Reconstructing evolutionary timescales using phylogenomics

K. Jun Tong*, Nathan Lo, and Simon Y. W. Ho

School of Life and Environmental Sciences, University of Sydney, Sydney NSW 2006, Australia

**Corresponding author, E-mail: jun.tong@sydney.edu.au*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2403v1 | CC BY 4.0 Open Access | rec: 31 Aug 2016, publ: 31 Aug 2016

Review

Reconstructing evolutionary timescales using phylogenomics

K. Jun Tong*, Nathan Lo, and Simon Y. W. Ho

School of Life and Environmental Sciences, University of Sydney, Sydney NSW 2006, Australia
**Corresponding author, E-mail: jun.tong@sydney.edu.au*

Abstract Reconstructing the timescale of the Tree of Life is one of the principal aims of evolutionary biology. This has been greatly aided by the development of the molecular clock, which enables evolutionary timescales to be estimated from genetic data. In recent years, high-throughput sequencing technology has led to an increase in the feasibility and availability of genome-scale data sets. These represent a rich source of biological information, but they also bring a set of analytical challenges. In this review, we provide an overview of phylogenomic dating and describe the challenges associated with analysing genome-scale data. We also report on recent phylogenomic estimates of the evolutionary timescales of mammals, birds, and insects.

Key words molecular clock, phylogenetic analysis, genomes, rate variation, placental mammals, birds, insects.

1 Introduction

The molecular clock is a useful tool that enables evolutionary timescales to be estimated using nucleotide sequences, amino acid sequences, and other products of the evolutionary process. Each 'tick' of the molecular clock represents a measurable unit of genetic change, such as a nucleotide or amino acid substitution (Zuckerkandl & Pauling, 1962). Even though the ticks occur stochastically rather than regularly, the outcome is that genetic change accumulates as a function of time (Zuckerkandl & Pauling, 1965). When the tick rate of the molecular clock has been relatively constant, the genetic distance between species is proportional to the time since their evolutionary divergence. The use of molecular clocks in biological research has provided valuable insights into the evolutionary timescales of animals and other organisms (Hedges *et al*., 2006).

Advances in sequencing technology have led to a proliferation of nucleotide sequence data, including the sequences of entire metazoan genomes. This has provided a wealth of information for phylogenetic studies using molecular clocks. Improvements in computational power have made it possible to perform phylogenomic analyses of these data sets, and the first generation of genome-scale dating studies have been published in recent years. These include phylogenomic estimates of the evolutionary timescales of birds (Jarvis *et al*., 2014; Mitchell *et al*., 2015; Prum *et al*., 2015), mammals (dos Reis *et al*., 2012), and insects (Misof *et al*., 2014; Tong *et al*., 2015). The date estimates produced by these studies have confirmed some of the previous views about metazoan evolutionary history, but they have also offered fresh insights and provided a scaffold for detailed studies of the taxa within these groups. However, the flood of genomic data has brought a new suite of analytical challenges. This has inspired some major recent developments in clock models and molecular dating methods (Ho, 2014; Kumar & Hedges, 2016).

In this review article, we describe the insights that have been gained from phylogenomic dating studies of animals. We describe recent developments in molecular clocks, including methods for dealing with evolutionary rate variation. We provide an overview of the computational and analytical challenges associated with analysing genome-scale data. Finally, we summarise the strategies that have been used in recent studies to handle large data sets.

2 Nucleotide sequences and clock calibrations

The first studies in molecular dating were based on comparisons of biochemical data and proteins (Sarich & Wilson, 1967; Wilson & Sarich, 1969; Brown *et al*., 1972). In the 1980s, however, the development of the polymerase chain reaction and Sanger sequencing allowed nucleotide sequences to be determined efficiently (Sanger *et al*., 1977a; Sanger *et al*., 1977b; Mullis & Faloona, 1987). By opening up a new source of information-rich data, these technologies greatly increased the power of molecular phylogenetics. Recent advances in

sequencing technology, often referred to using the umbrella term 'high-throughput sequencing' (HTS), mean that large-scale sequencing is now a far less labour-intensive exercise than it once was. HTS methods are able to sequence large segments of the genome quickly and with ever-decreasing cost (McCormack *et al*., 2013). Since the beginning of this millennium, the data sets used for phylogenetic dating analyses have grown from sequence alignments of single genes, to multiple genes, and now to hundreds or thousands of genes.

By incorporating a molecular clock into phylogenomic analysis, we can estimate timescales of evolutionary diversification. However, genetic data can only offer an estimate of the relative timing of evolutionary events. To obtain absolute date estimates, the molecular clock needs to be calibrated using a source of independent temporal information. Calibrations are usually applied in the form of an age constraint on at least one node in the tree. Such calibrations can come from the fossil record, whereby the age of a clade in the tree is constrained to be older than any fossils that are assigned to that clade (Benton & Donoghue, 2007). Less commonly, calibrations can be based on geological events that have had impacts on the evolutionary process, such as the formation or disappearance of islands, land bridges, riverine connections, and mountain ranges (Ho *et al*., 2015). Time calibrations are usually applied to internal nodes in the tree, but they can also be applied to the tips of the tree when the sequence data have been sampled from ancient specimens (Rambaut, 2000).

