

NEWS AND VIEWS

PERSPECTIVE

Do complex population histories drive higher estimates of substitution rate in phylogenetic reconstructions?

UMA RAMAKRISHNAN* and ELIZABETH A. HADLY†

*National Centre for Biological Sciences, TIFR, GKVK Campus, Bangalore 560065, India, †Department of Biology, Stanford University, Stanford, CA 94305-5020, USA

Our curiosity about biodiversity compels us to reconstruct the evolutionary past of species. Molecular evolutionary theory now allows parameterization of mathematically sophisticated and detailed models of DNA evolution, which have resulted in a wealth of phylogenetic histories. But reconstructing how species and population histories have played out is critically dependent on the assumptions we make, such as the clock-like accumulation of genetic differences over time and the rate of accumulation of such differences. An important stumbling block in the reconstruction of evolutionary history has been the discordance in estimates of substitution rate between phylogenetic and pedigree-based studies. Ancient genetic data recovered directly from the past are intermediate in time scale between phylogenetics-based and pedigree-based calibrations of substitution rate. Recent analyses of such ancient genetic data suggest that substitution rates are closer to the higher, pedigree-based estimates. In this issue, Navascués & Emerson (2009) model genetic data from contemporary and ancient populations that deviate from a simple demographic history (including changes in population size and structure) using serial coalescent simulations. Furthermore, they show that when these data are used for calibration, we are likely to arrive at upwardly biased estimates of mutation rate.

Keywords: Ancient genetic data, molecular evolution, mutation rate, serial coalescent

Received 1 July 2009; revision received 20 July 2009; accepted 23 July 2009

Estimates of mutation rate based on fossil calibrations of known genetic divergences revealed rates on the order of one substitution per base pair per million years. On the other hand, pedigree-based estimates have determined a mutation rate an order of magnitude higher. Ho *et al.*

Correspondence: Uma Ramakrishnan, Fax: 91-80-23636662; E-mail: uramakri@ncbs.res.in

(2005, 2007a,b) used multiple datasets and a relaxed clock-based Bayesian phylogenetic analysis to hypothesize that mutation rate changes in a time-dependent fashion: older calibrations result in lower mutation rate estimates and more recent calibrations result in higher mutation rate estimates. Why would such differences exist? Ho *et al.* (2005) suggest various statistical (e.g. incorrect or inadequate parameterization of models for DNA evolution), technical (e.g. sequencing errors), molecular (e.g. mutational hot-spots) and evolutionary (e.g. purifying selection) possibilities and they rule out the former three. The purifying selection explanation suggests that if mutations were slightly deleterious, they would be relatively short-lived in the population. This would result in a low fixation rate and hence a low substitution rate, but would be consistent with a high instantaneous mutation rate. Woodhams (2006) showed that for slightly deleterious mutations to explain discrepancy between mutation rate and substitution rate requires very high effective sizes (10^8 for birds and 10^5 for primates). Furthermore, purifying selection does little to account for perceived differences in rates in relatively neutral regions of the genome.

Use of sequence data from ancient samples was hypothesized to provide an excellent opportunity to explore this conundrum, as ancient DNA datasets include samples from multiple time points separated by hundreds to thousands of years. Estimates of mutation rate using Bayesian analyses of ancient genetic data from Adélie penguins (Lambert *et al.* (2002) were comparable to the higher, pedigree-based estimates. Re-analyses of several ancient DNA datasets by Ho *et al.* (2007a) using Bayesian Markov chain Monte Carlo (MCMC) methods also revealed relatively high substitution rate estimates from ancient samples, comparable to pedigree-based estimates. In fact, reanalysis by Ho *et al.* (2007b) of ancient genetic data of bison revealed decreasing estimates of substitution rate within the same dataset when comparing older samples with younger samples. Thus, ancient genetic data further complicated the problem.

