

## NEWS AND COMMENTARY

Apparent recent elevation of mutation rates

# Don't forget the ancestral polymorphisms

D Charlesworth

Heredity (2010) 105, 509–510; doi:10.1038/hdy.2010.14; published online 17 February 2010

Using estimates of sequence divergence, several taxa have been found to have sharply increased molecular evolutionary rates between the most recently diverged species studied (reviewed in Peterson and Masel, 2009). This 'acceleration of the molecular clock at short timescales' has been considered puzzling, and has been attributed to changes in mutation rates. However, it has also been suggested that this is only an apparent acceleration caused by the segregation of lineages present in the ancestral population of recently diverged species. The common ancestor of two alleles at a locus existed before two species became isolated, and so divergence times between alleles in two species exceed the species divergence times (Gillespie and Langley, 1979), which 'would exaggerate the genetic difference between the populations' (Nei, 1971). For recently diverged species, this can add substantially to the numbers of sites with sequence differences (Figure 1a). Peterson and Masel have now studied this possibility thoroughly, and show that this effect can explain the observations.

It is important to define clearly what is measured when sequence divergence is estimated between species. Under the neutral theory of molecular evolution, substitutions accumulate with evolutionary time. After two populations become isolated, fixed differences accumulate between them in proportion to the time since isolation. Limited samples cannot distinguish fixed differences between species from sites with polymorphisms in one or both species, but divergence estimates often use just a single allele per species, and such 'raw' divergence estimates include contributions from within-species polymorphisms: '...the mean number of site differences between two [alleles], one from each population, is equal to the mean number of differences in the ancestral population plus  $2vt$ , the amount of differentiation after separation' (Li, 1977). If sequences are available from multiple alleles, a rough

correction for polymorphisms can be made by subtracting nucleotide diversity estimates from the estimated divergence to obtain a 'net divergence' (Nei, 1987); this correction assumes that the ancestral population's diversity was similar to that in the populations studied.

If raw divergence values are used to estimate mutation rates over a known time from substitution rates at putatively neutral sites, this will clearly overestimate the rates for short times, making close relatives unsuited for such rate estimates (for example, Makova and Li, 2002). There is no major problem for species that diverged long ago, because within-species diversity values are commonly only around 0.01, making the correction for ancestral polymorphism minor for the kinds of divergence values that are usually involved when fossil datings are used in estimates of mutation rates (Figure 1b); these calibrations are rough and usually involve large times. The overestimation will be greatest for neutral sites and less for deleterious mutations, which rarely rise to high frequencies within species and contribute little diversity, so that counting observed inter-species differences as fixations, ignoring polymorphism, is less inaccurate. However, one cannot use amino-acid or non-synonymous differences to estimate mutation rates, because some of these substitutions will have been driven by positive selection for new variants.

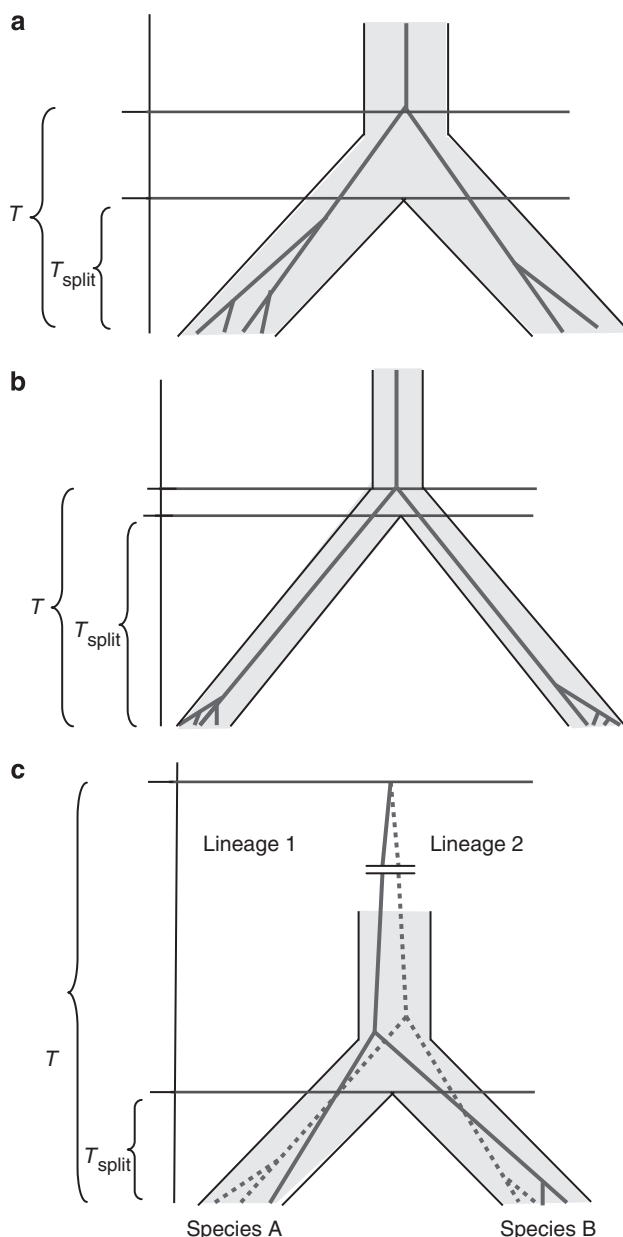
Although the effect of ancestral polymorphisms is well known, it was not clear whether it could account quantitatively for the mutation rate observations. A recent paper shows that it can, and that no actual acceleration in the molecular evolutionary rate at short timescales need be invoked (Peterson and Masel, 2009). In the models simulated, the estimated numbers of neutral differences accumulated between sequences sampled from two simulated species increased sharply as the number of generations since they split decreased, converging to the true long-

term rate after a number of generations equal to 5–10 times the population sizes. The model was capable of fitting the observations from two studies. One from cichlid fishes fitted well, using plausible effective population sizes. Results from human data could also, in principle, be fitted, although the simulations assumed that the sites were all unlinked, whereas the data were from the mitochondrial genome.

