Published in final edited form as:
Evolution. 2009 April ; 63(4): 807–812. doi:10.1111/j.1558-5646.2008.00600.x.

Why Does a Method That Fails Continue To Be Used: The Answer

Alan R. Templeton
Department of Biology, Campus Box 1137, Washington University, St. Louis, Missouri 63130

Abstract

It has been claimed that hundreds of researchers use nested clade phylogeographic analysis (NCPA) based on what the method promises rather than requiring objective validation of the method. The supposed failure of NCPA is based upon the argument that validating it by using positive controls ignored type I error, and that computer simulations have shown a high type I error. The first argument is factually incorrect: the previously published validation analysis fully accounted for both type I and type II errors. The simulations that indicate a 75% type I error rate have serious flaws and only evaluate outdated versions of NCPA. These outdated type I error rates fall precipitously when the 2003 version of single locus NCPA is used or when the 2002 multi-locus version of NCPA is used. It is shown that the treewise type I errors in single-locus NCPA can be corrected to the desired nominal level by a simple statistical procedure, and that multilocus NCPA reconstructs a simulated scenario used to discredit NCPA with 100% accuracy. Hence, NCPA is a not a failed method at all, but rather has been validated both by actual data and by simulated data in a manner that satisfies the published criteria given by its critics. The critics have come to different conclusions because they have focused on the pre-2002 versions of NCPA and have failed to take into account the extensive developments in NCPA since 2002. Hence, researchers can choose to use NCPA based upon objective critical validation that shows that NCPA delivers what it promises.

Keywords

nested clade analysis; phylogeography; statistics; computer simulation

Knowles (2008) asks the question “Why does a method that fails continue to be used?” The method in this case is nested clade phylogeographic analysis (NCPA), and she is particularly concerned about the many authors and coauthors who have used this method even though they cite a paper that she feels shows that the method is a failure (Knowles and Maddison 2002). She answers this question with another question, asking the entire phylogeography community “how long will a field cling to an ideal rather than requiring objective critical validation”. There is another answer to her question that does not denigrate the objectivity of hundreds of scientists; namely, that the method does not fail. To see how this alternative answer arises, I must first examine her arguments that NCPA is a failed method.

ACCURATE AND EXTENSIVELY VALIDATED?

The inferences obtained by NCPA have been extensively validated by applying the method to 150 cases of positive controls; that is, cases where prior information exists about actual historical events that occurred (Templeton 1998; Templeton 2004). This method validates NCPA in the most relevant way: how it behaves with real data and actual historical events. Knowles severely criticizes this method of validation as follows (Knowles 2008, pg. 2714):

E-mail: E-mail: temple_a@wustl.edu Telephone: 314−935−6868 FAX: 314−935−5125.

> What was not included in this tabulation/validation procedure was how many times processes other than those that were expected were also inferred, which is the most salient result of the simulation studies—NCPA repeatedly infers processes when no such events have occurred (Knowles and Maddison 2002; Panchal and Beaumont 2007). The argument that NCPA has been extensively tested and shown to be accurate (Templeton 2004, 2008) is based on blatantly confusing type I and type II errors (Sokal and Rohlf 1995). The simulation studies both clearly show that NCPA incorrectly identifies significant geographic associations at a disturbingly high rate, which leads to inferences about process that never occurred. This finding cannot be rebuffed by the argument that NCPA has a high rate of detecting an expected fragmentation or range expansion (i.e., a high rate of true positives) and a low rate of failing to detect an expected fragmentation or range expansion (i.e., a low rate of false negatives)(see Templeton 2004). This logic is fundamentally flawed.

