MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 499–501

NEWS AND VIEWS

PERSPECTIVE

Ancient hyaenas highlight the old problem of estimating evolutionary rates

BETH SHAPIRO* and SIMON Y. W. HO†

**Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 1156 High Street Mailstop:EEBiology/EMS, Santa Cruz, CA 95064, USA; †School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia*

Phylogenetic analyses of ancient DNA data can provide a timeline for evolutionary change even in the absence of fossils. The power to infer the evolutionary rate is, however, highly dependent on the number and age of samples, the information content of the sequence data and the demographic history of the sampled population. In this issue of *Molecular Ecology*, Sheng *et al.* (2014) analysed mitochondrial DNA sequences isolated from a combination of ancient and present-day hyaenas, including three Pleistocene samples from China. Using an evolutionary rate inferred from the ages of the ancient sequences, they recalibrated the timing of hyaena diversification and suggest a much more recent evolutionary history than was believed previously. Their results highlight the importance of accurately estimating the evolutionary rate when inferring timescales of geographical and evolutionary diversification.

Keywords: ancient DNA, calibration, substitution rate

Received 6 November 2013; revised 2 December 2013; accepted 2 December 2013

Identifying a reliable timescale for demographic and biogeographical processes is a valuable component of evolutionary research. For example, it allows us to evaluate the influence of environmental and anthropogenic factors on populations, species and higher taxa. Phylogenetic methods can be used to estimate timescales from genetic data based on the assumption of a molecular clock. The 'tick' rate of the molecular clock is usually unknown, but can be estimated by constraining the times of nodes in the phylogenetic tree (Fig. 1). This procedure, known as calibration, requires temporal information from an independent source.

Previously, estimates of the evolutionary rate of the mitochondrial cytochrome *b* gene were used to date the dispersal and colonization history of hyaenas. Rohland *et al.* (2005) assumed an age of 10 Myr for the divergence between the spotted hyaena and striped/brown hyaena to calibrate their evolutionary rate and suggested that there have been four separate dispersals of hyaenas from Africa into Eurasia within the last 3.5 Myr. In a similar analysis of hyaena mitochondrial sequences, Sheng *et al.* (2014) inferred a different and significantly less protracted evolutionary history. The new analysis suggests that the distribution of hyaena fossils, rather than representing a series of out-of-Africa dispersals, actually reflects a recent contraction of a previously widespread population. Using their new estimate of the evolutionary rate, the authors place the population contraction within the last 500 000 years.

Two methods are most often used to calibrate an evolutionary rate. When the data set includes serially sampled ancient DNA or fast-evolving viruses, the ages of the sequences can be used to apply calibrations to terminal nodes in the tree (Rambaut 2000; Fig. 1A). This 'tip calibration' allows the substitution rate to be estimated, provided that the rate is sufficiently high compared with the period of time over which the samples were isolated (Drummond *et al.* 2003). A second method, often referred to as 'fossil calibration', is to specify calibrations for internal nodes in the tree, based on palaeontological evidence (Fig. 1B). The latter method was used by Rohland *et al.* (2005), who timed of the first appearance of the two distinct hyaena lineages using the age of the Miocene fossils that had been ascribed to those lineages.

Fossil- and tip-calibration approaches can lead to very different estimates of the molecular rate, even for the same data set (Ho *et al.* 2008). These disparities are characteristic of time-dependent rates, whereby higher estimates of evolutionary rates are obtained when the calibrations span a shorter period of time (Ho *et al.* 2011a). However, both calibration methods can be appropriate, depending on the data set being analysed. When inferring divergence times between species, fossil-calibrated rates may be more appropriate because much of the information informing these rates comes from the genetic changes that have become fixed between lineages. In contrast, the faster, tip-calibrated rates are assumed to be more accurate for populations with short histories because they include mutations that are still drifting to fixation (or loss).

In their study of cave hyaenas, Sheng *et al.* (2014) compared the effects of using tip calibrations and fossil calibrations on their estimates of the evolutionary timescale. Their data represent an important special case that is often encountered in molecular ecological analyses. Although estimating evolutionary timescales at population (within-species) and phylogenetic (between-species) scales are often treated as separate issues (Arbogast *et al.* 2002), some data sets straddle these two timescales. In the case of the hyaenas, the data are sampled from a population with a