When Bayesian phylogenetic methods are used to estimate evolutionary timescales, calibrations are incorporated as prior distributions on the ages of nodes in the tree (Drummond *et al*., 2006). Each of these prior distributions reflects the uncertainty associated with the assignment of the calibration to the node, as well as the uncertainty in the age of the calibration itself (Ho & Phillips, 2009). Choosing a prior distribution that appropriately represents the relevant palaeontological or biogeographical information is a difficult exercise. Errors in the calibrations, including misrepresentation of their uncertainty, can lead to highly unreliable estimates of evolutionary timescales (Warnock *et al*., 2015). For this reason, a number of authors have proposed criteria for evaluating the quality of potential fossil calibrations and their impact on the resulting date estimates (Parham *et al*., 2012; Sauquet *et al*., 2012).

3 Evolutionary rate variation

Since the idea of a molecular clock was proposed more than half a century ago (Zuckerkandl & Pauling, 1962), there has been widespread evidence of evolutionary rate variation (Bromham, 2011). Genetic change can occur somewhat erratically, with different evolutionary rates across genes, species, and timescales (Lee & Ho, 2016). Therefore, to use the molecular clock effectively, these different forms of rate variation need to be taken into account. In the simplest model, often referred to as the strict clock, the rate of evolution is

assumed to be homogeneous throughout the tree (but not necessarily across different genes). The assumption of a constant rate throughout the tree is often violated, especially in genome-scale data sets, except when sequences have been samples from very closely related lineages (Brown & Yang, 2011).

Identifying the different forms and components of evolutionary rate variation is important because it allows us to incorporate them into the models used in phylogenetic analysis. Rate variation can be caused by gene effects, lineage effects, and gene-by-lineage effects (Fig. 1; Gaut *et al*., 2011). Gene effects cause rates to differ between genomic markers. These differences are largely due to the varying degree of selective constraint between regions of the genome. For example, slowly evolving genes probably have very important biological functions, such that many mutations within these genes are likely to be harmful to the organism. At a finer scale, evolutionary rates can vary across individual nucleotide sites. For example, nucleotides at third codon positions tend to have lower selective constraints, such that they evolve more quickly than the nucleotides at the first two codon positions. Among-site rate variation is commonly taken into account by assuming that the site rates follow a gamma distribution (Yang, 1993).

Some species evolve more quickly than others, leading to rate variation across lineages. The causes of these lineage effects include differences in life-history traits, such as generation length (Bromham, 2009). Organisms that have short generations generally have a higher rate of evolution because their genomes tend to be copied more frequently than those of organisms with long generations. Lineage effects can also be caused by differences in population size, metabolic rate, exposure to UV radiation, and the fidelity of DNA repair mechanisms. Rate variation across lineages can be taken into account using relaxed molecular clocks, which were first developed in the late 1990s (Sanderson, 1997; Thorne *et al*., 1998). These clock models allow a different evolutionary rate along each branch of the phylogeny (for a recent review, see Ho & Duchêne, 2014).

Gene effects and lineage effects can interact to produce complex patterns of rate variation, also known as residual effects (Gillespie, 1991). When there are residual effects, evolutionary rates vary across lineages but not in a consistent pattern across genes. As a result, the phylogenetic trees for different genes will have different sets of branch lengths (Muse & Gaut, 1997). In relatively small multi-locus datasets, residual effects can be taken into account by assigning separate relaxed-clock models to different loci. However, applying these principles to genome-scale datasets is likely to lead to substantial over-parameterisation. A more efficient approach is to focus on groups of genes that share similar patterns of among-lineage rate variation and to assign a separate relaxed-clock model to each of these groups (Duchêne *et al*., 2013). This can be done using rapid clustering methods, and can lead to improved estimates of evolutionary timescales (Duchêne & Ho, 2014).

4 New approaches for analysing genome-scale data

Many molecular-clock methods employ parameter-rich models of the evolutionary process. Owing to their large computational requirements, these methods cannot be readily applied to genome-scale data sets. Instead, there are two broad approaches that can be used to analyse large data sets using molecular clocks. The first of these is to use a data-filtering approach, whereby the analysis is carried out on a chosen subset of the data. For example, researchers might select the most informative genes or the genes that exhibit the smallest degree of rate variation across lineages. Data filtering aims to reduce the data set to a manageable size while preserving a useful part of the signal from the original data set. This allows complex and parameter-rich methods, such as Bayesian relaxed clocks, to be applied to the filtered data.

