

## NEWS AND VIEWS

### REPLY

### Nested clade analysis: an extensively validated method for strong phylogeographic inference

ALAN R. TEMPLETON

Department of Biology, Washington University, St Louis, Missouri 63130-4899, USA

*Keywords:* nested clade analysis, phylogeography, statistics, false positives, computer simulation, cross-validation

*Received 19 January 2008; revision received 26 January 2008; accepted 3 February 2008*

Petit (2008) speculates that three papers (Knowles & Maddison 2002; Petit & Grivet 2002; Panchal & Beaumont 2007) have delivered the coup de grâce for nested clade phylogeographic analysis (NCPA). As Petit (2008) notes, NCPA is one of the most frequently used analytical tools in the area of phylogeography; thus, this speculation must be evaluated carefully.

Templeton (2002a) pointed out that the criticisms of Petit & Grivet (2002) were based upon misunderstandings and misrepresentations of NCPA, and that the one situation that they identified as creating errors had already been identified by Templeton (1998). More importantly, the NCPA inference key in this situation leads to false negatives (Templeton 1998), *not* false positives as speculated by Petit (2008).

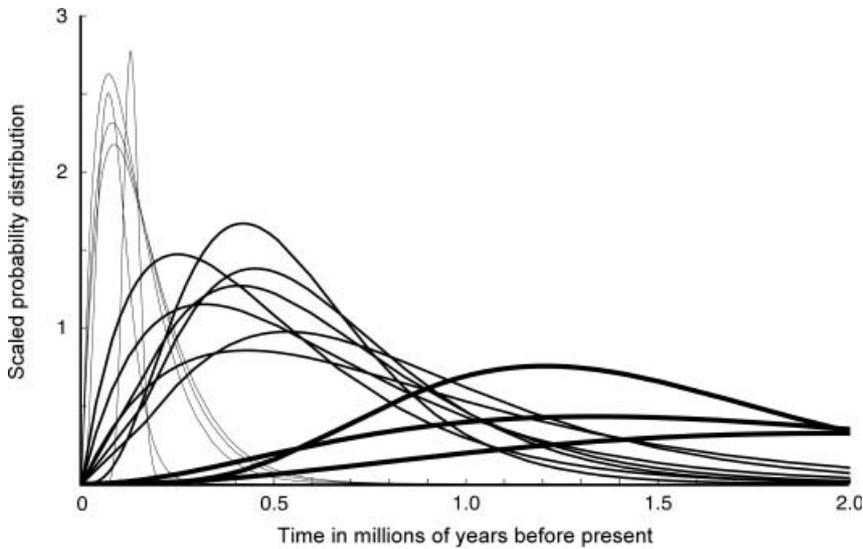
False positives have been evaluated in NCPA through positive controls: the analysis of real data with a known event(s). Templeton (2004b) validated NCPA with 150 a priori expectations; probably the most massive set of positive controls ever used to validate a procedure in evolutionary genetics, and certainly in intraspecific phylogeography. The salient results are: (i) false negatives ranged from 12% for fragmentation events to 38% for range expansion events; (ii) false positives ranged from a *maximum* (under the *conservative* assumption that all inferences not predicted *a priori* are false) of 3% for fragmentation events to a *maximum* of 23% for range expansion events; and (iii) multiple events can be inferred without interference from one another.

The 150 positive controls span a broad range of organisms, spatial distributions, and sampling designs, but all share two features: they are real evolutionary events, and they are real sampling designs. To counter this massive analysis of 150 positive controls, Petit (2008) cites two papers (Knowles, Maddison 2002; Panchal, Beaumont 2007), each describing a *single* artificial evolutionary scenario with an artificial sampling

