# Statistical phylogeography: methods of evaluating and minimizing inference errors

ALAN R. TEMPLETON

*Department of Biology, Campus Box 1137, Washington University, St Louis, Missouri 63130–4899, USA*

## Abstract

**Nested clade phylogeographical analysis (NCPA) has become a common tool in intraspecific phylogeography. To evaluate the validity of its inferences, NCPA was applied to actual data sets with 150 strong a priori expectations, the majority of which had not been analysed previously by NCPA. NCPA did well overall, but it sometimes failed to detect an expected event and less commonly resulted in a false positive. An examination of these errors suggested some alterations in the NCPA inference key, and these modifications reduce the incidence of false positives at the cost of a slight reduction in power. Moreover, NCPA does equally well in inferring events regardless of the presence or absence of other, unrelated events. A reanalysis of some recent computer simulations that are seemingly discordant with these results revealed that NCPA performed appropriately in these simulated samples and was not prone to a high rate of false positives under sampling assumptions that typify real data sets. NCPA makes a posteriori use of an explicit inference key for biological interpretation after statistical hypothesis testing. Alternatives to NCPA that claim that biological inference emerges directly from statistical testing are shown in fact to use an a priori inference key, albeit implicitly. It is argued that the a priori and a posteriori approaches to intraspecific phylogeography are complementary, not contradictory. Finally, cross-validation using multiple DNA regions is shown to be a powerful method of minimizing inference errors. A likelihood ratio hypothesis testing framework has been developed that allows testing of phylogeographical hypotheses, extends NCPA to testing specific hypotheses not within the formal inference key (such as the out-of-Africa replacement hypothesis of recent human evolution) and integrates intra- and interspecific phylogeographical inference.**

*Keywords*: cross-validation, haplotype trees, hypothesis testing, nested clade analysis, phylogeography, statistical inference

*Received 29 June 2003; revision received 25 September 2003; accepted 8 October 2003*

## Introduction

Intraspecific phylogeography deals with a species' evolutionary history over space and time. The era of modern phylogeographical studies began with the pioneering work of Avise *et al.* (1979) on mitochondrial DNA (mtDNA) variation in the mouse genus *Peromyscus*. In this and subsequent studies genetic variation in mtDNA was scored, and the resulting haplotypes were used to estimate an evolutionary tree that portrays the accumulation of mutations in DNA lineages over time to yield the current array of sampled haplotypes. These haplotype trees were then overlaid upon geography to make phylogeographical inferences.

Although visual overlays of haplotype trees upon geography can be suggestive of phylogeographical events or processes, such overlays do not constitute a formal estimation or hypothesis testing framework. There is no determination of whether or not enough individuals and geographical sites have been sampled to ensure that the observed patterns could not have arisen by chance alone. Even when one accepts the patterns as real, there is no formal, explicit interpretative framework for making biological conclusions from the observed patterns. These inadequacies were both addressed through the development of nested-clade, phylogeographical analysis (NCPA) (Templeton *et al.* 1995).

Correspondence: Alan R. Templeton. Fax: 1 314 935 4432; E-mail: temple_a@biology.wustl.edu

NCPA uses the haplotype tree to define a series of hierarchically nested clades (branches within branches) using a set of explicit nesting rules (Templeton *et al.* 1987; Templeton *et al.* 1992). Haplotypes are the lowest units of analysis, being nested together into mutationally close subsets called one–step clades. The one–step clades in turn are nested into two–step clades, and so on, until a nesting level is reached such that the next higher nesting level would result in only a single clade spanning the entire original haplotype network. The nested design captures the most reliable temporal information contained in the haplotype network. When the haplotype network is properly rooted, the oldest clade is known in any given nesting category. Even if the haplotype tree were unrooted, coalescent theory predicts that clades on the tips of the tree are highly likely to be younger than the interior clades to which the tips are connected within a population or set of populations well connected by gene flow (Castelloe & Templeton 1994). Within a nesting category, contrasts of interiors or the oldest clade vs. the tips or younger clades therefore constitute a temporal contrast that does not depend upon a molecular clock or any sort of rate calibration. There is a role for making use of a molecular clock in multilocus NCPA studies (Templeton 2002), but the NCPA of any individual DNA region uses the nesting design to make temporal contrasts of spatial information in a manner independent of a clock.

NCPA next quantifies the spatial distribution of haplotypes and clades through two distance measures (Templeton *et al.* 1995). The clade distance, $D_c$, measures the spatial spread of the clade. This distance, as all others used in NCPA, can be based either upon geographical distance or upon user-input distances (e.g. river distances for a riparian species). The nested clade distance, $D_n$, quantifies how far away a haplotype or clade is located from those haplotypes or clades with which it is nested into a higher level clade.

Because of sampling artefacts, it is dangerous to make biological inferences from a visual overlay of a haplotype tree upon geography or from just the observed values of quantitative distance measurements. For example, suppose a geographically widespread species is characterized by gene flow with isolation by distance. However, if one sampled local populations only from two separate geographical clusters that were very distant from one another and did not sample geographically intermediate populations, the result would be two subsets of local populations similar within but highly genetically differentiated between the geographical clusters. Such a pattern could be confused with fragmentation, and the pattern associated with isolation by distance would become apparent only when the geographically intermediate populations were sampled. Nested clade analysis therefore addresses the issue of sampling adequacy in two distinct phases: (1) is

sampling adequate to detect a significant association between clades and geography; and (2) is sampling adequate to interpret biologically any detected significant associations between clades and geography?

To address the first question, concerning sampling adequacy, the nested clade analysis quantifies the degree of confidence in the quantitative distance measures by testing the null hypothesis that the haplotypes or clades nested within a high-level nesting clade show no geographical associations given their overall sample numbers. This null hypothesis is tested by permuting the observations randomly within a nesting clade across geographical locations in a manner that preserves the overall clade frequencies and sample sizes per locality (Templeton *et al.* 1995). After each random permutation, the clade and nested clade distances are recalculated. The distribution of these distances under the null hypothesis of no geographical association for a fixed frequency is simulated by repeating this procedure 1000 or more times. The observed clade and nested clade distances are then contrasted to this null distribution, and the algorithm then infers which distances are statistically significant. Statistical power can be enhanced within a nesting clade by taking the average of the clade or nested clade distances for all the tips pooled together and subtracting the tip average from the corresponding average for the older interiors. The average interior–tip difference captures the temporal contrast of old vs. young within a nesting clade, but often has greater power to reject the null hypothesis of no geographical association with the distance measurements.

To address the second question, concerning sampling adequacy, the nested clade analysis executes an exhaustive examination of all the statistically significant associations detected in phase one with explicit, a priori criteria for biological interpretation and possible sampling artifacts. Statistical significance alone does not provide an interpretation for those geographical associations. Indeed, no single test statistic discriminates between recurrent gene flow, past fragmentation and past range expansion in NCPA; rather, it is a pattern formed from several statistics that allows discrimination. Also, many different patterns can sometimes lead to the same biological conclusion because a single evolutionary event or process can have multiple genetic impacts. Moreover, as pointed out above, a statistically significant pattern can still be biologically ambiguous because of inadequate geographical sampling. Finally, NCPA searches out multiple, overlaying patterns within the same data set. In light of these complexities in biological and sampling interpretation (which reflect reality), an inference key was provided as an appendix to Templeton *et al.* (1995), with the latest version being available at http://bioag.byu.edu/zoology/crandall_lab/geodis.htm along with the program GEODIS for implementing the nested clade analysis.

Although the nested clade approach to phylogeographical inference has many strengths, it does have limitations. First, inference is limited by sample size and sample sites. Because biological interpretation is limited to those distance statistics that result in a significant rejection of the null hypothesis of no geographical association, the ability to make inference in NCPA is obviously limited by sample size. In addition, even when significant geographical associations are detected, the inference key may lead to the conclusion that there has been an inadequate geographical sampling for unambiguous biological interpretation. When these sampling limitations are encountered, an investigator can circumvent them only by additional sampling, and the NCPA key provides specific guidance for future sampling efforts.

A second limitation is the possibility of insufficient genetic resolution to detect an event or process that actually occurred. For example, a known range expansion in *Drosophila buzzatii* was not detected by NCPA not because of inadequate sampling, but rather because an appropriate mutation had not occurred in the right place and time to mark the event (Templeton 1998a). Another example is provided by studies on a freshwater mussel that revealed NCPA could not make inference about recent gene flow among local populations due to a lack of genetic resolution (Turner *et al*. 2000). These examples show that no one DNA region can capture the totality of a species' population structure and recent evolutionary history because of the stochasticity of the mutational and coalescent process (Templeton 2002).

The third, and perhaps most serious, limitation arises when NCPA makes a false inference or biological misidentification. Templeton (1998a) validated the original 1995 inference criteria for range expansion by examining biological examples for which strong prior evidence existed for range expansion. Overall, NCPA did well with these a priori expectations, but these same worked examples reveal that NCPA sometimes fails to detect known range expansions and leads more rarely to false inferences. False inferences can arise from the evolutionary stochasticity of the coalescent process itself, from the haplotype tree being skewed or otherwise altered by natural selection or from inadequacies in NCPA and/or its inference key. One major inadequacy of the original NCPA was its failure to incorporate secondary contact and hybridization, and supplemental test statistics were designed to fill this gap (Templeton 2001). Other, more minor difficulties with the inference key have been discovered over the years that were corrected by minor modifications (hence the version of the inference key on the website should be used rather than previously published versions).

The first purpose of this paper is to examine the adequacy and accuracy of the inference key by using the same approach given in Templeton (1998a); namely, using actual data sets

with strong prior evidence for certain phylogeographical events. Previously, only range expansion (including both contiguous range expansion and long distance colonization) was examined by this validation technique, but now both fragmentation and range expansion are considered. Moreover, cases are examined in which both, one or none of these classes of events are expected to have occurred, allowing a full analysis of errors and the potential for different events to confound one another.

Another approach to validation is with simulated data sets. Recently, the validity of NCPA inferences has been challenged on the basis of 10 simulations of fragmentation (Knowles & Maddison 2002). The merits of this particular set of 10 simulations will be discussed and the simulated data sets of Knowles & Maddison (2002) will be reanalysed by changing just one sampling assumption — an assumption used *only* in the inference key given the simulated data and *not* related to the conditions under which the simulations were conducted. In addition, the simulation results of Irwin (2002) on discriminating fragmentation from isolation by distance will be examined.