A second way of performing phylogenomic dating is to use rapid molecular-clock methods. Large increases in computational speed can be achieved by using approximate likelihood functions in Bayesian methods (Thorne *et al*., 1998; dos Reis & Yang, 2011). Alternatively, faster maximum-likelihood or least-squares methods can be used (Kumar & Hedges, 2016). For instance, the recently developed program RelTime first estimates branch lengths using maximum likelihood, then infers the age of each node using smoothing and averaging techniques to account for rate variation (Tamura *et al*., 2012). In this way, the method avoids relying on an explicit model of rate variation. RelTime produces a chronogram with relative node ages, but these can be scaled to absolute time by applying calibrations to the tree. A similar method that relies on least squares has been developed for time-structured sequence data, such as those obtained from ancient samples (To *et al*., 2015). These new methods are much faster than Bayesian phylogenetic methods, but they can have comparable accuracy when there is low rate variation across lineages (Duchêne *et al*., submitted). However, rapid dating methods usually do not provide an indication of the uncertainty in the estimate of the evolutionary timescale.

5 Insights from phylogenomic dating: Mammals

The timescale of placental mammal diversification has been a major focus of molecular dating research. According to the fossil record, the evolution of placental mammals had a ‘long fuse’ (Archibald & Deutschman, 2001), whereby the ancestral lineages arose in the Cretaceous period before undergoing rapid diversification during the early Paleogene, after the Cretaceous-Paleogene (K-Pg) extinction event. This scenario is in sharp conflict with the results of molecular-clock analyses carried out in the late 1990s and the 2000s. Many of these studies placed the radiation of placental mammals in the Cretaceous period (Springer, 1997; Kumar & Hedges, 1998; Bininda-Emonds *et al*., 2007). More recently, Meredith *et al*. (2011) inferred a less protracted evolutionary timescale that aligned more closely with the fossil record, but their estimates had a large degree of uncertainty. The results of the molecular

studies collectively imply a substantial gap in the Cretaceous fossil record of placental mammals. However, the Cretaceous fossil record is well sampled (Benton, 1999) and shows that mammals were morphologically similar and uniform, unlike the diversity exhibited in the early Paleogene (Alroy, 1999).

A landmark phylogenomic dating study of placental mammals was carried out by dos Reis *et al*. (2012), who analysed a genome-scale data set of 14,632 genes from 36 mammal taxa. To account for rate variation across genes, the data set was partitioned into 20 equal subsets according to the relative evolutionary rate of each gene. Further analyses were conducted on a subset of 857 genes that had a smaller proportion of missing data than the full data set. This smaller data set was more finely partitioned, with genes being divided according to their branch-length patterns.

To analyse the data, dos Reis *et al*. (2012) used an approximate likelihood method in the Bayesian dating program MCMCtree (Yang, 2007). They estimated an evolutionary timescale that supported a much more recent diversification than those found in previous molecular studies. According to this estimate, the major crown groups of placental mammals originated after the K-Pg boundary, but these groups shared an ancestor in the late Cretaceous (Fig. 2). Thus, the findings of dos Reis *et al*. (2012) are consistent with the 'long fuse' model of evolutionary diversification.

6 Insights from phylogenomic dating: Birds

The evolutionary history of birds has been progressively revised as additional data are collected and as new methods are developed. Despite this large amount of research effort, the evolutionary relationships and timescale of birds have been difficult to resolve with confidence. This is largely due to a lack of informative fossils and because many of the major divergence events within the order are likely to have occurred in a short period of time. The long-standing consensus view is that the modern orders of birds diversified in a small window of time following the extinction of non-avian dinosaurs at the end of the Cretaceous. In contrast, many molecular-clock studies have placed the origin of Neoaves (all birds except the Palaeognathae and Galloanserae), or even the origin of the diverse order Passeriformes, about 10-40 million years before the K-Pg boundary (van Tuinen *et al*., 2006; Brown *et al*., 2008; Ericson *et al*., 2014).

The timescale of avian evolution has been investigated by phylogenomic studies in recent years. In an analysis by Jarvis *et al*. (2014), 1,156 genes were sampled from a total of 8,295 genes that were used for phylogenomic analysis. The subsample of genes was selected on account of their clocklike evolution, as determined using Bayesian phylogenetic analysis. The third codon positions were removed in order to reduce the impacts of mutational saturation

and nucleotide compositional heterogeneity (Jarvis *et al*., 2015). The subset of 1,156 genes was then analysed using MCMCtree (Yang, 2007) with approximate likelihood calculation.