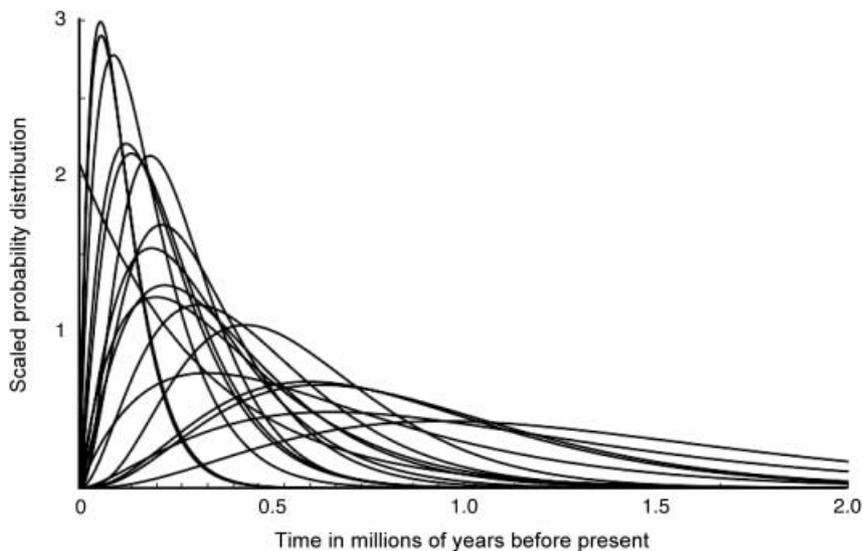
scheme. The first of these scenarios is microvicariance in which each local deme is isolated (Knowles, Maddison 2002). This type of event is specifically *excluded* from NCPA (p. 773, Templeton *et al.* 1995) and is covered by a separate test (Hutchison & Templeton 1999), making it irrelevant to NCPA. Knowles & Maddison (2002) assume exhaustive sampling of all populations. Given that none of the data sets in the massive validation study in Templeton (2004b) claimed such exhaustive sampling, the results in Knowles & Maddison (2002) were reanalysed *without* the assumption of exhaustive sampling. The false-positive rate dropped from 75–80% to 18% (Templeton 2004b), showing that their high false-positive rate is an artefact of an unrealistic assumption.

Panchal & Beaumont (2007) simulated a panmictic population. Petit (2008) ascribes their false positives to NCPA, but this study has three potential sources of false positives: (i) unrealistic assumptions in their simulations (as happened in Knowles and Maddison); (ii) flaws in their implementation of NCPA, and (iii) the failure of the original NCPA. The performance of NCPA with the 150 positive controls only involves cause (iii); therefore, if Petit is right, the false-positive rates should be homogeneous between these two analyses. To be conservative, the false-positive rate of 23% for range expansions, the highest false-positive rate in Templeton (2004b), is tested for homogeneity with the simulated false-positive rate of 50% for range expansions (Panchal, Beaumont 2007). The null hypothesis is rejected (chi-square goodness of fit is 8.53 with one degree of freedom, a probability  $\leq 0.0035$ ). Hence, the source of most of the false positives reported by Panchal & Beaumont (2007) is in their own simulations and/or in their novel implementation of NCPA.

Panchal & Beaumont (2007) claim that the inferences most subject to false positives are range expansion and isolation by distance, each with a false-positive rate of 50%. Templeton (2005) performed NCPA upon 24 gene regions (after excluding one, as discussed later) to reconstruct human evolutionary history. Some 15 inferences of out-of-Africa range expansions were collectively inferred, and there were 18 inferences of isolation-by-distance involving African and Eurasian populations. The results of Panchal & Beaumont (2007) would predict 12 false positives for each type of inference. Figures 1 and 2 show the estimated probability distributions for the timing of these inferences (Templeton 2005). Because range expansions are events, the times of the individual inferences should cluster around one or a few time points (if the event occurred more than once). In contrast, gene flow is a recurrent process with no expectation of temporal clustering. Hence, *if the inferences are biologically real*, the range expansion events should show temporal clustering and the gene flow processes need not. On the other hand, if most of these inferences are false positives, there should be no temporal patterning for either case, and hence homogeneity between these two cases of inference. The first test of the null hypothesis of homogeneity is based upon the fact that clustering increases the



**Fig. 1** The estimated time distributions for 15 out-of-Africa range expansions inferred by NCPA (modified from Templeton 2005).



**Fig. 2** The estimated time distributions for 18 NCPA inferences of gene flow between African and Eurasian populations restricted by isolation by distance (modified from Templeton 2005).

variance of the estimated times. The null hypothesis of equal variances is rejected with  $P = 0.043$  using the nonparametric Klotz test since the median times are homogeneous (all tests are performed with STATXACT 7 from Cytel Software).