However, most models considered to date assumed that the samples from a species were from a single population, often of constant size, and they attempted to explain discrepancies with variable mutation rates. But here is where the genetics of populations are very different than the genetics of species. The accumulation of mutations in populations relies not just on mutation rate and selection on those mutations, but also on the size of the population (drift) as well as its connectivity with other populations (gene flow). Navascués & Emerson (2009) attempt to account for these population-based variables using an elegant assembly of modelled populations. They compiled a collection of modelled populations with population

histories that included variable population structure and changes in population size. They then used the serial coalescent to generate genealogies that correspond to different nonideal population histories. On the basis of an assumed substitution rate and model of DNA evolution, they added mutations to these genealogies to generate genetic data for these demographic scenarios (such as in Fig. 1). And finally, they estimated mutation rates using the Bayesian MCMC approach implemented in BEAST. They found that population histories that include variable population structure and changes in population size resulted in higher estimates of substitution rate. While use of more complex models presents methodological and analytical challenges, these results demonstrate how critical it is to consider the structure and dynamics of populations in reconstructing their histories.

Navascués & Emerson (2009) provide a significant step towards unravelling the potentially complex impacts of population history. Their results are especially significant because the estimates of substitution rate from ancient DNA studies by Ho *et al.* (2007b) include many datasets where samples are drawn from a large geographical area and could hence be affected by population structure and changes in population size. In fact, some ancient DNA studies focused on reconstruction of population history suggest changes in population size (e.g. Valdiosera *et al.* 2008), whereas others hypothesize the presence or absence of gene flow in past populations (e.g. Hadly *et al.* 2004; Hofreiter *et al.* 2004). Other studies reveal replacement of ancient populations by modern ones (Belle *et al.* 2006). These and other studies using ancient genetic data suggest that there are many cryptic events in population histories that we may not detect using more conventional analyses

(Ramakrishnan & Hadly 2009). Thus, Navascués & Emerson (2009) clearly establish the importance of population size and migration between populations, as these processes can overwhelm the role of mutation rate in the generation of genetic diversity through time.

How then do we begin to face the challenge of integrating more than simplistic demographic histories with mutation rate in an estimation framework? One option, as Navascués & Emerson (2009) suggest, is to use summary-statistic approaches. Modified rejection algorithm approaches (Tishkoff *et al.* 2007) and model-testing approaches (Ramakrishnan & Hadly 2009) are other alternatives. Alternatively, the Bayesian MCMC facilitates the exploration of the sensitivity of estimates to population structure by analysing subsets of data, say from different geographical regions or different periods of time (Chan *et al.* 2006).

Our understanding of the ecology of populations indicates that plants and animals are keyed to resource availability and competition throughout their range, suggesting that population sizes and gene flow between populations will vary as we move across a species' geographical range. As environments and species themselves change through time, we can expect population size and gene flow to vary temporally as well. In fact, palaeontological and palaeoclimatic data attest to prehistoric changes in species distributions and population abundance patterns in response to climate (e.g. Grayson 1993; Hadly 1996; Barnosky 2004). Within species, patterns of genetic variation within and between populations are a result of interactions between micro-evolutionary processes (mutation, migration, selection and drift), all except mutation rate probably influenced by the environment. Reconstructing the interplay between

