Separating Population Structure from Population History: A Cladistic Analysis of the Geographical Distribution of Mitochondrial DNA Haplotypes in the Tiger Salamander, *Ambystoma tigrinum*

Alan R. Templeton, Eric Routman[1] and Christopher A. Phillips[2]

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

Manuscript received October 17, 1994
Accepted for publication February 16, 1995

ABSTRACT

Nonrandom associations of alleles or haplotypes with geographical location can arise from restricted gene flow, historical events (fragmentation, range expansion, colonization), or any mixture of these factors. In this paper, we show how a nested cladistic analysis of geographical distances can be used to test the null hypothesis of no geographical association of haplotypes, test the hypothesis that significant associations are due to restricted gene flow, and identify patterns of significant association that are due to historical events. In this last case, criteria are given to discriminate among contiguous range expansion, long-distance colonization, and population fragmentation. The ability to make these discriminations depends critically upon an adequate geographical sampling design. These points are illustrated with a worked example: mitochondrial DNA haplotypes in the salamander *Ambystoma tigrinum*. For this example, prior information exists about restricted gene flow and likely historical events, and the nested cladistic analyses were completely concordant with this prior information. This concordance establishes the plausibility of this nested cladistic approach, but much future work will be necessary to demonstrate robustness and to explore the power and accuracy of this procedure.

ONE of the principal aims of population genetics is the measurement of the amount and patterns of genetic variation found within and among subpopulations of interbreeding organisms to study gene flow, genetic drift, system of mating, mutation, and natural selection. Traditionally, inferences about microevolutionary forces have been based upon the number of alleles (or haplotypes), their frequencies, and their geographical distribution. For example, WRIGHT (1969) developed hierarchical *F* statistics as a tool to study gene flow, genetic drift and system of mating from data on allele and genotype frequencies and geography. After estimating these *F* statistics, microevolutionary parameters could be estimated by relating the *F* statistics to an underlying model. For example, in the "island model" of gene flow in which the population is subdivided into many "islands" of inbreeding effective size N with a rate of exchange of m per generation at random over all islands, Wright's F_{st} should be equal to $1/[4N(m + \mu) + 1]$ where μ is the mutation rate. Hence, if F_{st} and μ are estimated, it is possible to estimate Nm, the effective number of migrating individuals per generation.

One serious limitation of this approach is that the relation of F_{st} to underlying microevolutionary parameters changes with different models of population structure. For example, if the populations are in a one-dimensional habitat (such as in or along a river) and all dispersal is limited to exchanges between adjacent populations at a rate of $m/2$ per generation, then $F_{st} = 1/[4N(2m\mu)^{1/2} + 1]$ (KIMURA and WEISS 1964). Many other models exist, each with its own relationship between F_{st} and underlying inbreeding effective size, mutation, and gene flow parameters. Consequently, one major limitation of the use of F_{st} (and related statistics, *e.g.*, LYNCH and CREASE 1990; HUDSON *et. al.* 1992a) is that the data used to estimate the *F* statistics often do not indicate which model of gene flow is appropriate for the populations being studied. This is further complicated by the fact that the various models of gene flow are not necessarily alternatives; one part of a species range may be restricted to one-dimensional, stepping stone gene flow, whereas another part may fit the two-dimensional continuous, isolation by distance model. Even worse, the geographical genetic variation measured by F_{st} and related statistics may have nothing to do with current patterns of gene flow at all. For example, LARSON (1984) pointed out that when a population expands into or colonizes a new geographic area, genetic homogeneity could be created within the recently colonized area that does not reflect current patterns of gene flow. Similarly, two populations may have been fragmented in the past and currently have no gene flow whatsoever, yet their shared ancestry can create F_{st} values less than one that

Corresponding author: Alan R. Templeton, Department of Biology, Washington University, St. Louis, MO 63130-4899.
E-mail: templeton@wustlb.wustl.edu

[1] *Present address:* Department of Biology, San Francisco State University, San Francisco, CA 94132.

[2] *Present address:* Illinois Natural History Survey, Center for Biodiversity, 607 E. Peabody Drive, Champaign, IL 61820.

would erroneously imply gene flow. Hence, the geographical pattern of genetic variation is influenced by population structure, by population history, and by combinations of these structural and historical factors. This paper is concerned with separating population structure from population history as sources of geographical associations with genetic variants.

The effects of population history and structure have been studied by spatial autocorrelation (SOKAL *et al.* 1989a,b; SLATKIN and ARTER 1991; EPPERSON 1993), principal component analyses (*e.g.*, AMMERMAN and CAVALLI-SFORZA 1984), and multidimensional scaling (LESSA 1990). These approaches can be thought of as the analogue of "phenetic" approaches in systematics, as they are based upon some sort of genetic distance or identity measure. In contrast, we will outline a cladistic approach in this paper that uses a character-state based haplotype evolutionary tree. This paper will be limited to genetic surveys using DNA sequence or restriction site data with samples derived from a finite number of distinct sampling locations. With such genetic surveys, it is possible not only to estimate the number of alleles, their frequencies and their geographical distribution, but also their genealogical structure. With the rapid development of coalescent theory over the last decade (KINGMAN 1982a,b; EWENS 1990; HUDSON 1990), an increasingly rich theoretical framework within population genetics is arising for dealing with gene genealogies and allele frequency distributions in an integrated fashion. This gene genealogy approach has already proven to be a powerful tool for studying the relationship of genotype to phenotype (TEMPLETON *et al.* 1987, 1988, 1992; TEMPLETON and SING 1993), natural selection (ANTONARAKIS *et al.* 1984; GOLDING 1987; GOLDING and FELSENSTEIN 1990; HARTL and SAWYER 1991; O'BRIEN 1991), and—most relevant to the problem at hand—population structure (AVISE *et al.* 1988; SLATKIN 1989; SLATKIN and MADDISON 1989, 1990; NEIGEL *et al.* 1991; EXCOFFIER *et al.* 1992; HUDSON *et al.* 1992a; NEIGEL and AVISE 1993; TEMPLETON 1993; EXCOFFIER and SMOUSE 1994). When the genetic variation is organized into a genealogy, the resulting analysis of how geography overlays upon genealogy has been called "intraspecific phylogeography" (AVISE 1989). Such analyses commonly find a strong association between the geographical location of haplotypes and their evolutionary position within a gene tree, but the demonstration of such an association *per se* does not reveal the causes of the association.

Fortunately, much more information than mere association can be gathered from a geographical overlay upon gene trees; different causes of geographical association can yield qualitatively different patterns that can be assessed through rigorous statistical testing (TEMPLETON 1993). The purpose of this paper is to extend the statistical methodology outlined in TEMPLETON (1993). As this approach to separating population history and structure is novel, many questions obviously need to be raised about its accuracy, robustness, power and applicability to other types of data. It is not the intent of this paper to address all of these issues immediately. Rather, the purpose of this paper is to outline the basic methodology and test statistics and to demonstrate the plausibility of this approach by checking for concordance with prior knowledge on a worked example. The example will be an analysis of mitochondrial DNA (mtDNA) restriction site haplotypes in tiger salamanders (*Ambystoma tigrinum*) sampled throughout the central United States. As will be discussed later, these salamanders most likely have been affected by population fragmentation events during the Pleistocene, followed by range expansion into formerly glaciated areas, with an overlay of isolation by distance caused by limited dispersal capabilities. Hence, this case should have geographical associations due to multiple causes and accordingly represents a good test case for illustrating this methodology. Obviously, one worked example is insufficient to address such issues as the power, accuracy, and robustness of this novel approach. These critical issues will be addressed in future papers through a combination of computer simulations and additional worked examples that can be checked for concordance with prior knowledge.

STATISTICAL METHODOLOGY

Haplotype network estimation: TEMPLETON (1993) showed how geographical association tests could be performed with allele or haplotype genealogies by using the same statistical techniques developed for studying genotype-phenotype associations within a species (TEMPLETON *et al.* 1987; TEMPLETON *et al.* 1988). The first step of such an analysis is to estimate the haplotype phylogenetic tree. This is accomplished by the algorithm given in TEMPLETON *et al.* (1992) that not only estimates the unrooted haplotype tree but also simultaneously provides a 95% plausible set for all haplotype linkages in the unrooted tree. By using a finite-site model of DNA evolution, this algorithm first assesses the limits of parsimony. Haplotype networks are then constructed using parsimony only when parsimony has a probability of at least 0.95 of being true. When parsimony cannot be justified at this probability level, both parsimonious and nonparsimonious connections are allowed between haplotypes until their cumulative probability exceeds 0.95. Hence, the resulting 95% plausible set of alternative networks is not necessarily just a set of alternative maximum parsimony unrooted cladograms, but can include nonparsimonious alternatives as well. This estimation algorithm can also deal with the complexities caused by limited recombination, although this capability is not relevant for the salamander mtDNA data set to be analyzed in this paper because mtDNA does not undergo genetic recombination. An empirical

verification of this algorithm and a comparison of its performance to that of maximum parsimony with bootstrapping can be found in CRANDALL (1994).