A second test is to calculate the differences between adjacent ranked times for each inference type. If the inferences of range expansion are real, we expect many differences of small magnitude due to temporal clustering, with a few differences being large (the times between range expansion events). If almost all inferences are false positives, there should be no clustering for either range expansion or isolation by distance, and hence homogeneity. This null hypothesis of homogeneity is tested with a two-sample Kolmogorov–Smirnov test, and it is rejected with  $P = 0.015$ . Hence, the NCPA inferences significantly deviate from the signature expected under the high false-positive rate given in Panchal & Beaumont (2007). Moreover,

all three range expansions detected by NCPA are strongly corroborated by fossil, archaeological and palaeoclimatic data (Templeton 2005, 2007), making the high false-positive rates given in Panchal & Beaumont (2007) even more implausible.

The above results indicate that the high false-positive rate in Panchal & Beaumont (2007) is *not* due to NCPA, but rather due to their simulation procedure and/or implementation of NCPA. Hence, users of NCPA should *not* use the programs of Panchal & Beaumont (2007) until the cause of *their* high false-positive rate is known.

Because the false-positive rates of NCPA reported in Templeton (1998, 2004b) exceed the 5% level, Templeton (2002b, 2004a, b) developed a multilocus cross-validation procedure to reduce false negatives and positives and to provide a framework for testing specific phylogeographic hypotheses. The ability of cross-validation to exclude false positives was

inadvertently demonstrated by the inclusion of a published data set on the human *MX1* locus that contained a paralogous copy (pointed out by Justin Fay in a personal communication after the original analysis was published). Although *MX1* did yield significant inferences under NCPA, the inferences were not cross-validated and eliminated (Templeton 2005). As the entire field of phylogeography moves towards multilocus analysis, Petit's (2008) criticisms become increasingly irrelevant.

Currently, there is no multiple test correction for single-locus NCPA. This problem is easily corrected. Each nesting clade yields only a single inference in NCPA, so no multiple-test correction is needed for the tests *within* a nesting clade. The NCPA data structure is categorical (discrete clades and discrete sampling locations), and the tests across nesting categories are asymptotically independent under the null hypothesis (Prum *et al.* 1990). A simple Bonferroni correction can therefore be used in which the number of tests is the number of nesting clades analysed by the *GEODIS* program. The Bonferroni *P* value should be applied for the tests *within* nesting clades rather than the 5% criterion. For those wishing to correct for tests within nesting clades, a step-down procedure (Westfall & Young 1993) can be used, as has already been implemented for the program *TREESCAN* (Posada *et al.* 2005).

False positives can be corrected for in single- and multilocus NCPA. This is not the case in the alternatives favoured by Petit (2008) that only consider the relative fit of a finite number of alternative hypotheses obtained by computer simulation. Falsification of a hypothesis is strong inference in science (Popper 1959), whereas relative fit within a finite set of alternatives is weak inference when a finite set cannot ensure coverage of the universe of possible models. The model space of phylogeography is immense, and a false hypothesis can erroneously be supported if all the simulated alternatives are worse. Moreover, the simulation approach is subject to another type of logical error known as the 'ecological fallacy' (Templeton 2007).

The simulation methods preferred by Petit have been validated primarily by computer simulations, usually by the same group that developed the simulation method. Hence, there is no protection against common assumptions in both sets of simulations of inducing parallel, and thereby hidden, errors. Moreover, the simulated evolutionary scenarios are simple, and work well with the simulation-based analytical techniques that also deal with simple models. Biological reality is far more complex, and there is no guarantee that these simulation methods work well with the complexity of the real world. This can only be shown with positive controls, as was done for NCPA (Templeton 2004b). However, the finite simulation approach cannot be validated with positive controls. The *only* way for the finite simulation approach to obtain the correct answer is to simulate the model that contains what is known to be correct *a priori*. The trouble is, in many cases, we do not know what the correct model is *a priori*. Positive controls therefore ignore the primary source of inference error in the finite simulation approach: not including the true model in the finite set.

