MIPoD
Microevolutionary Inference from Patterns of Divergence
USER MANUAL

MIPoD 1.0 – Neutral Module
October 8, 2007
Updated August 11, 2008

About the Program

      MIPoD makes inferences about microevolutionary process from the pattern of divergence in multiple traits among a set of related taxa. It takes in data including a list of taxa, each with estimated means for a common set of traits, and a phylogeny with branch lengths calibrated to generation time. It also incorporates independent, direct estimates of the **G**-matrix for the traits and effective population size, if such estimates are available. The theoretical background for MIPoD 1.0 in the two-trait case is provided in Hohenlohe and Arnold (2008). Complications for more than two traits will be dealt with in a forthcoming paper, although MIPoD 1.0 conducts all the analyses for more than two traits.

      MIPoD is written in C using the MAC OSX programming tools. The user inteface is command-line-based; stay tuned for more user-friendly versions. The compiled version was compiled with the GNU C Compiler, and the compiled version should work on any Mac OSX computer. The source code is also available for use and modification, with the condition that any publication resulting from use of this program or a modified version thereof contains the following citation:

Hohenlohe, P.A. and S.J. Arnold 2008. MIPoD: A hypothesis-testing framework for microevolutionary inference from patterns of divergence. American Naturalist 171: 366-385.

To use MIPoD 1.0

      On a Mac OSX computer, place the compiled version and the input data file in your home directory. Double-click the application icon. This should open a terminal window.

      If the program prompts you for the name of the input file at this point, skip to the next paragraph below. If instead you get only a standard UNIX prompt, you most likely need to change permissions on the executable file. Do this by typing "chmod +x mipod" (without the quotation marks) and hit return. Then type "./mipod" (again without the quotation marks) and the program should begin.

      If the input file is not in your home directory, you will need to supply the path to the file; for instance, phylogeny/datafiles/inputfile.txt. Otherwise, just type the name of the text file and hit return. The analysis may take several minutes, depending on the number of taxa and number of traits. The output will appear in your home directory in a text file called Outfile.txt.

      If the permission fix above does not work, or if you are operating on Windows or another operating system, you nay need to modify and re-compile the C source code.

Good luck, and feedback would be greatly appreciated at
hohenlop@science.oregonstate.edu.

Input data file
        All the data to be input to MIPoD should be combined in a single text file. As a guide to the standard format, a sample dataset text file for two traits in garter snakes (Thamnophis_data.txt) is included with the program files. The data to be analyzed in MIPoD include several components, ordered as below. No extraneous text or formatting should be included in the input file – it will likely cause the file to be misread. The items below may be separated by tabs or carriage returns. The data file should contain:

- number of taxa (integer).
- number of traits (integer).
- independent estimate of effective population size. If none is available, use 0.
- number of generations per unit branch length on the phylogeny.
- test vector for comparison (e.g., major axis of an empirically estimated selection surface). Elements of the vector should be separated by tabs, and there must be as many elements as traits. The vector does not need to be normalized to unit length. MIPoD expects a vector here; if you have none to compare, use "1 0 0 ..." and the result will be a test of covariance among the traits.
- **G**-matrix. **G** should be a square, symmetric variance/covariance matrix (i.e. positive definite), matching the number of traits indicated above. If no independent **G**-matrix is available, put 0 on this line.
- the phylogeny. The phylogeny should be given in standard Newick format (see http://evolution.genetics.washington.edu/phylip/newicktree.html for more information), *with the exception* that interior nodes cannot be named. Each taxon name should be no more than 10 characters and contain no spaces.
- taxon names and trait measurements. Each taxon name should exactly match the name used in the phylogeny. The trait measurements for each taxon should follow the taxon name, separated by tabs, in the same order of traits as used in the **G**-matrix. Missing data for any trait in any taxon are not allowed.

Output file
        Results of the MIPoD analysis are output as a text file in a standard format as Outfile.txt. Note that subsequent runs of MIPoD will overwrite this file – if you want to keep the results of multiple runs, be sure to re-save each one with a different name before proceeding!
        A note on ordering of eigenvectors: Traditionally, eigenvectors are ordered by the absolute value of the associated eigenvalue. When MIPoD inputs the user-supplied **G** matrix, it orders the eigenvectors of **G** in this way. During subsequent estimation of parameters, eigenvectors are not re-ordered until the final step, when MIPoD estimates the best-fit **G** matrix. For instance, at the shape estimation step, the relative sizes of the eigenvalues may change but MIPoD will not re-order the eigenvectors. At the final step of the orientation tests, eigenvectors are again re-ordered by eigenvalue. The complete set of numbers below the summary table should make this clear.
        Below is a sample output file for *Thamnophis* for two traits, with annotation shown in red. Output differs slightly when estimates of N_e or **G** are not provided. For

more than two traits, parameter estimates are not shown in the summary statistics table (the number of parameters grows quickly with many traits). See Hohenlohe and Arnold (2008) for further details on how to interpret the results.

39 taxa
2 traits

Ne = 500.00

Direct estimate of G:
7.53   1.98
1.98   7.98
 [These are the user-supplied estimates]

Step	sig	eps	phi	lnL	LR	df	p
Direct	15.5050	0.6288	0.8425	-408.37	-	-	-

[Parameters and likelihood for the user-supplied direct estimates of **G** and Ne.]

| 1.Size | 0.3075 | 0.6288 | 0.8425 | -293.71 | 229.33 | 1 | 0.0000 |

[ML estimate for sigma, and results of likelihood ratio test comparison with previous step.]

| 2.Shape | 0.3381 | 0.7925 | 0.8425 | -290.57 | 6.28 | 1 | 0.0122 |

 [ML estimate for epsilon and sigma, with LRT comparison to previous step.]

| 3.Orientation | 0.3380 | 0.8705 | 0.5130 | -283.20 | 14.74 | 1 | 0.0001 |

 [ML estimate for all **G**-matrix parameters, with LRT comparison to previous step.]

| Test vector | 0.3380 | 0.8688 | 0.4636 | -283.42 | 0.44 | 1 | 0.5072 |

 [ML estimate of **G**, constrained by user-supplied test vector. Likelihood ratio test compares to previous step.]

Eigenvalues (eigenvectors) for G-matrix estimates:

[Eigenvalues and corresponding eigenvectors for each step in the table above. Eigenvalues are normalized to sum to 1, with sigma giving the total size of **G** (i.e., sum of eigenvalues). 95 percent confidence limits are given for each estimate of sigma.]
Direct:
Sigma = 15.5050
0.6288 (0.6656 , 0.7463)
0.3712 (0.7463 , -0.6656)

1.Size (with Ne = 500.00):
Sigma = 0.3075 [0.2282 to 0.4283]
0.6288 (0.6656 , 0.7463)
0.3712 (0.7463 , -0.6656)

2.Shape(1):
Sigma = 0.3381 [0.2509 to 0.4708]
0.7925 (0.6656 , 0.7463)
[0.6767 to 0.8618]   [Marginal confidence limits for epsilon; calculated only for 2-trait case]
0.2075 (0.7463 , -0.6656)

3.Orientation(1):
Sigma = 0.3380 [0.2509 to 0.4707]

0.8705 (0.8713 , 0.4908)
[0.7934 to 0.9142]
0.1295 (0.4908 , -0.8713)

Test vector:
Sigma = 0.3380 [0.2509 to 0.4707]
0.8688 (0.8944 , 0.4472)
[0.7908 to 0.9130]
0.1312 (-0.4472 , 0.8944)