Nested statistical design: After the 95% plausible set has been estimated, the set of plausible cladograms is converted into a nested design in which haplotypes ("0-step clades") separated by a single mutation are grouped together into "one-step clades" proceeding from the tips to the interior of the network, then one-step clades separated by a single mutation are grouped together in "two-step clades", *etc.*, until the next level of nesting would encompass the entire tree. The basic nesting rules are described in TEMPLETON *et al.* (1987), and additional rules that deal with the complications caused by ambiguities within the plausible set of haplotype networks are given in TEMPLETON and SING (1993). It is important to point out that the nested design does not require that the haplotype tree be accurately known, but rather explicitly allows ambiguities in the haplotype network estimator.

Categorical test for geographical association: Once the nested design has been determined by the topology of the haplotype networks in the 95% plausible set, the simplest test for geographical association treats each sample location as a categorical variable. An exact permutational contingency analysis of categorical variation is then performed in a manner similar to that described in HUDSON *et al.* (1992a), but in this case the exact permutational tests are implemented in a nested fashion (clade types within a nested category *vs.* geographical location) using the exact, nested, contingency test procedures given in TEMPLETON and SING (1993). Although this contingency analysis can detect significant geographical associations and localize them within the haplotype cladogram, its power is diminished by the fact that it does not incorporate any information about the geographical distances or positions among the sample locations. Hence, a more elaborate analysis that utilizes information on geographical distance will now be described.

Geographical measures and tests for association: This paper will be limited to geographical distances between sampling sites (measured in kilometers), although there is nothing about the statistical methodology that would prevent other distance measures from being used when appropriate (*e.g.*, using river distances for a riparian species). Two measures of geographical distance will be used in the context of a nested, cladistic design. The first is that described in TEMPLETON (1993). Suppose we are examining the geographical range of n-step clade X that is nested in $(n + 1)$-step clade Y. The first step is to calculate the geographical center of all individuals bearing haplotypes that fall within clade X. This is accomplished by averaging the latitude and longitude over all observations from clade X. Then, for each individual bearing a haplotype from clade X, the geographical distance of that individual from clade X's geographical center is determined by using the standard formula for great circle distances. Finally, these individual distances are averaged for all members of clade X; to determine the average distance of clade X haplotypes from their geographical center. This distance will be called the clade distance of clade X, and is symbolized by $D_c(X)$. The biological interpretation of this distance measure is straightforward—it measures how geographically widespread are the individuals that bear haplotypes from clade X.

The second distance measure is the nested clade distance of clade X, symbolized by $D_n(X)$. To calculate this distance, the geographical center of the nesting clade (clade Y) is first determined. Next, all the great circle distances of the individual observations from clade X are determined from the geographical center of the nesting clade Y. These distances from the geographical center of the nesting clade are then averaged over all observations from clade X to yield $D_n(X)$. The nesting clade distance does not measure how geographically widespread clade X is, but rather how far individuals bearing clade X haplotypes are from all individuals that bear clade Y haplotypes (which includes clade X and other clades that are evolutionarily close to clade X). For example, if all individuals bearing clade X haplotypes are found in only one location, $D_c(X) = 0$, but $D_n(X)$ could be quite large if these X-bearing individuals are found far from individuals bearing the other clades nested in Y.

Figure 1 provides a hypothetical example of the type of data needed to calculated these distances for a sample of three 0-step clades (haplotypes) found in a single nested category. In this hypothetical example, all copies of haplotype 1 were found in sampling area A, so $D_c(1) = 0$. One-third of the sample bearing haplotype 2 is found in area A, and two-thirds are found in area B, with area A being 2 distance units from the geographical center of all copies of haplotype 2, and B being 1 distance unit. These numbers yield a clade distance of $D_c(2) = (1/3)(2) + (2/3)(1) = 1.33$ distance units. Similarly, $D_c(3) = (1/3)(1.9) + (1/3)(1.9) + (1/3)(1.9) = 1.9$ distance units. Nested clade distances are calculated in a similar fashion, but in this case as the average distance of the observations falling into a particular haplotype class from the geographical center of the entire nesting clade indicated by the "N" in the octogon. Thus, $D_n(1) = (1)(1.6) = 1.6$ distance units, $D_n(2) = (1/3)(1.6) + (2/3)(1.5) = 1.53$ distance units, and $D_n(3) = (1/3)(1.6) + (1/3)(1.5) + (1/3)(2.3) = 1.8$ distance units.

Several additional geographical statistics could have been calculated (*e.g.*, the mean distance between members of a clade, the standard deviation of the distance between individuals within a clade, *etc.*). Many of these additional distance measures were calculated in preliminary analyses, but the information contained in them was redundant with that contained in the above mea-

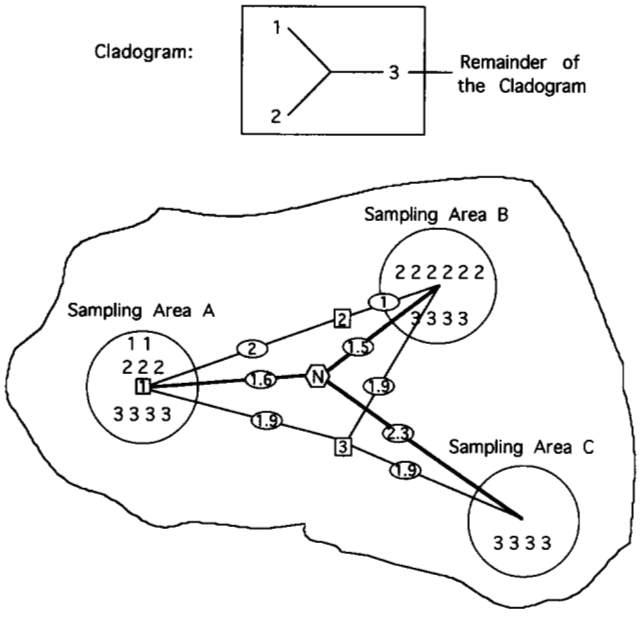

FIGURE 1.—An hypothetical example of clade and nested clade distances. Three sampling areas on an island are indicated by the letters A, B, and C. Within each sampling area, the number of times haplotypes of type 1, 2, and 3 are sampled is indicated by the number of times their respective haplotype numbers appear within the circle centered at the sampling location. These three haplotypes are all within a common nested clade, as indicated at the top of the figure, and note that haplotypes 1 and 2 are tips and that haplotype 3 is an interior haplotype. The geographical center of a particular haplotype is indicated by a box containing the number of the haplotype. The geographical center of the entire nested clade is indicated by the hexagon enclosing the letter N. The great circle distance between these geographical centers and the sampling locations are indicated by the numbers enclosed on an oval on the line connecting the geographical locations under consideration.

sures. Hence, the analyses will only use the D_c and D_n distances and their associated test statistics as these statistics are straightforward to compute and have obvious phylogeographical interpretations.

The distributions of these two distance measures under the null hypothesis of no geographical associations within the nesting clade are determined by recalculating both distances after each random permutation of clades against sampling location that was used for the nested contingency analysis described above. As discussed in TEMPLETON and SING (1993), the permutation procedure preserves all marginal values (that is, the clade frequencies and sample sizes are held constant) by using the algorithm of ROFF and BENTZEN (1989). A minimum of 1000 random permutations are needed to make statistical inference at the 5% level of significance (EDINGTON 1986). Such tests are only performed when there is both more than one clade and more than one sample location with nonzero observations in a nested category. These randomization procedures allows us to test for significantly large and small distances (both D_c and D_n) for each clade within a nested group of clades with respect to the null hypothesis of no geographical associations within the nested clade. These tests are asymptotically independent, both at the same clade level and across clade levels because of the nested design (PRUM *et al.* 1990; TEMPLETON and SING 1993).

The primary utility of having the D_c and D_n distance measures is that their joint analysis allows discrimination between short- *vs.* long-distance movement (either individual dispersal or population movements). Given that the null hypothesis has been rejected for one or both distances, we know that movement has been restricted to some extent. If that restriction is due exclusively to small distance dispersal/population events on a per generational basis, then all significantly restricted clades should be geographically close to their evolutionary neighbors. In contrast, if the significant restriction is due to rare but long-distance movements of individuals or populations, then clades can be found geographically far away from some of their evolutionary sister clades. This will cause the D_ns to be much larger than the D_cs. Hence, long-distance movements are inferred when there is a significant discrepancy in the patterns of clade *vs.* nested clade distances, whereas a pattern of concordance of these two distance measures implies short-distance movements. This prediction, however, assumes that the geographical range of the populations under study has been adequately sampled. If the sampling is sparse or if there are geographical gaps in the sampling process, the nested clade distances can become quite large even if the actual spread of haplotypes was accomplished by short-distance movements. Hence, the ability to discriminate between long- *vs.* short-distance movements increases with finer geographical sampling.

Predicted distance patterns under restricted gene flow: Rejecting the null hypothesis for clade and/or nested clade distances merely tells us that there is a significant association between haplotypes and geography. However, to discriminate between population history and population structure, we first need to have predictions for the population structure hypothesis that the geographical associations are caused by some sort of restricted gene flow. The patterns associated with restricted but nonzero gene flow have been investigated by NEIGEL *et al.* (1991) and NEIGEL and AVISE (1993) through computer simulations that included both short and long distance dispersal and equilibrium and nonequilibrium situations (with respect to time, population size effects, and cycles of fragmentation and range expansion) and by NATH and GRIFFITHS (1993) and SLATKIN (1991, 1993) using coalescent theory with several models of restricted gene flow. These simulations and theoretical results show that a robust attribute of restricted gene flow is that the geographical extent of a haplotype is strongly correlated with its age. Basically, the older the haplotype, the more widespread it tends to be under a

restricted gene flow model. This basic prediction is not obscured when confounding historical events are also overlaid upon the process, such as past fragmentation and range expansion (NEIGEL and AVISE 1993).