The NCPA inference key represents a formal separation of statistical testing vs. biological interpretation (but with interpretation being predicated upon statistical testing). Knowles & Maddison (2002) have also questioned the validity of such a separation, indeed arguing that it makes NCPA a nonstatistical approach. However, there is a separation of formal testing from interpretation in the entire area of statistics, and it will be shown that the approaches advocated by Knowles & Maddison (2002) are no exception.

The revised NCPA inference key has low error rates. However, as is common with any statistical procedure, errors can never fully be eliminated. Both false positives and the failure to detect events that occurred can be reduced by performing NCPA simultaneously upon many DNA regions rather than just one (Templeton 2002). Recently, a maximum-likelihood hypothesis-testing framework has been developed for cross-validation across DNA regions (Templeton 2003a). It will be shown that cross-validation provides a method for validating the specific inferences made for a particular species and geographical area.

Overall, there are many methods of validating inference in NCPA. These validation methods reveal that NCPA is reliable in general, and that specific inferences can be validated for particular species and geographical areas.

## Validating NCPA inferences using examples with prior expectations

There are many circumstances in which prior information generates strong phylogeographical expectations for a particular species and sets of locations. For example, if a species currently lives in an area that was uninhabitable during the Pleistocene, one can be confident that a range

expansion must have occurred into that area. Similarly, knowledge about a species' dispersal abilities and the existence of current or past barriers to dispersal can generate strong expectations of past or current fragmentation or its absence. Other information that can be used includes previous genetic studies, multispecies vicariance biogeographical studies, etc. Often, a single species will generate many prior predictions, although only those predictions that are well supported by other data or prior knowledge are used in this analysis. In some cases, research groups have applied NCPA specifically to organisms with explicit prior expectations to test the validity and accuracy of NCPA (Paulo *et al.* 2002; Masta *et al.* 2003), and the results of such studies are incorporated into the current survey. What is critical in all these cases is that the information used to generate prior expectations is exterior to the data set being subjected to NCPA.

The only data sets included in this survey were those with a statistically significant rejection of the null hypothesis of no association between the haplotype tree and geography as ascertained through the clade and/or nested clade distances (that is, phase one of sampling adequacy is satisfied). This limitation biases the analyses of those cases in which prior evidence indicates no range expansion and no fragmentation. Such cases are drawn primarily from species with good dispersal abilities over the sampled area, often with a continuously distributed population. In such cases, the acceptance of the null hypothesis of no association between clades and geography may well be the most appropriate biological outcome (indicating effective panmixia over the sampled area), but one could argue that the failure to reject the null hypothesis in such cases is instead due to inadequate sampling. By only including those analyses that reject the null hypothesis adequate sampling is ensured, but many potentially informative data sets with correct biological inferences are thereby excluded. The exclusion of those cases that indicate panmixia even when prior information is consistent with that inference will reduce the total number of cases considered (the denominator in the false positive rate) without affecting the number of false positives (the numerator in the rate), so the frequency of false positives in this analysis must be higher than the true false positive rate.

All together, 150 prior expectations were analysed by NCPA that involved fragmentation and range expansion events, the primary types of historical events dealt with by NCPA. This represents a substantial increase over the 18 data sets with prior information about range expansion used by Templeton (1998a) in the original validation of the NCPA inference key. The species involved, along with references, comments, prior expectations and NCPA inferences, are given in Appendix I. Of these 150 prior expectations, 82 came from data sets that had not previously been analysed by NCPA. In those cases, the NCPA is not in the references cited in Appendix I but is available upon request to the author. Also, 10 of the prior inferences were from data sets explicitly gathered to test the validity of NCPA (Masta *et al.* 2003; Paulo *et al.* 2002). Thus, about two-thirds of the prior expectations tested in this analysis (92 if 150) were chosen without any knowledge of what the NCPA results would be. The remaining third of the prior expectations came from papers that also performed NCPA, so the results of NCPA were known at the time they were chosen for inclusion. However, the sole criterion for inclusion in this study was the availability of evidence exterior to the NCPA to generate prior expectations and not the results of the NCPA itself.

The original inference key for NCPA deals with three basic types of events or processes: (1) fragmentation events, (2) range expansion events and (3) restricted gene flow processes. In addition, NCPA can lead to an inconclusive inference with no unambiguous biological interpretation. Appendix I 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 I 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. Sometimes the same event (e.g. a range expansion into a particular area) was inferred from more than one clade, but these are counted as a single event in Appendix I. For example, as discussed in Templeton (1998a), the expected range expansion was detected in 12 of the 13 data sets with prior evidence for range expansion in the original 1998 survey, yielding a 92% success rate at the data set level. However, range expansion was inferred in 35 nesting clades in these 12 data sets of a total of 99 nesting clades with significant geographical association (most inferences were restricted gene flow). Of these 35 inferences of range expansion, 34 of them were consistent with prior expectations, for a 97% success rate at the level of nesting clades. The only exception was the expansion of *Galaxias truttaceus* to the north coast of Tasmania that was mentioned above. In contrast, a single clade yielding an inference for range expansion in a data set without prior evidence for range expansion is counted as an error. For example, in the original survey Templeton (1998a), 24 nesting clades had statistically significant geographical associations for the six data sets without prior expectation of range expansion, but only one led to the inference of range expansion. This leads to a 20% error rate (1/5) at the data set level, but to only a 4% error rate at the nesting clade level. This is to be expected. When an event actually occurred, it can have an impact upon many haplotypes

**Table 1** Expected and observed inferences of population fragmentation with the 2001 key

| Prior expectation for fragmentation | Inference obtained with the 2001 NCPA key | | |
|---|---|---|---|
| | Yes | No | Misidentified |
| Yes | 26 | 5 | 3 |
| No | 1 | 31 | |

**Table 2** Expected and observed inferences of range expansion with the 2001 key

| Prior expectation for range expansion | Inference obtained with the 2001 NCPA key | |
|---|---|---|
| | Yes | No |
| Yes | 38 | 17 |
| No | 13 | 17 |

scattered throughout the cladogram in a geographically concordant fashion. In contrast, when the inference is an artefact of sampling, evolutionary stochasticity or errors in the inference key, it is not likely to yield multiple inferences of the same event in a geographically concordant fashion. Hence, by collapsing together multiple inferences of the same event within a single data set, Appendix I reports the successful inferences of NCPA in a highly conservative fashion. This is appropriate because the data set level is the level at which most biological conclusions are drawn.

Concerning other inferences, all the data sets in which some sort of restricted gene flow or inconclusive result was inferred are indicated in Appendix I, but each single instance is not presented. For example, if three different nesting clades in a single data set lead to the inference of isolation by distance, all that would be listed in Appendix I is that isolation by distance was inferred for that data set. Because gene flow is biologically plausible in virtually all of these intraspecific data sets, no attempt was made to test the validity of the inference of restricted gene flow. Hence, this approach to validating NCPA is limited to inferences of historical events and excludes the validation of restricted gene flow processes.

Table 1 shows the performance of the 2001 version of the NCPA inference key (the last revision prior to this paper) for inferring fragmentation events. For 34 cases with prior evidence for fragmentation, the 2001 NCPA key identified 26 of them correctly, failed to identify five (that is, either no event was detected at all, or the biological interpretation was ambiguous) and misidentified three (all inferring some sort of long distance movement). For 32 cases with prior evidence indicating fragmentation to be unlikely, NCPA inferred no fragmentation in 31 cases. The single 'false' positive of inferred fragmentation reveals another limitation of this approach to validations: prior evidence *for* an event (either fragmentation or range expansion) is generally more reliable than evidence that an event did *not* occur. Evidence for lack of fragmentation was generally based on organisms with good dispersal abilities over the sampled area with no obvious barriers to dispersal. However, human judgements on current and past dispersal barriers may be incorrect, in which case this is not a false

positive at all but merely an unexpected fragmentation event that could give us more insight into the species' dispersal abilities and recent evolutionary history. For the purposes of this analysis, all positive inferences made without prior expectations are counted as an error, which further inflates the false positive rate. Even so, the results shown in Table 1 indicate that the 2001 inference key is unlikely to result in false positives for fragmentation. The most common error is failure to detect an event. This inference pattern can be quantified by pooling the misidentifications with the not inferred column in Table 1 and performing a contingency test on the resulting two × two table. The null hypothesis of such a contingency test is that NCPA produces random inferences about fragmentation. The two-tailed Fisher's exact test for Table 1 rejects this null hypothesis with a probability level of less than 0.0001. This low probability value indicates that almost all NCPA inferences about fragmentation events are on the diagonal (i.e. concordant with prior expectations) in the two × two table.

Table 2 shows the performance of the 2001 version of the NCPA inference key for inferring range expansion events. The misidentification of a fragmentation event as a long-distance colonization event is classified as a false positive in this analysis, so this error is counted twice in these analyses (both in Table 1 and Table 2). The two-tailed Fisher's exact test of the null hypothesis that NCPA produces random inferences about range expansion events has a probability value of 0.0361. This low probability is due to the fact that most observations lie on the diagonal; that is, most inferences of range expansion or its absence are concordant with prior expectations. Thus, the 2001 inference key performs significantly well in dealing with both fragmentation and range expansion events.

This good performance is not due to how the data sets were chosen. As mentioned above, 58 of the 150 prior inferences come from data sets analysed originally with NCPA and the other 92 come from data sets from either papers with no NCPA or from papers in which the data set was gathered explicitly to test prior expectations with NCPA. Table 3 shows how the NCPA inferences are distributed in a concordant vs. discordant manner with prior expectations in these two types of data sets. The Pearson $\chi^2$ statistic for this table is 2.027 with one degree of freedom, yielding

**Table 3** The impact of studies analysed originally with NCPA vs. those in which NCPA was performed afterwards or to test explicitly NCPA upon the concordance and discordance with prior expectations with the 2001 key

| Study originally analysed with NCPA | Inferences obtained with the 2001 NCPA Key compared to prior expectations | |
|---|---|---|
| | Concordant | Discordant |
| Yes | 47 | 11 |
| No | 65 | 27 |

a nonsignificant *P*-value of 0.1545 under the null hypothesis that the data set types have no impact on the concordance of the inferences with prior expectations. Similarly, this null hypothesis has a *P*-value of 0.1802 when evaluated with a two-tailed Fisher exact test. Thus, the inclusion of some data sets that had originally been analysed with NCPA has no detectable impact on the rate of concordance with prior expectations.