Disadvantageous mutations are rarely fixed, but are nevertheless found as polymorphisms within species because they are not instantly eliminated. The ratio ( $K_a/K_s$ ) of non-synonymous (mostly disadvantageous) differences between species to synonymous differences (probably often close to neutral) should therefore also increase as divergence times get smaller (approaching the ratio for variants within species). The model shows, however, that neutral mutations probably cause most of the observed apparent increase in substitution rates.

These conclusions apply to the simplest divergence rate estimates (using a time-calibrated two-species split), but, clearly, the effects on substitution rates estimated for different branches in a tree should be studied—it seems unlikely that ignoring ancestral polymorphism will have no effects (Peterson and Masel, 2009). Highly elevated substitution (and mutation) rate estimates might thus arise in cases wherein balancing selection maintains variants within a species, perhaps explaining the estimated highly elevated mutation rates in the mitochondrial genomes of some plant taxa (Mower *et al.*, 2007). Some of these taxa, including *Silene* (Sloan *et al.*, 2009) and *Plantago*, have many gynodioecious species, with female individuals (male-steriles, owing to mitochondrial genome mutations), as well as hermaphrodites (Damme, 1983; Charlesworth and Laporte, 1998). Mitochondrial sequence diversity is indeed high in some gynodioecious *Silene* species (for example, Touzet and Delph, 2009). If different mitochondrial haplotypes have been maintained since before the split of the species studied in these genera, and have diverged because of infrequent recombination, then, if different lineages are sampled for sequencing, this could potentially lead to apparently accelerated substitutions (Figure 1c).

A further possible cause of apparent acceleration is recombination. This extends back in time the nodes between branches of a tree estimated, assuming



**Figure 1** Panel a of the figure shows that the common ancestor of two alleles at a locus, one from each of two recently diverged species (a and b), existed a considerable time (depending on the effective population size in the ancestral species) before the two species became isolated and so divergence times between alleles in two species ( $T$ ) exceed the species divergence times ( $T_{\text{split}}$ ). Sequence divergence values will scale with divergence times, under neutrality. Panel b illustrates the relatively lesser effect on divergence values if species A and B diverged longer ago (with  $T$  increased by a smaller fraction of the  $T_{\text{split}}$  value). (c) Illustrates the potential large effect of long-term balancing selection acting to maintain alleles at the locus under study. Here, two different lineages evolved before the species split, and have become diverged in sequence (as will occur if they rarely or never recombine). If sequences of alleles of such diverged lineages are compared in two species, the divergence time would appear to be much greater than  $T_{\text{split}}$  (whereas two alleles of the same lineage would more correctly reflect  $T_{\text{split}}$ ). If an analysis uses estimated species divergence times (or a phylogeny of the species constraining the divergence to have occurred after the studied species split), the substitution rates could potentially be overestimated.

that the sequences do not recombine (Schierup and Hein, 2000). If closely related taxa are analysed, recombination

in the ancestral population might have substantial effects. Plant mitochondrial genomes do recombine (McCauley and

Ellis, 2008), and so this may also contribute. These potential effects need to be better understood, so that reliable conclusions can be drawn from analyses of closely related taxa, which are becoming more common, including the use of recombining variants.

## Conflict of interest

The author declares no conflict of interest.

Dr D Charlesworth is at the Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, ICAPB Ashworth Laboratories, The King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK.

e-mail: Deborah.Charlesworth@ed.ac.uk

Charlesworth D, Laporte V (1998). The male-sterility polymorphism of *Silene vulgaris*. Analysis of genetic data from two populations, and comparison with *Thymus vulgaris*. *Genetics* **150**: 1267–1282.

Damme JMMv (1983). Gynodioecy in *Plantago lanceolata* L. II. Inheritance of three male sterility types. *Heredity* **50**: 253–273.

Gillespie JH, Langley CH (1979). Are evolutionary rates really variable? *J Mol Evol* **13**: 27–34.

Li W (1977). Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* **85**: 331–337.

Makova KD, Li WH (2002). Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**: 624–626.

McCauley DE, Ellis JR (2008). Recombination and linkage disequilibrium among mitochondrial genes in structured populations of the gynodioecious plant *Silene vulgaris*. *Evolution* **62**: 823–832.

Mower J, Touzet P, Gummow J, Delph L, Palmer J (2007). Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol Biol* **7**: 135.

Nei M (1971). Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. *Amer Nat* **105**: 385–398.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.

Peterson GL, Masel J (2009). Quantitative prediction of molecular clock and Ka/Ks at short timescales. *Mol Biol Evol* **26**: 2595–2603.

Schierup MH, Hein J (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**: 879–891.

Sloan D, Oxelman B, Rautenberg A, Taylor D (2009). Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. *BMC Evol Biol* **9**: 260.

Touzet P, Delph L (2009). The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics* **181**: 631–644.

## Editor's suggested reading

Noor MAF, Bennett SM (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**: 439–444.

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