PROCEEDINGS B

rspb.royalsocietypublishing.org

Comment

Cite this article: Sağlam İK, Baumsteiger J, Miller MR. 2017 Failure to differentiate between divergence of species and their genes can result in over-estimation of mutation rates in recently diverged species. *Proc. R. Soc. B* **284**: 20170021.
http://dx.doi.org/10.1098/rspb.2017.0021

Received: 4 January 2017
Accepted: 14 February 2017

Subject Category:
Evolution

Authors for correspondence:
İsmail K. Sağlam
e-mail: iksaglam@ucdavis.edu
Michael R. Miller
e-mail: micmiller@ucdavis.edu

The accompanying reply can be viewed at http://dx.doi.org/10.1098/rspb.2017.0537.

THE ROYAL SOCIETY PUBLISHING

Failure to differentiate between divergence of species and their genes can result in over-estimation of mutation rates in recently diverged species

İsmail K. Sağlam[1,3], Jason Baumsteiger[1,2] and Michael R. Miller[1,2]

[1]Department of Animal Science, and [2]Center for Watershed Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA
[3]Ecological Sciences Research Laboratories, Department of Biology, Hacettepe University, Beytepe, 06800, Ankara, Turkey

İKS, 0000-0003-3136-7334; MRM, 0000-0001-5874-8693

Devil's Hole pupfish (*Cyprinodon diabolis*, DHP) is an icon of population persistence due to having survived in complete isolation in a small pool in the Mojave Desert (Devil's Hole, southwestern USA) for thousands of years [1]. Although the exact time and mode of colonization is unclear, DHP are generally assumed to have been isolated for 10–40 kyr [2]. However, in a recent paper, Martin *et al.* [3] use an analysis with over 13 000 genomic loci to seriously challenge this notion. Based on demographic modelling with a genomic substitution rate of 5.37×10^{-7} mutations per site per year (m/s/y) (obtained from a phylogenetic analysis of Cyprinodontidae), they estimated the age of DHP to be 0.105–0.830 ka, and argue evolutionary time scales in DHP and other pupfish species in the region have been overestimated [3].

The recent divergence of DHP according to Martin *et al.* [3] can be linked to the extremely high genomic mutation rate used in their demographic analysis. Although some teleost fish genomes are believed to evolve slightly faster than the typical vertebrate rate of 1×10^{-8} m/s/g [4,5], there is no precedent for a rate of 1.79×10^{-7}–5.37×10^{-7} m/s/g, depending on the assumed average generation interval. Martin *et al.* [3] defend this rate because it was estimated using the only well-defined internal calibration event known for Cyprinodon: the 8000 ± 200 year age of the Laguna Chichancanab [6]. This date was used to put a lower bound on the divergence between species within Laguna Chichancanab and the coastal *C. artifrons*, probably conspecific with the ancestral species [3]. Although this date is well supported, Martin *et al.* [3] fail to account for the bias associated with determining mutation rates when conflating species divergence times with time since the most recent common ancestor (TMRCA) of sampled genes.

The phylogenetic tree presented by Martin *et al.* [3] reconstructs relationships using a single haplotype (16 567 concatenated 100 bp RAD-loci) per population, and therefore represents a haplotype (gene) tree and not a species tree. Although Martin *et al.* [3] acknowledge the potential of their phylogeny to return biased results due to incomplete lineage sorting (which could produce topological incongruences), they nevertheless do not distinguish between species and gene divergence times when calibrating the tree, which could cause temporal incongruences and have severe consequences for mutation rate estimates.

Figure 1*a* illustrates the distinction between a species divergence time (t) and a gene divergence time (T: TMRCA of haplotypes sampled from two descendant species). Population genetic theory predicts the TMRCA of a random pair of haplotypes in the ancestral population will be, on average, 2Ne generations (where Ne is the coalescent effective size of the ancestral population). Hence, gene divergence (T) will be 2Ne generations greater than species divergence (t) on average [7]. Since the phylogeny in Martin *et al.* [3] represents a haplotype tree, the calibration node depicting the divergence between coastal