Although Tables 1 and 2 indicate that NCPA results in inferences about historical events that are significantly concordant with prior expectations, these tables also reveal that NCPA is making more errors for range expansion inference than for fragmentation inference. The first type of error, the failure to detect an event, has comparable sample rates for both types of inferences (24% for fragmentation, 31% for range expansion). The difference lies in the sample false positive rate, which was 3.1% for fragmentation and 43% for range expansion (once again, recall that the sample false positive rate in these studies is an overestimate of the true false positive rate). One explanation for this discrepancy is that it is easier to be confident that a fragmentation event did not occur compared to an expansion event, implying that many of the false positives shown in Table 2 are true expansion events that were cryptic to the prior evidence (Templeton 1998a). Alternatively, the inference key may have a higher false positive rate for range expansions. Indeed, an inspection of the results reveals some potential difficulties with the 2001 inference key with respect to range expansions.

Masta *et al*. (2003) pointed out explicitly one difficulty with the 2001 key. This involves the misidentification of an expected fragmentation event for range expansion through long-distance colonization, an error that appears both in Tables 1 and 2. The scenario expected from prior evidence in this case was past fragmentation of the population into isolates followed by extinction of intermediate areas and lineage sorting within the fragments. This was then followed by a more recent range expansion that brought the isolates back into contact. This yields a pattern of some haplotypes appearing in a geographical area far away from some of

their evolutionary neighbours, and the 2001 key led to an inference of long-distance colonization in a species for which long-distance movements are implausible. This particular deficiency of the inference key has been addressed partially by Templeton (2001), who gives a supplemental test for secondary contact that is based upon the distances from a particular sampling site to the geographical centres of the clades found at that sampling site (for other worked examples of this supplementary test, see Byrne *et al*. 2002 and Pfenninger & Posada 2002). The supplemental test of Templeton (2001) clarifies the situation here by indicating secondary contact. A similar problem arose with the other misidentification given in Table 1, a fragmentation event identified as long-distance dispersal (Paulo *et al*. 2002). Here, the prior evidence had indicated past fragmentation followed by lineage sorting and extinction in intermediate areas, but in this case there was no secondary expansion into the intermediate areas. Accordingly, a warning is now placed in the key that when these situations arise it could be due to either past fragmentation or long-distance movement, as well as a recommendation to perform the supplemental test if secondary contact is a possibility.

The discussion in Masta *et al*. (2003) and personal communications with Dr Eric Routman reveal another problem in the 2001 inference key; namely, that fragmentation and long-distance colonization were treated as separate events. However, the distinction between the two is not always clear. First, a long-distance colonization event results in fragmentation as well; that is, in an isolated colony. Second, range expansion followed by fragmentation (extinction of geographically intermediate populations) can also result in an isolated colony. For example, the collared lizard, *Crotaphytus collaris*, is found in the American Southwest and northern Mexico. Populations are also found in the Ozarks of Missouri and Arkansas on glades, areas of exposed bedrock associated with a desert-like microclimate (Hutchison & Templeton 1999). The collared lizards probably 'colonized' the Ozarks by gradually spreading northeast during the ipsothermal maximum, a period of hot, dry weather in central North America about 8000–5000 years ago (Hutchison & Templeton 1999). With the return of wetter weather the intervening populations went extinct, leaving the Ozark populations isolated from the main distribution range. Is this a long-distance colonization? The 2001 inference key would regard this as a long-distance colonization of the Ozarks, but Masta *et al*. (2003) would not, because it did not involve the direct movement of animals from the American Southwest into the Ozarks but rather a gradual range expansion followed by fragmentation through extinction of intermediate populations. The issue is therefore the meaning of 'long-distance colonization'. Does such a colonization demand the direct movement of the organisms from the ancestral area to the new colony, or can long-distance colonization also occur

by gradual movement during a period of range expansion followed by the extinction of intermediate populations? In either case, the product is an isolated colony that will yield similar patterns in NCPA, and hence these two cases are indistinguishable in NCPA (assuming that extinct, intermediate populations cannot be sampled using some palaeo-DNA technique). Accordingly, the inference key is reworded to acknowledge that fragmentation and long distance colonization are not necessarily mutually exclusive, but rather are often coincident and genetically indistinguishable. If it is desired to make the distinction between the mechanisms of organismal movement used in establishing the genetic pattern of a distant colony, Masta *et al.* (2003) suggest the use of outside information. The revised key therefore incorporates these additional suggestions by Masta *et al.* (2003) to aid in making such a distinction.

Although an effort was made to make the original inference key conservative (Templeton *et al.* 1995), some of the false positives in Table 2 indicate that the 2001 key is overinterpreting under marginally informative conditions. One such condition involves the pattern in which every clade in a nesting category is found in a unique location or set of locations with no overlap whatsoever with any other clade in the same nesting category. Such a pattern could be due to fragmentation, but it could also be due to very sparse sampling, and questions are inserted into the key to clarify this situation. The key also makes use of contrasts of interiors vs. tips, under the assumption that this is a temporal contrast of old vs. young. As pointed out earlier, this is a good assumption when dealing with a single population or set of populations interconnected by gene flow, but these conditions are increasingly likely to break down with increasing level of nesting, particularly at the highest nesting level (that is, the clades that when nested together span the entire haplotype tree) as these are the levels most commonly affected by past fragmentation. There are now known cases of haplotype trees rooted through outgroups in which the outgroup connects to a clade that would be considered a tip based only on the ingroup haplotypes (e.g. Fullerton *et al.* 2000). Whenever reliable rooting information is available, either through outgroups or outgroup probabilities (Castelloe & Templeton 1994), it should be used. In the absence of reliable rooting, the tip/interior status of the highest level clades should be regarded as ambiguous, making the inference key more conservative.

The above changes were made in the 2001 inference key as well as the incorporation of many suggestions made by various users to clarify the meaning of some of the questions. The revised version is attached as Appendix II to this paper and is posted on the GEODIS website. Inferences altered in the new key are indicated in Appendix I. The new inferences are summarized in Tables 4 and 5, which can be contrasted with Tables 1 and 2. The new key improves slightly the accuracy of inference for fragmentation.

**Table 4** Expected and observed inferences of population fragmentation with the 2003 key

| Prior expectation for fragmentation | Inference obtained with the 2003 NCPA key | |
|---|---|---|
| | Yes | No |
| Yes | 30 | 4 |
| No | 1 | 31 |

**Table 5** Expected and observed inferences of range expansion with the 2003 key

| Prior expectation for range expansion | Inference obtained with the 2003 NCPA key | |
|---|---|---|
| | Yes | No |
| Yes | 34 | 21 |
| No | 7 | 23 |

All three misidentifications are now eliminated, with one being ambiguous (and therefore counted as a failure to detect) and the other two now being correct after use of the supplemental test in Templeton (2001). One other fragmentation event that was not detected previously is now detected.

The new key has a more substantial effect on range expansion inference. The two-tailed probability of Fisher's exact test for Table 5 now has a probability of 0.0013, indicating a substantial decrease in the off-diagonal errors relative to the 2001 key. An examination of Table 5 reveals that there has been a modest increase in the rate of failure to detect a range expansion event. This is to be expected, as some of the modifications in the new key were designed specifically to make it more conservative about inferring range expansion events and, hence, there is less power in the new key to detect range expansion. However, in compensation for decreased power, there is a substantial decrease in the number of false positives. Therefore, inferences from the new key will be used in all subsequent analyses.

One important feature of NCPA is its ability to search for multiple events or processes because rarely would a species be subject to just one event or process throughout its evolutionary history. Although NCPA can and does make multiple inferences, to date there has been no assessment of the impact of one event upon the ability to infer another. To see if range expansion affects the accuracy of NCPA inferences about fragmentation, the data in Table 4 were partitioned into those cases with prior evidence for range expansion in the same data set vs. those without such evidence. The results are given in Table 6. To test the impact

**Table 6** Expected and observed inference of population fragmentation as a function of prior evidence for range expansion

| | No prior evidence for range expansion Inference obtained with the 2003 NCPA key | | Prior evidence for range expansion Inference obtained with the 2003 NCPA key | |
| --- | --- | --- | --- | --- |
| Prior expectation for fragmentation | Yes | No | Yes | No |
| Yes | 12 | 3 | 16 | 3 |
| No | 0 | 17 | 1 | 14 |

**Table 7** Expected and observed inference of range expansion as a function of prior evidence for fragmentation

| | No prior evidence for fragmentation Inference obtained with the 2003 NCPA key | | Prior evidence for fragmentation Inference obtained with the 2003 NCPA key | |
| --- | --- | --- | --- | --- |
| Prior expectation for range expansion | Yes | No | Yes | No |
| Yes | 15 | 8 | 13 | 4 |
| No | 2 | 10 | 1 | 7 |

of range expansion on fragmentation inference, an exact homogeneity test was performed on the classes of no prior evidence for range expansion vs. prior evidence against the four inference categories (yes/yes; yes/no; no/yes and no/no, with the first word referring to prior expectations and the second to the NCPA inference). The resulting two × four contingency table has an exact probability value of 0.6927 under the null hypothesis of homogeneity. Therefore, the presence or absence of range expansions in the same data set has no detectable impact upon the ability of NCPA to make inference about fragmentation. Similarly, Table 7 shows the inferences of range expansion as a function of whether or not fragmentation occurred in the same data set. In this case, the exact probability level of the contingency test of homogeneity is 0.3461. Therefore, the presence or absence of fragmentation events in the same data set has no detectable impact upon the ability of NCPA to make inference about range expansion.

