

Inference Key for the Nested Haplotype Tree Analysis of Geographical Distances

Start with haplotypes nested within a 1-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 non-sexual populations, small sample size, or inadequate geographical sampling). In that case, move on to another clade at the same or higher level.

When performing a single-locus nested clade analysis, the significance of the test results should be corrected for multiple testing with the Dunn-Sidak correction, which is now incorporated into the program. When performing multi-locus nested clade analysis, the Dunn-Sidak correction is not needed, and the original probability values should be used to determine significance. In the case of multi-locus nested clade analysis, the false positive rate is corrected by cross-validation across loci.

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 non-significant.
 - b. The D_c 's for one or more tips are significantly small or non-significant 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 non-significant
 - 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 non-significant.
 - NO - Go to step 4.
 - YES - Go to step 5.

4. Are both of the following conditions satisfied?
- The clades (or 2 or more subsets of them) with significantly small D_c or D_n values have ranges that are completely or mostly non-overlapping with the other clades in the nested group (particularly interiors).
 - The pattern of completely or mostly non-overlapping ranges in the above condition represents a break or reversal from lower level trends within the nested clade 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?
- The clades (or 2 or more subsets of them) with significantly small D_c values have ranges that are completely or mostly non-overlapping with the other clades in the nested group (particularly interiors).
 - The pattern of completely or mostly non-overlapping ranges in the above condition represents a break or reversal from lower level trends within the nested clade 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 two or more geographically concordant subsets.
- 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/Colonization and Restricted Dispersal/Gene Flow** - Proceed to step 7 to determine if the geographical sampling is sufficient to discriminate between short versus 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 non-sampled areas?
- NO - **Sampling Design Inadequate to Discriminate between Isolation by Distance (Short Distance Movements) versus 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 non-sampled 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 by at least partially non-overlapping 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 center of the other clades by intermediate geographical areas that were sampled?
- NO - Go to step 14.
 - YES – 1) **Long Distance Colonization, Past Larger Range Coupled with Subsequent Extinction in Some Intermediate Geographical Areas, or Past Range Expansion, All of Which Can Possibly Be Coupled with Subsequent Fragmentation** (subsequent fragmentation is indicated if the clades displaying restricted but at least partially non-overlapping distributions are mutationally

connected to one another by a larger than average number of steps) or 2) **Past Fragmentation Followed by Range Expansion**. To see if secondary contact is involved in scenario 2), perform the supplementary tests given in Templeton, *Molecular Ecology* **10**: 779-791, 2001. To discriminate the type of movement leading to this pattern in scenario 1), 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 non-overlapping 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 non-overlapping 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 non-overlapping 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 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.**
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 non-overlapping 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. If the case against long-distance movement is compelling, then the inference is **past gradual range expansion followed by fragmentation or a past larger range followed by extinction in intermediate areas.****