Purifying selection can reduce intraspecific mitochondrial gene variation to that of nuclear rRNA

Tshifhiwa G. Matumba[1], Jody Oliver[2], Nigel P. Barker[3], Christopher D. McQuaid[1], Peter R. Teske[2*]

[1]Department of Zoology and Entomology, Rhodes University, Grahamstown 6140, South Africa

[2]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, Auckland Park 2006, South Africa

[3]Department of Plant and Soil Sciences, University of Pretoria, Hatfield 0028, South Africa

*Correspondence: Peter Teske: pteske101@gmail.com

Abstract

Mitochondrial DNA (mtDNA) has long been used to date the divergence between species, and to explore the time when species' effective population sizes changed. The idea that mitochondrial DNA is useful for molecular dating rests on the premise that its evolution is neutral. This premise was questionable to begin with, and even though it has long been challenged, the evidence against clock-like evolution of mtDNA is usually ignored. Here, we present a particularly clear and simple example to illustrate the implications of violations of the assumption of selective neutrality. DNA sequences were generated for the mtDNA COI gene and the nuclear 28S rRNA of two closely related and widely distributed rocky shore snails whose geographical ranges are defined by different thermal preferences. To our knowledge, this is the first study to use nuclear rRNA sequence for studying species-level genealogies instead of phylogenetics, presumably because this marker is considered to be uninformative at this taxonomic level. Even though the COI gene evolves at least an order of magnitude faster, which was reflected in high inter-specific divergence, intraspecific genetic variation was similar for both markers. As a result, estimates of population expansion times based on mismatch distributions were completely different for the two markers. Assuming that 28S evolves effectively clock-like, these findings likely illustrate variation-reducing purifying selection in mtDNA at the species level, and an elevated divergence rate caused by divergent selection between the two species. Although these two selective forces together make mtDNA suitable as a DNA barcoding marker because they create a 'barcoding gap', estimates of demographic change can be expected to be highly unreliable. Our study

contributes to the growing evidence that the utility of mtDNA beyond DNA barcoding is limited.

Keywords: purifying selection, diversifying selection, mismatch distribution, mitochondrial DNA, molecular dating, nuclear DNA, population expansion

Introduction

Mitochondrial DNA (mtDNA) has long been a marker of choice for investigating concepts as diverse as estimating genetic diversity and effective population sizes, reconstructing species' evolutionary histories, exploring spatial genetic subdivisions, and identifying cryptic species. All these methods assume that mtDNA variation conforms to the neutral model of molecular evolution (Kimura 1983), but violations of this premise have long been recognised (Ballard and Kreitman 1995). Over the past decades, much evidence has accumulated that mtDNA can be strongly affected by selective sweeps and background selection (Blier et al. 2001; Ballard and Rand 2005; Meiklejohn et al. 2007; Stewart et al. 2008). As a result, the usefulness of the marker in assessing genetic diversity (Bazin et al. 2006) and exploring spatial genetic structure in continuously distributed populations (Teske et al. 2018) has been questioned, and corrections of the mitochondrial molecular clock that account for selection have been proposed (Soares et al. 2009).

The implications of reduced genetic diversity at the species or population levels due to purifying selection has so far received little attention. When mutations in mitochondrial genes occur at fewer sites than expected under the neutral model (Lawrie et al. 2013), molecular dating of historical demographic events by means of evolutionary rate estimates that are typically based on inter-specific divergence (Knowlton and Weigt 1998; Schubart et al. 1998) will result in considerable underestimates. This is particularly likely because divergence between species can be strongly affected by diversifying selection that is driven by different environmental conditions (Lamb et al. 2018; Sun et al. 2018), resulting in a faster accumulation of mutations characterising each species than is expected under the neutral model.

Here, we explore this issue using mitochondrial and nuclear DNA sequence data from two common southern African snails of the genus *Afrolittorina*. The finding that data from two

gene regions whose mutation rates are assumed to differ by at least an order of magnitude have similar levels of intraspecific variation challenges the usefulness of mitochondrial DNA sequences for studying historical demographic changes.

Materials and Methods

The two study species, *Afrolittorina africana* (Philippi, 1847) and *A. knysnaensis* (Philippi, 1847) are the most widespread and abundant littorinid snails on rocky shores along the southern African coast (McQuaid 1992). They were formerly believed to comprise a single species whose morphological appearance gradually changes along its range (Hughes 1979), but their species status is now confirmed using both molecular and morphological data (Reid and Williams 2004).

