

NEWS AND VIEWS

COMMENT

On the validity of nested clade phylogeographical analysis

MARK A. BEAUMONT and
MAHESH PANCHAL

School of Biological Sciences, University of Reading, Reading RG6 6BX, UK

Received 11 February 2008; revision received 03 March 2008; accepted 07 April 2008

As will be apparent to any reader of *Molecular Ecology*, a large number of genetic analyses of geographically structured populations are routinely carried out. Researchers make promises to their sponsors that they can uncover aspects of the demographic history of these populations. Inevitably, there is demand for statistical methods that will provide credible answers, and consequently, many methods have been supplied. Nested clade phylogeographical analysis (NCPA) is a technique that continues to be quite popular (Petit 2008). Naturally, there is curiosity to see whether the method works. Essentially, there is a disagreement between Templeton (2004, 2008), who suggests the method works well, and three independent groups (Knowles & Maddison 2002; Petit & Grivet 2002; Panchal & Beaumont 2007), who believe that they have demonstrated that it does not. As far as we are aware, there are currently no publications other than those of Templeton and co-workers to support the accuracy or efficacy of NCPA.

In this article, we address some comments by Templeton (2008) on the study by Panchal & Beaumont (2007). We begin by highlighting a recommendation in Templeton (2004) that NCPA should be tested using data simulated from a single panmictic population of constant size, which we have carried out. We argue that our automated implementation is as close as is feasible to the published versions, and that the false-positive rate that we observe is not a consequence of our implementation. With respect to validation, we suggest that it is not possible to draw comparisons between the outcome of our simulation-based tests and the inferences drawn from empirical data sets. We also note that many recent modifications of NCPA have had little validation. The difficulty of correcting for multiple statistical tests in NCPA is then discussed. Finally, we address the more general issues raised by Templeton (2008) concerning the method of scientific enquiry, and we recommend that a more model-based approach should be taken to infer demographic history.

Correspondence: Mark A. Beaumont, Fax: +44 118931 0180; E-mail: m.a.beaumont@reading.ac.uk

The tests discussed by Templeton (2008) were primarily applied to data assumed to have histories with range expansion and fragmentation. As Templeton (2004) noted, there is also a persuasive case for testing NCPA under scenarios of panmixia, using simulated data sets:

‘One method to obtain the true type I error rate is through computer simulation. Simulations are ideal for this purpose because the null model, a single panmictic population with no history of fragmentation or range expansion, is well defined and simple to simulate. Applying NCPA to such simulated panmictic populations would indicate the true type I error rate under the null hypothesis of no association between clades and geography’ (Templeton 2004).

We have undertaken such a study and have concluded that NCPA is highly prone to false-positives. To carry this out, one of us (M.P.) developed a computer package for automating the procedure (ANECA; Panchal 2007). Of course, it could be argued that we have implemented NCPA in some way that is ‘wrong’. There are a number of general observations to make here. First is that the NCPA procedure has evolved through time, as documented in detail in Panchal & Beaumont (2007), and if all incarnations of the method other than the most recent are regarded as ‘wrong’, so too will all papers that used these procedures. Second, there are many steps in the procedure where it is highly likely that different researchers would make slightly different decisions, and these are also discussed in Panchal & Beaumont (2007). The results of a questionnaire (Supplementary data in Panchal & Beaumont 2007) reflected this variability among researchers in their use of NCPA, but also provided support for the decisions taken in implementing ANECA. Similarly, there was a high concordance between the inferences obtained by ANECA and those obtained manually by the authors of published data sets. We would like to make clear here that we have provided the ANECA software not as a resource for phylogeographical analysis, but so that independent groups can evaluate our implementation of NCPA, compare it with manual applications, and generally form their own conclusions about the accuracy of our implementation and of the method itself.

Templeton (2008) suggests that the method described in Panchal & Beaumont (2007) may be prone to false-positives. It is important to be clear here that the false-positives arise from the output of the GEODIS program (Posada *et al.* 2000). If a statistic from GEODIS is deemed ‘significant’ from the permutation tests, one then consults the inference key. Our method uses the program *racs* (Clement *et al.* 2000) to construct haplotype networks. We have implemented an algorithm based on the rules provided by Templeton *et al.* (1987) for nesting clades (with some additions from later papers by Templeton

and colleagues). This information is then given to GEODIS. Thus, the only place in our implementation that could influence the false-positive rate is in the nesting algorithm. The published rules are relatively straightforward to implement, as detailed in Panchal & Beaumont (2007). Variations to these rules have been suggested when dealing with loops in the networks (also discussed in Panchal & Beaumont 2007). However, our simulated data sets typically had a low amount of homoplasy and hence, few loops.

