NEWS AND VIEWS

COMMENT

Babies and bathwater: a comment on the premature obituary for nested clade phylogeographical analysis

R. C. GARRICK,* R. J. DYER,*
L. B. BEHEREGARAY† and P. SUNNUCKS‡
*Department of Biology, Virginia Commonwealth University, Richmond, VA 23284, USA, †Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia, ‡Australian Centre for Biodiversity, School of Biological Sciences, Monash University, Clayton, Vic. 3800, Australia

A recent commentary in *Molecular Ecology* by Petit (2008) paints a rather grim picture of the utility of nested clade phylogeographical analysis (NCPA) for inferring population history. Drawing on simulation studies based on single locus data sets, including the recent work by Panchal & Beaumont (2007), the potential fallibility of NCPA was characterized as being so dire that the method should be abandoned until further evidence in support of its legitimacy emerges. Here, we reconsider the arguments presented by Petit in light of practical approaches for validating or strengthening inferences drawn from NCPA. As with any method that attempts to distinguish processes and events that shaped spatial-genetic structuring throughout complex evolutionary histories of natural populations, we propose that treatment of NCPA inferences should be set in the context of corroborating evidence (or lack thereof) that support those inferences. Indeed, results from computer simulation, studies lend no support to the idea that NCPA should not be employed for generating plausible hypotheses (i.e. consistent with species biology and landscape history) that can be further tested using other methods. Moreover, cross-validation of NCPA inferences via assessment of multiple independent loci, complementary analyses, and/or prior expectations, should at least partly — perhaps considerably — counter high false-positive rates reported for some inferences. NCPA uniquely offers the ability to explore patterns relating to complex, historical scenarios: an over-reaction to Panchal & Beaumont (2007) should not precipitate throwing out an approach currently with no computationally feasible substitute.

In recent years, use of the inference key in nested clade phylogeographical analysis (NCPA) to identify historical

Correspondence: Ryan Garrick, Fax: +1804-828-0503; E-mail: rcgarrick@vcu.edu

processes has attracted criticism (Knowles & Maddison 2002). A simulation study (Panchal & Beaumont 2007) has prompted Petit to suggest that '... the method be no longer used until it has been more thoroughly and critically evaluated ...'. In this brief communication, we hope to provide a more pragmatic view of where NCPA fits into the phylogeographer's toolbox by making the argument that there is a false dichotomy between NCPA and alternative 'statistical' methods.

The study by Panchal & Beaumont (2007) showed that clade statistics can be significantly biased, yielding more false-positives (type I error) than expected. Specifically, it yielded inferences of either restricted gene flow or contiguous range expansion under a simulated scenario of panmixia. While these results are alarming, it is important to point out that to date, all computer simulation studies that tested NCPA have been conducted using single locus data sets (e.g. Irwin 2002; Knowles & Maddison 2002; Panchal & Beaumont 2007).

There are known problems with historical inferences drawn from single locus data sets. For example, Kuo & Avise (2005) showed that while individual gene genealogies can exhibit haphazard phylogeographical breaks, spatially concordant breaks across multiple loci tend to emerge only in the presence of historical barriers to gene flow. Accordingly, any phylogeographical inferences based on a single locus must be interpreted with caution (Templeton 2002; Knowles 2004), and inclusion of other genes can substantially improve accuracy when reconstructing organismal history from molecular data (Ballard & Whitlock 2004). NCPA is no exception, and a number of studies have applied the method to multiple independent loci as means of accommodating coalescent stochasticity (e.g. Banke & McDonald 2005; Zhang *et al*. 2005; Garrick *et al*. 2007). To dismiss NCPA on the basis of its performance with simulated data sets representing just one 'snapshot' of population history seems to be at odds with well-documented difficulties associated with accurately reconstructing past events from a single locus. To date, no simulation studies have attempted to ascertain how the incorporation of corroborating genetic evidence might influence the frequency with which false-positives are mistakenly accepted. In general, we believe that as additional loci are added, the bias in type I error will be reduced.

As noted by Templeton (2004), using NCPA to the exclusion of all other analyses and without regard to any prior knowledge is not advised. Findings from a literature search presented by Petit overemphasizes the pervasiveness that reported high false-positive rates are likely to have on inferences drawn from empirical data. NCPA is routinely used in conjunction with complementary analyses and strong prior expectations. For example, range expansion is often cross validated via supplementary tests of exponential population growth using mismatch distributions, or coalescent-based methods (e.g. FLUCTUATE, Kuhner *et al*. 1998). Similarly, inferences of gene flow restricted by distance can be re-evaluated via simple linear regression, or when multilocus allelic data are available, with other more

geographically explicit approaches (e.g. Wartenberg 1985; Amos & Manica 2006; Dyer 2007).

Increasingly, empirical studies employ NCPA in the context of well-defined prior expectations based on information from Earth science disciplines (e.g. DeChaine & Martin 2004; Beheregaray *et al*. 2004; Sunnucks *et al*. 2006; Bell *et al*. 2007). In this regard, integration between phylogeographers and Earth scientists is critical to better address expectations about scenarios and temporal axes of diversification. This integration, together with basic knowledge of the ecology and life history of the organism(s) under study, permits researchers to recognize potential sources of error in the estimation of historical processes (see Masta *et al*. 2003; Morando *et al*. 2004).