Specimens of *Afrolittorina* were collected from 34 localities across the species' distribution ranges along the southern African coastline (supplemental online material, Table S1). Neither species exhibits genetic structure throughout its range (Matumba 2013). Laboratory procedures are described in the supplemental online material, and the primers used for PCR amplification of the mtDNACOI gene and the nuclear 28S rRNA are listed in Table S2.

Genealogical relationships between COI haplotypes and 28S alleles were reconstructed using the median joining algorithm (Bandelt et al. 1999) in popArt v1.7 (Leigh and Bryant 2015). Sites containing gaps were excluded. Variable sites in the mtDNA sequences were explored for evidence of intra- and inter-specific selection using a McDonald-Kreitman test (McDonald and Kreitman 1991) in MKT (Egea et al. 2008). The invertebrate mitochondrial code was specified, and the Jukes-Cantor model (Jukes and Cantor 1969) was used to correct for divergence. To explore the effect of using interspecific evolutionary rates to estimate when intraspecific demographic changes occurred, we calculated population expansion time under the sudden-expansion model (Rogers and Harpending 1992) using Arlequin v3.5 (Excoffier and Lischer 2010). For each marker type, we selected the slowest and the fastest published marine gastropod rates from the literature. For COI, these were $0.5\%.my^{-1}$ (Malaquias and Reid 2009) and $2.6\%.my^{-1}$ (Williams and Reid 2004), respectively. For 28S, a slow rate of $0.01\%.my^{-1}$ (Malaquias and Reid 2009) and a fast rate of $0.05\%.my^{-1}$ (Williams and Reid 2004) were used.

Results

Species-specific genetic clusters were highly distinct (Fig. 1a), with a minimum number of 44 nucleotide differences between the two species' most closely related haplotypes. In contrast to the COI sequences, differentiation between 28S sequences (Fig. 1b) was an order of magnitude smaller (4 nucleotide differences).
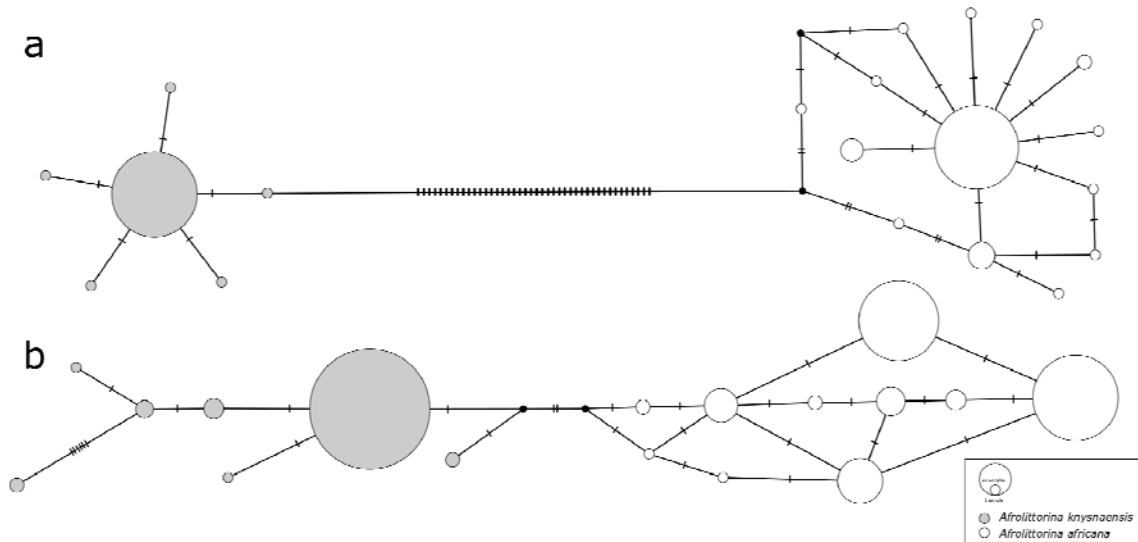

Fig. 1. Median-joining haplotype networks constructed from a) COI sequences of b) 28S rRNA sequences of *Afrolittorina knysnaensis* (grey) and *A. africana* (white). Low intra-specific variation and high inter-specific variation of COI potentially illustrate stabilising and divergent selection, respectively. The size of circles is proportional to the frequency of each haplotype, cross-bars represent nucleotide differences, and black dots are missing haplotypes not found in the samples.