Jarvis *et al*. (2014) compared several tree topologies, with between 17 and 20 calibrations being used for the dating analysis. The study focussed on a tree that had 18 fossil calibrations, most of which were applied as minimum age constraints. A minimum of 66 million years and maximum of 99.6 million years were also specified for the divergence between Palaeognathae, the clade containing ratites and tinamous, and Neognathae, containing all other extant birds. Although the minimum age constraints were all informed by direct fossil evidence, the maximum age bound was based on the absence of crown fossil taxa towards the beginning of the Upper Cretaceous. There has been some debate about the validity of this maximum age constraint (Cracraft *et al*., 2015; Mitchell *et al*., 2015), underscoring the important role of fossil calibrations in the phylogenomic dating analysis.

A more recent study by Prum *et al*. (2015) used a dataset that contained fewer genes (259) but a greater number of taxa (200). These genes were partitioned into 75 subsets to estimate a tree topology that was fixed for the subsequent dating analysis. Of these subsets, 36 were used in the molecular-clock analysis. These subsets of the data were found to maintain their phylogenetic informativeness towards the root of the tree. Each data subset was analysed separately in the Bayesian phylogenetic program BEAST, which is able to estimate the topology and timescale concurrently (Drummond *et al*., 2012). The results of these separate analyses were summarised in a single time-scaled tree, which revealed a rapid diversification of avian lineages in the early Paleogene (Fig. 2).

The studies by Prum *et al*. (2015) and Jarvis *et al*. (2014) used similar approaches to their dating analyses. Both filtered the sequence data with the aim of reducing noise and maximising signal. Despite differences in their methods of choosing fossil calibrations, the two phylogenomic analyses produced similar estimates of divergence times in birds. Both studies placed the age of crown Neoaves near the end of the Cretaceous period, with a rapid radiation of orders occurring in the very early Paleogene.

7 Insights from phylogenomic dating: Insects

Insects form the major part of metazoan diversity, but the timescale of their evolutionary history remains uncertain. As in birds, a deficient fossil record has hindered the palaeontological reconstruction of insect evolution. The oldest insect fossil is that of *Rhyniognatha*, a pair of jaws found in a Scottish deposit dated to the early Devonian over 400 million years ago (Grimaldi & Engel, 2005). This suggests that the origin of insects could have occurred in the Silurian or earlier. Indeed, molecular-clock studies have estimated that the origin of crown insects occurred as early as the Ordovician (Rota-Stabelli *et al*., 2013) or even in the Precambrian (Pisani *et al*., 2004). In their pioneering study of the insect

evolutionary timescale, Gaunt and Miles (2002) inferred that insects arose as late as the Devonian, although this study was published prior to the description of *Rhyniognatha* as an insect. Notable molecular-clock analyses of insects have reconstructed the diversification of holometabolous insects (Wiegmann *et al*., 2009) and flies (Wiegmann *et al*., 2011); and estimated the evolutionary rate for insect mitochondrial DNA (Papadopoulou *et al*., 2010).

In a landmark phylogenomic study, Misof *et al*. (2014) estimated the timescale of insect evolution from 1,478 single-copy protein-coding genes. This was the first study to use genome-scale data across all of the major insect orders. These data were partitioned into 85 subsets that each had a distinct model of amino acid substitution. Each subset was analysed separately using a Bayesian phylogenetic approach in BEAST, with 37 fossil-based calibrations. Most of these fossils satisfied the criteria recommended by Parham *et al*. (2012).

Misof *et al*. (2014) modelled 20 of the 37 fossil calibrations using lognormal prior distributions. The specific use of these prior distributions was disputed by Tong *et al*. (2015), who suggested that a more conservative approach was more appropriate. They reanalysed the data using uniform distributions for the calibrations and using MCMCtree with approximate likelihood calculation. This yielded a more protracted timescale of insect evolution (Fig. 2), with Diptera and Lepidoptera estimated to be around 100 million years older than in the analysis by Misof *et al*. (2014). The clade Polyneoptera was shifted by 80 million years into the past. The date estimates obtained by Tong *et al*. (2015) shared biological interpretations with a number of other studies (Grimaldi & Engel, 2005; Garwood & Sutton, 2010; Smith *et al*., 2011; Wiegmann *et al*., 2011).

8 Future directions

The phylogenomic age offers great opportunities for resolving the timescale of the Tree of Life. With access to genome-scale sequence data, there is considerable potential for improving the precision of molecular date estimates. In turn, this increases the statistical power of analytical methods to test evolutionary hypotheses. Of course, advances in computational power will also be highly beneficial to phylogenomic studies of evolutionary timescales. When quantum computing becomes available for biological research, the application of intensive Bayesian methods will become feasible for phylogenomic dating. Better computation will also enable the analysis of large datasets using complex evolutionary models such as the fossilised birth-death process (Heath *et al*., 2014); the Dirichlet process prior (Heath *et al*., 2012); total-evidence dating (Ronquist *et al*., 2012); and Bayesian dating using full likelihood calculations (Drummond *et al*., 2012) and graphical models (Höhna *et al*., 2016). These represent promising and exciting directions for phylogenomic dating using molecular clocks.

