

Calibrating molecular estimates of substitution rates and divergence times in birds

Simon Y. W. Ho

S. Y. W. Ho, *Department of Zoology, University of Oxford, United Kingdom. E-mail: simon.ho@zoo.ox.ac.uk*

Many evolutionary studies of birds rely on the estimation of molecular divergence times and substitution rates. In order to perform such analyses, it is necessary to incorporate some form of calibration information: a known substitution rate, radiometric ages of heterochronous sequences, or inferred ages of lineage splitting events. All three of these techniques have been employed in avian molecular studies, but their usage has not been entirely satisfactory. For example, the ‘traditional’ avian mitochondrial substitution rate of 2% per million years is frequently adopted without acknowledgement of the associated uncertainty. Similarly, fossil and biogeographic information is almost always converted into an errorless calibration point. In both cases, the resulting estimates of divergence times and substitution rates will be artificially precise, which has a considerable impact on hypothesis testing. In addition, using such a simplistic approach to calibration discards much of the information offered by the fossil record. A number of more sophisticated calibration methods have recently been introduced, culminating in the development of probability distribution-based calibrations. In this article, I discuss the use of this new class of methods and offer guidelines for choosing a calibration technique.

The estimation of molecular divergence times and substitution rates is one of the most important exercises in evolutionary studies of birds, as it is able to provide illumination on the tempo and time scale of molecular evolutionary processes. Through analysis of molecular data, we can obtain estimates of genetic relatedness among taxa, either in the form of pairwise distances or as branch lengths in a phylogenetic tree. Each distance estimate, measured in average substitutions per site, is the product of two components: the substitution rate and the time elapsed. With molecular data alone, we are unable to separate the contributions of these two components, so the distance estimates by themselves can only impart information about relative ages of lineage divergence events. In order to establish an absolute (geological) time scale, it is necessary to introduce some form of additional calibrating information. This can take one of several forms (Drummond et al. 2006): (i) Importation of a known substitution rate (Fig. 1a). If an independent study has estimated a substitution rate for the taxon in question, it is possible to use this rate in order to convert genetic distances into time units. For example, many ornithological studies

have employed the ‘traditional’ mitochondrial divergence rate of 2% per million years, which was obtained in a number of molecular studies of birds (Shields and Wilson 1987, Randi 1996, Fleischer et al. 1998). (ii) Including heterochronous sequences with known ages in the analysis (Fig. 1b). This is possible with ancient DNA extracted from subfossil or museum specimens (Lambert et al. 2002, Paxinos et al. 2002). The ages of the specimens must be known, either through historical records or by radiocarbon dating. Provided that the ages of the specimens are adequately distinct relative to the total depth of the phylogenetic tree, they can provide sufficient information for calibration (Drummond et al. 2002, Drummond et al. 2003). (iii) Calibration points based on palaeontological or other evidence (Fig. 1c–1g). The fossil record, certain biogeographic events (such as continental break-up or island formation), or an analysis of independent molecular data can offer an estimate of the age of an internal node in a tree, which represents a lineage splitting event. In a phylogenetic analysis, the age of the node can be fixed, restricted, or assumed to follow some parametric distribution (i.e.,

