

Test of Host-Parasite Coevolution: Program PARAFIT User's Guide

Pierre Legendre
Département de sciences biologiques
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal, Québec H3C 3J7, Canada

March 2002

E-mail: Pierre.Legendre@umontreal.ca

What does program PARAFIT do?

This program tests the hypothesis of coevolution between a clade of hosts and a clade of parasites. The null hypothesis (H_0) of the global test is that the evolution of the two groups, as revealed by the two phylogenetic trees and the set of host-parasite association links, has been independent. The method, which is described in detail in Legendre et al. (2002), requires some estimates of the phylogenetic trees or phylogenetic distances, and also a description of the host-parasite associations (H-P links) observed in nature.

Two types of test are produced by the program: a global test of coevolution and a test on each H-P link.

The principle of the global test is the following (H_0 : independent evolution of the hosts and parasites):

1. Read in a matrix of links between hosts and parasites (matrix **A** containing 0 and 1 data), as well a coordinate representation of the phylogenetic trees or phylogenetic distances among the parasites (matrix **B**) and among the hosts (matrix **C**). The input data matrices are described in a separate section below.
2. Compute matrix $\mathbf{D} = \mathbf{C} \mathbf{A}' \mathbf{B}$. **D** is a “fourth-corner matrix” (Legendre et al. 1997).
3. Compute the statistic $ParaFitGlobal = \sum (d_{ij}^2)$ which is the sum of squares of all values d_{ij} in matrix **D**.
4. Permute at random, separately, each row of matrix **A**, obtaining matrix \mathbf{A}^* . Compute $\mathbf{D}^* = \mathbf{C} \mathbf{A}^{*'} \mathbf{B}$, and from it, compute a permuted value $ParaFitGlobal^*$ for the statistic. Save this value in a vector $trace^*$ for the tests of individual links (below).
5. Repeat step 4 a large number of times.
6. Add the reference value of $ParaFitGlobal$ to the distribution of $ParaFitGlobal^*$ values. Calculate the permutational probability associated to $ParaFitGlobal$.

The test of each individual H-P link is carried out as follows (H_0 : this particular link is random). The data have been read into the program at the beginning of the global test (above).

1. Remove one link (k) from matrix **A**.
2. Compute matrix $\mathbf{D} = \mathbf{C} \mathbf{A}' \mathbf{B}$.
- 3a. Compute $trace(k) = \sum (d_{ij}^2)$ which is the sum of squares of all values d_{ij} in matrix **D**.
- 3b. Compute the statistic $ParaFitLink1 = (trace - trace(k))$ where $trace$ is the $ParaFitGlobal$ statistic.
- 3c. Compute the statistic $ParaFitLink2 = (trace - trace(k)) / (tracemax - trace)$ where $tracemax$ is the maximum value that can be taken by $trace$.

4. Permute at random, separately, each row of matrix **A**, obtaining **A***; use the same sequences of permutations as were used in the test of *ParaFitGlobal*. Using the values of trace and trace* saved during the global test, compute the permuted values of the two statistics, *ParaFit1** and *ParaFit2**.
5. Repeat step 4 a large number of times.
6. Add the reference value of *ParaFit1* to the distribution of *ParaFit1** values; add the reference value of *ParaFit2* to the distribution of *ParaFit2** values. Calculate the permutational probabilities associated to *ParaFit1* and *ParaFit2*.

Input files

There are three input files called A, B and C.

File A contains matrix **A**($n1 \times n2$) whose dimensions are $n1$ = number of parasites (rows of the matrix) and $n2$ = number of hosts (columns of the matrix). The file contains 1's when a host-parasite link has been observed in nature between the host in the column and the parasite in the row, and 0's otherwise. Note the dimensions of this matrix, $n1$ and $n2$: these numbers will be requested during the program run.

File B contains a matrix of principal coordinates ($n1 \times n4$) computed from either a matrix of phylogenetic distances or a matrix of patristic distances among the parasites. A matrix of patristic distances represents exactly the information in a phylogenetic tree. Note the dimension $n4$ of this file. Since there are at most $(n - 1)$ principal coordinates for n objects, $n4$ is at most equal to $(n1 - 1)$.

