xenarios I, Reymond P (2017) Low number of fixed somatic mutations in a long-lived oak tree. Nat Plants 3:926–929

Seligmann H (2010) Positive correlations between molecular and morphological rates of evolution. J Theor Biol 264:799–807

Shapiro B, Rambaut A, Drummond AJ (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol 23:7–9

Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A (2011) A Bayesian phylogenetic method to estimate unknown sequence ages. Mol Biol Evol 28:879–887

Smeds L, Qvarnström A, Ellegren H (2016) Direct estimate of the rate of germline mutation in a bird. Genome Res 26:1211–1218

Smith DR (2015) Mutation rates in plastid genomes: they are lower than you might think. Genome Biol Evol 7:1227–1234

Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89

Smith NGC, Eyre-Walker A (2003) Partitioning the variation in mammalian substitution rates. Mol Biol Evol 20:10–17

Smith AB, Peterson KJ (2002) Dating the time of origin of major clades: molecular clocks and the fossil record. Annu Rev Earth Planet Sci 30:65–88

Snir S, Wolf YI, Koonin EV (2012) Universal pacemaker of genome evolution. PLOS Comput Biol 8:e1002785

Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, Cooper A (2012) The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol Biol Evol 29:3345–3358

Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci USA 100:1056–1061

Stebbins GL, Lewontin RC (1972) Comparative evolution at the levels of molecules, organisms and populations. In: Le Cam LM, Neyman J, Scott EL (eds) Proceedings of the sixth Berkeley symposium on mathematical statistics and probability. Volume V: Darwinian, neo-Darwinian, and non-Darwinian evolution. University of California Press, Berkeley, CA

Sturtevant AH (1937) Essays on evolution. I. On the effects of selection on mutation rate. Q Rev Biol 12:467–477

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M (2012) Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. Proc Natl Acad Sci USA 109:19339–19344

Takahata N (1987) On the overdispersed molecular clock. Genetics 116:169–179

Takahata N (2007) Molecular clock: an *anti*-neo-Darwinian legacy. Genetics 176:1–6

Thomas JA, Welch JJ, Lanfear R, Bromham L (2010) A generation time effect on the rate of molecular evolution in invertebrates. Mol Biol Evol 27:1173–1180

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15:1647–1657

Tian D, Wang Q, Zhang P, Araki H, Yang S, Kreitman M, Nagylaki T, Hudson R, Bergelson J, Chen J-Q (2008) Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature 455:105–108

Tong KJ, Lo N, Ho SYW (2016) Reconstructing evolutionary timescales using phylogenomics. Zool Syst 41:343–351

Wang Z, Zhang J (2009) Why is the correlation between gene importance and gene evolutionary rate so weak? PLOS Genet 5:e1000329

Wang M, Jiang Y-Y, Kim KM, Qu G, Jo H-F, Mittenthal JE, Zhang H-Y, Caetano-Anollés G (2011) A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. Mol Biol Evol 28:567–582

Wang L, Ji Y, Hu Y, Hu H, Jia X, Jiang M, Zhang X, Zhao L, Zhang Y, Jia Y, Qin C, Yu L, Huang J, Yang S, Hurst LD, Tian D (2019) The architecture of intra-organism mutation rate variation in plants. PLOS Biol 17:e3000191

Webster AJ, Payne RJH, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. Science 301:478

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–2328

Weller C, Wu M (2015) A generation-time effect on the rate of molecular evolution in bacteria. Evolution 69:643–652

Wilson AC, Sarich VM (1969) A molecular time scale for human evolution. Proc Natl Acad Sci USA 63:1088–1093

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Annu Rev Biochem 46:573–639

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Wray GA, Levinton JS, Shapiro LH (1996) Molecular evidence for deep Precambrian divergences among metazoan phyla. Science 274:568–573

Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitutions in rodents than in man. Proc Natl Acad Sci USA 82:1741–1745

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11:367–372

Yang Z (2014) Molecular evolution: a statistical approach. Oxford University Press, Oxford, UK

Yue J-X, Li J, Wang D, Araki H, Tian D, Yang S (2010) Genome-wide investigation reveals high evolutionary rates in annual model plants. BMC Plant Biol 10:242

Zhang J, Yang J-R (2015) Determinants of the rate of protein sequence evolution. Nat Rev Genet 16:409–420

Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci USA 111:E2310–E2318

Zuckerkandl E (1978) Multilocus enzymes, gene regulation, and genetic sufficiency. J Mol Evol 12:57–89

Zuckerkandl E, Pauling L (1962) Molecular disease, evolution, and genic heterogeneity. In: Kasha M, Pullman B (eds) Horizons in biochemistry. Academic, New York, pp 189–225

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic, New York, pp 97–166