## Validating NCPA inferences with simulated populations

As shown in Tables 1 and 4, NCPA does well with respect to inferences involving fragmentation events. Most events are detected, and false positives are unlikely. These results seem to be contrary to the results reported in Knowles & Maddison (2002) that NCPA detected fragmentation in simulated populations less than 8% of the time in which fragmentation was the only phylogeographical event occurring and inferred the wrong historical process between 75% and 80% of the time. One potential reason for this

discrepancy is that Knowles & Maddison (2002) simulated a situation in which each local population is isolated completely from all others due to past fragmentation. As discussed on page 773 of Templeton *et al.* (1995), there are many types of fragmentation, and the type of fragmentation in which each local population is an isolate (microvicariance) was excluded explicitly from the NCPA inference key along with the statement that it would 'be dealt with in a subsequent paper'. Indeed, this type of fragmentation was considered in a subsequent paper (Hutchison & Templeton 1999), but the nature of the data (microsatellite data) and the testing procedure (a correlation test on genetic distance and its local variance vs. geographical distance) is different from NCPA. Because the simulated case is excluded explicitly from the inference key, there is no expectation for the inference key to work in this situation. The results in Tables 1 and 4 show that the inference key works well when dealing with fragmentation events that split the species into allopatric subsets of local populations. Hence, one contributor to the discrepancy is the nature of the fragmentation event in the real vs. simulated data sets.

Although NCPA is not applicable to microvicariance, the inference key was designed to yield ambiguous inference in such cases to protect users against misidentifications. Consequently, a high false positive rate under microvicariance would represent a serious flaw in the key. To see if the key needs further revision, a more detailed examination of the assumptions of the Knowles & Maddison (2002) simulations is required.

Many of the assumptions of the Knowles & Maddison (2002) simulations were not published in their paper nor

were they available in the addendum on their website at the time of publication. The assumptions were therefore obtained by request to Dr Knowles (pers. comm.). Each local isolate had an inbreeding effective size of 10 000 and the time between fragmentation events was 5000 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 of 10 000 is 40 000 generations for an autosomal gene or 20 000 for mtDNA (assuming that 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 retention of ancestral polymorphisms across isolates is difficult for any technique. Indeed, Knowles & Maddison (2002) reported that their own phylogeographical tests also had 'poor performance' with these simulated data sets. This difficult inference situation was made even more difficult by their sampling assumptions. Despite the large local inbreeding effective population sizes of 10 000 the sample sizes from each isolate were only 10, and only four isolates were sampled for a total sample size of 40 individuals. Of all the NCPA ever performed of which the author has knowledge, only the analysis of the human gene *PDHA1* (Templeton 2002) involved a smaller sample: four pooled geographical locations and a total sample size of 35. However, the *PDHA1* analysis was performed as a *part* of a cross-validation analysis based upon 10 different DNA regions. Hence, there is no stand-alone analysis of real data that has such sparse sampling as these simulated data. Because NCPA is a statistical procedure, it requires adequate sample sizes and sample locations for accurate inference (Templeton 2002). Another potential reason for the discrepancy of Tables 1 and 4 with the results of Knowles & Maddison (2002) is that the simulated data sets were sampled more sparsely than the real data sets in Appendix I. Hence, the results of Knowles & Maddison (2002) indicate that NCPA should not be applied to cases sampled so poorly.

There is one additional sampling assumption that was not published in Knowles & Maddison (2002), but rather appeared in a document entitled 'readme.doc' at their website. This document at the time of publication of their paper consisted of a single sentence: 'In each of the example data files (i.e. NoBot0 to NoBot9) there are four populations with no individuals inhabiting the intervening areas between the populations'. Thus, these simulated data sets sample each local population poorly, but sample every local population exhaustively within the study area. The assumption of exhaustive local population sampling has no impact upon the execution of the simulation nor the resulting simulated data sets. This assumption plays a role only when the inference key is applied to the GEODIS output. As stated previously, the inference key contains many

**Table 8** Inferences from the 10 simulated data sets of Knowles & Maddison (2002) using the 2001 inference key without the assumption of exhaustive sampling of all local populations

| Inconclusive | Inadequate sampling | Inadequate genetic resolution | Isolation by distance | Range expansion |
|---|---|---|---|---|
| 1 | 20 | 2 | 3 | 2 |

places where it screens for inadequate geographical sampling. The assumption of exhaustive sampling of all local populations circumvents those screens. To investigate the impact of this single assumption, all 10 simulated data sets and their GEODIS outputs were downloaded from the website http://mesquiteproject.org/knowles and reanalysed using the 2001 inference key (the same key used by Knowles & Maddison) but *without* the assumption of exhaustive sampling of all local populations. The inferences for all nesting clades with statistically significant results are shown in Table 8. Rather than obtaining 75–80% of the inferences being false positives as reported in Knowles & Maddison (2002), 82% of the inferences under the assumption of nonexhaustive sampling of all local populations are inconclusive, with the most common reason being inadequate sampling (71%). Without the assumption of exhaustive sampling of all local populations, the 2001 inference key is indeed giving the most appropriate interpretation to the simulated results: inadequate sampling. Given that none of the data sets reported in Appendix I claimed exhaustive sampling of all local populations, the sampling assumptions made by Knowles & Maddison (2002) are not relevant to most real data sets nor to the results reported in Tables 1 and 4. Consequently, no changes in the inference key are needed in response to the simulation results of Knowles & Maddison (2002). Under sampling conditions that typify real data sets, the inference key performs as it should when dealing with microvicariance (Table 8).

The inference key also deals well with another potential difficulty indicated by recent computer simulations: discriminating between fragmentation and isolation by distance. Irwin (2002) performed computer simulations that purport to show that strong phylogeographical breaks can arise without geographical barriers to gene flow. In these simulations, the 10 closest individuals were sampled from each of six locations that were distributed evenly across a linear range of fixed absolute value. Different simulations varied the amount of dispersal over this absolute distance. Note that by having the sampling locations at a fixed, absolute distance, the dispersal distances (that is, distances measured in units of the standard deviation of dispersal) between sampling points become larger and larger as the standard deviation of dispersal becomes smaller and smaller. When the standard deviation of dispersal is large,

the fixed sampling locations provide excellent coverage of the total geographical space. However, when the standard deviation of dispersal is small, the geographical sampling becomes extremely sparse with large, unsampled gaps between sample sites as measured by dispersal distance. In a population genetic model, it is distance in dispersal units that is relevant, not some arbitrary absolute units. Thus, as Irwin (2002) varied dispersal distances, he also inadvertently varied the amount of geographical sparseness in his simulated sampling design. When sampling is sparse, isolation by distance can yield apparent phylogeographical breaks that mimic those associated with true fragmentation (Templeton *et al*. 1995). An example of this is provided in Fig. 2B in Templeton (1998a), which shows an apparent phylogeographical break in the impala *Aepyceros melampus*. However, when the NCPA inference key was applied to the impala example, the inference for this apparent break was 'inadequate geographical sampling to discriminate between isolation by distance and fragmentation' — exactly the confoundment noted by Irwin (2002). Hence, the results of Irwin (2002) are a well-known artefact of sparse geographical sampling in a population characterized by isolation by distance. The original inference key for NCPA already dealt with that problem, so once again, no revisions are needed in the inference key in response to the simulation results reported by Irwin (2002).

## A priori versus a posteriori interpretation

Knowles & Maddison (2002) object to the a posteriori use of the inference key to make biological interpretations from statistically significant geographical associations. However, the methods advocated by Knowles & Maddison (2002) also have an implicit inference key, but it is generated a priori. For example, consider the phylogeographical analysis of the montane grasshopper *Melanoplus oregonensis* (Knowles 2001). These grasshoppers currently have a fragmented distribution, inhabiting mountaintops in western North America, and showing much genetic differentiation. One hypothesis for their origin is that they arose from a single widespread ancestral population. Alternatively, regional groups of these mountaintop populations could have come from different ancestral-refugial source populations. The alternative hypotheses in this case can all be portrayed as population trees with different topologies. Knowles (2001) then simulated a neutral coalescent process under these alternative population trees and measured the discord between the reconstructed haplotype tree in the simulations to the hypothesized population trees. Statistics measuring the discord between two evolutionary trees are nothing new (e.g., Templeton 1983), and can and do have a variety of biological interpretations depending upon the context of their use. There is nothing inherent in a statistic that measures tree discordance that inherently has only a single

phylogeographical interpretation: discordance could arise from different patterns of fragmentation as assumed in this example, or tree discordance could arise from alternatives not even considered, such as gene flow (Templeton 1998b). The biological interpretation of the tree discord statistic in this case is driven entirely by the finite number of a priori hypotheses that were specified. Without such an a priori universe of finite possibilities, a tree discordance statistic has no clear biological interpretation. Consequently, in this example and the others of 'statistical phylogeography' given by Knowles & Maddison (2002), there is an implicit inference key for the biological interpretation of the statistics being calculated that is defined by the set of a priori hypotheses. Both the phylogeographical methods advocated by Knowles & Maddison (2002) and NCPA distinguish among alternative interpretations by finding a statistic or set of statistics that deviate significantly from some well-defined model coupled with an interpretative key. The main difference between these approaches is that the interpretative key is applied a priori and implicitly in statistical phylogeography *sensu* Knowles & Maddison (2002), whereas it is applied a posteriori and explicitly by Templeton *et al*. (1995).

The realm of biological possibilities being considered is explicit in the a posteriori inference key, and it is implicit in the statistical phylogeography *sensu* Knowles & Maddison (2002) by examining the alternatives that were simulated. Because all conceivable alternatives can never be simulated, there is no general evaluation of the best fitting phylogeographical model. Consequently, despite the claims of Knowles & Maddison (2002), their a priori approach does *not* provide a general method of evaluating the statistical validity of the resulting inferences; it evaluates those inferences only with respect to the small number of alternatives that were simulated. Similarly, NCPA does not consider all possible biological truths; for example, the inference key does not include microvicariance. Neither approach can claim an exhaustive coverage of all possible biological events or processes. This is why the inference key presented in Appendix II should always be regarded as a work in progress. Just as past versions of this key were modified because of ambiguities and biological lacunae, future work will undoubtedly reveal additional deficiencies.