To look for the restricted gene flow pattern described above, it is necessary to have some degree of temporal polarity in the haplotype network, which up to now has been assumed to be unrooted. If the haplotype network can be rooted reliably, the polarity of the clades is then known. Unfortunately, obtaining reliable roots for intraspecific haplotype trees is extremely difficult and traditional methods (such as the use of outgroups) usually do not work (MADDISON *et al.* 1992; TEMPLETON 1992, 1993; CASTELLOE and TEMPLETON 1994). Fortunately, neutral coalescent theory indicates that there is considerable information about age polarity from the topology of the unrooted tree itself and from the haplotype frequencies (CASTELLOE and TEMPLETON 1994). CASTELLOE and TEMPLETON (1994) distinguished between two topological classes of clades: tip clades that are connected to the remainder of the cladogram through only one connecting branch of mutational changes, and interior clades that are connected to the remainder of the cladogram through two or more connecting branches and that therefore represent interior nodes. CASTELLOE and TEMPLETON (1994) then showed that tip clades are almost always younger than the interior clades to which they are connected. This prediction of tips being younger than interior clades is also expected to be robust to deviations from neutrality (CASTELLOE and TEMPLETON 1994). One limitation on the use of coalescent theory for assessing relative ages of clades is that CASTELLOE and TEMPLETON (1994) also found that although the tip/interior contrast was very strong, there is much less ability to discriminate among the relative ages of tips as a group or of interiors as a group. Fortunately, this is not a problem for the nested analysis. All distances and statistical assessments are performed within nested groups of evolutionarily close clades: that is, tips are contrasted to clades that are interior specifically to them and are not contrasted with other tip clades. Hence, contrasts of interior *vs.* tip clades within a nested clade for a given clade level will strongly tend to be contrasts of older *vs.* younger clades and *at most* can be violated only once (given one tip is the oldest, all other tips must necessarily be younger than the interior clades to which they are connected). Accordingly, within each nested category and for both types of distances (D_c and D_n), we also calculate and determine the statistical significance of the average interior distance minus the average tip distance, provided of course that unambiguous variability in tip/interior status exists among the clades within the nesting clade. For example, consider the hypothetical situation illustrated in Figure 1. Haplotypes 1 and 2 are both tips, with clade distances of 0 and 1.33, respectively, and nested clade distances of 1.6 and 1.53, respectively. Haplotype 3 is the only interior clade within this nested category, with a clade distance of 1.9 and a nested clade distance of 1.8. Hence, the average interior clade distance minus the average tip clade distance is $1.9 - (0 + 1.33)/2 = 1.23$. Similarly, the average interior nested clade distance minus the average tip nested clade distance is $1.8 - (1.6 + 1.53)/2 = 0.23$. Thus, in this hypothetical example the interior clade is more widespread geographically than the tips by both distance measures.

A second type of temporal polarity is found in the nested design itself. Obviously, the age of a clade is the maximum of the ages of the smaller clades nested within it. Hence, on the average, the larger the clade-level, the older the clade within a nested series of clades. Once again, note that all contrasts are within a nested series, so no assumptions about the relative ages of clades in different nested groups are being made.

Given the above indicators of temporal polarity, we can now translate the basic result that older clades are more widespread under recurrent, restricted gene flow into an expected pattern for the geographical distance statistics defined earlier. Because younger clades are expected to be less widespread relative to older clades under restricted gene flow, we predict that tip clades should be less widespread than the clades interior to them; a prediction with empirical support (CRANDALL and TEMPLETON 1993). Hence, significantly small clade distances should be associated with tip clades whereas significantly large clade distances, if they occur, should be associated with interior clades. When more than one interior clade exists within a nesting clade, some—but not all—interior clades may also display significantly small clade distances. These predictions under restricted gene flow are summarized in Table 1 (pattern 1.a). A second way of detecting this same effect of restricted gene flow is through having the average interior clade distance minus the average tip clade distance be significantly large (pattern 1.b in Table 1).

A third pattern predicted from the relative ages of clades under restricted gene flow models is for the average clade distances to increase with increasing clade level. If the rate of gene flow is high enough to uniformly distribute the older clades throughout the population's geographical range, then the average clade distances should level off at the higher level clades. At this point, the null hypothesis of no geographical association should be accepted even if this null hypothesis had been rejected at one or more of the lower clade levels. Hence, with restricted gene flow, we either expect the average clade distances to increase with clade level within a nested series, or to level off while simultaneously accepting the null hypothesis of no geographical association even when this null hypothesis had been rejected for lower clade levels within the nested series (pattern 1.c in Table 1).

When the restricted gene flow is due only to short

TABLE 1

Expected patterns under the different models of population structure and historical events

Pattern 1. Restricted gene flow
- a. Significantly small D_cs, primarily for tip clades. Some interior clades with significantly large D_cs.
- b. $\bar{D}_c(I) - \bar{D}_c(T)$ significantly large, where $\bar{D}_c(I)$ is the average clade distance of interior clades within the nested category, and $\bar{D}_c(T)$ is the average clade distance of tip clades.
- c. Average D_cs should increase (and occasionally level off) with increasing clade level in a nested series of clades. If the distances level-off, the null hypothesis of no geographical association should no longer be rejected even though rejected at lower clade levels.
- d. The above patterns also hold for the D_ns unless some gene flow is due to long-distance dispersal events, then significant reversals of the above pattern can occur with the D_ns.

Pattern 2. Range expansion
- a. Significantly large D_cs and D_ns for tip clades, and sometimes significantly small for interior clades under contiguous range expansion, but some tip clades should show significantly small D_cs under long-distance colonization.
- b. $\bar{D}_i(I) - \bar{D}_i(T)$ significantly small for i = c, n for contiguous range expansion and for $i = n$ for long-distance colonization.
- c. The above patterns are not recurrent in the cladogram or are geographically congruent.

Pattern 3. Allopatric fragmentation
- a. Significantly small D_cs, primarily at the higher clade levels. The D_ns at this clade level may suddenly increase rapidly while the D_cs remain restricted, depending upon the geographical configuration of the isolates.
- b. The pattern of distances described in (a) should represent a break or a reversal of the distance pattern established by the lower level nested clades.
- c. Clades showing pattern (a and b) should tend to be connected to the remainder of the cladogram by a larger-than-average number of mutational steps.
- d. The above patterns are not recurrent in the cladogram or are geographically congruent.

dispersal distances (isolation by distance), we expect the above pattern for clade distances to also hold for nested clade distances, as previously discussed (pattern 1.d in Table 1). In contrast, when long-distance dispersal occurs significant reversals of the above pattern should be found for the nested clade distances (e.g., a pattern of tip clades with significantly small clade distances but significantly large nested clade distances). This is also indicated in pattern 1.d in Table 1. We repeat our warning, however, that even an isolation by distance model will yield significantly large nested tip clade distances if the geographical sampling is sparse or has gaps. Thus, with inadequate sampling, it may not be possible to discriminate between isolation by distance and the long distance gene flow models.

Definition of "historical events": The null hypothesis of no geographical association can also be rejected due to the effects of nonrecurrent, historical events. In this paper, we confine our attention to two types of historical events: range expansions and range contractions (fragmentation events). To be considered as a nonrecurrent historical event, such a range alteration must have occurred only once or sporadically during the time interval marked by the coalescence time of the gene region under investigation. Because our ability to detect geographical associations depends upon having mutational resolution in the haplotype tree, an historical event cannot be older than the coalescence time for the gene region being investigated. Note that what is considered to be a nonrecurrent historical event depends in part upon the gene region being investigated, which in turn determines the effective time scale through coalescence time and rate of mutational accumulation.

Predicted distance patterns under population range expansion: The pattern predicted under a population range expansion has been described by CANN *et al.* (1987) as being one in which some of the older haplotypes are confined to the ancestral, preexpansion area while some of the younger haplotypes that arose in the expanding populations are widespread geographically or distantly located from their ancestral (interior) haplotypes. Because the expansion in this case is an event by definition as given above, the patterns predicted by CANN *et al.* (1987) should not be recurrent, but they can occur in more than one nested category due to ancestral polymorphism in the expanding populations. However, in this case, the significant patterns should be geographically congruent since they reflect the same historical event. If the range expansion is achieved through an expanding population front via individual short-distance dispersal (contiguous range expansion), concordance of the D_c and D_n patterns is predicted, but significant discordance can arise if range expansion is achieved through long-distance colonization (once again, assuming an adequate sampling design). These predictions in terms of the distance statistics used here are given in pattern II of Table 1.