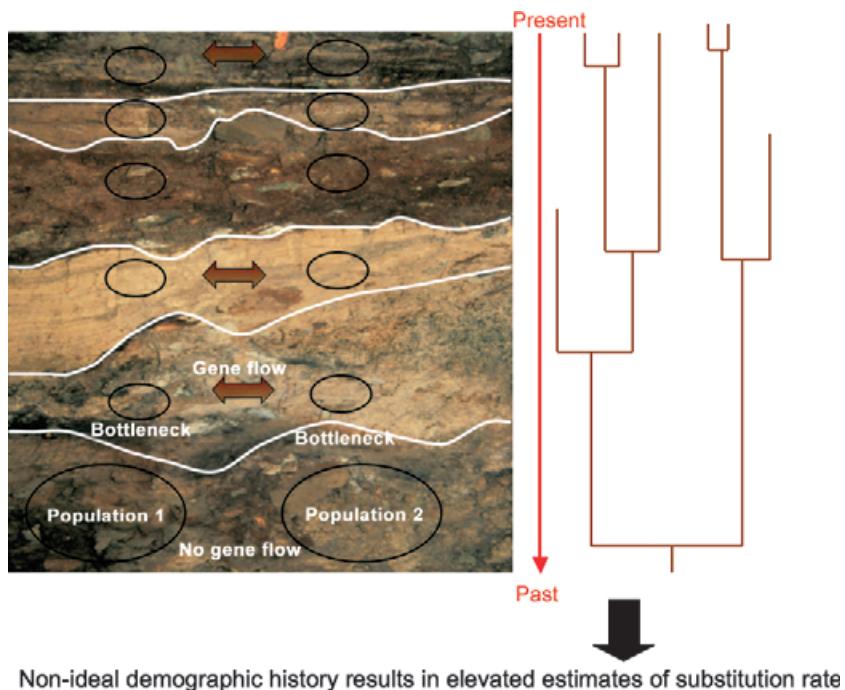


Fig. 1 Ancient genetic material preserved in a deposit (stratigraphy and possible population history shown in the left panel) maybe coincident with nonideal demographic histories, in this case two populations with changing levels of gene flow. Bayesian estimates from such data using the serial coalescent (right) might result in upwardly biased estimates of substitution rate.

these factors is a challenge that population genetics faces in the coming decade as each process has sculpted the genetic diversity of populations and species. Given the rapid progress in genome technology, it is possible that we will soon be generating ancient population genomic datasets, providing more statistical power to discriminate multi-event population histories, disentangled from variation in mutational processes. And as we step closer to reconstructing the histories of populations, we will better understand the evolution and fate of species on Earth.

References

Barnosky AD (2004) *Biodiversity Response to Climate Change in the Middle Pleistocene: The Porcupine Cave Fauna from Colorado*. University of California Press, Berkeley, CA.

Belle E, Ramakrishnan U, Mountain JL, Barbujani G (2006) Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proceedings of the National Academy of Sciences, USA*, **103**, 8012–8017.

Chan YL, Anderson CNK, Hadly EA (2006) Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *Public Library of Science, Genetics*, **2**, e59.

Grayson DK (1993) *The Desert's Past: A Natural Prehistory of the Great Basin*. Smithsonian Institution Press, Washington, DC.

Hadly EA (1996) Influence of Late Holocene Climate on Northern Rocky Mountain Mammals. *Quaternary Research*, **46**, 298–310.

Hadly EA, Ramakrishnan U, Chan YL *et al.* (2004) Genetic response to climatic change: insights from ancient DNA and phylochronology. *Public Library of Science, Biology*, **2**, e29.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Ho SYW, Kolokotronis S-O, Allaby RG (2007a) Elevated substitution rates estimated from ancient DNA sequences. *Biology Letters*, **3**, 202–205.

Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond A (2007b) Evidence for time dependency of molecular rate estimates. *Systematic Biology*, **56**, 515–522.

Hofreiter M, Rabeder G, Jaenicke-Despres V *et al.* (2004) Evidence for reproductive isolation between cave bear populations. *Current Biology*, **14**, 40–43.

Lambert DM, Ritchie PA, Millar CD *et al.* (2002) Rates of evolution in ancient DNA from Adélie penguins. *Science*, **295**, 2270–2273.

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

Ramakrishnan U, Hadly EA (2009) Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology*, **18**, 1310–1330.

Tishkoff SA, Gonder MK, Henn BM *et al.* (2007) History of click-speaking populations of Africa inferred from mtDNA and Y chromosome variation. *Molecular Biology and Evolution*, **24**, 2180–2195.

Valdiosera CE, Garcia-Garitagoitia JL, Garcia N *et al.* (2008) Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proceedings of National Academy of Sciences, USA*, **105**, 5123–5128.

Woodhams M (2006) Can deleterious mutations explain the time dependency of molecular rate estimates? *Molecular Biology and Evolution*, **23**, 2271–2273.

doi: 10.1111/j.1365-294X.2009.04334.x