The analysis of sequence data is not the only challenging frontier in molecular dating. When sequence data are abundant, the performance of molecular dating relies on the accuracy of the calibrations and the model of rate variation (Rannala & Yang, 2007; dos Reis & Yang, 2013; Zhu *et al*., 2014). For example, the academic disagreements seen in the phylogenomic analyses of birds and insects were largely due to conflicting interpretations and modelling of palaeontological evidence. This is likely to be an ongoing feature of molecular dating as the fossil record is updated, revised, and reinterpreted.

Molecular dating has expanded considerably and is now a multidisciplinary exercise. Studies of large groups of organisms can involve experts from computation and statistics, molecular evolution, and genetics for sequence analysis; palaeontology and biogeography for time calibrations; and ecology and systematics for species sampling. For studies concerned with more recent timescales, there is also a need for archaeological input (ancient DNA studies) and clinical and epidemiological expertise (viral studies). In this sense, the effective synthesis of knowledge and the ease with which collaborations can form between researchers is another limitation and barrier to overcome when attempting to read the molecular clock.

Funding This work was supported by an Australian Postgraduate Award to KJT and by the Australian Research Council (grant DP160104173 to NL and SYWH).

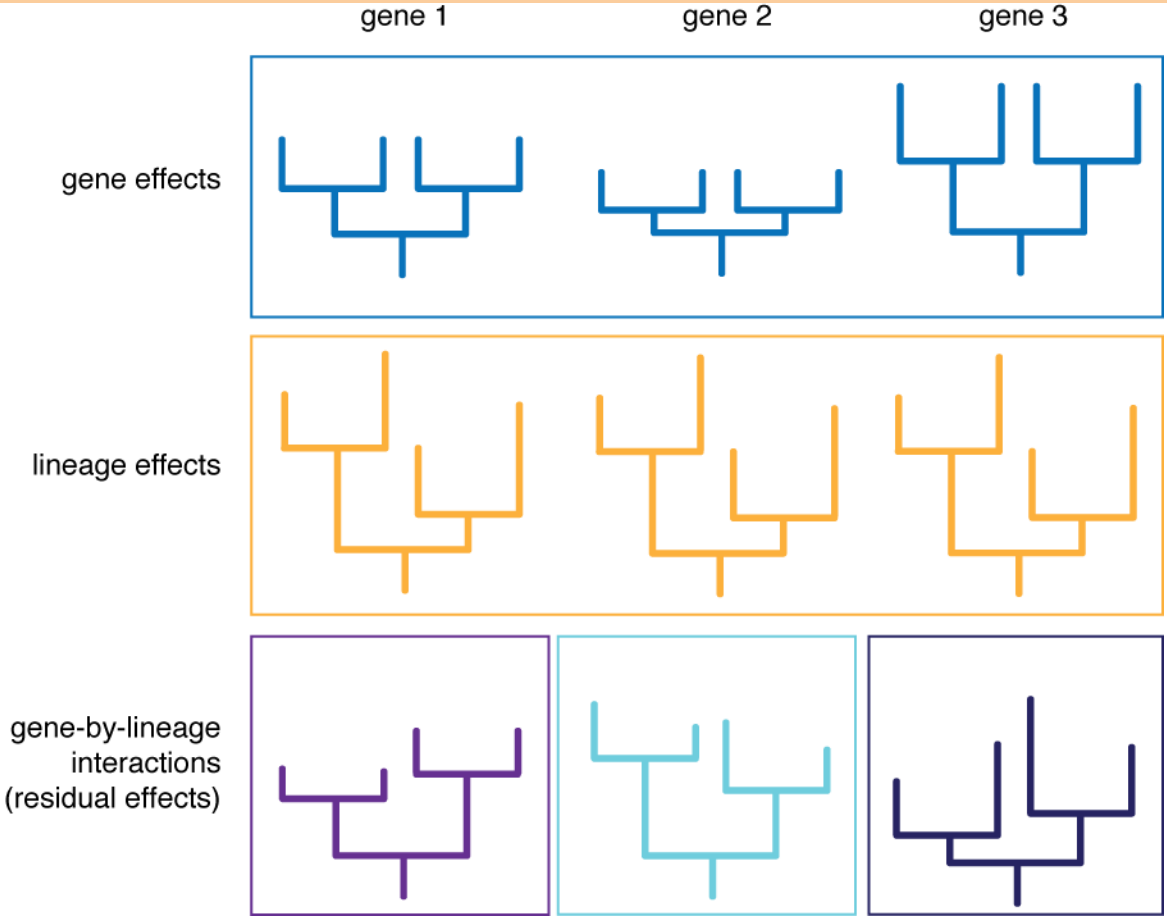

Fig. 1. An illustration of gene effects, lineage effects, and their interactions (residual effects). (a) When there are gene effects, each gene has a distinct rate of evolution, probably as a result of varying selective pressures. (b) When there are lineage effects, the evolutionary rate varies across branches of the tree. This can be caused by differences in life-history characteristics, such as generation length. (c) When there are gene-by-lineage interactions, or residual effects, rates vary across lineages in a gene-specific manner.