Predicted distance patterns under population fragmentation: We now consider fragmentation events in which an ancestral population had become subdivided into two or more subpopulations with no gene flow among them. Immediately after the fragmentation event, the isolates will simply reflect the prefragmented population structure and hence be indistinguishable from ordinary population structure. As the generations

go by, mutations will occur independently in the isolates, causing them to become genetically differentiated (HUDSON 1990; HEY 1991). Fragmentation can occur over both short and long distances. For example, a climatic change may cause a species' populations to survive in isolated habitat islands that extend throughout much of its former range (see TEMPLETON *et al.* 1990 for examples). The resulting habitat islands can be very close to one another, a phenomenon known as microvicariance. In contrast, other vicariant events can sever an ancestral population into two or more allopatric isolates that are clearly separated geographically. Only the latter case will be dealt with in this paper and the microvicariance situation will be dealt with in a subsequent paper. Another complication of allopatric isolates occurs when subsequent range expansion events bring the former isolates back into contact. As long as gene flow has been sufficiently restricted since secondary contact, a genetic signal of former fragmentation events should persist and influence the present-day distribution of generic variation by causing the clades found in the different former isolates to be only partially overlapping in geographical range.

Fragmentation, by definition, imposes strict limits on the geographical range of a clade. Hence, the primary expectation is for significantly small clade distances, both for tips and interiors. However, unlike models of restricted gene flow in which clade distances are expected to increase or level-off with increasing clade level, a significant restriction of clade distances at a high clade level should occur under fragmentation. The clade level will tend to be high because of the tendency to accumulate fixed mutational differences after fragmentation (HUDSON 1990; HEY 1991). This restriction will often represent a reversal of the distance patterns observed at the lower clade levels that have been affected by recurrent gene flow within the fragmented subpopulations. Although the clade distances should be significantly small at the clade levels that mark the fragmentation event, the nested clade distances at this clade level may suddenly increase rapidly if the isolates have well separated geographical centers, although this will not occur in those cases in which the geographical centers of the different isolates are close to one another. The accumulation of mutational differences between the clades that are restricted to different isolates also implies that the significant restrictions of clade distances should coincide with larger-than-average branch lengths in the cladogram. Finally, because the fragmentation is an event by definition, the above patterns should either not be recurrent in the cladogram, of if they affect more than one topological section of the cladogram due to ancestral polymorphism, these different topological subsets of the cladogram should be associated with geographically congruent subpopulations. These expectations are also summarized in Table 1 as pattern III.

Because of the complexity of the patterns described in Table 1, the APPENDIX provides a key to interpreting the distance statistics when the null hypothesis is rejected and to identifying situations when the sampling design is inadequate for inference. It is important to emphasize that the patterns summarized in Table 1 and incorporated into the key are not necessarily mutually exclusive. One great advantage of the nested cladistic design is that it performs many local analyses. Hence, one pattern may appear in one nested group or at some nesting level, while a different pattern occurs in a second nested group or nesting level. Thus, when geographical associations are detected, this analysis allows those associations to have arisen by any mixture of the forces summarized in Table 1.

MATERIALS AND METHODS

The collecting techniques and DNA methodologies for the work on *A. tigrinum* are given in ROUTMAN (1993). Figure 2 shows the sampling sites for this species.

RESULTS

The restriction site states of the 23 mtDNA haplotypes found in this species as well as the numbers of each haplotype that were found at each of the 53 sampling sites are given in ROUTMAN (1993). Using the formulae given in TEMPLETON *et al.* (1992) with the data given in ROUTMAN (1993), mtDNA haplotypes separated by up to seven mutational steps have a probability ≥ 0.95 of being connected in a parsimonious fashion. Using maximum parsimony within these limits, two disjoint networks are obtained, each with no internal ambiguities, as shown in Figure 3. These two haplotype networks are connected by a minimum of 14 mutational steps, well beyond the confidence limits for parsimony. Hence, we cannot be sure as to how these two disjoint networks are connected. Figure 3 also shows the nested design, using the rules given in TEMPLETON *et al.* (1987) and TEMPLETON and SING (1993). Despite the ambiguity in how the two disjoint portions of the cladogram are connected, there is no ambiguity in the nested design.

Table 2 presents the results of the nested contingency analyses for the mtDNA data set from *A. tigrinum*. This analysis simply treats the sampling locations as categorical variables and does not incorporate any geographical distance information. Nevertheless, the contingency analysis indicates that strong associations exist between clades and geographical location.

Figure 4 presents the results of the nested cladistic analysis of geographical distance for the mtDNA data set from *A. tigrinum*. As an aid in interpreting these results, Figure 5 presents a rough overlay of the cladogram upon the geography. Table 3 presents the results obtained when the key given in the APPENDIX is applied to the statistical results given in Figure 4 along with resulting inferences about population structure and history.

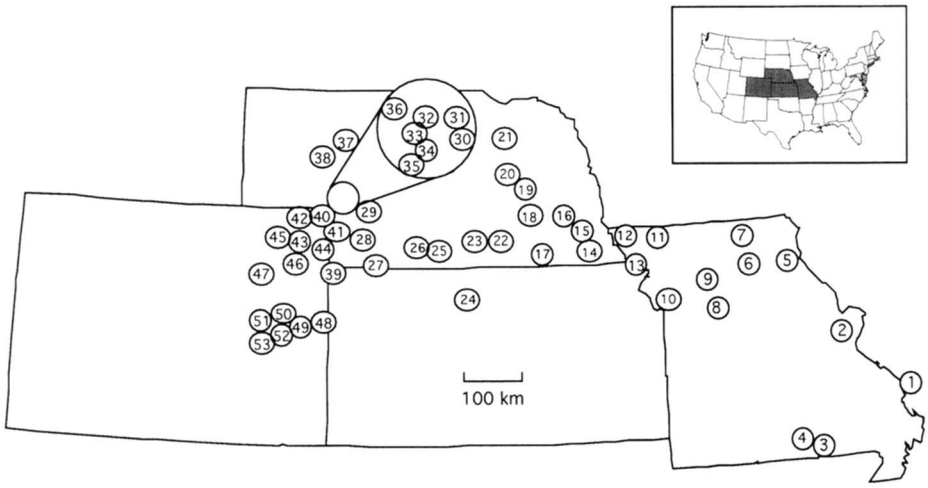

FIGURE 2.—The geographical locations of the sample sites for the study on *Ambystoma tigrinum*. From ROUTMAN (1993).

DISCUSSION

The results shown in Table 3 indicate that both population structure and population history have played an important role in determining the geographical associations of mtDNA haplotypes for tiger salamanders. This raises the question of just how reliable are these conclusions? To examine this question, we first consider the nature of the inference structure of the procedures given above.

The null hypothesis is that there is no association between the position of a haplotype in a cladogram with geographical position. There are three reasons why this null hypothesis would fail to be rejected: the populations under study have sufficient gene flow to be virtually panmictic and have not experienced expansion or fragmentation events, the samples are inadequate (either in terms of sample sizes per location or of the number and geographical positions of the locations sampled), or there is insufficient genetic variation in the sampled populations. As the null hypothesis was rejected repeatedly, the first explanation is not applicable. However, some of the lower level nesting results illustrate the importance of sampling and levels of genetic variation. Note from Figure 4 that the null hypothesis was only rejected at the haplotype level in nesting clades 1–1 and 1–3. These two nesting clades contain

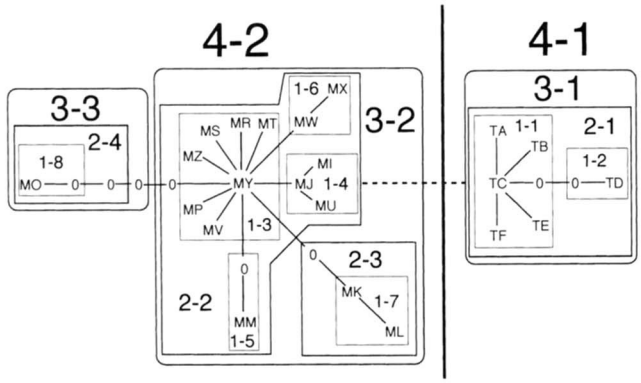

FIGURE 3.—The estimated 95% plausible set of cladograms and associated nested design for the mtDNA haplotypes found in *A. tigrinum*. Haplotypes are indicated by letter designations, as found in ROUTMAN (1993). Zeros indicate haplotype states that are necessary intermediates between observed haplotypes but that were not present in the sample. Each solid line represents a single mutational change that interconnects two haplotype states that has a probability greater than 95%. The thick dashed line indicates a multiple step mutational connection for which the exact interconnections are uncertain and for which parsimony is not supported at the 95% level. Narrow-lined boxes enclose one-step clades, which are designed by "1-*x*" where *x* is a number assigned to identify the clade; wide-lined boxes and polygons enclose two-step clades ("2-*x*"); rounded boxes enclose three-step clades ("3-*x*"); and a thick solid line separates the two four-step clades, 4–1 and 4–2. Technically, these last two clades are greater than four-step clades because of the large mutational distance (a minimum of 14 steps) that separates them, but these clades are the only nondegenerate categories at the level of four-step or above.

TABLE 2

Nested contingency analysis of geographical associations

Clade	Permutational chi-square statistic	Probability
1-1	148.88	0.000
1-3	784.21	0.000
1-4	98.93	0.025
1-6	0.88	1.000
1-7	1.07	0.612
2-1	128.00	0.000
2-2	376.82	0.000
3-1	312.93	0.000
4-1	36.48	0.490
Entire cladogram	496.00	0.000

Associations are for the mtDNA haplotype data gathered from *Ambystoma tigrinum*. Clades with no genetic and/or geographical variation within them are not given as no test is possible within such nested categories. The "Clade" column refers to the nesting clade, and the permutational chi-square probability was calculated by randomly permuting the lower level clade categories within the nesting clade *vs.* geographical locality one thousand times (*i.e.,* haplotypes were permuted versus locality within the designated one-step clades, one-step clades were permuted within the designated two-step clades, *etc.*).