I agree completely with Knowles that a legitimate validation analysis must deal with the inferences with no prior expectations and must estimate both type I and type II errors. My sole disagreement is with the characterization of what I did in my validation analysis. Here is what I *really* did (Templeton 2004, pg. 792):

> Appendix 1 presents *all* the inferences from NCPA with respect to historical events, both fragmentation and range expansion, whether or not they were predicted *a priori*. For example, Appendix 1 shows that the NCPA of the fish *Galaxias truttaceus* inferred two range expansion events, one predicted by outside evidence (their current range includes lakes created by melting Pleistocene glaciers) and one that did not (an unpredicted range expansion to the north coast of Tasmania). All events inferred by NCPA that were not predicted by outside information are regarded as false positives. No inferred historical events of any sort are excluded from this analysis.

The “false positives” in the above quotation are the type I errors. The procedure used in Templeton (2004) provides an *upper* bound to the type I error rate because some of the unexpected inferences may actually be true. As to “blatantly confusing type I and type II errors”, there are eight tables in Templeton (2004); Table 8 refers to the simulation results of Knowles and Maddison (2002) and the other seven tables give various results from the validation analysis. Six of these seven tables tabulate the following four categories: true positives, false negatives, false positives, true negatives. Hence, this analysis deals extensively, exhaustively and explicitly with both type I (false positive) and type II (false negative) errors. Ironically, on page 2715, Knowles quotes my type I error rate of 23%, which comes from Table 5 in Templeton (2004) – the very same paper that Knowles on page 2714 claimed had no such tabulation.

When this misunderstanding of how I actually performed the validation analysis is corrected, it becomes apparent that the validation analysis was performed in an appropriate fashion according to the criteria given in Knowles (2008). There is no other method of phylogeographic inference that has been subjected to such a thorough analysis of both type I and type II errors, so I stand by my claim that NCPA is accurate and extensively validated, now knowing that this claim is fully consistent with the criteria articulated so well by Knowles (2008).

This validation analysis also corrects another factual error in Knowles (2008). One of the great strengths of NCPA is that it can generate complex phylogeographic histories from multiple simple inferences, with each inference treated as a testable null hypothesis and with quantified statistical support (Templeton 2008b). Knowles (2008, pg. 2717) states “There is not a single analysis, simulation or otherwise, that has shown that NCPA can accomplish what it is purported to do—infer multiple historical processes that may characterize a species’ history.”

This is not true. I tested the null hypothesis that multiple historical events in a species' history do not interfere with one another in the NCPA statistical inference structure (Tables 6 and 7 in Templeton, 2004) and found that there was indeed no detectable interference, which falsifies Knowles' claim.

ERRONEOUS ERROR RATES OF 75%?

The second part of Knowles's argument that NCPA is a failed method is that the type I error rates are 75% or more, according to computer simulations (Knowles and Maddison 2002; Panchal and Beaumont 2007). First consider Panchal and Beaumont (2007). This study confounds three sources of false positive errors. The first source of error arises from unrealistic simulations. The simulation program that they used, SIMCOAL, only allows the use of unrealistic mutational models Templeton (2008b). Simulations with unrealistic mutational models are known to generate false positive phylogeographic inferences (Palsbøll et al. 2004), so there is no justification for ignoring this potential source of error. Moreover, in half of their simulations they assumed exhaustive sampling of every local deme in their species, and in the other half they assumed 50% coverage of all local demes. Of all the hundreds of data sets that I am familiar with, both from my own lab and from other labs, no one has every claimed 100% deme sampling, or even 50%. These are exceedingly unrealistic assumptions. These assumptions eliminate or minimize geographical sampling as a source of error, thereby creating artificial power. Eliminating geographical sampling error by assumption induces artificial type I error rates between 55−60% (Templeton 2004), and Panchal and Beaumont (2007) themselves showed a highly significant effect on type I error rates due to the two sampling proportions that they used. Hence, there is no doubt that the 75% figure is inflated by these unrealistic sampling assumptions. Second, Panchal and Beaumont (2007) do not use the inference key that has been legitimately validated by the criteria given in Knowles (2008); rather, they use their own unvalidated inference algorithm. I know from personal experience that the inferences emerging from their algorithm can be discrepant with inferences drawn from the validated inference key, and I know from direct communication that other users of NCPA have also encountered discrepancies. Hence, there is no justification for ignoring an unvalidated inference algorithm as a source of additional error. Finally, the errors could be due to NCPA itself, which are certainly much higher than the nominal rate of 5%, as shown in Templeton (1998, 2004). The design used by Panchal and Beaumont (2007) confounds these three sources of error, and it is logically impossible to attribute all of the errors to NCPA alone.