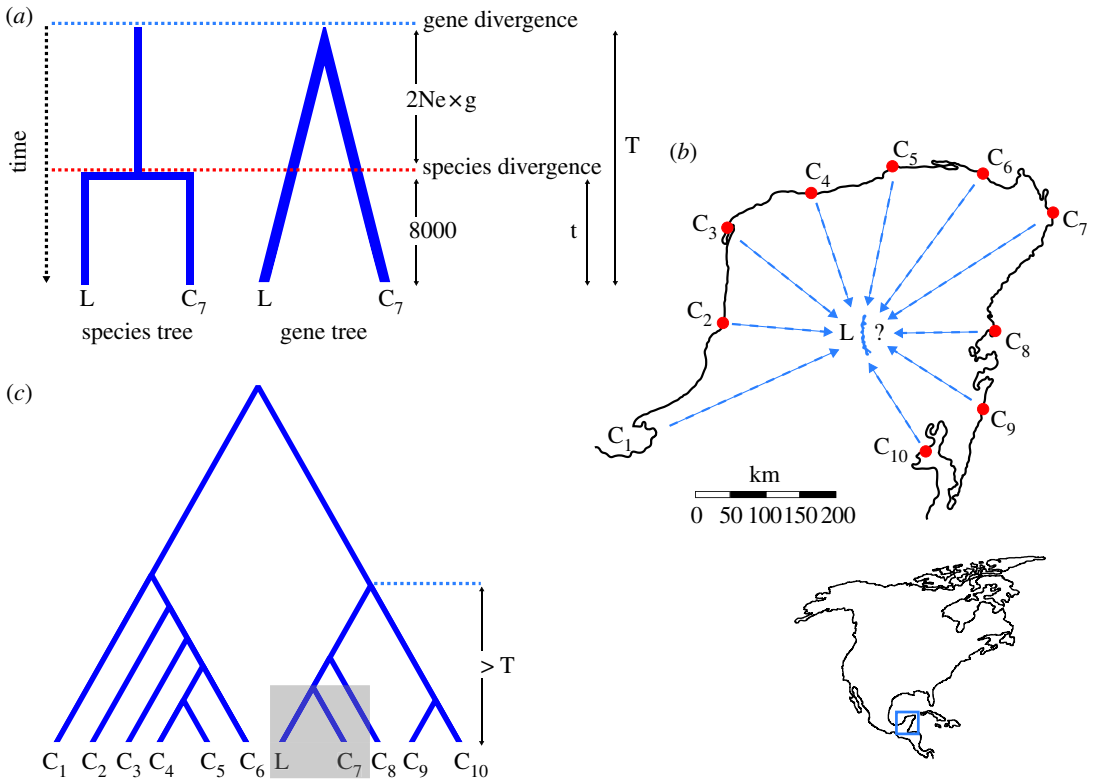

Figure 1. Differences in divergence times of Yucatan Peninsula pupfish and their genes. (*a*) Schematic diagram representing the difference between gene (triangular trees) and species/population divergence times (rectangular trees) for hypothetical coastal (C) and lake (L) populations. (*b*) Hypothetical locations of *C. artifrons* populations around the Yucatan Peninsula. The ancestral origin of Laguna Chichancanab taxa are unknown. (*c*) Hypothetical scenario where C_7 is the true ancestral population but C_9 or C_{10} were sampled for phylogenetic analysis. Laguna Chichancanab is shown for representative purposes and not drawn to scale. (Online version in colour.)

and inland species reflects gene divergence and not species divergence. Thus, the appropriate calibration for this node is not t (i.e. 8000) but T (8000 + 2Ne × g, g = generation interval in years). This difference is often ignored when species have been separated for long periods of time as the difference between T and t is proportionally less [7]. For example, if species divergence occurred 10 million years ago and ancestral Ne was 50 000, the difference between gene and species divergence would be minor (10.1 million versus 10 million, when g = 1). However, when divergence times between species are small, the difference between t and T can be substantial, especially if ancestral Ne valves are large.

The ancestral species in Martin *et al.* [3], *C. artifrons*, is broadly distributed throughout coastal/brackish areas around the Yucatan Peninsula (figure 1*b*). Although its exact coalescent Ne is unknown, similar species have coalescent Ne on the order of tens to hundreds of thousands [8]. If we conservatively assume the ancestral population had a coalescent Ne of 50 000, expected T would be 108 000 (assuming g = 1), even though t is only 8000. Thus, unless coalescent Ne of ancestral *C. artifrons* was very small, the calibration information used by Martin *et al.* [3] represents a substantial underestimation of T and hence a substantial overestimation of mutation rates.