It is inappropriate to regard NCPA and statistical phylogeography *sensu* Knowles & Maddison (2002) as mutually exclusive alternatives, with one better or worse than the other. When the research question is focused upon a small number of a priori alternatives and there is much prior confidence that these alternatives cover the appropriate universe of biological possibilities, the approach of Knowles & Maddison (2002) is both appropriate and powerful, as it focuses statistical inference upon a narrow set of alternatives. However, if one does not have strong prior knowledge about the universe of possibilities, or when one

suspects that processes or events other than the ones with prior knowledge may also be occurring, the a posteriori inference approach of NCPA with its broader coverage of biological possibilities is more appropriate (Turner *et al.* 2000). There is thus a legitimate role in the field of statistical phylogeography for both a priori and a posteriori interpretative frameworks, and the field is diminished by falsely casting one approach or the other as not being 'statistical'. The a priori and a posteriori approaches to statistical phylogeography are complementary, not contradictory.

## Multilocus cross-validation

Tables 4 and 5 establish the validity of NCPA inferences in general, but they do not directly provide a method for evaluating the specific inferences from a single sample of populations. Cross-validation is a common tool in the general statistical literature (Good 1999) for achieving validation of specific inferences. Templeton (2002) used cross-validation in a nested clade phylogeographical analysis of recent human evolution. In this case, cross-validation is achieved by scoring the sampled populations for multiple loci or DNA regions, ideally independently segregating, and performing separate NCPA upon each DNA region. Templeton (2002) then accepted only those inferences that were corroborated by two or more of the 10 DNA regions in that study. There are several benefits to this multilocus, cross-validation approach. First, as shown in Tables 4 and 5, the most common error in NCPA is failure to detect an event. This occurs because NCPA requires a mutation or mutations that are appropriately placed in space and time in order to detect an event, in addition to the statistical requirements of adequate sample sizes and number of sample sites. As a result, any one DNA region will miss some events or processes. By studying multiple DNA regions, the resolution of NCPA is greatly enhanced simply because inferences missed by one DNA region can be detected by another.

Second, there is extreme variance in coalescent times from one DNA region to another, even including the same type of DNA (such as autosomal loci) within a single population (Templeton 2002). NCPA is most informative only in the time period of the most recent half of the total coalescent time for a given DNA region (Templeton 2002). Under a neutral coalescent model in a single population, all the haplotype diversity is expected to collapse to just two segregating lineages at the halfway point to total coalescence to a single ancestral DNA molecule. This is usually too little genetic diversity for most inferences. As a result, NCPA (and other techniques as well) effectively sample the temporal period defined by the time of the latest mutations that have occurred going back to the first half of the total coalescent time for a given DNA region. All phylogeographical information is lost at the time of ultimate coalescence to a common ancestor. Hence, by sampling a variety of DNA regions, and thus a variety of coalescent times, the temporal breadth of NCPA inference can be greatly broadened.

Third, cross-validation protects against false positives, the focus here. There are many reasons for false positives, including the error associated with any statistical inference and the error caused by natural selection operating upon a specific DNA region to distort its geographical pattern, but typically in a locus-specific fashion. To protect against such errors, Templeton (2002) limited inferences to those cross-validated by two or more DNA regions. Cross-validation in this case requires that two or more DNA regions yield the same qualitative type of inference (e.g. a range expansion) in a geographically concordant fashion (e.g. a range expansion out of Africa into Eurasia), and in a temporally concordant fashion (e.g. a range expansion out of Africa into Eurasia around 100 000 years ago).

Given that the type I error rate is overestimated by the incidence of false positives in Tables 4 and 5, as mentioned earlier, it is unlikely that two or more DNA regions will yield the same false inference in a manner that is both geographically and temporally concordant. Concordance for type of inference (e.g. fragmentation or contiguous range expansion) and geographical position can be judged qualitatively, but temporal concordance should be judged in a quantitative fashion. However, no formal statistical framework was provided in Templeton (2002) for judging temporal concordance. Recently, Templeton (2003a) developed a maximum likelihood framework for estimation and hypothesis testing of temporal concordance under a null model of a neutral coalescent in a population with no long-term fragmentation. In particular, the log-likelihood ratio test of the hypothesis that *j* DNA regions detected separate events (already concordant by type and geography) vs. the hypothesis that all *j* regions detected the same event at the same time is:

$$G = -2 \sum_{i=1}^{j} (1 + k_i) \left( 1 - \frac{t_i}{\hat{T}} + \ln t_i - \ln \hat{T} \right) \qquad (1)$$

where G is distributed asymptotically as a $\chi^2$ with $j - 1$ degrees of freedom under the null hypothesis of a single event, $t_i$ is the estimated age of the event detected by DNA region $i$ (see Templeton 2002), $k_i$ is the average pairwise nucleotide divergence of the clade used to age the event from DNA region $i$, and

$$\hat{T} = \frac{\sum_{i=1}^{j} t_i (1 + k_i)}{\sum_{i=1}^{j} (1 + k_i)} \qquad (2)$$

is the maximum likelihood estimator of the event under the hypothesis of a single event. Small values of *G* favour

the hypothesis of a single event, whereas large values favour the hypothesis of more than one event. For example, five different DNA regions were concordant with the inference of an out-of-Africa range expansion (Templeton 2002). Because NCPA of all 10 DNA regions revealed no evidence for long-term fragmentation among human populations in Africa and Eurasia (Templeton 2002), equation 1 is used to test the hypothesis that there was a single out-of-Africa expansion event, yielding $G = 27.63$ with 4 degrees of freedom and a $P$-value of 0.000015 (Templeton 2003a). Therefore, the hypothesis of a single out-of-Africa expansion event is strongly rejected. Furthermore, equation 1 can be used to show that there were two out-of-Africa expansion events, with two DNA regions (mtDNA and Y-DNA) detecting an event about 127 000 years ago with $G = 0.44$ and 1 degree of freedom, yielding a $P$-value of 0.51; and a second event detected with three autosomal DNA regions at 703 400 years ago with $G = 0.1233$ and 2 degrees of freedom, yielding a $P$-value of 0.94. The low $G$-values within these two sets of DNA regions indicates strong concordance within each set for these two out-of-Africa expansion events.

These likelihood ratio tests allow much flexibility in hypothesis testing that go beyond the standard NCPA inferences. For example, one of the more contentious issues concerning recent human evolution is the hypothesis that the last out-of-Africa expansion event was a total replacement event in which the populations expanding out-of-Africa drove to complete genetic extinction all other human populations already living in Eurasia: the out-of-Africa replacement hypothesis (Stringer 2002). Range expansion coupled with replacement is not a formal inference in the NCPA key, but the out-of-Africa replacement hypothesis can still be formally tested with this maximum likelihood framework and inferences based on NCPA (Templeton 2003a). If the replacement hypothesis were true, there should be no inferred events or processes involving Eurasian populations that are older than the most recent out-of-Africa expansion event detected by mtDNA and Y-DNA. All other eight DNA regions in the study of Templeton (2002) detected such older events and processes, so the hypothesis of an out-of-Africa replacement event can be conservatively tested by using equation 1 to test the hypothesis that the ages of these eight other events or processes are homogeneous with the ages of the out-of-Africa expansion event detected by mtDNA and Y-DNA. The resulting test is $G = 77.27$ with 9 degrees of freedom and a $P$-value is $6 \times 10^{-13}$ (Templeton 2003a), indicating a strong rejection of the replacement hypothesis. Moreover, mtDNA and Y-DNA have shallow coalescent times compared to the eight autosomal or X-linked DNA regions, and the out-of-Africa expansion is the oldest event they can detect, meaning that mtDNA and Y-DNA have no information concerning replacement of earlier populations.

Therefore, every DNA region that is informative about replacement (that is, those DNA regions with older coalescent times than mtDNA and Y-DNA) cross-validate the rejection of the replacement hypothesis. Thus, the genetic evidence strongly supports the idea that the most recent expansion of humans out of Africa was not a total replacement event.

## Discussion

Tables 1 and 2 show that the 2001 inference key (which has only minor differences from the original 1995 key) works well. However, an examination of the failures of the key, particularly the false positives, revealed some deficiencies. The key was altered in light of these deficiencies in a manner that reduces the incidence of false positives with only a modest reduction in power. Thus, NCPA is strengthened by this review of its performance against actual data sets with a *priori* expectations. As more such data sets are gathered in the future, the inference key can probably be further improved, but the results reported here indicate that NCPA already generates reliable phylogeographical inference.

One major limitation of testing the validity of the inference criteria against actual data sets with a priori expectations is that some events scored as false positives might have actually occurred. This biases false positive error rates upwards, and this bias is further augmented by the exclusion of data sets in this analysis that had no statistically significant geographical associations. 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. However, care should still be taken to simulate sampling assumptions that fall within the bounds of actual data sets. 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. It is not difficult to find situations with prior information about historical events such as range expansion or fragmentation, but it has proven difficult to find many examples with strong prior expectations about specific patterns of gene flow other than the general expectation that gene flow is plausible when dealing with intraspecific data sets. One can test for concordance of gene flow inferences made from multiple types of data and analyses. For example, the most common inference from NCPA of 10 DNA regions for humans was gene flow restricted by isolation by distance — an inference concordant with other types of analysis on human gene flow

patterns (Templeton 1998b). However, simulations provide another alternative. Various patterns of gene flow could be simulated and then tested with NCPA. This would not only provide a method for validating gene flow inferences, but it would also provide an assessment of false positive rates for historical events under the hypothesis that there are associations between clades and geography, but the associations are due solely to gene flow patterns. Hence, simulations could be a valuable addendum to the analyses performed on actual data with a priori expectations. Both approaches to verification have a role to play in validating phylogeographical inference.

Similarly, there is a need in phylogeography for having both a priori and a posteriori frameworks of inference. NCPA is an a posteriori framework; it allows great flexibility in the events or processes that have shaped the spatial/temporal spread of haplotypic variation, and infers these events or processes after an initial statistical analysis of geographical associations. Tables 6 and 7 reveal that there is no detectable difference in how NCPA does with a posteriori inference of an event regardless of the presence or absence of other, unrelated events. Thus, NCPA provides a method of reconstructing complex phylogeographical histories with little or no prior information. However, in some cases an investigator has prior evidence that only a restricted universe of possibilities is likely. NCPA makes no use of such prior evidence so other, complementary statistics need to be developed for that purpose.

