DimSI
Dimensionality of Sexual Isolation
USER MANUAL

DimSI 1.0 – Binomial and Poisson models
[latest update 8/16/2012]

About the Program

DimSI uses data from mating trials among multiple populations to estimate the dimensionality of sexual isolation. For more details on the background and analysis, please consult Hohenlohe & Arnold (2010) and Nosil & Hohenlohe (2012).

DimSI is written in C and run from the command line in a unix terminal window. 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 2010. The dimensionality of mate choice, sexual isolation, and speciation. Proc Nat Acad Sci USA 107: 16583-16588.

Questions and feedback are welcome at hohenlohe@uidaho.edu.

To use DimSI 1.0

DimSI is split into two executable versions for the binomial and Poisson models (see Hohenlohe & Arnold 2010). Which one is appropriate depends on the data being analyzed. They can be run from a MacOSX terminal window with the following commands [with options]:

./binomial datafile.txt [-o outfile.txt] [-v] [-c 20] [-summary] [-input 1] [-max spp]

./Poisson datafile.txt [-o outfile_name] [-v] [-c 20] [-summary] [-max spp]

Options
- [datafile.txt] Specify the name of the data file. This file MUST be in the format below.
- [-o outfile.txt] If no outfile name is supplied, the program will write the output to a file called outfile.txt. Note that if a file by this name already exists in the directory, it will be overwritten without warning – specifying a new outfile name is highly recommended.
- [-v] The option -v provides verbose output to the terminal window – this allows you to keep track of the progress. On each line appear the number of dimensions d, the maximum lnL reached so far, and the lnL of the most recent iteration. At

the end appear the number of taxa and the matrix of mating trial results (this is useful to check that the program is correctly reading in the datafile).

- [-c 20] The -c option allows you to specify a number of cycles used at various steps in estimating the best fit of points in d dimensions. The number following "-c" must be a positive integer. The default value is 20; lower numbers will run faster, and higher numbers may improve precision of the estimates.
- [-summary] If this option is specified, the output file will only contain the summary statistics for each value of d and not the coordinates in latent trait space (see below).
- [-input 1] Specify the input format below for the binomial model. Default is format 1, and format 2 is the alternative.
- [-max spp] Specify the maximum value of d to be tested; the program will test up to this number minus 1. Default is the number of species, and cannot exceed twice the number of species.

Input data file

Data input files must be text files, conforming to the following format. Comments or extraneous text are allowed only at the end of the file, after all of the information described below. Example datafiles (those analyzed for Hohenlohe & Arnold 2010 and Nosil & Hohenlohe 2012) are available on the website. The format differs slightly for the binomial vs. Poisson models.

For the binomial model, the data file should contain:
- Number of populations/taxa (integer).
- Number of sets of pairwise mating trials (integer). Multiple sets of trials are permitted for any given pair of populations; the numbers will simply be added together for each pair of populations as the data are read.
- Names of populations, exactly as they appear below (case-sensitive).
- Mating trial data as follows. The program will read as many lines as the number of sets of mating trials given above, and text after this point will be ignored. For **input type 1** (default), each line should contain the following, separated by tabs:
  - Name of population A
  - Name of population B
  - k_{AA} (number of successful matings for A females and A males)
  - n_{AA} (total number of trials for A females and A males; must be greater than or equal to k_{AA})
  - k_{AB} (number of successful matings for A females and B males)
  - n_{AB} (total number of trials for A females and B males)
  - k_{BA} (number of successful matings for B females and A males)
  - n_{BA} (total number of trials for B females and A males)
  - k_{BB} (number of successful matings for B females and B males)
  - n_{BB} (total number of trials for B females and B males)

  For **input type 2**, each line should contain the following, separated by tabs:
  - Name of population A
  - Name of population B

- k_{AB} (number of successful matings for A females and B males)
- n_{AB} (total number of trials for A females and B males)

For the Poisson model, the data file should contain:

➢ Number of populations/taxa (integer).
➢ Number of pairwise mating counts (integer). Multiple counts are permitted for any given pair of populations; the numbers will simply be added together for each pair of populations as the data are read.
➢ Names of populations, exactly as they appear below (case-sensitive).
➢ Mating trial data as follows. The number of lines should equal the number of mating counts given above; each line should contain the following, separated by tabs:
  - Population of female
  - Population of male
  - Number of successful matings between population A and population B

Output file

Results of the DimSI analysis are output as a text file in a standard tab-separated format. Below is a sample output file for the Kurokawa et al (1982) dataset for the *Drosophila auraria* complex. See Hohenlohe and Arnold (2010) for further details on how to interpret the results. The first line gives column headings for the summary data at each dimensionality. At each value of d, the subsequent lines give coordinates for the female (F) and male (M) means of each population. These coordinates incorporate rotation analogous to principal components analysis, so that the first dimension (column) includes the maximum variance among points and so forth.

Dim	lnL	AIC	params	c	n_D
1	-279.09504183	574.19008367	8	0.8948	1.0000
A F	1.4360				
A M	0.9412				
B F	-1.0135				
B M	-1.5413				
T F	-0.4311				
T M	0.1447				
Q F	0.2571				
Q M	0.2068				
2	-53.68705094	135.37410189	14	1.0000	1.7505
A F	-0.3597	1.1382			
A M	-0.2851	0.6329			
B F	1.0532	0.2240			
B M	1.1186	-0.3850			
T F	-0.9518	-0.6104			
T M	-0.2879	-0.5225			
Q F	-0.1912	-0.0771			
Q M	-0.0961	-0.4001			
3	-53.60398992	145.20797984	19	1.0000	1.7713
A F	-0.3266	1.1425	0.0227		
A M	-0.2664	0.6348	0.0348		

```
B F      1.0359  0.1675  -0.1937
B M      1.1222  -0.3629 0.1019
T F      -0.9384 -0.5870 -0.1602
T M      -0.3250 -0.5124 0.1014
Q F      -0.1945 -0.0783 0.0518
Q M      -0.1071 -0.4042 0.0413
```