Panchal and Beaumont (2007) also claimed that their simulation results mimicked and could explain the frequency and pattern of inferences observed in the NCPA of real data sets; indeed, they devote an entire section of their paper to this claim. However, they did not actually test their claim as a null hypothesis. I (Templeton 2008a) first tested the null hypothesis that their frequency of type I errors was homogeneous with that observed in the data sets used in the validation analysis (Templeton 2004), and I strongly rejected their claim (probability ≤ 0.0035). Knowles (2008) dismisses this by saying that it shows only a difference in error rates, which indeed is all this test was intended to show. However, the principle argument made by Knowles against NCPA is based on a difference in error rates, and this test result falsifies her claims about error rates. Moreover, I (Templeton 2008a) presented the results of two other statistical tests of the claim made by Panchal and Beaumont (2007) that the false positives in their simulations can explain the inferences observed in real data sets with NCPA. These two additional statistics test null hypotheses related to the expected spatial/temporal patterns of type I errors irrespective of the frequency of type I errors. These two pattern tests also strongly falsified the claim of Panchal and Beaumont (2007). Beaumont and Panchal (2008) reacted to this triple falsification by rejecting the premise of the entire section of Panchal and Beaumont (2007) that compared their simulated results to actual results, now arguing that actual data and simulated data are so different that no valid comparison can be made. I fully accept this new

position of Beaumont and Panchal (2008) because it is exactly the same conclusion that I drew from the three statistical tests of the claim in Panchal and Beaumont (2007). Hence, all agree now that the 75% figure does not reflect the behavior of NCPA with real data.

The second set of simulations referred to by Knowles (2008) is that given in Knowles and Maddison (2002), which reported a type I error rate of 75−80%. I have already examined this claim in detail (Templeton 2004), so I will give only a brief summary here. First, it is based upon only 10 simulation runs, a number that is inadequate to estimate error rates with statistical credibility. Second, they simulated a situation in which every local population was fragmented from all other local populations, and this case was specifically *excluded* from the inference key for NCPA (pg. 773, Templeton et al. 1995), a fact not mentioned in Knowles and Maddison (2002) nor by Knowles (2008). Third, Knowles and Maddison (2002) assumed that they sampled just 40 individuals from a total population of 40,000, but that this sparse sampling represented an exhaustive sampling of every local population in the species. As mentioned above, exhaustive deme sampling is unrealistic in that it does not correspond to the sampling situation of a single real data set. When one assumes the more realistic case that there were local populations that were not sampled, the type I error rate drops to 18%, a result consistent with the 23% error rate (Templeton 2004). Knowles (2008) counters these results by claiming that I "assumed a new evolutionary history that differed from the one used to simulate the data" (pg. 2716, Knowles 2008). This is not true. I used the *output* files from their simulations, and I did not change in any way the simulations themselves – only the sampling assumption that had no impact whatsoever upon the actual simulations.