In contrast to the high inter-specific differentiation between COI haplotypes, intra-specific genetic differentiation was comparatively low for this marker, and similar to that of 28S. In *A. knysnaensis*, six COI haplotypes and seven 28S haplotypes were found, while the maximum differentiation between the COI haplotypes was only two nucleotide differences, but 10 for 28S. The number of haplotypes was greater for *A. africana*, where 14 were found for COI and 10 for 28S. Maximum nucleotide differences for this species were seven in the COI network and five for 28S.

All five variable sites in *A. knysnaensis*, and 11 out of 12 variable sites in *A. africana*, were found at synonymous third character positions. The remaining site in *A. africana* was at a first character position but was also synonymous (CUG vs. UUG, which both code for Leu using the invertebrate mitochondrial code). For the complete dataset, which comprised 58 mutations of which the majority was found between species, 57 mutations were synonymous and one was non-synonymous. The McDonald-Kreitman test had the highest possible proportion of adaptive substitutions (α) of 1.0, but was nonetheless non-significant ($\chi^2 = 0.33$, P = 0.57), supposedly because of the lack of non-synonymous polymorphism.

The practical implications of two markers with very different evolutionary rates based on inter-specific divergence having similar levels of intraspecific variation are illustrated in Table 1. Using published rates, estimates of population expansion times were several orders of magnitude greater for the 28S data than for the COI data.

Table 1. Estimates of population expansions of the two species of *Afrolittorina* under the sudden expansion model. The moment estimator τ is equal to 2ut, where u equals 2 μk (μ is the mutation rate and k is the length of the sequence), and *t* is the time of expansion.

Species	τ	Marker	μ (%.my^{-1})	*t* (my)
Afrolittorina knysnaensis	2.00	COI	0.50[1]	0.40 (0.00 – 0.41)
			2.60[2]	0.07 (0.00 – 0.08)
	3.25	28S	0.01[1]	32.1 (18.5 – 61.3)
			0.05[2]	6.41 (3.69 – 12.3)
Afrolittorina africana	2.50	COI	0.50[1]	0.50 (0.30 – 0.79)
			2.60[2]	0.10 (0.06 – 0.15)
	2.75	28S	0.01[1]	27.1 (19.1 – 51.5)
			0.05[2]	5.42 (3.81 – 10.3)

[1]Malaquias & Reid 2009; [2]Williams & Reid 2004. A generation time of one year was used.

Discussion

The usefulness of the mtDNA COI gene to uncover overlooked biodiversity is undisputed because of the marker's tendency to have a well-defined barcoding gap (as was also found

here). Many researchers explore their mtDNA data for additional information, but variation-reducing selection at the population or species levels, and adaptive selection at the inter-species level that together create the barcoding gap (Stoeckle and Thaler 2014) make the usefulness of mtDNA for other applications questionable (Bazin et al. 2006; Teske et al. 2018). The finding that only synonymous mutations were found at the intra-specific level in *Afrolittorina*, whereas inter-specific mutations included a non-synonymous mutation, is by no means strong evidence for intra-specific neutrality and inter-specific selection. Recent evidence shows that selection also affects synonymous mutations, and evidence for purifying selection at the species level typically comes from a smaller number of synonymous mutations than expected under the neutral model, rather than the presence of non-synonymous mutations (Lawrie et al. 2013). Of the different gene regions that make up the mitochondrial genome, COI and COII are under particularly strong purifying selection (Stewart et al. 2008). In the present study, we have highlighted a largely unexplored problem resulting from selection in mtDNA data: the fact that the reconstruction of demographic events that are potentially affected by purifying selection are typically dated using molecular clocks that were calibrated using variation-increasing inter-specific divergence.

The differences in variation found between the mtDNA COI sequences and the nrDNA 28S rRNA sequences of two closely related snails of the genus *Afrolittorina* likely reflect differences in the extent of selection pressure. The snails' COI sequences were much more strongly differentiated than their 28S sequences, potentially reflecting diversifying selection as a result of adaption to different thermal environments. In contrast, there was comparatively little genetic variation at the intraspecific level for both markers. To our knowledge, this is the first study to use 28S to explore genetic variation at the species level, perhaps because it is commonly assumed that this marker evolves so slowly that it would only be informative at deeper taxonomic levels. The finding that the barcoding gap is not strongly developed in 28S rRNA suggests that the evolution of this marker is more likely to conform to the expectations of the neutral model of evolution.