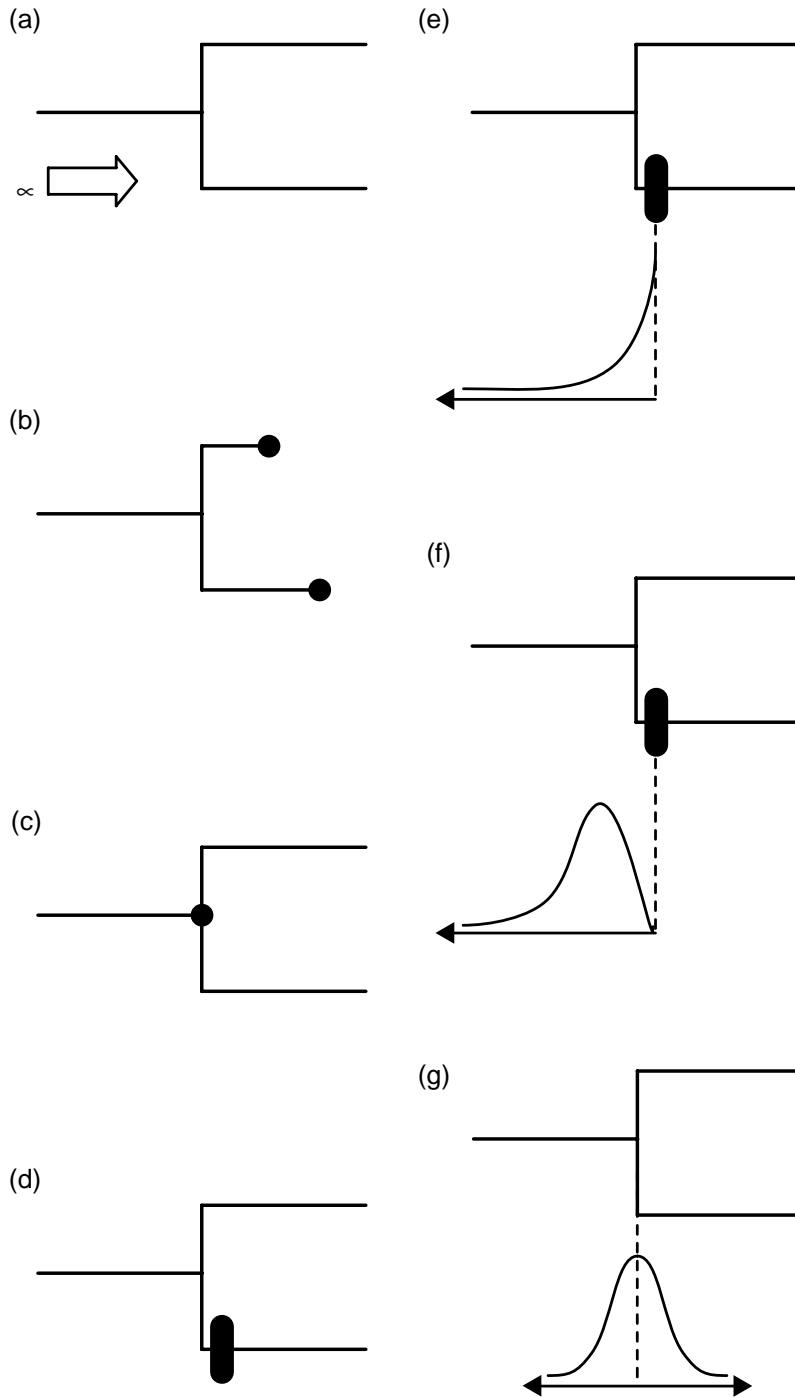


Fig. 1. Different methods for calibrating estimates of substitution rates and divergence times. (a) Importation of an independently derived rate, μ . (b) Inclusion of sequences with distinct, known ages. (c) Point calibration: fixing the age of a node to a specific value based on independent evidence. (d) Lower calibration bound: using independent evidence to place a minimum constraint on the age of a node. (e-g) Assuming that the probability of the node being a certain age follows some parametric distribution: (e) exponential distribution, with a rigid minimum bound, (f) lognormal distribution, with a rigid minimum bound, or (g) normal distribution.

a statistical distribution governed by a small number of parameters, such as a normal distribution).

All three of these methods have been used in phylogenetic studies of birds. In general, however, the suitability and rigorousness of their application has not been entirely satisfactory. In this article, I discuss the use of calibration information in evolutionary studies of birds, and I strongly encourage a movement towards more sophisticated and informative calibration techniques.

In order to survey recent usage of calibration techniques, I examined the 75 studies listed by Peterson (2006) in his recent investigation of avian molecular clocks. These studies, all presenting analyses of avian sequence data, were selected at random from papers published in the period spanning 2000–2006. Of the 75 studies, 47 performed divergence time estimation (Table 1). Two of the three calibration methods were represented among the studies: 32 studies imported a substitution rate that was estimated in an independent study, whereas 15 studies used calibration points to fix or constrain the ages of internal nodes in the tree. The second calibration method mentioned above, the use of heterochronous sequences with known ages (Fig. 1b; Drummond et al. 2002), was not used in any of the 75 studies surveyed. It has, however, been used elsewhere in studies of ancient DNA from Adélie penguins (Lambert et al. 2002) and moa (Ho et al. 2007a).