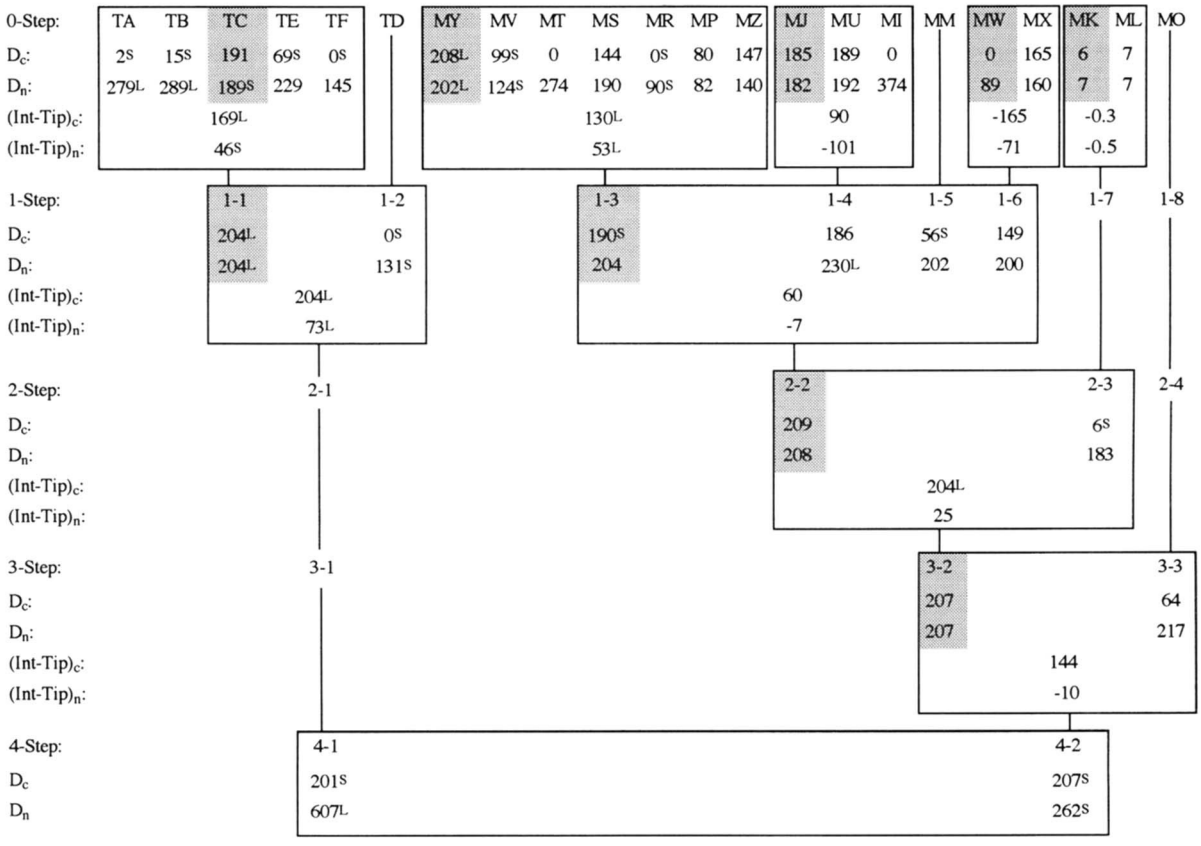

FIGURE 4.—Results of the nested cladistic analysis of geographical distance for the mtDNA haplotypes of *A. tigrinum*. The haplotype designations are given at the top and are boxed together to reflect the one-step nested design given in Figure 3. Higher level clade designations are given as one moves down the figure, with boxed groupings indicating the nesting structure. Immediately below each clade designation are the clade and nested clade distances respectively. An "S" superscript indicates the distance is significantly small at the 5% level, and an "L" indicates that it is significantly large. For nested clades in which the tip/interior status is known and for which both tips and interiors exist within the same nesting group, the clade name and distances are shaded for interior clades and are left unshaded for tip clades. At the bottom of the boxes that indicate the nested groups containing both tip and interior clades, the lines indicated by the symbols "$(\text{Int-Tip})_c$" and "$(\text{Int-Tip})_n$" give the average difference in distances between interior clades and tip clades within the nested group for clade distances and nested clade distances, respectively.

most of the genetic variation at the haplotype level found in the entire sample, cover virtually all sampling locations and have the largest sample sizes. In contrast, nesting clades 1–4 and 1–6 have smaller levels of haplotype variation and smaller numbers of observations. Nesting clade 1–7 in addition suffers from having few sample locations included within it. Hence, the failure to reject the null hypothesis in these cases has undoubtedly been affected by a lack of statistical power. However, one great benefit of the nested design is that it results in an automatic pooling procedure, so that statistical power can be recovered at higher levels of analysis (TEMPLETON *et al.* 1987). Indeed, even the haplotypes in clade 1–7 result in a significant rejection of the null hypothesis when pooled together within nesting clade 2–3 (see Figure 4).

Despite these localized instances of lack of power, overall the data analysis revealed strong geographical associations, so that the null hypothesis can clearly be rejected for this species (Table 2 and Figure 4). Given a rejection of the null hypothesis, our inference structure now enters its second stage by attempting to identify the likely causes for the observed geographical associations via the predictions outlined in Table 1 and incorporated into the key given in the APPENDIX.

The first cause considered in this paper is restricted gene flow. The predictions given in Table 1 and the APPENDIX are not tied to a specific model of gene flow, but rather are based on two qualitative attributes of gene flow: gene flow is a recurrent evolutionary force (which is true by definition) and older clades tend to be more geographically widespread than younger ones. This later attribute is well supported by both computer simulations and population genetic theory (NEIGEL *et al.* 1991; Slatkin 1991, 1993; NATH and GRIFFITHS 1993; NEIGEL and AVISE 1993). However, what is less clear is how often this tendency for older clades to be more widespread is violated, and how often our tip/interior criterion for relative age is violated. With respect to the later, the computer simulations of CASTELLOE and

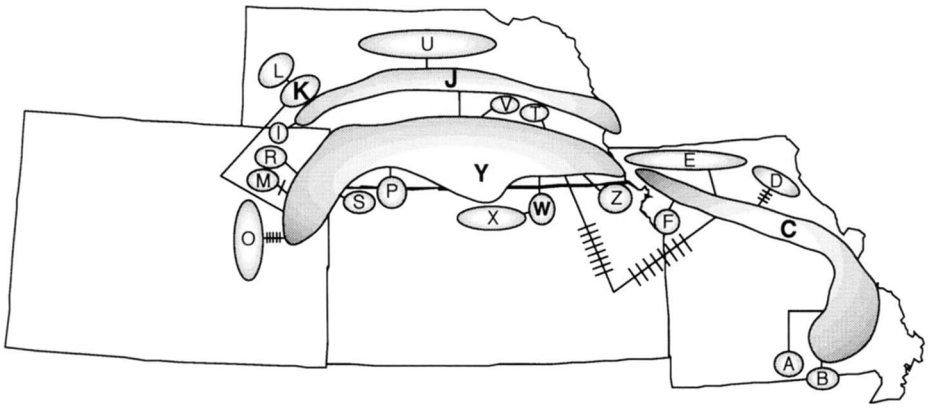

FIGURE 5.—A rough geographical overlay of the *A. tigrinum* mtDNA cladogram (Figure 3) upon the sampling locations (Figure 2). Many of these haplotypes are found as polymorphisms in the same ponds, but are shown as nonoverlapping in this figure for ease of pictorial representation. Consequently, the distributions indicated in this figure are only approximate. The "M" and "T" prefixes to the haplotypes (as seen in Figures 3 and 4 and that indicate the formal subspecies "*mavortium*" and "*tigrinum*") are not given here in order to save space. Haplotypes A–F are found in *A. t. tigrinum*, and haplotypes I–Z are found in *A. t. mavortium*. Haplotypes in bold are interior haplotypes; all other haplotypes are tips.

TEMPLETON (1994) indicate that the tip/interior criterion is only rarely violated. Moreover, when it is violated for a given clade level, at most only one tip/interior contrast will be affected. The only way a violation can occur is when one of the tips is the oldest clade and therefore serves as an outgroup to the remainder of the haplotype network. This automatically insures that all other tips are evolutionarily younger than the interior clades to which they are connected because these interior clades lie between the other tips and the "outgroup" tip in a topological sense. Hence, even when a tip is the oldest clade (an unlikely event in itself), it will always have a limited impact on the assumed temporal polarities in the various nested categories. Nevertheless, if the the tip/interior criterion for temporal polaritity is indeed violated, and the clade affected by this violation is associated with a statistically significant rejection of the null hypothesis, the application of the inference key provided in the APPENDIX could indeed lead to an incorrect biological inference. It still remains to be determined how serious or frequent this problem will be.

Another concern is the question of how often do younger clades become more widespread under restricted gene flow than the interior clades to which they are connected. If this is a frequent occurence in nature, it would seriously undermine the validity of the criteria used to discriminate between population structure and population history. This important issue therefore needs to be thoroughly examined in future studies. One approach that will be used is computer simulations to quantify the likelihood of this possibility under different gene flow models. Another strategy is empirical verification using groups for which there is some prior evidence for restricted flow, a strategy already initiated in CRANDALL and TEMPLETON (1993) and for which the current worked example can be regarded as a continuation.