This finding has important implications for the interpretation of species-level mtDNA data. Both species of *Afrolittorina* in the haplotype network have a 'star-phylogeny', where a dominant haplotype has given rise to a large number of comparatively rare haplotypes (Castelloe and Templeton 1994). Star-phylogenies are often interpreted as signatures of a population bottleneck that resulted in a reduction of genetic diversity until a single haplotype was left, for instance during a previous glacial phase (Magoulas et al. 1996), followed by a

rapid increase in population size. The finding that the corresponding but similarly variable 28S haplotype networks are less star-like (although the *A. knysnaensis* network is more star-like than that of *A. africana*) suggests that species-level purifying selection could be an alternative explanation for the patterns found in the mtDNA data. In this case, the apparent population bottleneck would instead be attributed to a selective sweep that has reduced genetic diversity, which is expected to be common in species with very large population sizes (Gillespie 2001), such as littorinid snails.

Even if the assumption that demographic expansions have taken place is correct, the dating of these may be strongly affected by selection. Estimation of demographic trends by means of mismatch distributions using only mtDNA sequence data continue to be widely used in the recent literature (Low et al. 2017; Diringer et al. 2019; Gao et al. 2019; Iván Pérez-Quiñonez et al. 2019) although not all have used the method for molecular dating. A similar method, Bayesian skyline plots (Drummond et al. 2005), is also frequently used with mtDNA data only, and findings tend to be similar (Grant 2015).

The finding that intraspecific mtDNA variation can be as low as that of slowly-evolving nuclear rRNA cautions against the continued use of mtDNA for exploring historical demographic changes. In our opinion, it is time to discontinue the use of fixed mtDNA rates based on divergence dating of closely related taxa, such as the closure of the Central American Seaway to date phylogenies of marine species (Knowlton and Weigt 1998; Schubart et al. 1998) or the 2% rule in birds (Shields and Wilson 1987). The very large datasets generated using next-generation sequencing have considerable potential to facilitate more accurate dating by identifying nuclear markers that conform to the assumptions of a molecular clock. Curiously, fixed rates based on mtDNA data are still being used to calibrate such datasets when no suitable fossil calibration points exist (Trucchi et al. 2014). A possible solution in such cases may involve the identification of a suite of neutral markers that can be used to assess divergence between the species used in the original molecular dating studies, and 28S rRNA may be a suitable candidate.

Acknowledgements

References

Ballard, J. W., and M. Kreitman. 1995. Is mitochondrial DNA a strictly neutral marker? Trends Ecol. Evol. 10:485–488.

Ballard, J. W. O., and D. M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annu. Rev. Ecol. Evol. Syst. 36:621–642.

Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572.

Blier, P. U., F. Dufresne, and R. S. Burton. 2001. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends Genet. 17:400–406.

Castelloe, J., and A. R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. 3:102–113.

Diringer, B., K. Pretell, R. Avellan, C. Chanta, V. Cedeño, and G. Gentile. 2019. Genetic structure, phylogeography, and demography of *Anadara tuberculosa* (Bivalvia) from East Pacific as revealed by mtDNA: implications to conservation. Ecol. Evol. 9:4392–4402.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22:1185–1192. Oxford Academic.

Egea, R., S. Casillas, and A. Barbadilla. 2008. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res. 36:W157–W162.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Gao, T.-X., T.-Y. Yang, T. Yanagimoto, and Y.-S. Xiao. 2019. Levels and patterns of genetic variation in Japanese whiting (*Sillago japonica*) based on mitochondrial DNA control region. Mitochondrial DNA Part DNA Mapp. Seq. Anal. 30:172–183.

Gillespie, J. H. 2001. Is the population size of a species relevant to its evolution? Evol. Int. J. Org. Evol. 55:2161–2169.

Grant, W. S. 2015. Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. J. Hered. 106:333–346. Oxford Academic.

Hughes, R. N. 1979. On the taxonomy of *Littorina africana* (Mollusca: Gastropoda). Zool. J. Linn. Soc. 65:111–118.

Iván Pérez-Quiñonez, C., C. Quiñonez-Velázquez, and F. J. García-Rodríguez. 2019. Genetic homogeneity of the Pacific thread herring (*Opisthonema libertate*) (Günther, 1867) in the Eastern Pacific, inferred from mtDNA sequences. Mitochondrial DNA Part DNA Mapp. Seq. Anal. 30:517–524.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Academic Press, New York.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

9

Knowlton, N., and L. A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. Proc. R. Soc. B Biol. Sci. 265:2257.