The recent commentary also depicts phylogeographical analyses focusing on 'populations' as being more desirable than individual-based approaches. However, when analysing DNA sequence data from continuously distributed species, it is often unclear where population boundaries lie and how permeable such boundaries are to gene flow. Indeed, in the absence of complementary genetic data (e.g. nuclear genotypes scored for multiple individuals), an objective identification of the number, composition, and distribution of biologically meaningful genetic units can be challenging (Guillot *et al*. 2005). But it is precisely this a priori delineation of clearly demarcated populations (e.g. Figure 2 in Petit 2008) that underpins many of the alternative statistical approaches to phylogeographical analysis (see Hey & Machado 2003). To our knowledge, no studies to date have explicitly examined the impact that incorrectly defining population boundaries might have on inferences about population history. Thus, NCPA remains a valuable component in the analytical toolbox, even if just for comparison with other methods, because it is not subject to the assumption of crisp, panmictic populations, as made by popular model-based approaches (e.g. Beerli & Felsenstein 2001; Hey & Nielsen 2004). Concerns raised by Petit about the confounding influence that nonhistorical processes (e.g. kin clustering) might have on NCPA inferences are also relevant to analyses employing traditional population-genetic models.

Integrative analytical approaches to interpreting spatial genetic patterns seen in organisms from landscape settings with well-understood biogeographical histories will help pave the way for further refinement of NCPA. Rather than abandoning the approach in response to indications that the method can be prone to spurious inferences on some occasions, a more constructive course of action is to encourage its validation. The application of NCPA as part of multifaceted battery of analyses should help provide a better appreciation of the amount of data needed for sound inferences and the situations in which inferences may be unreliable. Based on the presently limited number of null demographic models that have been considered in simulation studies (i.e. allopatric fragmentation, one-dimensional isolation by distance, panmixia), and taking into account Templeton's (2004) concerns about methodological oversights pertinent to some of those studies, a more tempered reaction to the reported high false-positive rates for specific NCPA inferences is warranted. While it is clear that NCPA should not be applied blindly, as if merely appealing to authority for interpretation of data, we do not believe that NCPA is attractive only to 'researchers eager to identify statistical support for their interpretation of the data'. We suspect that, in the vast majority of applications, the method has not been employed as a prospecting tool in a relentless pursuit of some form of statistical significance. Rather, it has been used as one of several analytical approaches to understand organismal histories.

References

Amos W, Manica A (2006) Global genetic position: evidence for early human population centers in coastal habitats. *Proceedings of the National Academy of Sciences, USA*, **103**, 820–824.

Ballard JW, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Banke S, McDonald BA (2005) Migration patterns among global populations of the pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology*, **14**, 1881–1896.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.

Beheregaray LB, Gibbs JP, Havill N *et al*. (2004) Giant tortoises are not so slow: rapid diversification and biogeographic consensus in the Galápagos. *Proceedings of the National Academy of Sciences, USA*, **101**, 6514–6519.

Bell KL, Moritz C, Moussalli A, Yeates DK (2007) Comparative phylogeography and speciation of dung beetles from the Australian wet tropics rainforest. *Molecular Ecology*, **16**, 4984–4998.

DeChaine EG, Martin AP (2004) Historic cycles of fragmentation and expansion in *Parnassius smintheus* (Papilionidae) inferred using mitochondrial DNA. *Evolution*, **58**, 113–127.

Dyer RJ (2007) The evolution of genetic topologies. *Theoretical Population Biology*, **71**, 71–79.

Garrick RC, Sands CJ, Rowell DM, Hillis DM, Sunnucks P (2007) Catchments catch all: long-term population history of a giant springtail from the southeast Australian highlands — a multigene approach. *Molecular Ecology*, **16**, 1865–1882.

Guillot G, Estoup A, Mortier F, Cosson J-F (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

Hey J, Machado CA (2003) The study of structured populations — new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 636–643.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, **56**, 2383–2394.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Kuo C-H, Avise JC (2005) Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. *Genetica*, **124**, 179–186.

Masta SE, Laurent NM, Routman EJ (2003) Population genetic structure of the toad *Bufo woodhousii*: an empirical assessment of the effects of haplotype extinction on nested cladistic analysis. *Molecular Ecology*, **12**, 1541–1554.

Morando M, Avila LJ, Baker J, Sites JW (2004) Phylogeny and phylogeography of the *Liolaemus darwinii* complex (Squamata: Liolaemidae): evidence for introgression and incomplete lineage sorting. *Evolution*, **58**, 842–861.

Panchal M, Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution*, **61**, 1466–1480.

Petit RJ (2008) The coup de grâce for the nested clade phylogeographic analysis? *Molecular Ecology*, **17**, 516–518. [Correction added after publication 13 February 2008: the reference Petit RJ (in press) On the falsifiability of the nested clade phylogeographic analysis method. *Molecular Ecology*, doi: 10.1111/j.1365-294X.2008.03692.x was replaced.]

Sunnucks P, Blacket MJ, Taylor JM *et al.* (2006) A tale of two flatties: different responses of two terrestrial flatworms to past environmental climatic fluctuations at Tallaganda in montane southeastern Australia. *Molecular Ecology*, **15**, 4513–4531.

Templeton AR (2002) Out of Africa again and again. *Nature*, **416**, 45–51.

Templeton AR (2004) Statistical phylogeography: methods for evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.

Wartenberg D (1985) Canonical trend surface analysis: a method for describing geographic patterns. *Systematic Zoology*, **34**, 259–279.

Zhang AB, Kubota K, Takami Y *et al.* (2005) Species status and phylogeography of two closely related *Coptolabrus* species (Coleoptera: Carabidae) in South Korea inferred from mitochondrial and nuclear gene sequences. *Molecular Ecology*, **14**, 3823–3841.

Ryan Garrick is interested in phylogeography and population structure of terrestrial invertebrates and their host plants. Rodney Dyer's research focuses on understanding how genes are dispersed in both space and time. Luciano Beheregaray has broad interests in conservation genetics, phylogeography and speciation. Paul Sunnucks works on the molecular population biology and ecology of diverse organisms at a range of spatial and temporal scales.