An additional flaw in Knowles and Maddision (2002) is that is it out-of-date. The original NCPA inference key was published in 1995, the first major changes were made in 1998, and the next set of major changes were made in 2003. Minor changes are constantly made and updated at the GEODIS (the program that implements NCPA) website. One major difference between the 1998 key and its minor updates versus the 2003 key and its minor updates was that many more ways of inferring fragmentation were incorporated into the 2003 key, due mostly to user suggestions. As a result, the 2003 key and thereafter could deal more effectively with the type of micro-fragmentation that was *specifically excluded* from the key used by Knowles and Maddison (2002). The 2002 analysis by Knowles and Maddison is therefore irrelevant to Knowles' (2008) criticism of current usage of NCPA. Table 1 presents an updated analysis of the simulation results of Knowles and Maddison (2002) using the 2003 inference key (the same results are obtained with the current key) with nothing else altered from the original analysis, including the sampling assumptions of Knowles and Maddison (2002). Hence, there is no possibility of this analysis being subject to the criticism that it alters the scenario simulated by Knowles and Maddison (2002); rather, it merely makes their simulations relevant to the version of NCPA that has existed for over the past five years. The results of using the 2003 key are shown in Table 1. The two range expansions in Table 1 are clearly type I errors, and the 18 inferred fragmentation events are clearly true positives. The inferences of restricted gene flow require a more detailed examination. In their simulations, Knowles and Maddision (2002) start with a single ancestral population, which then splits into two, with each isolate subsequently splitting again at a later time (see Figure 2 in Knowles 2008). During the time between these two splits, there is the opportunity for gene flow, but it is restricted to being between populations 1 and 2 and between populations 3 and 4. All five of the inferences of restricted gene flow are consistent with these geographical constraints, but four of them are also consistent with patterns of gene flow that violate these constraints. Hence, these four will be counted as type I errors. Type I errors therefore occurred in 21% of the significant results in the post-2002 version of single locus NCPA: a figure substantially lower than that given in Knowles and Maddison (2002).

The 75% figure has no logical validity as it stems solely from the results of Panchal and Beaumont (2007) that confound three sources of error and is significantly inconsistent with real data in both frequency and pattern. The 75% figure is not supported by the simulations of Knowles and Maddision (2002) for any version of the NCPA inference key used since 2003. Hence, the only credible type I error rate for single locus NCPA as used today is around 23%, as shown by both simulated and real data.

THE 23% SOLUTION

Knowles (2008) argues that even the 23% type I error rate is still too high to inspire confidence in NCPA. I agree with her completely. Indeed, I noticed that the type I error rates were much higher than the nominal 5% rate when I did the initial validation analysis on a smaller number of real data sets (Templeton 1998). I therefore published a solution to this problem (Templeton 2002) *before* the publication of Knowles and Maddision (2002). However, I will first discuss a more recent solution that I have suggested (Templeton 2008a).

The unit of statistical analysis in NCPA is the nesting clade and not the total haplotype tree. Hence, the nominal type I error rate technically refers to a single nesting clade and not to all the clades in a tree. When the nominal rate is set to 5%, the type I error rate is 4% for nesting clades when validated from actual data (Templeton 2004); that is, the type I error is at the expected level for nesting clades. The problem of an inflated error rate is therefore not at the level of nesting clades, but rather at the level of analyzing an entire haplotype tree with multiple clades embedded within it. One of the strengths of a nested design is that all of the nested categories are independent under the null hypothesis of no associations of haplotypes with geography (Templeton 2008a). Therefore, a treewise type I error rate of α is achieved when one regards as significant only those tests *within* a nesting clade that have a probability less than α' where (Sokal and Rohlf 1995):

$$\alpha' = 1 - (1 - \alpha)^{1/k} \qquad 1$$

and where k is the number of nesting clades analyzed with GEODIS for a given haplotype tree. Beaumont and Panchal (2008) feel that some correction is necessary for the multiple tests within a nesting clade, and indeed I showed how that problem could also be solved with another test correction (Templeton 2008a). However, I did argue that this additional correction is not necessary because each nesting clade yields only a single inference. Beaumont and Panchal (2008) dismissed equation (1) without actually testing it with real data, so I do so here. Because the output of GEODIS gives the exact probabilities for all test statistics, it is easy to apply equation 1 retroactively. I did so with the data sets used in the validation study (Templeton 2004), setting the nominal level to 5%. With this test correction, the treewise type I error rate in the validation study is 2.0% – a value not significantly different from the nominal level of 5%. Hence, the problem of an inflated type I error at the level of the haplotype tree is a solved problem for single-locus NCPA.