A scenario where the ancestral population could have a low coalescent Ne would be if *C. artifrons* exhibits substantial population structure. In this case, specific locations along the coast (small brackish inlets for example) would contain a relatively small number of individuals isolated from other areas. However, in structured species, the TMRCA between random haplotype copies sampled from two distinct populations will depend on how closely related the populations are, and as the number of populations increases, so will the variance in coalescent times [7] (figure 1*c*). Under this scenario, phylogenetic analyses must use *C. artifrons* samples that originated from the correct ancestral population because the branch length describing divergence between a random *C. artifrons* haplotype and a random inland haplotype will equal T (8000 + 2Ne × g) only if haplotypes of *C. artifrons* came from the true ancestral population that colonized Laguna Chichancanab (C_7, figure 1*c*). If the sampled haplotype belonged to a distantly related population (C_9 or C_{10}, figure 1*c*), the branch length would be greater than T (8000 + 2Ne × g), resulting in a severe and unpredictable overestimation of mutation rates. Unfortunately, since the route of colonization and original location of founding individuals is unknown, it is difficult to determine the true level of error caused by calibrating T at 8000 in Martin *et al.* [3]. To do this would require information from multiple coastal *C. artifrons* individuals to determine overall phylogenetic structure and the true sister population of the inland species group.

The issues we raise here are hardly new, as the errors of calibrating gene trees using species level information was discussed as recently as 2011, by McCormack *et al.* [9]. Put simply, without adherence to proper population genetic principles (detailed above), the extremely high genomic mutation rate estimated by Martin *et al.* [3] is likely to be a severe overestimate of the true mutation rate. Therefore, when this mutation rate was used in demographic modelling, the resulting divergence time is likely a serious underestimate. Given this issue, a reasonable assumption is that the mutation rate of DHP is similar to typical vertebrates (1×10^{-8} m/s/g).

In fact, the difference between the typical vertebrate mutation rate and the one estimated by Martin *et al.* [3] (approximately 20–50-fold) is consistent with the difference between species and gene divergence expected from a short divergence time and large ancestral Ne. If Martin *et al.* [3] had used the typical vertebrate mutation rate in their demographic analysis, their divergence estimate would have been at least an order of magnitude greater than presented in their paper. Therefore, the estimate by Martin *et al.* [3] that DHP is as young as 0.105–0.830 ka and the conclusion of 'a surprisingly rapid timescale for speciation, genetic assimilation and the evolution of intrinsic reproductive incompatibilities in this group' should be considered with utmost caution.

Data accessibility. This article has no additional data.
Competing interests. We declare we have no competing interests.
Funding. We received no funding for this study.

References

1. Reed JM, Stockwell CA. 2014 Evaluating an icon of population persistence: the Devil's Hole pupfish. *Proc. R. Soc. B* **281**, 20141648. (doi:10.1098/rspb.2014.1648)
2. Echelle AA. 2008 The western North American pupfish clade (Cyprinodontidae: *Cyprinodon*): mitochondrial DNA divergence and drainage history. *Geol. Soc. Am. Spec. Pap.* **439**, 27–38. (doi:10.1130/2008.2439(02))
3. Martin CH, Crawford JE, Turner BJ, Simons LH. 2016 Diabolical survival in Death Valley: recent pupfish colonization, gene flow and genetic assimilation in the smallest species range on earth. *Proc. R. Soc. B* **283**, 20152334. (doi:10.1098/rspb.2015.2334)
4. Lynch M. 2010 Evolution of the mutation rate. *Trends Genet.* **26**, 345–352. (doi:10.1016/j.tig.2010.05.003)
5. Jaillon O *et al.* 2004 Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957. (doi:10.1038/nature03025)
6. Hodell DA, Curtis JH, Brenner M. 1995 Possible role of climate in the collapse of Classic Maya civilization. *Nature* **375**, 391–394. (doi:10.1038/375391a0)
7. Edwards SV, Beerli P. 2000 Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839–1854. (doi:10.1111/j.0014-3820.2000.tb01231.x)
8. Duvernell DD, Lindmeier JB, Faust KE, Whitehead A. 2008 Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*. *Mol. Ecol.* **17**, 1344–1360. (doi:10.1111/j.1365-294X.2007.03648.x)
9. McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL. 2011 Calibrating divergence times on species trees versus gene trees: implications for speciation history of Aphelocoma jays. *Evolution* **65**, 184–202. (doi:10.1111/j.1558-5646.2010.01097.x)