Cross-validation using data from multiple DNA regions represents a powerful way to increase phylogeographical resolution and temporal breadth, to protect against false positive errors, and to validate a posteriori inference for a specific species and set of populations. The statistical framework for cross-validation also extends NCPA to testing hypothesis not within the framework of the formal inference key. Cross-validation in addition provides the key to integrating intraspecific and interspecific phylogeography (Templeton 2003a). Finally, there is information beyond the haplotype tree in multi-DNA region studies because of the potential for both assortment and recombination. Such information can add a valuable new component to phylogeographical analysis, such as the detection of past hybridization events (Templeton 2003b). These properties of integrated, multilocus studies greatly augment the ability to make reliable, validated phylogeographical inference that goes well beyond the studies based upon a single, nonrecombining DNA molecule that have up to now dominated the field of intraspecific phylogeography.

## Acknowledgements

## References

Avise JC, Lansman RA, Shade RO (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**, 279–295.

Byrne M, MacDonald B, Coates D (2002) Phylogeographical patterns in chloroplast DNA variation within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. *Journal of Evolutionary Biology*, **15**, 576–587.

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution*, **3**, 102–113.

Fullerton SM, Clark AG, Weiss KM *et al.* (2000) Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *American Journal of Human Genetics*, **67**, 881–900.

Good PI (1999) *Resampling Methods: a Practical Guide to Data Analysis*. Birkhäuser, Boston.

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.

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

Knowles LL (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshopprers. *Molecular Ecology*, **10**, 691–701.

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

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.

Paulo OS, Jordan WC, Bruford MW, Nichols RA (2002) Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian Lizard. *Molecular Ecology*, **11**, 809–820.

Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution*, **56**, 1776–1788.

Stringer C (2002) Modern human origins: progress and prospects. *Philosophical Transactions of the Royal Society of London B*, **357**, 563–579.

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution*, **37**, 221–244.

Templeton AR (1998a) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR (1998b) Human races: a genetic and evolutionary perspective. *American Anthropologist*, **100**, 632–650.

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, **10**, 779–791.

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

Templeton AR (2003a) A maximum likelihood framework for cross validation of phylogeographic hypotheses. In: *Evolutionary Theory and Processes: Modern Horizons* (ed. Wasser SP). Kluwer Academic Publishers, the Netherlands, in press.

Templeton AR (2003b) Using haplotype trees for phylogeographic and species Inference in fish populations. *Environmental Biology of Fishes*, in press.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of Alcohol Dehydrogenase activity in *Drosophila. Genetics*, **117**, 343–351.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

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.

Turner TF, Trexler JC, Harris JL, Haynes JL (2000) Nested cladistic analysis indicates population fragmentation shapes genetic diversity in a freshwater mussel. *Genetics*, **154**, 777–785.

I first developed the nested clade analysis for detecting phenotypic associations at candidate loci for coronary artery disease, and I continue to use and extend nested clade analysis and related haplotype tree-based techniques in clinical genetic studies. I also have long been interested in the process of speciation, population structure, human evolution and conservation biology. The extension of nested clade analysis to intraspecific phylogeography was motivated by my interests in basic evolutionary biology and conservation. The development and extension of nested clade analysis reflects my general interest in using haplotype trees as a tool for many problems in biology, including many applications that seemingly fall outside the domain of evolutionary biology.

## Appendix 1

Data sets with strong prior expectations and with statistically significant associations between the haplotype tree and geography. 'MI' indicates an inference in which an event was detected but it was misidentified for another type of event or process. Following the 'MI' in parenthesis is the nature of the misidentified event: RE being range expansion, and LDD being long distance dispersal. The other inferences include gene flow restricted by isolation by distance (IBD), gene flow with some long distance dispersal (LDD), range expansion with secondary contact (2nd) inadequate geographical sampling (IGS), and ambiguous or inconclusive (IN)

| Ref. | Organism | Comments | Exp. Frag. | Inf. Frag. | Exp. Range Exp. | Inf. Range Exp. | Other Inf. |
|---|---|---|---|---|---|---|---|
| [1] | *Ambystoma tigrinum tigrium* | Current range includes areas uninhabitable in Pleistocene | No | No | Yes | Yes | IBD |
| | *A. t. mavortium* | Current range includes areas uninhabitable in Pleistocene | No | No | Yes | Yes | IBD |
| | *A. t. tigrinum and mavortium* | Subspecies with known genetic differences and narrow overlap of geographical ranges | Yes | Yes | | | |
| [2] | *Etheostoma blennioides blennioides* | Current range includes areas uninhabitable in Pleistocene; populations separated by Kanawha Falls | Yes | Yes | Yes | Yes | IGS |
| | *E. b. pholidotum* | Current range includes areas uninhabitable in Pleistocene; not adapted to large rivers, with some populations in rivers draining into Missouri and others into Mississippi | Yes | Yes | Yes | Yes | LDD |
| [3] | *Trimerotropis saxatilis* | Ozarks colonized during the xerothermic maximum; currently highly separated populations in Oklahoma vs. the Ozarks and Illinois | Yes | Yes | Yes | Yes | IBD IN |
| [4] | *Geomys bursarius* | Current range includes areas uninhabitable in Pleistocene; found on both sides of the Mississippi River | Yes | Yes | Yes | Yes* | IBD IN IN‡ |
| [5] | *Galaxias truttaceus* | Current range includes lakes created by melting Pleistocene glaciers that later became land-locked. An unpredicted range expansion to the north coast of Tasmania also detected. | Yes | Yes | Yes No | Yes Yes | None |
| [6] | *Drosophila melanogaster* | Current global distribution due to human activities | ? | No | Yes | Yes | IBD IN |
| [7] | *Drosophila buzzatii* | Introduction to Europe from South America via humans | Yes | No | Yes | No | IBD |
| [8] | *Canis latrans* | Historical range expansion since 1900 | No | No | Yes | Yes* | IBD IN IN‡ |
| [9] | *Macaca fascicularis* | Introduction to Mauritius in the 1500s | Yes | No* | Yes | Yes | IGS |
| [10] | *Homo sapiens* | Human settlement of Pacific islands: mtDNA | No | No | Yes | Yes | IBD LDD IN |
| [11, 12] | *Homo sapiens: mtDNA* | Human settlement of Siberia and Americas | No Yes | No Yes | Yes Yes | Yes Yes | IBD LDD IN |
| [13] | *Homo sapiens: nuclear* DNA, MS205 | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | Yes | IBD |
| | |   Expansion into N. Eurasia | | | Yes | Yes | |
| | |   Expansion into Pacific | | | Yes | Yes | |
| | |   Expansion into Americas | | | Yes | Yes | |
| | MX1 | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | No | IBD |
| | |   Expansion into N. Eurasia | | | Yes | No | IN |
| | |   Expansion into Pacific | | | Yes | Yes | |
| | |   Expansion into Americas | | | Yes | Yes | |

**Appendix 1** *Continued*

| Ref. | Organism | Comments | Exp. Frag. | Inf. Frag. | Exp. Range Exp. | Inf. Range Exp. | Other Inf. |
|---|---|---|---|---|---|---|---|
| | MC1R | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | Yes | |
| | | Expansion into N. Eurasia | | | Yes | Yes* | IN‡ |
| | | Expansion into Americas | | | Yes | No | |
| | EDN | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | No | IBD |
| | | Expansion into Americas | | | Yes | Yes* | |
| | Hbβ | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | Yes | IBD |
| | | Expansion into Pacific | | | Yes | Yes* | IGS‡ |
| | | Expansion into Americas | | | Yes | No | |
| | | Expansion out of Asia | | | No | Yes* | IN‡ |
| | Xq13.3 | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | No | IBD |
| | | Expansion into N. Eurasia | | | Yes | No | IN |
| | | Expansion into Pacific | | | Yes | No | |
| | | Expansion into Americas | | | Yes | No | |
| | ECP | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | No | IBD |
| | | Expansion into Americas | | | Yes | No | |
| | PDHA1 | Out of Africa expansion, no frag. within Eurasia | No | No | Yes | No | IBD IGS |
| [14, 15] | *Homo sapiens: Y-DNA* | Expansion out of Africa, no frag. within Eurasia | No | No | Yes | Yes | IBD |
| | | Expansion into N. Eurasia | | | Yes | Yes | LDD |
| | | Expansion into Pacific | | | Yes | Yes | |
| | | Expansion into Americas | | | Yes | Yes | |
| | | Expansion within Africa | | | No | Yes | |
| | | Expansion out of Asia | | | No | Yes | |
| [16] | *Linckia laevigata* | High dispersal starfish in the Indo-West Pacific | No | No | No | No | IBD IN |
| [17] | *Syncerus caffer* | Bovids in Eastern Africa with strong dispersal abilities, but including populations on both sides of the Rift | No | No | No | No | IBD IGS |
| | *Aepyceros melampus* | Valley, with *S. caffer* also inhabiting the Rift Valley and the other two not | Yes | Yes | No | Yes* | IBD IGS |
| | *Connochaetes taurinus* | | Yes | Yes | No | No | IBD IGS |
| [18] | *Drosophila melanogaster* | High dispersal fly, limited to eastern United States: *Adh* locus | No | No | No | No | IBD |
| [19] | *Drosophila melanogaster* | High dispersal fly, limited to eastern United States: *Amy* locus | No | No | No | No | IBD IN |
| [20] | *Drosophila melanogaster* | High dispersal fly, limited to eastern United States: *Ddc* locus | No | No | No | No | IBD |
| [21] | *Lacerta schreiberi* | Several explicit a priori predictions to test NCA | Yes | Yes | Yes | No | IBD |
| | | | Yes | Yes | | | LDD |
| | | | Yes | MI** (LDD) | | | IN‡ |
| [22] | *Bufo woodhousii* | Several explicit a priori predictions to test NCA | Yes | MI† (RE) | Yes | Yes | IBD |
| | | | | | Yes | Yes | LDD |
| | | | | | Yes | Yes | IN |
| | | | | | Yes | No | |
| | | | | | Yes | No | |
| [23] | *Opsanus tau* | Marine toadfish on Atlantic coast with no obvious barriers | No | No | No | Yes* | IN‡ |
| [24] | *Crassostrea virginica* | Atlantic vs. Gulf vicariance detected with | Yes | Yes | | | |
| | | multiple species, but no barriers within | No | No | No | No | IBD |
| | | Atlantic and within Gulf | No | No | No | No | IN |
| [25] | *Heliocidaris erthrograma* | Sea urchin with no obvious dispersal barriers | No | No | No | No | IBD |