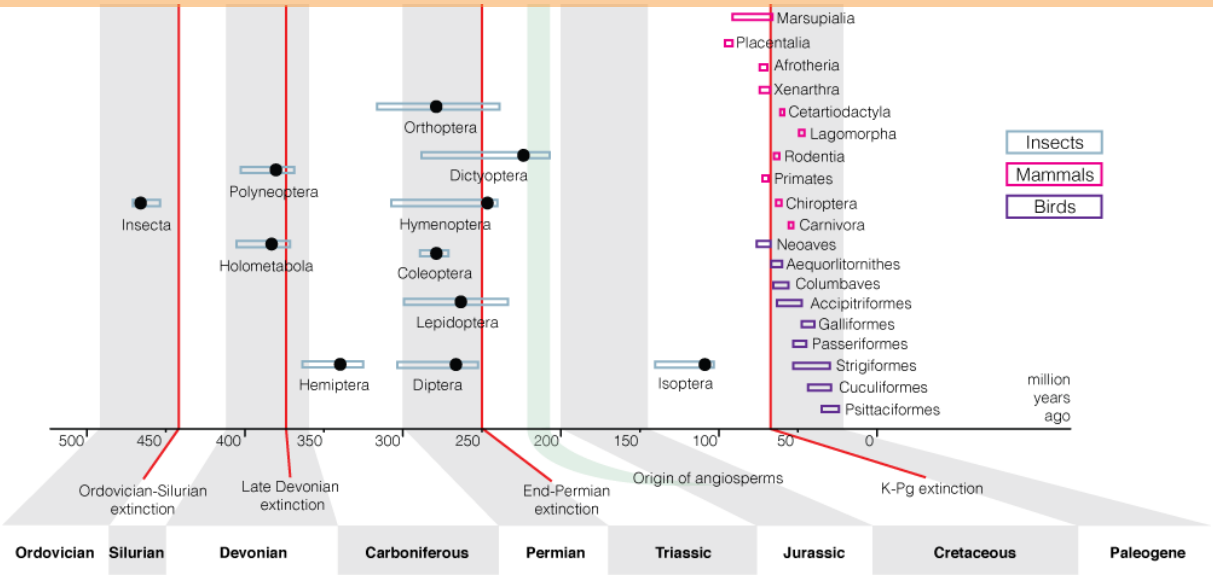

Fig. 2. Phylogenomic estimates of the crown ages of major groups within mammals, birds, and insects. Black circles indicate median age estimates, whereas horizontal bars indicate the associated 95% credibility intervals. Ages for insect groups are according to Tong *et al*. (2015); ages of bird groups are according to Prum *et al*. (2015); and ages of placental mammal groups are according to dos Reis *et al*. (2012). The timings of four mass extinction events are also shown.

References

Alroy, J. 1999. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Systematic Biology*, 48: 107–118.

Archibald, J. D., Deutschman, D. H. 2001. Quantitative analysis of the timing of the origin and diversification of extant placental orders. *Journal of Mammalian Evolution*, 8: 107–124.

Benton, M. J. 1999. Early origins of modern birds and mammals: molecules vs. morphology. *BioEssays*, 21:1043-1051.

Benton, M. J., Donoghue, P. C. J. 2007. Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, 24: 26–53.

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D., Beck, R. M., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., Purvis, A. 2007. The delayed rise of present-day mammals. *Nature*, 446: 507–512.

Bromham, L. 2009. Why do species vary in their rate of molecular evolution? *Biology Letters*, 5: 401–404.

Bromham, L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society of London B*, 366: 2503–2513.

Brown, J. W., Rest, J. S., García-Moreno, J., Sorenson, M. D., Mindell, D. P. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biology,* 6: 6.

Brown, R., Richardson, M., Boulter, D., Ramshaw, J., Jefferies, R. 1972. The amino acid sequence of cytochrome *c* from *Helix aspersa* Müller (garden snail). *Biochemical Journal,* 128: 971–974.

Brown, R. P., Yang, Z. 2011. Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology*, 11: 271.

Cracraft, J., Houde, P., Ho, S. Y. W., Mindell, D. P., Fjeldså, J., Lindow, B., Edwards, S. V., Rahbek, C., Mirarab, S., Warnow, T. 2015. Response to Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science*, 349: 1460–1460.

dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. J., Yang, Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society of London B,* 279: 3491–3500.

dos Reis, M., Yang, Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Molecular Biology and Evolution*, 28: 2161–2172.

dos Reis, M., Yang, Z. 2013. The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution*, 51: 30–43.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4: e88.