The basic pattern predicted for restricted gene flow is for tips to be more geographically restricted than

TABLE 3

Inference chain on the results given in Figure 4

Clade	Chain of inference	Inference
Haplotypes nested in 1-1	1-2-3-5-6-13-14 NO	Range expansion, but cannot discriminate between contiguous range expansion and long-distance colonization
Haplotypes nested in 1-2	1-2-3-4 NO	Restricted gene flow via isolation by distance
One-step clades nested in 2-1	1-2-3-4 NO	Restricted gene flow via isolation by distance
One-step clades nested in 2-2	1-2-11-12 NO	Contiguous range expansion
Two-step clades nested in 3-2	1-2-3-4 NO	Restricted gene flow via isolation by distance
>Four-step clades nested in entire cladogram	1-2-3-5-9 NO and associated with longest branch length	Allopatric fragmentation

The chain uses the key in the APPENDIX on the results given in Figure 4. Only those clades that resulted in a rejection of the null hypothesis are included in this table.

interiors, for tips to be scattered throughout the range of the interior clades, and for these patterns to keep occurring at higher and higher clade levels unless geographical homogeneity is achieved. All of these patterns are certainly evident in Figure 5, and indeed restricted gene flow via isolation by distance was inferred for many of the clades (Table 3 and Figure 4). These statistical inferences are concordant with what is known of the biology of these salamanders. First, this species is an obligate pond breeder. This means that they cannot disperse in their larval aquatic phase as riparian species can. Rather dispersal is accomplished through the adult, terrestrial phase, which has only limited physical abilities for terrestrial movement. These limited dispersal abilities are quite evident in an F_{st} analysis of nine ponds in western Nebraska that are separated by no more than 36 km. Despite close geographical proximity, $F_{st} = 0.433$ for mtDNA and 0.405 for nuclear allozyme loci (both significantly different from 0), indicating highly restricted movements among these ponds (ROUTMAN 1993).

A second reason for predicting restricted gene flow is that many adults become paedomorphic; that is, they become sexually mature in the aquatic phase (ROUTMAN 1993). As long as adults remain in this facultative paedomorphic condition, dispersal between ponds is impossible. Third, PHILLIPS and SEXTON (1989) have shown that the closely related *Ambystoma maculatum* displays strong orientation behavior such that many individuals tend to return to the same pond in successive breeding seasons. This phenomenon is thought to be common in ambystomid salamanders and would further restrict effective gene flow. Hence, an isolation by distance model of gene flow is to be expected for this species, and indeed, such was detected. Thus, the current example can be regarded as supporting the validity of the fundamental premise of this analysis that tip clades are more geographically restricted than the interior clades to which they are connected under restricted gene flow, a premise that is also empirically supported by the analyses given in CRANDALL and TEMPLETON (1993). Given the critical role that this premise plays in the inference scheme outlined in this paper, much more verification (both empirically and through simulation) is warranted and will be performed.

A second cause of geographical association considered in this paper is range expansion. The predictions given in Table 1 and incorporated into the APPENDIX are based upon the arguments found in CANN *et al.* (1987) that range expansion causes some older (interior) haplotypes to be left in the ancestral area while younger (tip) haplotypes that originated in the expanding population can be geographically widespread and/or distant from their ancestral haplotypes (which have to be interior by definition). This pattern does not as yet have the backing of computer simulation models, extensive theory, or initial empirical surveys, as did the fundamental gene flow model predictions. Fortunately, however, these predictions are amenable to extensive empirical verification. For example, there are a large number of species that are known to have expanded their geographical ranges into formerly glaciated areas. Hence, there is a plethora of opportunities for studying biological examples that had to have experienced range expansion to verify these predictions and to quantify how likely they are to be violated. The current example should be regarded as initiating this program of empirical verification.

Two separate instances of population range expansion were inferred for these *A. tigrinum* samples; one in clade 1–1 and the second in clade 2–2, each within a separate allopatric fragmented subpopulation, as will be discussed shortly. The range expansion detected within clade 1–1 involves haplotypes found in southern Missouri, an area never subjected to glaciation, and in northern Missouri, an area that was glaciated at least once and that was heavily impacted climatically by other glacial advances (NIGH *et al.* 1992). Obviously, salamander range expansion must have occurred into northern Missouri only during the last 18,000 years, and the present analysis detected this expansion.

The range expansion in the western fragment of *A. tigrinum* identified within clade 2–2 involves individuals found in a broad arc extending from Colorado through southeastern Nebraska and into northwestern Missouri (Figure 5). This is also consistent with a post-Pleistocene expansion of salamander populations from the southern, eastern edge of the Rockies expanding into formerly glaciated areas of the northern Great Plains. In summary, the current analysis leads to the inference of range expansion in both cases in which salamanders are living in formerly glaciated areas and in no other cases. This implies that the predictions of CANN *et al.* (1987) made with regard to human populations have at least some generality to other species. Obviously, more examples are necessary to establish the extent of this generality and the robustness of this expected pattern for range expansion, and these examples will be forthcoming in future papers.

The two inferences of range expansion also illustrate the importance of the geographical sampling design. The range expansion found in the western subpopulation (clade 2–2) was specifically inferred to be a contiguous range expansion. In contrast, it is not possible to discriminate between contiguous range expansion and long-distance colonization for the expansion occurring from southern to northern Missouri within clade 1–1 (Table 3). These differences reflect the fact that the western subpopulation was sampled thoroughly throughout its geographical range whereas the eastern subpopulation was sampled sparsely in central Missouri, the geographical center of the expanding clade 1–1 (see Figure 2). When sampling is sparse in an area, it becomes impossible to discriminate between short and

long-distance movements. Hence, strong inferences about the forces that explain the geographical distribution of genetic variation require adequate geographical sampling, a seemingly obvious point that has sometimes been ignored (*e.g.*, see TEMPLETON 1993 and 1994 for a discussion of this point with respect to the out-of-Africa replacement hypothesis of human evolution and the spread of mtDNA across the world).

The third cause of geographical association considered in this paper is allopatric fragmentation. Fragmentation is predicted to impose strict geographical limits on how widespread all clades can become and to result in the accumulation of mutational differences among clades found in different isolates. Many of the predictions used to detect fragmentation follow simply from the definition of allopatric fragmentation as a nonrecurrent historical event involving subpopulations that are completely or nearly completely nonoverlapping geographically. The predictions with respect to high clade level and larger-than-average branch lengths are well established theoretically (HUDSON 1990; HEY 1991). Indeed, the criteria used to detect fragmentation events in this paper have already been extensively used in the literature (see AVISE 1994 and references therein), albeit usually implicitly and with no statistical assessment of whether or not the samples are sufficient to reject the null hypothesis of no geographical association. Hence, Table 1 and the APPENDIX make the commonly used criteria for inferring fragmentation explicit and more objective by coupling it with significance testing.

Population fragmentation was clearly inferred for *A. tigrinum* by satisfying all of the criteria associated with fragmentation (Table 1.III). The clades that identify the fragmentation event separate the species into western (clade 4–2) and eastern (clade 4–1) subpopulations that only narrowly overlap in the northwestern corner of Missouri (locations 10, 11 and 13 in Figure 2). Because of the large number of mutational steps that separate these two clades (a minimum of 14 mutational steps—by far the longest branch length in the cladogram), these clades are not contrasted until the highest clade level. They are symbolized with the step number "4," although technically this is higher than a four-step contrast because of the many mutational steps that separate these two clades.

This inference of eastern and western isolates can also be subjected to verification by the use of prior information. First, the two clades identified as fragmented by the analysis in this paper correspond exactly to two named subspecies based upon morphological differentiation, *A. t. mavortium* in the west, and *A. t. tigrinum* in the east. The morphological differences are observed at all life stages, including egg mass size, larval gill rakers, and adult color pattern (COLLINS *et al.* 1980). Second, these two subspecies are also ecologically differentiated, with *A. t. tigrinum* being obligately metamorphic whereas *A. t. mavortium* is facultatively neotenic, and with *mavortium* frequently having cannibalistic larvae whereas *A. t. tigrinum* only has cannibalistic larvae in the far northwestern part of its range (COLLINS *et al.* 1993) where the work presented here indicates that it is likely to be in contact with *mavortium*. Third, the two subspecies are distinct in their nuclear DNA as measured by protein electrophoresis, with each subspecies having several unique alleles at many loci and some diagnostic loci (ROUTMAN 1993). All these indicate a long-standing separate evolutionary history of precisely the populations marked by the clades that identify the fragmentation event according to the nested analysis.