Imported substitution rate

Of the 32 studies that used an independently derived substitution rate (Fig. 1a), 28 employed the ‘traditional’ divergence rate of 2% per million years or some similar value (broadly defined as falling in the range of 1.5–3%). This rate has long been regarded as a standard across diverse avian taxa, but there are several reasons why its use is questionable and probably invalid. First, its universal applicability has been seriously questioned (García-Moreno 2004, Lovette 2004, Ho et al. 2005, Pereira and Baker 2006a). The fact that the assumption

Table 1. Summary of calibration methods used in 75 recent molecular studies of birds.

Calibration method	No. of studies
Imported substitution rate	
2% rate or similar	28
Other rate	4
Independent calibration	
Palaeontological	5
Biogeographic	3
Secondary (molecular)	7
No date estimation	28
Total	75

of a molecular clock was rejected for two-thirds of the data sets tested (Peterson 2006) indicates that significant rate variation is widespread among birds.

Second, the 2% rate has almost always been used without due consideration of the associated uncertainty. All rate estimates come with some amount of estimation error, which results from uncertainty in estimates of genetic distances and in calibration points. Therefore, any studies adopting the 2% rate as an errorless value will produce divergence time estimates that are artificially precise. This has important implications for studies investigating the correlation between evolutionary time scales and biogeographic, climatic, or other events. A re-analysis of these studies, taking calibration rate uncertainty into account, could reveal that many of them lack sufficient statistical power to distinguish among competing evolutionary hypotheses.

A third problem, potentially the most serious, is that the 2% rate is a substitution rate derived from interspecific comparisons. Observed rates of molecular change within species are theoretically expected to be higher than those observed between species, because many of the polymorphisms seen at the population level are lost over longer time frames. This is supported by empirical evidence of higher intraspecific rates in a range of birds and mammals (Lambert et al. 2002, Howell et al. 2003, Ho et al. 2005, 2007a, b, Mao et al. 2006). For example, Lambert et al. (2002) estimated a short-term mutation rate of 95% per million years from Adélie penguins. Consequently, employment of the 2% interspecific rate in analyses of intraspecific data, which was done in 13 of the studies surveyed (representing almost half of the 28 studies using the 2% rate), is likely to produce overestimates of the age of intraspecific divergence events (Ho et al. 2005, Ho and Larson 2006). Conversely, it is inappropriate to use an intraspecific rate to study molecular evolution between species; this was done in one of the studies (Zink 2002).

Point calibrations

Of the 15 studies that used calibration points to fix or constrain the ages of internal nodes, seven used point calibrations based on palaeontological or biogeographic information. In all instances, fossil information was represented as a single, errorless value (Fig. 1c). This treatment of fossil calibrations is inappropriate because fossils can only provide minimum age estimates for divergence events, as the appearance of a fossil necessarily postdates the origin of the clade of which it is a member (Marshall 1990, Hedges and Kumar 2004, Benton and Donoghue 2007). Moreover, the geological age and taxonomic affinity of the calibration fossil will have some amount of uncertainty, which is often considerable and must be taken into account (Graur and Martin 2004).

In addition to the seven studies using fixed calibration points, seven studies used secondary calibration points (i.e., those based on molecular date estimates). In none of the cases was there acknowledgement of the associated uncertainty. The practice of using secondary calibration points has been questioned (Shaul and Graur 2002, Graur and Martin 2004), but they may be useful when primary calibration points are not available (Hedges and Kumar 2004). For example, some groups of birds have conspicuously poor fossil records, such as those evolving on the southern continents (Feduccia 2003). Secondary calibrations must be employed very cautiously, however, because they often have large estimation errors due to the compounding of uncertainties in genetic distance estimates and in primary calibrations; it is clearly inappropriate to co-opt them as errorless calibrations (Graur and Martin 2004).

Age constraint calibrations

Only one of the surveyed studies used fossil information in an immediately justifiable manner. Pereira and Baker (2006b) set five minimum calibration bounds using palaeontological evidence (Fig. 1d). The fact that only one study implemented age constraints is surprising because divergence dating methods with this facility were published several years prior to the publications surveyed here (Sanderson 1997, Thorne et al. 1998). Given that the molecular clock model was rejected so frequently (Peterson 2006), it is also remarkable that only six of the 75 studies used relaxed-clock techniques, which allow substitution rates to vary among lineages (for reviews of these methods, see Renner 2005, Welch and Bromham 2005, Rutschmann 2006).