The modelling assumptions that were made when simulating the data seem reasonable, given the aims of the project. The mutation rates are based on estimates from mammalian mitochondrial DNA. A range of sample sizes was investigated, reflecting those observed in empirical data. The sampling scheme was conducted on a lattice of demes, and was designed so that all answers to questions in the inference key regarding sampling scheme could be investigated (e.g. whether a deme is present or absent between two clades). It is also important to emphasize that AMOVA, implemented in ARLEQUIN (Excoffier *et al.* 2005), showed the data sets to be panmictic — indeed the coverage of the test was close to perfect (5% false-positives at the 5% threshold).

A general feature of the comments in Templeton (2008) is to compare results from analyses of empirical data sets with the results in Panchal & Beaumont (2007), and apply statistical tests to highlight these differences. We fail to see how any valid comparison can be drawn between these two very different groups of data, irrespective of the sophistication of the statistical tests that are applied. For example, attention is drawn to the results obtained in Templeton (2005) concerning inferred human population history, which are compared with those obtained by Panchal & Beaumont (2007). However, it is difficult to see a connection between the two studies since the human data are clearly structured whereas our simulations were not.

Templeton (2008) describes the method as 'extensively validated', but, for example, as noted in Templeton (2004), 'simulations could also help in validating inferences related to gene flow, which is a gap in the procedure of validating using data sets with prior expectations'. A number of procedures have been modified or added since the inception of NCPA, such as various methods for resolving loops in networks (e.g. Brisson *et al.* 2005), and the 'cross-validation' method of Templeton (2002, 2005). Indeed, this latter method is not cross-validation as typically implemented in statistical analyses but a procedure in which an inference is reported if more than one locus infers the same historical process involving the same geographical locations, with some additional calculation to determine the timing of demographic events. We are unaware that there has been any statistical evaluation of these changes to the original NCPA procedure.

Templeton (2008) correctly points out that there is currently no multiple test correction for NCPA. Panchal & Beaumont (2007) observed that 75% of data sets simulated from a panmictic population of constant population size had at least one clade with an informative inference (i.e. giving rise to a statement such as 'contiguous range expansion'). Posada *et al.* (2006) noted that 265 papers had cited GEODIS, presumably having performed an NCPA analysis. Assuming a bias against publishing negative results, it is therefore quite possible that

virtually, all these papers contain spurious inferences. An interesting feature of the results published in Panchal & Beaumont (2007) is that there is quite a strong and statistically significant relationship between the frequency of different categories of inference in data simulated under complete panmixia and those observed in published data sets. In particular, restricted gene flow with isolation by distance and contiguous range expansion were the most commonly inferred scenarios in both cases.

We are not as sanguine as Templeton (2008) that a correction for multiple tests will be straightforward. Templeton (2008) states that each nesting clade yields only a single inference in NCPA, but omits also to state that each nesting clade can yield a very large number of statistics, only one of which needs to be significant for an inference to be made using the key. Panchal & Beaumont (2007) observed that the relationship between the probability, P , of obtaining at least one significant statistic in any clade tested and the number of summary statistics in the clade, n , is complex. In comparison to the naïve expectation of $P = 1 - (1 - \alpha)^n$, where α is the significance level used, we find that P is too low at small n , and too high for large n (Fig. 1).

Templeton (2008) cites Popper (1959) in support of the NCPA approach against model-based statistical analysis. However, we would suggest that although NCPA consists of a large number of hypothesis tests based on permutation methods, in the end it follows an inductivist paradigm of trying to derive a general explanation directly from the data ('NCPA does *not* require an a priori model', Templeton 2008). The method purports to 'read' demographic history from phylogenies. A verbal, reasoned, argument is presented in Templeton *et al.* (1995) to justify the method, and the inferences it makes, not dissimilar in style and authority to the *Corpus Aristotelicum*. The authors of 265 papers that have used NCPA are, in a sense, appealing to this authority. One needs to ask: is this science? Is the hypothesis 'NCPA works' falsifiable? We would argue that it is not. The whole structure of the technique is set up in such a way that only its proponents can decide whether it has been falsified or not. There have been no extensive analyses, other than those by the original authors, that report in favour of the method. The only independent groups that have studied the method have concluded that it is prone to false-positives (Knowles & Maddison 2002; Petit & Grivet 2002; Panchal & Beaumont 2007).