Drummond, A. J., Suchard, M. A., Xie, D., Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29: 1969–1973.

Duchêne, S., Geoghegan, J. L., Holmes, E. C., Ho, S. Y. W. 2016. Estimating evolutionary rates using time-structured data: a general comparison of phylogenetic methods. *Bioinformatics*, submitted.

Duchêne, S., Ho, S. Y. W. 2014. Using multiple relaxed-clock models to estimate evolutionary timescales from DNA sequence data. *Molecular Phylogenetics and Evolution*, 77: 65–70.

Duchêne, S., Molak, M., Ho, S. Y. W. 2013. ClockstaR: Choosing the number of relaxed-clock models in molecular phylogenetic analysis. *Bioinformatics*, 30: 1017–1019.

Ericson, P. G., Klopfstein, S., Irestedt, M., Nguyen, J. M. T., Nylander, J. A. A. 2014. Dating the diversification of the major lineages of Passeriformes (Aves). *BMC Evolutionary Biology*, 14: 1.

Garwood, R., Sutton, M. 2010. X-ray micro-tomography of Carboniferous stem-Dictyoptera: new insights into early insects. *Biology Letters*, 6: 699–702.

Gaunt, M. W., Miles, M. A. 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution*, 19: 748–761.

Gaut, B., Yang, L., Takuno, S., Eguiarte, L. E. 2011. The patterns and causes of variation in plant nucleotide substitution rates. *Annual Review of Ecology, Evolution, and Systematics,* 42: 245–266.

Gillespie, J. H. 1991. The Causes of Molecular Evolution. Oxford University Press, Oxford.

Grimaldi, D., Engel, M. S. 2005. Evolution of the Insects. Cambridge University Press, Cambridge.

Heath, T. A., Holder, M. T., and Huelsenbeck, J. P. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. *Molecular Biology and Evolution*, 29: 939–955.

Heath, T. A., Huelsenbeck, J. P., Stadler, T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 111: E2957–E2966.

Hedges, S. B., Dudley, J., Kumar, S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*, 22: 2971–2972.

Ho, S. Y. W., Phillips, M. J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58: 367–380.

Ho, S. Y. W., Tong, K. J., Foster, C. S. P., Ritchie, A. M., Lo, N., Crisp, M. D. 2015. Biogeographic calibrations for the molecular clock. *Biology Letters*, 11: 20150194.

Ho, S. Y. W. 2014. The changing face of the molecular evolutionary clock. *Trends in Ecology and Evolution*, 29: 496–503.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., Ronquist, F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65: 726–736.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., da Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M. S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W. C., Ray, D., Green, R. E., Bruford, M. W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E. P., Bertelsen, M. F., Sheldon, F. H., Brumfield, R. T., Mello, C. V., Lovell, P. V., Wirthlin, M., Schneider, M. P. C., Prosdocimi, F., Samaniego, J. A., Velazquez, A. M. V., Alfaro-Núñez, A., Campos, P. F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D. M., Zhou, Q., Perelman, P., Driskell, A. C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F. E., Braun, M., Fjeldså, J., Orlando, L., Barker, F. K., Jønsson, K. A., Johnson, W., Koepfli, K.-P., O'Brien, S., Haussler, D., Ryder, O. A., Rahbek, C., Willerslev, E., Graves, G. R., Glenn, T. C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S. V., Stamatakis, A., Mindell, D. P., Cracraft, J., Braun, E. L., Warnow, T., Wang, J., Gilbert, M. T. P., Zhang, G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346: 1320–1331.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., da Fonseca, R. R., Alfaro-Núñez, A., Narula, N., Liu, L., Burt, D., Ellegren, H., Edwards, S. V., Stamatakis, A., Mindell, D. P., Cracraft, J., Braun, E. L., Warnow, T., Wang, J., Gilbert, M. T. P., Zhang, G., The Avian Phylogenomics Consortium. 2015. Phylogenomic analyses data of the avian phylogenomics project. *GigaScience*, 4: 4.

Kumar, S., Hedges, S. B. 1998. A molecular timescale for vertebrate evolution. *Nature*, 392: 917–920.

Kumar, S., Hedges, S. B. 2016. Advances in time estimation methods for molecular data. *Molecular Biology and Evolution*, 33: 863–869.

Lee, M. S. Y., Ho, S. Y. W. 2016. Molecular clocks. *Current Biology*, 26: R399–R402.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., Brumfield, R. T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66: 526–538.

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., Goodbla, A., Eizirik, E., Simão, T. L., Stadler, T. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, 334: 521–524.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T. R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L. S., Kawahara, A. Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D. D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J. L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B. M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N. U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M. G., Wiegmann, B. M., Wilbrandt, J., Wipfler, B., Wong, T. K. F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D. K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K. M., Zhou, X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346: 763–767.