**Appendix 1** *Continued*

| Ref. | Organism | Comments | Exp. Frag. | Inf. Frag. | Exp. Range Exp. | Inf. Range Exp. | Other Inf. |
|---|---|---|---|---|---|---|---|
| [26] | *Drosophila ananassae* | Continuous distribution over sampled area | No | No | No | No | IGS IN |
| [27] | *Dineutus assimilis* | Beetle with good dispersal abilities sampled on small scale | No | No | No | Yes | IBD |
| [28] | *Agelaius phoeniceus* | Bird with excellent dispersal abilities over sampled range | No | No | No | No | IBD IN |
| [29] | *Caretta caretta* | Fragmentation expected between Atlantic and Pacific populations | Yes | Yes | | | |
| | | Expect gene flow within Atlantic | No | No | No | No | IBD IN |
| | | Expect gene flow within Pacific | No | No | No | Yes* | IGS‡ |
| [30] | *Plesiastrea versipora* | Widespread coral with excellent dispersal | No | No | No | Yes | IBD IN |
| [31] | *Drosophila buzzatii* | Continuous distribution over sampled area | No | No | No | No | IBD IN |
| [32] | *Anopheles gambiae* | Continuous distribution over sampled area | No | No | No | Yes* | IN‡ |
| [33] | *Drosophila silvestris* | Populations on either side of Hawaii are morphologically and behaviourally differentiated | Yes | Yes | No | No | None |
| [34] | *Spalax ehrenbergi* | Frag. four chromosomal races with many distinctions | Yes | Yes | | | |
| | | Range expansion into Golan Heights | Yes | Yes | Yes | Yes | IBD |
| | | Range expansion from North to South in Israel | Yes | No | No | Yes | IN |
| [35] | *Chioglossa lusitanica* | Previous work indicated episodes of fragmentation and range expansion | Yes | Yes | Yes | No* | 2nd‡ |
| | | | Yes | Yes | | | IBD |
| [36] | *Piriqueta caroliniana* | Fragmentation expected between Bahamas and Florida | Yes | Yes | No | No | LDD |
| | | Fragmentation between ecotypes | Yes | Yes | | | |
| [37] | *Brachionus plicatilis* | Independent evidence for Pleistocene fragmentation | Yes | Yes | Yes | Yes | IBD |
| | | Long distance colonization of isolated habitats | | | | | |
| | | Unexpected contiguous range expansion on coast | | | No | Yes | |
| [38] | *Plasmodium azurophilum* | Two species, Red and White, found on different Caribbean Islands | Yes | Yes | No | No | None |
| | | | Yes | Yes | | | |
| [39] | *Oncorhynchus gorbuscha* | Current range includes areas uninhabitable in Pleistocene; in two temporally isolated broods | Yes | Not tested | Yes | Yes | IBD |
| | | | | | Yes | Yes | IGS |
| [40] | *Ambystoma maculatum* | Current range includes areas uninhabitable in Pleistocene | No | Yes | Yes | Yes | IBD IGS |
| [41] | *Nesticus* species | Several fragmented populations due to post-Pleistocene climatic changes | Yes | Yes | No | No | IBD |
| | | | Yes | No | | | IN |
| | | | Yes | Yes | | | |
| | | | Yes | Yes | | | |
| | | | Yes | No | | | |
| [42] | *Oliarus polyphemus* | Lava-tube adapted planthopper living in old and recent lava flows, with one population well separated from the others | Yes | MI† (RE) | Yes | Yes | IN |
| | | | | | Yes | Yes | |

*Change from Yes to No or vice versa under the 2003 key.
**Change from misidentification to ambiguous under the 2003 key.
†Change from misidentification to correct inference under the 2003 key.
‡New inference under the 2003 key.

# References

**1** 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.

**2** Wilson P (1997) *Mitochondrial DNA Variation and Biogeography Among Some Etheostomid Darters of the Central Highlands*. Division of Biomedical and Biological Sciences. Washington University, St Louis.

**3** Gerber A (1997) *The Semiotics of Subdivision: an Empirical Study of the Population Structure of* Trimerotropis saxatilis *(Acrididae)*. Division of Biomedical and Biological Sciences. Washington University, St Louis.

**4** Davis SK (1994) *Population Structure and Patterns of Speciation in* Geomys *(Rodentia: Geomyidae): an Analysis Using Mitochondrial and Ribosomal DNA*. Division of Biomedical and Biological Sciences, Washington University, St Louis.

**5** Ovenden JR, White RWG (1990) Mitochondrial and allozyme genetics of incipient speciation in a landlocked population of *Galaxias truttaceus* (Pisces, Galaxiidae). *Genetics*, **124**, 701–716.

**6** Hale LR, Singh RS (1991) A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history vs. selection in the genetic structure of geographical populations. *Genetics*, **129**, 103–117.

**7** Rossi MS, Barrio E, Latorre A (1996) The evolutionary history of *Drosophila buzzatii*. Mitochondrial DNA polymorphism in original and colonizing populations. *Molecular Biology and Evolution*, **13**, 314–323.

**8** Lehman N, Wayne RK (1991) Analysis of coyote mitochondrial-DNA genotype frequencies: estimation of the effective number of alleles. *Genetics*, **128**, 405–416.

**9** Lawler SH, Sussman RW, Taylor LL (1995) Mitochondrial DNA of the Mauritian Macaques (*Macaca fascicularis*): an example of the founder effect. *American Journal of Physical Anthropology*, **96**, 133–141.

**10** Sykes B, Leibuff A, Low-Beer J (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics*, **57**, 1463–1475.

**11** Torroni A, Schurr TG, Cahell MF (1993) Asian affinities and continental radiation of the four founding native American mtDNAs. *American Journal of Human Genetics*, **53**, 563–590.

**12** Torroni A, Sukernik RI, Schurr TG (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *American Journal of Human Genetics*, **53**, 591–608.

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

**14** Hammer MF, Karafet TM, Redd AJ (2001) Hierarchical patterns of global human *y*-chromosome diversity. *Molecular Biology and Evolution*, **18**, 1189–1203.

**15** Karafet T, Liping X, Ruofu D (2001) Paternal population history of East Asia: sources, patterns and microevolutionary processes. *American Journal of Human Genetics*, **69**, 615–628.

**16** Williams ST, Benzie JAH (1997) Indo-West Pacific patterns of genetic differentiation in the high-dispersal starfish *Linckia laevigata*. *Molecular Ecology*, **6**, 559–573.

**17** Templeton AR, Georgiadis NJ (1996) A landscape approach to conservation genetics: conserving evolutionary processes in the African Bovidae. In: *Conservation Genetics: Case Histories From Nature* (eds Avise JC, Hamrick JL), pp. 398–430. Chapman & Hall, New York.

**18** Aquadro CF, Desse SF, Bland MM (1986) Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics*, **114**, 1165–1190.

**19** Brown CJ, Aquadro CF, Anderson WR (1990) *DNA sequence evolution of the amylase multigene family in Drosophila pseudoobscura. Genetics*, **126**, 131–138.

**20** Aquadro CF, Jennings RM, Bland MM (1992) Patterns of naturally occurring restriction map variation, dopa decarboxylase activity variation and linkage disequilibrium in the ddc gene region of *Drosophila-melanogaster*. *Genetics*, **132**, 443–452.

**21** Paulo OS, Jordon WC, Bruford NW (2002) Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian lizard. *Molecular Ecology*, **11**, 809–820.

**22** 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.

**23** Avise JC, Reeb CA, Saunders NC (1987) Geographical population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning toadfishes (Batrachoididae). *Evolution*, **41**, 991–1002.

**24** Reeb CA, Avise JR (1990) A genetic discontinuity in a continuously distributed species: mitochondrial-DNA in the American oyster, *Crassostrea virginica. Genetics*, **124**, 397.

**25** McMillan, WO, Raff PA, Palumbi SR (1992) Population genetic consequences of developmental evolution in sea urchins (genus Heliocidaris). *Evolution*, **46**, 1299–1312.

**26** Chen Y, Marsh BJ, Stephen W (2000) Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. *Genetics*, **155**, 1185–1194.

**27** Nurnberger B, Harrison RG (1995) Spatial population structure in the whirligig beetle *Dineutus assimilis*: evolutionary inferences based on mitochondrial DNA and field data. *Evolution*, **49**, 266–275.

**28** Ball RM, Freeman S, James FC (1988) Phylogeographical population structure of red-winged blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences USA*, **85**, 1558–1562.

**29** Bowen BW, Kamezaki N, Limpus CJ (1994) Global phylogeography of the loggerhad turtle (*Caretta caretta*) as indicated by mitochondrial DNA haplotypes. *Evolution*, **48**, 1820–1828.

**30** Rodriguez-Lanetty M, Hoegh-Guldberg O (2002) The phylogeography and connectivity of the latitudinally widespread scleractinian coral *Plesiastrea versipora* in the Western Pacific. *Molecular Ecology*, **11**, 1177–1191.

31 de Brito RA, Manfrin MH, Sene FM (2002) Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genetics and Molecular Biology*, **25**, 161–171.

32 Besansky NJ, Lehmann T, Fahey GT (1997) Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *A. arabiensis*, suggest extensive gene flow. *Genetics*, **147**, 1817–1828.

33 Templeton AR (1998) Species and speciation: geography, population structure, ecology, and gene trees. In: *Endless Forms: Species and Speciation* (eds D.J. Howard DJ, Berlocher SH), pp. 32–43. Oxford University Press, Oxford.

34 Templeton AR (1999) Using gene trees to infer species from testable null hypothesis: cohesion species in the Spalax ehrenbergi complex. In: *Evolutionary Theory and Processes: Modern Perspectives, Papers in Honour of Eviatar Nevo* (ed. Wasser SP), pp. 171–192. Kluwer Academic Publishers, Dordrecht.

35 Alexandrino J, Arntzen JW, Ferrand N (2002) Nested clade analysis and the genetic evidence for population expansion in the phylogeography of the golden-striped salamander, *Chioglossa lusitanica* (Amphibia: Urodela). *Heredity*, **88**, 66–74.

36 Templeton AR, Maskas SD, Cruzan MB (2000) Gene trees: a powerful tool for exploring the evolutionary biology of species and speciation. *Plant Species Biology*, **15**, 211–222.