By contrast to NCPA, in model-based analysis, such as that of Fagundes *et al.* (2007), one model is pitted against another in the face of the data, and this, surely, is a more valid scientific approach (Tarantola 2006). These models do not arise from data, but from current arguments. Of course, the results will depend on the models that are compared. Often, genetic data will not be able to distinguish between complicated models of demographic history, and it may be helpful to use idealizations (Wakeley 2004). However, we always need to bear in mind that there is a real demographic history that explains the data, but we do not yet have sufficient information to elucidate what it is. There may well be arguments about precisely what is the best way to compare models: whether through the use of likelihood-ratio tests, or Bayesian posterior probability. Similarly, there are arguments about the numerical techniques

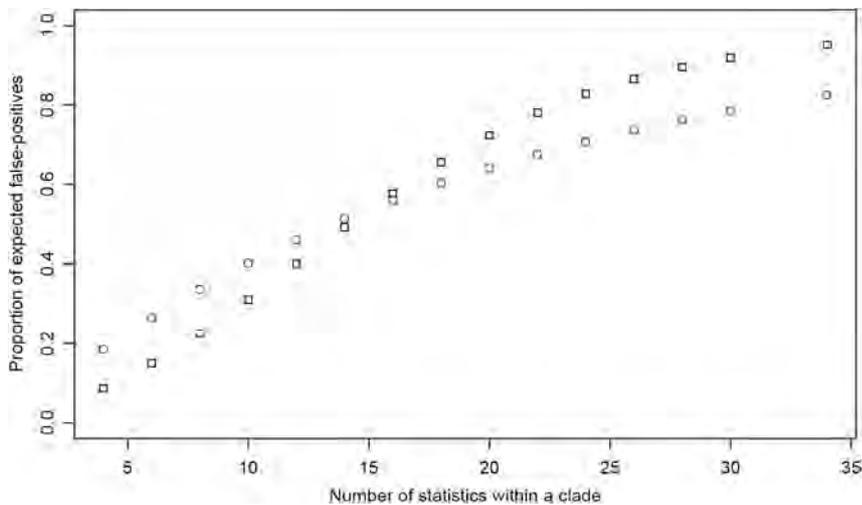


Fig. 1 A plot showing the proportion of expected false-positives as a function of the number of statistics within a clade. These false-positives are clades that will give rise to an inference under NCPA, although this inference may not be informative (i.e. 'inconclusive outcome', 'sampling design inadequate ...'). The open circles are the naive expectations from $P = 1 - (1 - \alpha)^n$, where $\alpha = 0.05$ is the critical P value, and n is the number of statistics. The squares are those observed in the simulations described in Panchal & Beaumont (2007; p. 1474).

to be used: genealogical Markov chain Monte Carlo (MCMC), composite likelihood, and approximate Bayesian computation (ABC). These arguments can be resolved by objective comparison of different algorithms, and the efficacy of methods can be compared through computer simulations. In our opinion, it is these methods that represent the future for phylogeographical analysis.

Acknowledgements

We thank Ferran Palero for helpful comments.

References

- Brisson JA, de Toni DC, Duncan I, Templeton AR (2005) Abdominal pigmentation variation in *Drosophila polymorpha*: geographic variation in the trait, and underlying phylogeography. *Evolution*, **59**, 1046–1059.
- Clement M, Posada D, Crandall KA (2000) tcs: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN (ver. 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **Online 1**, 47–50.
- Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences, USA*, **104**, 17614–17619.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- Panchal M (2007) The automation of nested clade phylogeographic analysis. *Bioinformatics*, **23**, 509–510.
- Panchal M, Beaumont MA (2008) 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.
- Petit RJ, Grivet D (2002) Optimal randomization strategies when testing the existence of a phylogeographic structure. *Genetics*, **161**, 469–471.
- Popper KR (1959) *The Logic of Scientific Discovery*. Hutchinson, London.
- Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.
- Posada D, Crandall KA, Templeton AR (2006) Nested clade analysis statistics. *Molecular Ecology Notes*, **6**, 590–593.
- Tarantola (2006) Popper, Bayes and the inverse problem. *Nature Physics*, **2**, 492–494.
- Templeton AR (2002b) Out of Africa again and again. *Nature*, **416**, 45–51.
- Templeton AR (2004b) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.
- Templeton AR (2005) Haplotype trees and modern human origins. *Yearbook of Physical Anthropology*, **48**, 33–59.
- Templeton AR (2008) Nested clade analysis: an extensively validated method for strong phylogeographic inference. *Molecular Ecology*, **17**, 1877–1880.
- Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. 1. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–351.
- Templeton AR, Routman E, Phillips C (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.
- Wakeley J (2004) Recent trends in population genetics: more data! More math! Simple models? *Journal of Heredity*, **95**, 397–405.

Mark Beaumont is interested in using genetic data to study local selection and the demographic history of populations. Mahesh Panchal has recently completed a PhD on phylogeographical analysis.

doi: 10.1111/j.1365-294X.2008.03786.x