Lamb, A. M., H. M. Gan, C. Greening, L. Joseph, Y. P. Lee, A. Morán-Ordóñez, P. Sunnucks, and A. Pavlova. 2018. Climate-driven mitochondrial selection: a test in Australian songbirds. Mol. Ecol. 27:898–918.

Lawrie, D. S., P. W. Messer, R. Hershberg, and D. A. Petrov. 2013. Strong purifying selection at synonymous sites in D. melanogaster. PLoS Genet. 9:e1003527.

Leigh, J. W., and D. Bryant. 2015. popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116.

Low, V. L., Y. Norma-Rashid, A. Yusoff, W. Y. Vinnie-Siow, B. K. Prakash, T. K. Tan, M. Noorhidayah, C. D. Chen, and M. Sofian-Azirun. 2017. Pleistocene demographic expansion and high gene flow in the Globe Skimmer dragonfly *Pantala flavescens* Fabricius (Odonata: Libellulidae) in Peninsular Malaysia. Zool. Anz. 266:23–27.

Magoulas, A., N. Tsimenides, and E. Zouros. 1996. Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). Mol. Biol. Evol. 13:178–190.

Malaquias, M. A. E., and D. G. Reid. 2009. Tethyan vicariance, relictualism and speciation: evidence from a global molecular phylogeny of the opisthobranch genus *Bulla*. J. Biogeogr. 36:1760–1777.

Matumba, T. G. 2013. Genetics and thermal biology of littorinid snails of the genera *Afrolittorina*, *Echinolittorina* and *Littoraria* (Gastropoda: Littorinidae) from temperate, subtropical and tropical regions. Rhodes University, Grahamstown, South Africa.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654.

McQuaid, C. D. 1992. Stress on the high shore: a review of age-dependent causes of mortality in *Nodilittorina knysnaensis* and *N. africana*. Proceedings of the third international symposium on littorinid biology.

Meiklejohn, C. D., K. L. Montooth, and D. M. Rand. 2007. Positive and negative selection on the mitochondrial genome. Trends Genet. TIG 23:259–263.

Reid, D. G., and S. T. Williams. 2004. The subfamily Littorininae (Gastropoda: Littorinidae) in the temperate Southern Hemisphere: the genera *Nodilittorina*, *Austrolittorina* and *Afrolittorina*. Rec. Aust. Mus. 56:75–122.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9:552–569.

Schubart, C. D., R. Diesel, and B. Hedges. 1998. Rapid evolution of terrestrial life in Jamaican crabs. Nature 393:363–365.

Shields, G. F., and A. C. Wilson. 1987. Calibration of mitochondrial DNA evolution in geese. J. Mol. Evol. 24:212–217.

Soares, P., L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, A. Salas, S. Oppenheimer, V. Macaulay, and M. B. Richards. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet. 84:740–759.

Stewart, J. B., C. Freyer, J. L. Elson, and N.-G. Larsson. 2008. Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease. Nat. Rev. Genet. 9:657–662.

Stoeckle, M. Y., and D. S. Thaler. 2014. DNA barcoding works in practice but not in (neutral) theory. PloS One 9:e100755.

Sun, J.-T., P.-Y. Jin, A. A. Hoffmann, X.-Z. Duan, J. Dai, G. Hu, X.-F. Xue, and X.-Y. Hong. 2018. Evolutionary divergence of mitochondrial genomes in two *Tetranychus* species distributed across different climates. Insect Mol. Biol. 27:698–709.

Teske, P. R., T. R. Golla, J. Sandoval-Castillo, A. Emami-Khoyi, C. D. van der Lingen, S. von der Heyden, B. Chiazzari, B. J. van Vuuren, and L. B. Beheregaray. 2018. Mitochondrial DNA is unsuitable to test for isolation by distance. Sci. Rep. 8:8448.

Trucchi, E., P. Gratton, J. D. Whittington, R. Cristofari, Y. Le Maho, N. C. Stenseth, and C. Le Bohec. 2014. King penguin demography since the last glaciation inferred from genome-wide data. Proc. R. Soc. B Biol. Sci. 281:20140528.

Williams, S. T., and D. G. Reid. 2004. Speciation and diversity on tropical rocky shores: a global phylogeny of snails of the genus *Echinolittorina*. Evolution 58:2227–2251.