One disadvantage of using equation (1) is that although it decreases type I errors, it also increases type II errors. Therefore, when I realized in 1998 that the treewise type I error rate was too high, I devised an alternative correction that would make use of multi-locus data, as it was apparent to me that the future of phylogeography was going to be with multi-locus data sets. A multi-locus, cross-validation procedure was created that reduces both type I and type II errors (Templeton 2002; Templeton 2004). Multi-locus NCPA has many optimal statistical properties that are discussed elsewhere (Templeton 2008b), so here I will give only one example to illustrate its properties: the simulations of Knowles and Maddison (2002). The ten replicates that they simulated can be regarded as ten different loci in a single multi-locus NCPA.

This example is actually a very difficult case. In these simulations each local isolate had an inbreeding effective size of 10,000, and the time between fragmentation events was 5,000 generations, with the total time of the entire simulated process being 10,000 generations. Given that the expected coalescence time within isolates of inbreeding effective size 10,000 is 40,000 generations for an autosomal gene or 20,000 for mtDNA (assuming the 10,000 is the effective size of females), the parameter choices of Knowles & Maddison (2002) ensure the retention of much ancestral polymorphism across isolates. Inferring temporally shallow fragmentation events with extensive retention of ancestral polymorphisms across isolates is difficult for any technique. Indeed, Knowles & Maddison (2002) reported that their own phylogeographic tests had "poor performance" with these simulated data sets.

The first step in multi-locus NCPA is to retain only those inferences that are cross-validated by two or more loci in type of inference (disregarding such inferences as "inconclusive", "inadequate sampling", etc.). Table 1 gives all the inferences obtained across all 10 "loci" when all the simulation and sampling assumptions given by Knowles and Maddison (2002) are retained without any alterations at all. The inferences in Table 1 that are cross-validated by type are fragmentation, restricted gene flow, and range expansion – all with two or more loci supporting these types of inference. Second, one next retains within each inference-type category only those inferences that are geographically cross-validated by two or more loci. For the two cases of range expansion in Table 1, one is inferred from locus 6 in the simulation and involves expansion from population 1 into sites 2 and 3. The second range expansion event is inferred from locus 7 to be an expansion from population 3 into sites 1 and 4. Hence, these two range expansion events are not geographically concordant and therefore are excluded. The geographical concordance of the fragmentation and gene flow events is more difficult to gauge because of the extensive retention of ancestral polymorphism in these simulations. For fragmentation, one establishes geographical cross-validation by tabulating how many inferred fragmentation events imply that a single specific population is isolated from one or more of the others, how many pairs of populations are inferred to behave as a single isolate from one or more of the other populations, and how many triplets of populations (the largest logical group in a 4-unit system) are inferred to be isolated from the remaining population. These tabulations are given in Table 2 (note, the total numbers are greater than 18 because not all of these categories are mutually exclusive). As can be seen, the only geographically cross-validated fragmentation inferences are that each of the four populations acts as an isolate, and the pairs (1, 2) and (3, 4) have also acted as isolates. Table 3 gives a similar tabulation of which populations were inferred to have been interconnected by gene flow. Obviously, gene flow involves at least two populations, so single populations are impossible with this inference type. Since all inferences are of restricted gene flow, all four populations are also excluded. As can be seen from Table 3, there are only two inferences of cross-validated gene flow; between populations 1 and 2, and between populations 3 and 4.