37 Gomez A, Carvalho GR, Hunt DH (2000) Phylogeography and regional endemism of a passively dispersing zooplankter: mitochondrial DNA variation in rotifer resting egg banks. *Proceedings of the Royal Society of London B*, **267**, 2189–2197.

38 Perkins SL (2001) Phylogeography of Caribbean lizard malaria: tracing the history of vector-borne parasites. *Journal of Evolutionary Biology*, **14**, 34–45.

39 Churikov D, Gharrett AJ (2002) Comparative phylogeography of the two pink salmon broodlines: an analysis based on a mitochondrial DNA genealogy. *Molecular Ecology*, **11**, 1077–1102.

40 Phillips CA (1989) *Population Structure, Breeding Pond Fidelity and Phylogeography in the Spotted Salamander,* Ambystoma maculatum. Division of Biology and Biomedical Sciences, Washington University, St Louis.

41 Hedin MC (1995) *Speciation and Morphological Evolution in Cave Spiders (Araneae: Nesticidae: Nesticus) of the Southern Appalachians.* Washington University, St Louis.

42 Shingleton KE (2001) *Population Structure and Evolutionary History of a Cave-Adapted Planthopper in the Hawaiian Islands.* Division of Biology and Biomedical Sciences, Washington University, St Louis.

## Appendix 2

*Inference key for the nested haplotype tree analysis of geographical distances*

Start with haplotypes nested within a one-step clade and work up to clades nested within the total tree. If the tree is not rooted through an outgroup or if none of the clades nested at the total tree level have the sum of the outgroup probabilities of their haplotypes greater than or equal to 0.95, regard all clades nested at the total tree level as tips. When rooting is deemed reliable, interiors should also refer to the older clades in a nesting category, and tips to their evolutionary descendants.

This key is applied only if there are some significant values for $D_c$, $D_n$ or I–T within the nesting clade. If there are no statistically significant distances within the clade, the null hypothesis of no geographical association of haplotypes cannot be rejected (either panmixia in sexual populations, extensive dispersal in nonsexual populations, small sample size or inadequate geographical sampling). In that case, move on to another clade at the same or higher level.

**1** Are all clades within the nesting clade found in separate areas with no overlap?
- NO — go to step 2.
- YES — go to step 19.

**2** Is at least one of the following conditions satisfied?
 a. The $D_c$s for one or more tips are significantly small and the $D_c$s for one or more of the interiors are significantly large or nonsignificant.
 b. The $D_c$s for one or more tips are significantly small or nonsignificant and the $D_c$s for some but *not* all of the interiors are significantly small.
 c. The $D_c$s for one or more interiors are significantly large and the $D_c$s for the tips are either significantly small or nonsignificant.
 d. The I–T $D_c$ is significantly large.
- NO — go to step 11.
- YES — go to step 3.
- Tip/interior status cannot be determined — **inconclusive outcome**.

**3** Is at least one of the following conditions satisfied?
 a. Some $D_n$ and/or I–T $D_n$ values are significantly reversed from the $D_c$ values.
 b. One or more tip clades show significantly large $D_n$s.
 c. One or more interior clades show significantly small $D_n$s.
 d. I–T has a significantly small $D_n$ with the corresponding $D_c$ value nonsignificant.
- NO — go to step 4.
- YES — go to step 5.

**4** Are both of the following conditions satisfied?
 a. The clades (or two or more subsets of them) with significantly small $D_c$ values have ranges that are completely or mostly nonoverlapping with the other clades in the nested group (particularly interiors).

b. The pattern of restricted ranges represents a break or reversal from lower level trends within the nested series (applicable to higher-level clades only).

- NO — **restricted gene flow with isolation by distance (restricted dispersal by distance in non-sexual species).** This inference is strengthened if the clades with restricted distributions are found in diverse locations, if the union of their ranges roughly corresponds to the range of one or more clades (usually interiors) within the same nested group (applicable only to nesting clades with many clade members or to the highest level clades regardless of number), and if the $D_c$ values increase and become more geographically widespread with increasing clade level within a nested series (applicable to lower level clades only).
- YES — go to step 9.

5 Are both of the following conditions satisfied?
  a. The clades (or two or more subsets of them) with significantly small $D_c$ values have ranges that are completely or mostly nonoverlapping with the other clades in the nested group (particularly interiors).
  b. The pattern of restricted ranges represents a break or reversal from lower level trends within the nested series (applicable to higher-level clades only).
- NO — go to step 6.
- YES — go to step 15.

6 Are either of the following conditions satisfied?
  a. Clades (or haplotypes within them) with significant reversals or significant $D_n$ values without significant $D_c$ values define two or more geographically concordant subsets.
  b. Clades (or haplotypes within them) with significant reversals or significant $D_n$ values without significant $D_c$ values are geographically concordant with other haplotypes/clades showing similar distance patterns?
- No — go to step 7.
- YES — go to step 13.
- TOO FEW CLADES ($\leq$ 2) TO DETERMINE CONCORDANCE – **insufficient genetic resolution to discriminate between range expansion/colonization and restricted dispersal/gene flow** — proceed to step 7 to determine if the geographical sampling is sufficient to discriminate between short- vs. long-distance movement.

7 Are the clades with significantly large $D_n$s (or tip clades in general when $D_n$ for I–T is significantly small) separated from the other clades by intermediate geographical areas that were sampled?
- NO — go to step 8.

- YES — restricted gene flow/dispersal but with some long-distance dispersal.

8 Is the species absent in the nonsampled areas?
- NO — sampling design inadequate to discriminate between isolation by distance (short-distance movements) vs. long-distance dispersal.
- YES — restricted gene flow/dispersal but with some long-distance dispersal over intermediate areas not occupied by the species; or past gene flow followed by extinction of intermediate populations.

9 Are the different geographical clade ranges identified in step 4 separated by areas that have not been sampled?
- NO — **allopatric fragmentation.** (If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially non-overlapping distributions are mutationally connected to one another by a larger than average number of steps.)
- YES — go to step 10.

10 Is the species absent in the nonsampled areas?
- NO — geographical sampling(s) inadequate to discriminate between fragmentation and isolation by distance.
- YES — **allopatric fragmentation.** (If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially nonoverlapping distributions are mutationally connected to one another by a larger than average number of steps.)

11 Is at least one of the following conditions satisfied?
  a. The $D_c$ value(s) for some tip clade(s) is/are significantly large.
  b. The $D_c$ value(s) for all interior(s) is/are significantly small.
  c. *The* I–T $D_c$ is significantly small.
- NO — go to step 17
- YES — **range expansion**, go to step 12.

12 Are the $D_n$ and/or I–T $D_n$ values significantly reversed from the $D_c$ values?
- NO — contiguous range expansion.
- YES — go to step 13.

13 Are the clades with significantly large $D_n$s (or tip clades in general when $D_n$ for I–T is significantly small) separated from the geographical centre of the other clades by intermediate geographical areas that were sampled?
- NO — go to step 14.
- YES — **long-distance colonization possibly coupled with subsequent fragmentation** (subsequent frag-

mentation is indicated if the clades displaying restricted but at least partially nonoverlapping distributions are mutationally connected to one another by a larger than average number of steps) **or past fragmentation followed by range expansion**. To see if secondary contact is involved, perform the supplementary tests given in Templeton, Molecular Ecology **10**: 779–791, 2001. To discriminate the type of movement leading to this pattern, go to step 21.

14 Is the species present in the intermediate geographical areas that were not sampled?
   - YES — sampling design inadequate to discriminate between contiguous range expansion, long-distance colonization and past fragmentation.
   - NO — **long-distance colonization and/or past fragmentation** (not necessarily mutually exclusive). If inferred at a high clade level, fragmentation rather than colonization is inferred if the clades displaying restricted but at least partially nonoverlapping distributions are mutationally connected to one another by a larger than average number of steps. If the branch lengths are short, a colonization event is inferred, perhaps associated with recent fragmentation. To discriminate the type of movement leading to this pattern, go to step 21.

15 Are the different geographical clade ranges identified in step 5 separated by areas that have not been sampled?
   - NO – **past fragmentation and/or long-distance colonization** (not necessarily mutually exclusive). If inferred at a high clade level, fragmentation rather than colonization is inferred if the clades displaying restricted but at least partially nonoverlapping distributions are mutationally connected to one another by a larger than average number of steps. If the branch lengths are short, a colonization event is inferred, perhaps associated with recent fragmentation. To discriminate the type of movement leading to this pattern, go to step 21.
   - YES — go to step 16.

16 Is the species present in the intermediate geographical areas that were not sampled?
   - YES — go to step 18.
   - NO — **allopatric fragmentation.** If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially nonoverlapping distributions are mutationally connected to one another by a larger than average number of steps.

17 Are either of the following conditions satisfied?
   a. The $D_n$ values for tip or some (but not all) interior clades are significantly small.
   b. The $D_n$ for one or more interior clades is/are significantly large.
   c. *The* I–T $D_n$ value is significantly large.
   - NO — inconclusive outcome.
   - YES — go to step 4.

18 Are the clades found in the different geographical locations separated by a branch length with a larger than average number of mutational steps.
   - NO — geographical sampling(s) inadequate to discriminate between fragmentation, range expansion and isolation by distance.
   - YES — geographical sampling(s) inadequate to discriminate between fragmentation and isolation by distance.

19 Is the species present in the areas between the separated clades?
   - NO — **allopatric fragmentation.** If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially nonoverlapping distributions are mutationally connected to one another by a larger than average number of steps.
   - YES — go to step 20.

20 Was the species sampled in the areas between the separated clades?
   - NO — inadequate geographical sampling.
   - YES — go to step 2.

21 Are all of the following true?
   a. Is it biologically realistic that the organism could have undergone long-distance movement?
   b. Are the nested haplotypes that mark a potential long-distance colonization event within a clade that shows evidence of population growth by other methods (such as mismatch distributions)?
   c. At the level of the entire cladogram, does the clade *not* inferred to have produced long-distance colonization *not* show evidence of past population growth with other methods?
   - YES — long-distance movement.
   - NO — insufficient evidence to discriminate between long-distance movements of the organism and the combined effects of gradual movement during a past range expansion and fragmentation.