Inference Key for the Nested Haplotype Tree Analysis of Geographical Distances

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 non-sexual 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. 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 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. Are any $D_n$ and/or I-T $D_n$ values significantly reversed from the $D_c$ values?
   b. Do one or more tip clades show significantly large $D_n$’s with the corresponding $D_c$ values being non-significant?
   c. Do one or more interior clades show significantly small $D_n$’s with the corresponding $D_c$ values being non-significant?
   d. Does I-T have a significantly small $D_n$ with the corresponding $D_c$ values being non-significant?
   - NO - Go to step 4.
   - YES - Go to step 5.

4. Are both of the following conditions satisfied (when no lower level nested clades have significant effects, answer only part a.)?
   a. Do the clades (or 2 or more subsets of them) with restricted geographical distributions have ranges that are completely or mostly non-overlapping with the other clades in the nested group (particularly interiors)?
   b. 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 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. Do the clades (or 2 or more subsets of them) with restricted geographical distributions have ranges that are completely or mostly non-overlapping 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.
6. Do clades (or haplotypes within them) with significant reversals or significant \( D_n \) values without significant \( D_c \) values (identified in step 3) satisfy one or more of the following conditions:
   a. They define two or more geographically concordant subsets (that is, they have similar geographic distributions).
   b. They are geographically concordant with other haplotypes/clades showing similar distance patterns.

   • No - Go to step 7.
   • YES - Go to step 13.
   • TOO FEW CLADES (\( \leq 2 \)) TO DETERMINE CONCORDANCE - Insufficient Genetic Resolution to Discriminate between Range Expansion/Colonization and Restricted Dispersal/Gene Flow - Proceed to step 7 to determine if the geographical sampling is sufficient to discriminate between short 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.

9. Are the different geographical clade ranges identified in step 4 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 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 any of 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 - 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 geographical clade ranges identified in step 5 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 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 16.

16. Is the species absent in the non-sampled areas?
• NO - Go to step 18.
• 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.)

17. Is at least one of the following conditions satisfied?
   a. Are the $D_n$ values for tip or some (but not all) interior clades significantly small?
   b. Are the $D_n$ for one or more interior clades significantly large?
   c. 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.