Mitchell, K. J., Cooper, A., Phillips, M. J. 2015. Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science*, 349: 1460.

Mullis, K. B., Faloona, F. A.. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 155: 335–350.

Muse, S. V., Gaut, B. S.. 1997. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. *Genetics*, 146: 393–399.

Papadopoulou, A., Anastasiou, I., Vogler, A. P. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, 27: 1659–1672.

Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patané, J. S. L., Smith, N. D., Tarver, J. E., van Tuinen, M., Yang, Z., Angielczyk, K. D., Greenwood, J. M., Hipsley, C. A., Jacobs, L., Makovicky, P. J., Müller, J., Smith, K. T., Theodor, J. M., Warnock, R. C. M., Benton, M. J. 2012. Best practices for justifying fossil calibrations. *Systematic Biology*, 61: 346–359.

Pisani, D., Poling, L. L., Lyons-Weiler, M., Hedges, S. B. 2004. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. *BMC Biology,* 2: 1.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., Lemmon, A. R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526: 569–573.

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. *Systematic Biology*, 56: 453–466.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., Rasnitsyn, A. P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, 61: 973–999.

Rota-Stabelli, O., Daley, A. C., Pisani, D. 2013. Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Current Biology*, 23: 392–398.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, 14: 1218–1231.

Sanger, F., Air, G., Barrell, B., Brown, N., Coulson, A., Fiddes, J., Hutchison, C., Slocombe, P., Smith, M. 1977a. Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, 265: 687–695.

Sanger, F., Nicklen, S., Coulson, A. R. 1977b. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463–5467.

Sarich, V. M., Wilson, A. C. 1967. Immunological time scale for hominid evolution. *Science*, 158: 1200–1203.

Sauquet, H., Ho, S. Y. W., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., Bayly, M. J., Bromham, L., Brown, G. K., Carpenter, R. J., Lee, D. M., Murphy, D. J., Sniderman, J. M. K., Udovicic, F. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of *Nothofagus* (Fagales). *Systematic Biology*, 61: 289–313.

Smith, V. S., Ford, T., Johnson, K. P., Johnson, P. C., Yoshizawa, K., Light, J. E. 2011. Multiple lineages of lice pass through the K–Pg boundary. *Biology Letters*, 7: 782–785.

Springer, M. S. 1997. Molecular clocks and the timing of the placental and marsupial radiations in relation to the Cretaceous–Tertiary boundary. *Journal of Mammalian Evolution*, 4: 285–302.

Tamura, K., Battistuzzi, F. U., Billing-Ross, P., Murillo, O., Filipski, A., Kumar, S. 2012. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America*, 109: 19333–19338.

Thorne, J. L., Kishino, H., Painter, I. S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, 15: 1647–1657.

To, T.-H., Jung, M., Lycett, S., Gascuel, O. 2015. Fast dating using least-squares criteria and algorithms. *Systematic Biology*, 65: 82–97.

Tong, K. J., Duchêne, S., Ho, S. Y. W., Lo, N. 2015. Comment on "Phylogenomics resolves the timing and pattern of insect evolution". *Science*, 349: 487.

van Tuinen, M., Stidham, T. A., Hadly, E. A. 2006. Tempo and mode of modern bird evolution observed with large-scale taxonomic sampling. *Historical Biology*, 18: 209–225.

Warnock, R. C. M., Parham, J. F., Joyce, W. G., Lyson, T. R., Donoghue, P. C. J. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society of London B*, 282: 20141013.

Wiegmann, B. M., Trautwein, M. D., Kim, J. W., Cassel, B. K., Bertone, M. A., Winterton, S. L., Yeates, D. K. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biology*, 7: 34.

Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J.-W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Pape, T., Sinclair, B. J., Skevington, J. H., Blagoderov, V., Caravas,, J., Kutty, S. N., Schmidt-Ott, U., Kampmeier, G. E., Thompson, F. C., Grimaldi, D. A., Beckenbach, A. T., Courtney, G. W., Friedrich, M., Meier, R., Yeates, D. K. 2011. Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, 108: 5690–5695.

Wilson, A. C., Sarich, V. M. 1969. A molecular time scale for human evolution. *Proceedings of the National Academy of Sciences of the United States of America,* 63: 1088–1093.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution*, 10: 1396–1401.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24: 1586–1591.

Zhu, T., dos Reis, M., Yang, Z. 2014. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Systematic Biology*, 64: 267–280.

Zuckerkandl, E., Pauling, L. 1962. Molecular disease, evolution and genic heterogeneity. *In*: Kasha, M., Pullman, B. (eds), *Horizons in Biochemistry*. Academic Press, New York, pp. 189–255.

Zuckerkandl, E., L. Pauling. 1965. Evolutionary divergence and convergence in proteins. *In*: Bryson, V., Vogek, H. J. (eds), *Evolving Genes and Proteins*. Academic Press, New York. pp. 97–166.