In light of the above, the present example offers an initial empirical validation of the criteria used to identify fragmentation. Once again, many more examples and simulations will be required before the robustness of these criteria is established. In addition, as with range expansion, the validity of these criteria to infer fragmentation depends critically upon the geographical sampling design. For example, consider a situation in which a species is continuously or nearly continuously distributed over a geographical area that is very large relative to typical individual dispersal distances. Suppose further that there has been no fragmentation event, but strong geographical associations have arisen due to gene flow restricted by isolation by distance. If the entire geographical distribution is sampled uniformally, the patterns associated with restricted gene flow would be expected to be observed. But if only geographically distant locations were sampled with no intermediate sampling locations, the effective amount of gene flow between those distant locations could be so low as to be regarded as sporadic on the time scale defined by the coalescent process of the genetic system being assayed. When the effective amount of gene flow between two locations is sporadic on the coalescent time scale, the gene tree pattern is similar to that given here for fragmentation (TAJIMA 1983; TAKAHATA and SLATKIN 1990). This is not surprising because if the rate of gene flow is so low as to become sporadic on the coalescent time scale, it satisfies our definition of an historical event and not a recurrent process. An example of this situation is provided by the cladistic analysis of geographical distribution of mtDNA haplotypes in African impala (*Aepyceros melampus*) given in TEMPLETON and GEORGIADIS (1995). Impala were sampled from several nearby locations in central Eastern African and in Southern Africa, but with no intermediate samples. The resulting cladistic analysis revealed that the two clades at the highest level of nesting were significantly restricted because one clade was found exclusively in central Eastern Africa, and the other exclusively in Southern Africa. In terms of the patterns outlined in Table 1, this is the expectation associated with fragmentation. However, because of the theoretical work of TAJIMA (1983) and TAKAHATA and SLATKIN (1990), the effect of inade-

quate geographical sampling were anticipated in this case, and the key in the APPENDIX accordingly incorporates questions about the geographical sampling design before making the biological inference of fragmentation. In the case of the impala, the chain of inference lead to questions 15 and 16 in the APPENDIX concerning the sampling scheme, and given that the intermediate locations had not been sampled, the answer to those questions lead to question 18 and the ultimate inference that it was impossible to discriminate with these samples between restricted gene flow, range expansion, or fragmentation despite a highly significant rejection of the null hypothesis. Given that relatively little theory has dealt with the problem of the interaction between geographical sampling and evolutionary processes upon gene tree geographical patterns, it is anticipated that future theoretical, simulation, and empirical studies will reveal additional pattern artifacts that can arise from various inadequacies of geographical sampling. As these potential artificats are identified, additional questions about geographical sampling will probably need to be incorporated into the inference key. Hence, the key given in the APPENDIX must be regarded only as a first attempt to integrate the expected patterns given in Table 1 with the geographical sampling scheme to produce biological inferences from statistical rejections of the null hypothesis. This key will undoubtedly undergo many refinements as our knowledge increases about the effects of sampling upon the expected patterns outlined in Table 1.

In the current example, the geographical sampling scheme was adequate to result in the biological inference of isolation by distance within each subspecies and the biological inferences of three population historical events (two range expansions and one fragmentation event). All of these inferences are concordant with prior data. Obviously, more examples will be needed to judge the accuracy and robustness of the predictions in Table 1, but this first test case indicates that the predictions are biologically reasonable under at least some real conditions.

The inferences made by this quantitative analysis are also concordant with the overlay of the haplotype network upon sampling locations shown in Figure 5; so much so that the question arises as to what is gained by performing this laborious statistical analysis as opposed to simply inspecting Figure 5 and making inferences. Indeed, such pictorial overlays of haplotype networks upon geography are currently the standard inference tool in the area of intraspecific phylogeography (see AVISE 1994 and references cited therein). Such pictorial representations are an excellent exploratory tool for formulating hypotheses, but are inappropriate inference tools. First, such pictures provide no assessment of whether or not the sample sizes were sufficient to truly discriminate among alternative hypotheses or even to reject the null hypothesis of no association. For example, note from Figure 4 that six haplotypes had clade distances of 0, but only half of these are identified as being significantly small. The reason is that all of the haplotypes with nonsignificant clade distances of 0 occur only once in the sample. Hence, they have to have a clade distance of 0, and that observation is therefore without statistical significance. In contrast, there are clade distances of up to 207 km that are significantly small, despite this being close to the largest clade distance observed in the entire table (209 km, which was not significant). Once again, this is a function in part of sample size and in part of the exact comparison being made. The clade distance of 207 is significantly small when the distribution of clade 4–2 is tested within the entire sample range because this clade is not found in most Missouri locations despite being very abundant in the sample as a whole. In contrast, the clade distance of 207 for clade 3–2 (just one nesting level below 4–2) is consistent with the null hypothesis of no geographical association when the distribution of clade 3–2 is considered only within the range of the western subspecies because this clade is uniformly distributed within the range of *A. t. mavortium*. These types of subtleties are not easily portrayed in a figure, yet they are critical for appreciating just what inferences the samples can or cannot justify. Hence, the absolute values in Figure 4 or the observed geographical ranges depicted in Figure 5 are not reliable guides to the conclusions supported by the data.

Another danger of making inferences from a figure is that the criteria for inference tend to be implicit and subjective. In the current analysis, when the null hypothesis is rejected the criteria for inference are explicit (Table 1) and can be applied in a manner that minimizes subjective biases (the key in the APPENDIX). Moreover, the APPENDIX forces the user to examine the nature of the geographical sampling scheme, and disallows biological inference whenever known inadequacies are found. Thus, patterns that superficially appear quite strong and straightforward (such as the apparent "fragmentation" in the impala discussed in TEMPLETON and GEORGIADIS 1995) are often judged inadequate for biological inference—a judgement that is not readily made from a mere pictorial overlay. The identification of such sampling inadequacies also provides a precise guide to the researcher for how to collect future samples to make sound biological inference.

The use of coalescent theory and cladistic analysis in this paper differs considerably, but in a complementary fashion, from the use of cladistics to study gene flow by HUDSON *et al.* (1992b); NEIGEL *et al.* (1991), SLATKIN (1989); SLATKIN and BARTON (1989), and SLATKIN and MADDISON (1989, 1990). The statistical inference structure presented in this paper is one of hypothesis testing and discrimination and is designed to identify and localize (in a geographical and cladistic sense) the effects of restricted gene flow and historical events on geo-

graphical associations of haplotypes. In contrast, the works cited above assume that the geographical associations are due to some sort of restricted gene flow, and given that assumption, attempt to estimate the amount of gene flow. The best use of these gene flow estimation algorithms would be after testing with the procedures outlined here. For example, the results of our analysis would indicate that estimating a gene flow parameter for the entire *A. tigrinum* data would be inappropriate because much of the association is due to past population fragmentation and range expansion events that do not reflect ongoing patterns of gene flow. Instead, the present analysis indicates that gene flow parameters can be estimated in a biologically meaningful fashion only if applied separately to the southern populations of *A. t. tigrinum* and the southwestern populations of *A. t. mavortium*. Extending the gene flow estimation procedure beyond these samples would violate the underlying assumptions of the gene flow model. Thus, the methods presented in this paper strengthen the biological validity of inferences made from gene flow estimation algorithms and prevent inappropriate uses of those algorithms.

We thank the graduate students and postdoctoral associates of the Templeton laboratory and three anonymous reviewers for their excellent comments upon an earlier draft of this work. This work was supported by National Institute of General Medical Sciences grant R01-GM-31571.

LITERATURE CITED

AMMERMAN, A. J., and L. L. CAVALLI-SFORZA, 1984 *The Neolithic Transition and the Genetics of Populations in Europe.* Princeton University Press, Princeton, NJ.

ANTONARAKIS, S. E., C. D. BOEHM, G. R. SERJEANT, C. E. THEISEN, G. J. DOVER *et al.*, 1984 Origin of the β^S-globin gene in Blacks: the contribution of recurrent mutation or gene conversion or both. Proc. Natl. Acad. Sci. USA **81:** 853–856.

AVISE, J. C., 1989 Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution **43:** 1192–1208.

AVISE, J. C., 1994 *Molecular Markers, Natural History and Evolution.* Chapman & Hall, New York.

AVISE, J. C., R. M. BALL and J. ARNOLD, 1988 Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. Mol. Biol. Evol. **5:** 331–344.

CANN, R. L., M. STONEKING and A. C. WILSON, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CASTELLOE, J., and A. R. TEMPLETON, 1994 Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylog. Evol. **3:** 102–113.

COLLINS, J. P., J. B. MITTON and B. A. PIERCE, 1980 *Ambystoma tigrinum*: a multispecies conglomerate? Copeia **1980:** 938–941.

COLLINS, J. P., K. E. ZERBA and M. J. SREDL, 1993 Shaping intraspecific variation: development, ecology and the evolution of morphology and life history variation in tiger salamanders. Genetica **89:** 167–183.

CRANDALL, K. A., 1994 Intraspecific cladogram estimation: accuracy at higher levels of divergence. Syst. Biol. **43:** 222–235.

CRANDALL, K. A., and A. R. TEMPLETON, 1993 Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics **134:** 959–969.

EDGINGTON, E. S., 1986 *Randomization Tests,* Ed. 2. Marcel Dekker, New York.

EPPERSON, B. K., 1993 Recent advances in correlation studies of spatial patterns of genetic variation. Evol. Biol. **27:** 95–155.

EWENS, W. J., 1990 Population genetics theory—the past and the future, pp. 177–227 in *Mathematical and Statistical Developments of Evolutionary Theory,* edited by S. LESSARD. Kluwer Academic Publishers, Dordrecht, The Netherlands.

EXCOFFIER, L., and P. E. SMOUSE, 1994 Using allele frequencies and geographic subdivision to reconstruct gene trees within a species—molecular variance parsimony. Genetics **136:** 343–359.