For example, the NCPA of the 24 human genome regions yield a cross-validated, statistically significant inference of three out-of-Africa expansion events (Fig. 1). None of the standard models of human evolution over the past three decades had

proposed the middle expansion that corresponds to the spread of the Acheulean culture out of Africa. NCPA was able to detect this novel aspect of human evolution, strongly corroborated by palaeontological, archaeological, and palaeoclimatic data (Templeton 2005, 2007), precisely because NCPA does *not* require an *a priori* model. The hypothesis that the most recent out-of-Africa expansion resulted in the complete replacement of Eurasian populations was also falsified by NCPA ( $P < 10^{-17}$ ) (Templeton 2005). Fagundes *et al.* (2007) simulated eight models of human evolution, none of which incorporated the Acheulean expansion. Thus, *all* eight simulated models have already been falsified. Among these eight falsified alternatives, the out-of-Africa replacement model had the best relative fit. Eswaran *et al.* (2005) also simulated the out-of-Africa replacement model and some alternatives, among which was a model with isolation by distance that is more consistent with NCPA inferences (Fig. 2). Eswaran *et al.* (2005) strongly refuted the out-of-Africa replacement model. Thus, *both* NCPA and computer simulation/goodness-of-fit analysis indicate that the support for the replacement model reported by Fagundes *et al.* (2007) is a false positive. There is no way of detecting this false positive within the simulation framework until someone simulates more appropriate models.

There is a legitimate role for the simulation approach. Once a null hypothesis based inference technique, such as NCPA, has sufficiently restricted the class of appropriate models, simulations can add much further detail and estimate relevant parameters (for example, Strasburg *et al.* 2007). Thus, NCPA plus simulation approaches is a powerful combination, whereas simulations alone are flawed when the model space is large and unknown.

NCPA is the only method of phylogeographic analysis that has been validated by a massive data set of positive controls that cover a broad range of species, geographic scales, and sampling designs. Petit & Grivet (2002) rediscovered a situation known to give false negatives, not false positives. Knowles & Maddison (2002) generated a high false-positive rate as an artefact of an unrealistic sampling assumption for a situation irrelevant to NCPA. Statistical analysis shows that the high false-positive rate in Panchal & Beaumont (2007) is due to their simulation and/or their nonvalidated implementation of NCPA. The alternatives favoured by Petit (2008) can generate false positives with seemingly strong support. Therefore, perhaps Petit (2008) was right; the simulation artefacts of Knowles & Maddison (2002) and the discrepancies with real data in the simulations of Panchal & Beaumont (2007) have indeed delivered a coup de grâce to a phylogeographic technique; Petit just had the target wrong.

### Acknowledgements

This work was supported in part by NIH grant P50-GM65509. I thank David Posada and Keith Crandall for their suggestions on an earlier draft of this paper

### References

- Eswaran V, Harpending H, Rogers AR (2005) Genomics refutes an exclusively African origin of humans. *Journal of Human Evolution*, **49**, 1–18.

- Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, **104**, 17614–17619.
- Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- 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.
- 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, Maxwell TJ, Templeton AR (2005) TREESCAN: a bioinformatic application to search for genotype/phenotype associations using haplotype trees. *Bioinformatics*, **21**, 2130–2132.
- Prum B, Guilleud-Bataille M, Clerget-Darpoux F (1990) On the use of  $\chi^2$  tests for nested categorized data. *Annals of Human Genetics*, **54**, 315–320.
- Strasburg J, Kearney M, Moritz C, Templeton A (2007) Combining phylogeography with distribution modeling: multiple Pleistocene range expansions in a parthenogenetic gecko from the Australian arid zone. *PLoS ONE*, **2**, e760.
- Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.
- Templeton AR (2002a) 'Optimal' randomization strategies when testing the existence of a phylogeographic structure: a reply to Petit and Grivet. *Genetics*, **161**, 473–475.
- Templeton AR (2002b) Out of Africa again and again. *Nature*, **416**, 45–51.
- Templeton AR (2004a) A maximum likelihood framework for cross validation of phylogeographic hypotheses. In: *Evolutionary Theory and Processes: Modern Horizons* (ed. Wasser SP), pp. 209–230. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- 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 (2007) Perspective: genetics and recent human evolution. *Evolution*, **61**, 1507–1519.
- 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.
- Westfall PH, Young SS (1993) *Resampling-Based Multiple Testing*. John Wiley & Sons, New York.

---

Alan Templeton has both a Master's Degree in Statistics and a Doctoral degree in Human Genetics. This combination of statistical and genetical training has allowed him to develop many innovative statistical methods in human genetics and evolutionary biology, including the statistical procedure of nested clade analysis.

---

doi: 10.1111/j.1365-294X.2008.03731.x