Most actual data sets subjected to multi-locus NCPA also have outgroup data for each locus; e.g., humans (Templeton 2002, 2005, 2007), African elephants (Templeton 2008b), and lizards (Gifford and Larson 2008). [Note, Knowles (2008) incorrectly claims that there is not a single multi-locus NCPA with cross-validation in the literature – but such papers have been in the literature since 2002, with the first one being a full-length article in *Nature*.] The use of outgroup data allows a formal testing of temporal concordance for all inferences. Moreover, the same maximum likelihood framework developed for testing temporal concordance allows the statistical strength of all cross-validated inferences to be quantified and provides great flexibility in testing a wide variety of other phylogeographic null hypotheses (Templeton 2008b). Unfortunately, no outgroup data were simulated in this case, so the simulated data of Knowles and Maddison (2002) cannot show the full strength of multi-locus NCPA. However, even in the absence of outgroup data, some temporal inferences are possible. Sometimes a single locus contains two or more inferences of fragmentation in different nesting clades. For

example, locus 8 at the total cladogram level inferred the (1, 2) isolate, and locus 8 at the clade 2−1 level inferred populations 1 and 2 to be isolated from one another. Because the inference of 1 and 2 being isolated from one another is nested within the inference of (1, 2) being an isolate, the temporal polarity of the fragmentation events had to have been (1, 2) fragmented from (3, 4), followed later by 1 fragmented from 2. Similar nesting patterns exist at other loci (e.g, locus 0) that show that the (1, 2) vs. (3, 4) fragmentation event is older than the 3 vs. 4 fragmentation event. Similarly, for the restricted gene flow inferences, the gene flow between 1 and 2 is indicated by one locus at the total cladogram level, and for the other 2 loci at the 3-step level. Gene flow between populations 3 and 4 is indicated at the total cladogram level for two loci, and only for one locus at the 2-step level. These high-order nesting associations imply that these gene flow events are old ones.

All these cross-validated inferences can be placed into a single phylogeographic history: the population pair (1, 2) was fragmented from the population pair (3, 4); followed by a period of time in which gene flow was restricted to be between 1 and 2 and between 3 and 4 but not between (1, 2) and (3, 4); followed by the fragmentation of 1 from 2 and the fragmentation of 3 from 4. This is a 100% accurate reconstruction of the events simulated by Knowles and Maddison (2002): there are no type I errors and no type II errors in this multi-locus NCPA of simulated data. Hence, even though this is an information-poor simulated data set compared to those found in real multi-locus data sets, multi-locus NCPA could not have performed better.

THE ANSWER

Knowles asks “why does a method that fails continue to be used?” and “how long will a field cling to an ideal rather than requiring objective critical validation?” My answer to these questions is that NCPA has not been shown to fail. To the contrary, NCPA has been subjected to extensive, objective critical validation both through the analysis of actual data sets in a manner that satisfies the requirements for legitimacy as stated by Knowles (2008) and by the very same computer simulations of Knowles and Maddison (2002) that Knowles (2008) cites as an argument against NCPA. Her argument that these simulations undermine NCPA is based on using a version of NCPA that has not been used for over 5 years; yet she applies this outdated analysis to current use. The problem of inflated type I errors was first recognized in 1998 (Templeton 1998) and was solved by the development of multi-locus cross-validation *before* the publication of any of the critiques of type I error in NCPA. The critics of NCPA have steadfastly chosen to focus only on the pre-2002 version of NCPA and have never addressed the current, multi-locus version. The critics of NCPA also make many claims against NCPA, but do not test these claims with real data. When their claims are tested, they are consistently and strongly falsified. Finally, mutli-locus NCPA has been shown to have many optimal statistical properties and provides a robust framework for testing null, phylogeographic hypotheses (Templeton 2008b). Hence, I continue to use NCPA precisely because I have demanded and produced objective critical validations in a manner that fully satisfy the published criteria for legitimacy given by Knowles (2008) and Knowles and Maddison (2002).

ACKNOWLEDGMENTS

I thank Allan Larson, Jared Strasburg, Paul Sunnucks, and the graduate students in my lab for their many useful comments. This work was supported in part by NIH grant P50-GM65509.