Placing minimum bounds on the ages of nodes is possible in a range of available relaxed-clock methods (Sanderson 1997, Thorne et al. 1998, Sanderson 2002, Drummond et al. 2006). While this technique is strictly valid, the distillation of fossil data into simple minimum bounds is not ideal because it discards potentially useful information (Yang and Rannala 2006). Equal probability is effectively assigned to all values above this bound, implying a uniform distribution from the bound to infinity (Fig. 1d). For example, placing a minimum bound of 10 million years on the age of a node implies that there is a probability of 0 for ages that are more recent than this value, and a probability of 1 for ages above this value. In other words, there is an equal probability of the node being 10, 100, or 1000 million years. This is patently a poor expression of the available palaeontological information; the age of the node is more likely to be close to the age of the oldest known fossil, and less likely to be significantly older. For this reason, employing a uniform distribution with a mini-

imum bound can be regarded as an excessively conservative approach to modelling the fossil record. Other difficulties associated with rigid age constraints have been discussed elsewhere (Yang and Rannala 2006).

Calibration using parametric distributions

Recently, a Bayesian relaxed-clock method was published that allows the implementation of sophisticated calibration methods (Drummond et al. 2006). Using information acquired from independent evidence, priors can be specified for the ages of internal nodes. These priors take the form of parametric statistical distributions, such as the lognormal, exponential, and normal distributions (Fig. 1e–1g). Each distribution is appropriate under a different set of circumstances, and I outline these below.

Exponential distribution

This distribution is suitable for modelling fossil calibrations, because the probability decreases with a growing discrepancy between estimated nodal age and the age of the calibrating fossil (Fig. 1e). An exponential distribution is most appropriate in situations where we are reasonably confident that a fossil falls very close to the actual divergence event. This may be the case when the fossil record is relatively complete for the group in question. An exponential distribution requires the specification of two parameters: a lower bound, representing the age of the fossil evidence, as well as the mean. The mean can be chosen so that 95% of the probability is contained between the rigid lower bound and some 'soft' maximum bound (*sensu* Yang and Rannala 2006).

Lognormal distribution

This is perhaps the most appropriate for modelling palaeontological information (Fig. 1f; Hedges and Kumar 2004, Drummond et al. 2006). It explicitly assumes that the actual divergence event is most likely to have occurred some time prior to the earliest appearance of fossil evidence. One of the drawbacks of using the lognormal distribution is the number of parameters that need to be specified (rigid lower bound, mean, and standard deviation). Selection of the mean is somewhat subjective, but, as with the exponential distribution, the parameters can be chosen so that 95% of the probability lies within some interval. If there is insufficient information to describe a lognormal distribution, an exponential distribution can be used instead.

Normal distribution

The nature of this distribution perhaps makes it inappropriate for summarising fossil information, but it is particularly suitable in several other settings (Fig. 1g). First, it can be used to model certain biogeographic events. For example, Fleischer et al. (1998) calibrated their rate estimates using the dates of formation of the Hawaiian islands. An explicit assumption is that the extant, inhabitant lineages diversified at some unknown time subsequent to island formation. It is possible, however, that lineage diversification actually occurred prior to colonisation, or that the age of the island has been underestimated for some reason (Heads 2005). It is, therefore, unwise to impose a strict minimum bound on the age of the divergence. Instead, a normal distribution would be appropriate for modelling the non-directional uncertainty. A similar argument could be used for the formation of the Panamanian isthmus, which provided a minimum calibration bound of 3.1 million years in an analysis of woodpeckers (Fleischer et al. 2006); this specific geological event might not necessarily have an immediate functional role for avian evolution, so it is advisable to allow for bi-directional uncertainty. In their study of the evolution of horses in the Americas, Weinstock et al. (2005) modelled the age of the formation of the Panamanian isthmus as a normal distribution with a mean of 3.0 million years and standard deviation of 0.5 million years.

Second, a normal distribution can be used to reflect uncertainty in imported rate or date estimates, including secondary calibration points. For example, the mutation rate estimated from Adélie penguins by Lambert et al. (2002) could be used to provide calibration information in an analysis of a closely related species. In a study of European brown bears, Saarma et al. (2007) used a rate estimated from a different population of brown bears, incorporating it as a normally-distributed parameter with the standard deviation appropriately reflecting uncertainty on the original rate estimate.