EXCOFFIER, L., P. E. SMOUSE and J. M. QUATTRO, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial-DNA restriction data. Genetics **131:** 479–491.

GOLDING, G. B., 1987 The detection of deleterious selection using ancestors inferred from a phylogenetic history. Genet. Res. **49:** 71–82.

GOLDING, B., and J. FELSENSTEIN, 1990 A maximum likelihood approach to the detection of selection from a phylogeny. J. Mol. Evol. **31:** 511–523.

HARTL, D. L., and S. A. SAWYER, 1991 Inference of selection and recombination from nucleotide sequence data. J. Evol. Biol. **4:** 519–532.

HEY, J., 1991 The structure of genealogies and the distribution of fixed differences between DNA-sequence samples from natural populations. Genetics **128:** 831–840.

HUDSON, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.

HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992a A statistical test for detecting geographical subdivision. Mol. Biol. Evol. **9:** 138–151.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992b Estimation of levels of gene flow from DNA-sequence data. Genetics **132:** 583–589.

KIMURA, M., and G. H. WEISS, 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49:** 561–576.

KINGMAN, J. F. C., 1982a The coalescent. Stochast. Proc. Appl. **13:** 235–248.

KINGMAN, J. F. C., 1982b On the genealogy of large populations. J. Appl. Probab. **19A:** 27–43.

LARSON, A., 1984 Neontological inferences of evolutionary pattern and process in the salamander family Plethodontidae. Evol. Biol. **17:** 119–217.

LESSA, E. P., 1990 Multidimensional analysis of geographic genetic structure. Syst. Zool. **39:** 242–252.

LYNCH, M., and T. J. CREASE, 1990 The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. **7:** 377–394.

MADDISON, D. R., M. RUVOLO and D. L. SWOFFORD, 1992 Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. Syst. Biol. **41:** 111–124.

NATH, H. B., and R. C. GRIFFITHS, 1993 The coalescent in two colonies with symmetrical migration. J. Math. Biol. **31:** 841–851.

NEIGEL, J. E., and J. C. AVISE, 1993 Application of a random-walk model to geographic distributions of animal mitochondrial DNA variation. Genetics **135:** 1209–1220.

NEIGEL, J. E., R. M. BALL and J. C. AVISE, 1991 Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. Evolution **45:** 423–432.

NIGH, T. A., W. L. PFLIEGER, J. P. L. REDFEARN, W. A. SCHROEDER, A. R. TEMPLETON *et al.*, (order alphabetical), 1992 *The Biodiversity of Missouri.* Missouri Department of Conservation, Jefferson City, MO.

O'BRIEN, S. J., 1991 Ghetto legacy. Curr. Biol. **1:** 209–211.

PHILLIPS, C. A., and O. J. SEXTON, 1989 Orientation and sexual differences during the breeding migration of the spotted salamander, *Ambystoma maculatum.* Copeia **1989:** 17–22.

PRUM, B., M. GUILLOUD-BATAILLE and F. CLERGET-DARPOUX, 1990 On the use of c^2 tests for nested categoried data. Ann. Hum. Genet. **54:** 315–320.

ROFF, D. A., and P. BENTZEN, 1989 The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. **6:** 539–545.

ROUTMAN, E., 1993 Population structure and genetic diversity of metamorphic and paedomorphic populations of the tiger salamander, *Ambystoma tigrinum.* J. Evol. Biol. **6:** 329–357.

SLATKIN, M., 1989 Detecting small amounts of gene flow from phylogenies of alleles. **121:** 609–612.

SLATKIN, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. **58:** 167–175.

SLATKIN, M., 1993 Isolation by distance in equilibrium and nonequilibrium populations. Evolution **47:** 264–279.
SLATKIN, M., and H. E. ARTER, 1991 Spatial autocorrelation methods in population genetics. Am. Nat. **138:** 499–517.
SLATKIN, M., and N. H. BARTON, 1989 A comparison of 3 indirect methods for estimating average levels of gene flow. Evolution **43:** 1349–1368.
SLATKIN, M., and W. P. MADDISON, 1989 A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123:** 603–613.
SLATKIN, M., and W. P. MADDISON, 1990 Detecting isolation by distance using phylogenies of genes. Genetics **126:** 249–260.
SOKAL, R. R., R. M. HARDING and N. L. ODEN, 1989a Spatial patterns of human gene frequencies in Europe. Am. J. Phys. Anthropol. **80:** 267–294.
SOKAL, R. R., G. M. JACQUEZ and M. C. WOOTEN, 1989b Spatial autocorrelation analysis of migration and selection. Genetics **121:** 845–855.
TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.
TAKAHATA, N., and M. SLATKIN, 1990 Genealogy of neutral genes in two partially isolated populations. Theor. Popul. Biol. **38:** 331–350.
TEMPLETON, A. R., 1992 Human origins and analysis of mitochondrial DNA sequences. Science **255:** 737.
TEMPLETON, A. R., 1993 The "Eve" hypothesis: a genetic critique and reanalysis. Amer. Anthropol. **95:** 51–72.
TEMPLETON, A. R., 1994 "Eve": hypothesis compatibility versus hypothesis testing. Amer. Anthropol. **96:** 141–147.
TEMPLETON, A. R., and N. J. GEORGIADIS, 1995 A landscape approach to conservation genetics: conserving evolutionary processes in African bovids, in *Conservation Genetics: Case Histories from Nature*, edited by J. AVISE and J. HAMRICK. Chapman and Hall, New York (in press).
TEMPLETON, A. R., and C. F. SING, 1993 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics **134:** 659–669.
TEMPLETON, A. R., E. BOERWINKLE and C. F. SING, 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, A. R., C. F. SING, A. KESSLING and S. HUMPHRIES, 1988 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. Genetics **120:** 1145–1154.
TEMPLETON, A. R., K. SHAW, E. ROUTMAN and S. K. DAVIS, 1990 The genetic consequences of habitat fragmentation. Ann. MO Bot. Gard. **77:** 13–27.
TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 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.
WRIGHT, S., 1969 *Evolution and the Genetics of Populations, Vol. 2. The Theory of Gene Frequencies.* University of Chicago Press, Chicago.

Communicating editor: M. SLATKIN

APPENDIX

Start with haplotypes nested within a 1-step clade:

1. Are there any significant values for D_c, D_n, or *I-T* within the clade?
   - NO: 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). Move on to another clade at the same or higher level.
   - YES: go to step 2.
2. Are the D_c values for tip or some (but not all) interior clades significantly small or is the *I-T* D_c distances significantly large?
   - NO: go to step 11.
   - YES: go to step 3.
   - Tip/interior status cannot be determined—Inconclusive Outcome.
3. Are any D_n and/or *I-T* D_n values significantly reversed from the D_c values, and/or do one or more tip clades show significantly large D_ns or interior clades significantly small D_ns or *I-T* significantly small D_n with the corresponding D_c values being nonsignificant?
   - NO: go to step 4.
   - YES: go to step 5.
4. Do the clades (or 2 or more subsets of them) with restricted geographical distributions have ranges that are completely or mostly nonoverlapping with the other clades in the nested group (particularly interiors), and does the pattern of restricted ranges represent 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 nonsexual 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. Do the clades (or 2 or more subsets of them) with restricted geographical distributions have ranges that are completely or mostly nonoverlapping with the other clades in the nested group (particularly interiors), and does the pattern of restricted ranges represent 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. Do clades (or haplotypes within them) with significant reversals or significant D_n values without significant D_c values define geographically concordant subsets, or are they geographically concordant with other haplotypes/clades showing similar distance patterns?

- NO: go to step 7.
- YES: go to step 13.
- Too few clades (<2) to determine concordance—Insufficient genetic resolution to discriminate between range expansion/colinization 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_ns (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.
9. Are the different geographically concordant clade ranges separated by areas that have not been sampled?
   - NO: past fragmentation. (If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted but at least partially nonoverlapping geographical 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 scheme 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 but at least partially nonoverlapping geographical distributions are mutationally connected to one another by a larger than average number of steps.)
11. Are the D_c values for some tip clades significantly large, and/or the D_cs for all interiors significantly small, and/or the *I-T* D_c 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_ns (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 14.
    - YES: long distance colonization.
14. Is the species absent in the non-sampled areas?
    - NO: sampling design inadequate to discriminate between contiguous range expansion and long-distance colonization.
    - YES: long-distance colonization.
15. Are the different geographically concordant areas separated by areas that have not been sampled?
    - NO: past fragmentation. (If inferred at a high-clade level, additional confirmation occurs if the clades displaying restricted but at least partially nonoverlapping geographical distributions are mutationally connected to one another by a larger than average number of steps.)
    - YES: go to step 16.
16. Is the species absent in the nonsampled areas?
    - NO: go to step 18.
    - YES: allopatric fragmentation. (If inferred at a high-clade level, additional confirmation occurs if the clades displaying restricted but at least partially nonoverlapping geographical distributions are mutationally connected to one another by a larger than average number of steps.)
17. Are the D_n values for tip or some (but not all) interior clades significantly small, or the D_n for one or more interior clades significantly large, or is the *I-T* D_n value 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 scheme inadequate to discriminate between fragmentation, range expansion, and isolation by distance.
    - YES: geographical sampling scheme inadequate to discriminate between fragmentation and isolation by distance.