LITERATURE CITED

Beaumont MA, Panchal M. On the validity of nested clade phylogeographical analysis. Mol. Ecol 2008;17:2563–2565. [PubMed: 18482264]

Gifford ME, Larson A. *In situ* genetic differentiation in a Hispaniolan lizard (*Ameiva chrysolaema*): A multilocus perspective. Mol. Phylogenet. Evol 2008;49:277–291. [PubMed: 18611442]

Knowles LL. Why does a method that fails continue to be used? Evol 2008;62:2713–2717.

Knowles LL, Maddison WP. Statistical phylogeography. Mol. Ecol 2002;11:2623–2635. [PubMed: 12453245]

Palsbøll PJ, Berube M, Aguilar A, Notarbartolo-Di-Sciara G, Nielsen R. Discerning between recurrent gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean Sea fin whale (*Balaenoptera physalus*) populations. Evol 2004;58:670–675.

Panchal M, Beaumont MA. The automation and evaluation of nested clade phylogeographic analysis. Evol 2007;61:1466–1480.

Sokal, RR.; Rohlf, FJ. Biometry. W. H. Freeman & Company; San Francisco: 1995.

Templeton AR. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol 1998;7:381–397. [PubMed: 9627999]

Templeton AR. Out of Africa again and again. Nat 2002;416:45–51.

Templeton AR. Statistical phylogeography: methods of evaluating and minimizing inference errors. Mol. Ecol 2004;13:789–809. [PubMed: 15012756]

Templeton AR. Haplotype trees and modern human origins. Yearbook of Physical Anthropology 2005;48:33–59.

Templeton, AR. Population biology and population genetics of Pleistocene Hominins.. In: Henke, W.; Tattersall, I., editors. Handbook of Palaeoanthropology. Springer-Verlag; Berlin: 2007. p. 1825-1859.

Templeton AR. Nested clade analysis: an extensively validated method for strong phylogeographic inference. Mol. Ecol 2008a;17:1877–1880. [PubMed: 18346121]

Templeton AR. Statistical hypothesis testing in intraspecific phylogeography: NCPA versus ABC. Mol. Ecol. 2008bin press

Templeton AR, Routman E, Phillips C. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum*. Genetics 1995;140:767–782. [PubMed: 7498753]

Table 1

The type and number of inferences obtained from the simulation output files of Knowles and Maddison with the 2003 inference key under the original sampling assumptions of Knowles and Maddison (2002).

Inference:	Inconclusive	Inadequate Genetic Resolution	Fragmentation	Restricted Gene Flow	Range Expansion
Number:	1	2	18	5	2

Evolution. Author manuscript; available in PMC 2009 June 9.

Table 2

The populations inferred to be isolates relative to one or more other populations for the 18 fragmentation events given in Table 1. The geographically validated inferences are indicated in bold.

Single Isolates		Pairs Acting as Isolates		Triplets Acting as Isolates	
Population	**No. of Inferences**	**Populations**	**No. of Inferences**	**Populations**	**No. of Inferences**
		(1, 2)	**3**		
1	**7**	(1, 3)	1	(1, 2, 3)	1
2	**6**	(1, 4)	1	(1, 2, 4)	0
3	**6**	(2, 3)	1	(1, 3, 4)	1
4	**6**	(2, 4)	1	(2, 3, 4)	0
		(3, 4)	**3**		

Table 3

The populations inferred to have gene flow for the 5 restricted gene flow events given in Table 1. The geographically validated inferences are indicated in bold.

Gene Flow Between Pairs of Populations		Gene Flow Among Three Populations	
Populations	**No. of Inferences**	**Populations**	**No. of Inferences**
(1, 2)	**3**		
(1, 3)	1	(1, 2, 3)	1
(1, 4)	1	(1, 2, 4)	1
(2, 3)	1	(1, 3, 4)	0
(2, 4)	1	(2, 3, 4)	0
(3, 4)	**3**		