Third, a normal distribution can be used to reflect the non-directional uncertainty on radiometric dating. This would be suitable for radiometric dating of fossiliferous strata as well as for radiocarbon dates for ancient DNA sequences (Lambert et al. 2002, Paxinos et al. 2002).

All three of the distributions described above have diminishing tails of probability as the estimated age of the node increases. A major benefit of this characteristic is that it removes the need for including maximum bounds, which are otherwise necessary in most relaxed-clock analyses to allow dating algorithms to obtain unique solutions (Sanderson 2003, Benton and Donoghue 2007). Under normal circumstances, it is difficult to choose maximum bounds without being exceedingly conservative because one cannot rule out the possibility

of the true split occurring much earlier than the fossil evidence (Yang and Rannala 2006).

Several difficulties in constructing parametric distributions are evident. Apart from needing to select from the range of available distributions, the specification of the level of uncertainty is subjective. For example, it is difficult to formulate an objective, quantitative function to generate probability distributions for calibration from the fossil record; moreover, many of the factors affecting the quality of fossil data, such as taxonomic uncertainty and sampling effort, are not easily quantifiable. While the simple distributions presented above are probably unable to capture comprehensively the information in the fossil record, they certainly represent a significant improvement upon the use of errorless calibration points and rigid calibration bounds.

Conclusions

A vast amount of molecular and palaeontological data has been produced by research on birds, offering abundant opportunities for investigating evolutionary processes. To exploit the full potential of these data, it is paramount to conduct phylogenetic analyses in a considered and rigorous manner. If this is not possible, it may be preferable to avoid performing such analyses in order to prevent the production of misleading results. As I have outlined in this article, sophisticated techniques are now available for the estimation of divergence times. It is the responsibility of the ornithological research community to utilise them.

Acknowledgements – J. Haile, A. T. Peterson, and anonymous reviewers provided helpful comments on the manuscript, and the ideas presented here benefited from discussions with M. J. Phillips, R. A. Fortey, and A. Rambaut. This work was supported by the Leverhulme Trust.

References

- Benton, M. J. and Donoghue, P. C. 2007. Paleontological evidence to date the tree of life. – *Mol. Biol. Evol.* 24: 26–53.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J. and Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. – *PLoS Biol.* 4: e88.
- Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. and Solomon, W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. – *Genetics* 161: 1307–1320.
- Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. and Rodrigo, A. G. 2003. Measurably evolving populations. – *Trends Ecol. Evol.* 18: 481–488.