Correspondence: Beth Shapiro, Fax: +1 831 459 3383;
E-mail: bashapir@ucsc.edu

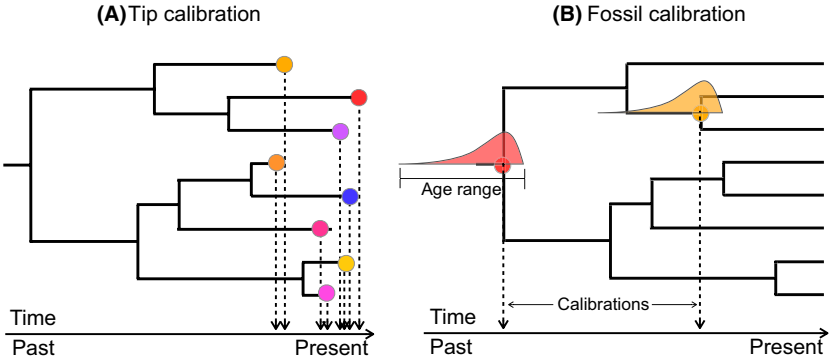

Fig. 1 The 'tick' rate of the molecular clock can be calibrated by fixing the age of nodes in the tree. In (A), the age of each DNA sequence can be fixed and used as tip calibrations to estimate the evolutionary rate. In (B), information from the fossil record can be used to either fix (circles) or model (density plots) the ages of nodes in the tree. In both approaches, the ages of the remaining nodes in the tree are free to vary.

very deep evolutionary history compared to the oldest sample, and consequently, neither fossil calibration nor tip calibration is ideal.

Identifying and justifying a particular approach to calibrate the evolutionary rate for such data sets remains a significant challenge. For data sets sampled from populations with deep evolutionary histories, or from closely related subspecies, fossils containing diagnostic features capable of confidently differentiating lineages within these data sets are rare. Including older fossil calibrations often requires the inclusion of (distant) outgroup taxa, which would introduce potential problems such as rate variation among species. Alternatively, tip calibrations might yield rate estimates that are too high, in particular if the sampling interval is small relative to the overall depth of the population's evolutionary history.

Sheng *et al.* (2014) presented several interesting and novel approaches to evaluate and select the most appropriate calibration for their evolutionary timescale. Because tip calibrations are often more readily available than are diagnostic fossils, a common approach is to assume that the tip-calibrated estimate of the rate is reasonably reliable. This assumption can be tested by investigating the temporal structure of the sample ages. Sheng *et al.* (2014) used a date-randomization test to demonstrate that the radiocarbon-dated sequences in their tree contained sufficient temporal signal to estimate the evolutionary rate (Ramsden *et al.* 2009; Ho *et al.* 2011b). In this test, the mean of the original rate estimate should not be contained within the 95% credibility intervals of any of the date-randomized replicates (this is the case for the hyaena data), although the results are more convincing if the credibility intervals do not overlap. In addition, the rate estimated from the original data set should be reasonably precise, without a tail that extends towards zero. Importantly, the date-randomization test can identify data sets that have insufficient calibrating information, but passing the test does not ensure that a data set will yield reliable estimates of rates and dates.

To investigate their choice of calibration, Sheng *et al.* (2014) compared the use of tip calibrations, fossil calibrations and a combination of the two. They also performed a series of simulation experiments to model the pattern of decay between the fossil-calibrated (slow) rates near the root of the tree and the tip-calibrated (fast) rates along the terminal branches of the tree. Importantly, Sheng *et al.* (2014) were able to take advantage of the hyaena fossil record in this test, asking explicitly which model best explained both the genetic and palaeontological data. The best-fitting model was one that used both fossil and tip calibrations, and assumed a linear decay in the evolutionary rate.

Modelling the time-dependent rate variation among lineages using a relaxed molecular clock (e.g. Korsten *et al.* 2009; Sheng *et al.* 2014) is likely to be widely useful in molecular ecological studies, as it takes advantage of all of the available data. Unlike the hyaenas, however, most populations and species lack a sufficiently robust fossil record to allow a detailed comparison between tip- and fossil-calibrated estimates of rates and timescales. In addition, alignments of ancient DNA sequence data rarely have sufficient information to support a relaxed-clock model, in particular when these alignments include only partial sequences of mitochondrial loci. Such an approach might become increasingly feasible, however, with the growing application of next-generation sequencing to studies of ancient DNA.

Rate estimation is an important aspect of molecular evolutionary studies, but it remains very difficult to determine whether we have done it well. Studies based on simulated data have highlighted various sources of error (Navascués & Emerson 2009), but the impact of these factors on real data is poorly understood. This therefore remains an open area of research.

References

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707–740.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology and Evolution*, **18**, 481–488.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One*, **3**, e1615.

Ho SYW, Lanfear R, Phillips MJ *et al.* (2011a) Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology*, **60**, 366–375.

Ho SYW, Lanfear R, Bromham L *et al.* (2011b) Time-dependent rates of molecular evolution. *Molecular Ecology*, **20**, 3087–3101.

Korsten M, Ho SYW, Davison J *et al*. (2009) Sudden expansion of a single brown bear lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Molecular Ecology*, **18**, 1963–1979.

Navascués M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Molecular Ecology*, **18**, 4390–4397.

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

Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Molecular Biology and Evolution*, **26**, 143–153.

Rohland N, Pollack JL, Nagel D *et al*. (2005) The population history of extant and extinct hyenas. *Molecular Biology and Evolution*, **22**, 2435–2443.

Sheng G-L, Soubrier J, Liu J-Y *et al*. (2014) Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, *Crocuta crocuta*. *Molecular Ecology*, **23**, 522–533.

The authors contributed equally to this work.

---

doi: 10.1111/mec.12621