- Feduccia, A. 2003. 'Big bang' for tertiary birds? – *Trends Ecol. Evol.* 18: 172–176.
- Fleischer, R. C., Kirchman, J. J., Dumbacher, J. P., Bevier, L., Dove, C., Rotzel, N. C., Edwards, S. V., Lammertink, M., Miglia, K. J. and Moore, W. S. 2006. Mid-Pleistocene divergence of Cuban and North American ivory-billed woodpeckers. – *Biol. Lett.* 2: 466–469.
- Fleischer, R. C., McIntosh, C. E. and Tarr, C. L. 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. – *Mol. Ecol.* 7: 533–545.
- García-Moreno, J. 2004. Is there a universal mtDNA clock for birds? – *J. Avian Biol.* 35: 465–468.
- Graur, D. and Martin, W. 2004. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. – *Trends Genet.* 20: 80–86.
- Heads, M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. – *Cladistics* 21: 62–78.
- Hedges, S. B. and Kumar, S. 2004. Precision of molecular time estimates. – *Trends Genet.* 20: 242–247.
- Ho, S. Y. W. and Larson, G. 2006. Molecular clocks: when times are a-changin'. – *Trends Genet.* 22: 79–83.
- Ho, S. Y. W., Heupink, T. H., Rambaut, A. and Shapiro, B. 2007a. Bayesian estimation of sequence damage in ancient DNA. – *Mol. Biol. Evol.* in press
- Ho, S. Y. W., Phillips, M. J., Cooper, A. and Drummond, A. J. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. – *Mol. Biol. Evol.* 22: 1561–1568.
- Ho, S. Y. W., Shapiro, B., Phillips, M., Cooper, A. and Drummond, A. J. 2007b. Evidence for time dependency of molecular rate estimates. – *Syst. Biol.* 56: in press.
- Howell, N., Smejkal, C. B., Mackey, D. A., Chinnery, P. F., Turnbull, D. M. and 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.
- Lambert, D. M., Ritchie, P. A., Millar, C. D., Holland, B., Drummond, A. J. and Baroni, C. 2002. Rates of evolution in ancient DNA from Adélie penguins. – *Science* 295: 2270–2273.
- Lovette, I. J. 2004. Mitochondrial dating and mixed support for the "2% rule" in birds. – *Auk* 121: 1–6.
- Mao, L., Zabel, C., Wacker, M. A., Nebrich, G., Sagi, D., Schrade, P., Bachmann, S., Kowald, A. and Klose, J. 2006. Estimation of the mtDNA mutation rate in aging mice by proteome analysis and mathematical modeling. – *Exp. Gerontol.* 41: 11–24.
- Marshall, C. R. 1990. The fossil record and estimating divergence times between lineages: Maximum divergence times and the importance of reliable phylogenies. – *J. Mol. Evol.* 30: 400–408.
- Paxinos, E. E., James, H. F., Olson, S. L., Sorenson, M. D., Jackson, J. and Fleischer, R. C. 2002. mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (*Branta canadensis*). – *Proc. Natl Acad. Sci. USA* 99: 1399–1404.
- Pereira, S. L. and Baker, A. J. 2006a. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. – *Mol. Biol. Evol.* 23: 1731–1740.
- Pereira, S. L. and Baker, A. J. 2006b. A molecular timescale for galliform birds accounting for uncertainty in time estimates and heterogeneity of rates of DNA substitutions across lineages and sites. – *Mol. Phylogenet. Evol.* 38: 499–509.
- Peterson, A. T. 2006. Application of molecular clocks in ornithology revisited. – *J. Avian Biol.* 37: 541–544.
- Randi, E. 1996. A mitochondrial cytochrome B phylogeny of the *Alectoris* partridges. – *Mol. Phylogenet. Evol.* 6: 214–227.
- Renner, S. S. 2005. Relaxed molecular clocks for dating historical plant dispersal events. – *Trends Plant Sci.* 10: 1360–1385.
- Rutschmann, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. – *Diversity Distrib.* 12: 35–48.
- Saarma, U., Ho, S. Y. W., Pybus, O. G., Kaljuste, M., Tumanov, I. M., Kojola, I., Vorobiev, A. A., Markov, N. I., Saveljev, A. P., Valdmann, A., Lyapunova, E. A., Abramov, A. V., Männil, P., Korsten, M., Vulla, E., Pazetnov, S. V., Pazetnov, V. S., Putschkovskiy, S. V. and Rõkov, A. M. 2007. Mitogenetic structure of brown bears (*Ursus arctos* L.) in north-eastern Europe and a new time-frame for the formation of brown bear lineages. – *Mol. Ecol.* 16: 401–413.
- Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. – *Mol. Biol. Evol.* 14: 1218–1231.
- Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. – *Mol. Biol. Evol.* 19: 101–109.
- Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. – *Bioinf.* 19: 301–302.
- Shaul, S. and Graur, D. 2002. Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. – *Gene* 300: 59–61.
- Shields, G. F. and Wilson, A. C. 1987. Calibration of mitochondrial DNA evolution in geese. – *J. Mol. Evol.* 34: 212–217.
- Thorne, J. L., Kishino, H. and Painter, I. S. 1998. Estimating the rate of evolution of the rate of molecular evolution. – *Mol. Biol. Evol.* 15: 1647–1657.
- Weinstock, J., Willerslev, E., Sher, A., Tong, W., Ho, S. Y. W., Rubinstein, D., Storer, J., Burns, J., Martin, L., Bravi, C., Prieto, A., Froese, D., Scott, E., Xulong, L. and Cooper, A. 2005. Evolution, systematics, and phylogeography of Pleistocene horses in the New World: a molecular perspective. – *PLoS Biol.* 3: e241.
- Welch, J. J. and Bromham, L. 2005. Molecular dating when rates vary. – *Trends Ecol. Evol.* 20: 320–327.
- Yang, Z. and Rannala, B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. – *Mol. Biol. Evol.* 23: 212–226.
- Zink, R. M. 2002. A new perspective on the evolutionary history of Darwin's finches. – *Auk* 119: 864–871.