File C contains the transpose ($n3 \times n2$) of a matrix of principal coordinates of size ($n2 \times n3$) computed from either a matrix of phylogenetic distances or a matrix of patristic distances among the hosts. Note the dimension $n3$ of this file. Since there are at most $(n - 1)$ principal coordinates for n objects, $n3$ is at most equal to $(n2 - 1)$.

Principal coordinates can be computed using several programs. Two programs are available from our WWW site:

- The Principal Coordinate module of THE R PACKAGE (for Macintosh only, at the moment).
- The program DISTPCOA, available for either Macintosh or 32-bit DOS.

To obtain either one of these programs, go to the lab's WWW site <<http://www.fas.umontreal.ca/biol/legendre/indexEnglish.html>>, click on *Computer Programs and Datasets*, choose a program, and click on the blue hyperlink to download.

Output file

See example below.

Note: a value of -99.00000 is printed when the statistic ParaFitLink2 cannot be calculated because its denominator is zero.

Disclaimer

This program is provided without any explicit or implicit warranty of correct functioning. It has been developed as part of a university-based research program. If, however, you should encounter problems with this program, the author will be happy to help solve them. Researchers may use this program for scientific purposes, but the source code remains the property of Pierre Legendre. Users of the program may refer to the present user's manual as follows:

Legendre, P. 2001. Test of host-parasite coevolution: program PARAFIT user's guide. Département de sciences biologiques, Université de Montréal. 10 pages.

Program distribution

The program is written in FORTRAN77 in order to facilitate diffusion. Indeed, there is a compiler, GNU FORTRAN (or g77; see <http://www.gnu.org/software/fortran/>), that is freely available for this level of FORTRAN, for the DOS/Windows, MacOS X, Unix, and Linux families of operating systems. The PARAFIT program in a variety of forms:

- FORTRAN source code for the Macintosh Language Systems Fortran compiler (file PARAFIT.f). Users may modify the Parameter statement at the beginning of the program, which fixes the size of the largest matrices that can be analysed (n1max = maximum number of parasites, n2max = maximum number of hosts). Recommendations for translation to other Fortran compilers are available on the WWW page of Philippe Casgrain <<http://www.fas.umontreal.ca/biol/casgrain/en/fortran.html>>.
- GNU FORTRAN source code for Mac OS X, Unix/Linux, or Windows (file PARAFIT.FOR), which can be compiled using a GNU FORTRAN77 compiler. Users may modify the Parameter statement at the beginning of the program, which fixes the size of the largest matrix that can be analysed (n1max = maximum number of parasites, n2max = maximum number of hosts).
- Compiled version of the program for any Macintosh computer (fat binary for 68xxx and PPC: file PARAFIT/FAT). The maximum size of the data matrix is 100 parasites and 100 hosts. The program requires 960 Kb RAM for running.
- Compiled version for Windows-compatible PC (file PARAFIT.EXE). The maximum size of the data matrix is 100 parasites and 100 hosts. The program has been compiled for 32-bit operating systems (i.e. Windows95 or WindowsNT) and requires about 1 Mb RAM for running.

References

Legendre, P., Y. Desdevises and E. Bazin. 2002. A statistical test for host-parasite coevolution. *Systematic Biology* 51(2): (in press).

A preprint is available on the WWW site <<http://www.fas.umontreal.ca/biol/legendre/>>.

Legendre, P., R. Galzin & M. L. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* 78: 547-562.

DOS user's notes prepared by Philippe Casgrain

How to start a DOS program

1. To open a DOS window: from the Start menu Programs Accessories Console Application (or MS-DOS).
2. At the DOS prompt (e.g., C:\WINDOWS\>), type "cd c:\path\to\the\program" where "\path\to\the\program" represents the directory where the program is found. Examples: c:\tmp, or c:\windows\desktop, etc. Press the *Return* key.
3. Type the name of the program to start it. Example: parafit.exe. Press the *Return* key.
4. Follow the on-screen instructions.

Appendix: Test run

Three matrices are used to compute ParaFit for the Gopher and Lice example from Hafner et al. (1994, Science 265:1087-1090), presented in the Legendre et al. (2002) paper. Files containing these matrices are available to carry out test runs of the program.

Matrix A: File A(17x15).txt contains the host-parasite association links.

```

0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0

```

Here is another presentation of the same matrix, showing the names of the parasites (at the beginning of the rows) and of the hosts (under the matrix). This format is not compatible with the ParaFit program.

T.minor	0	1	0	0	0	0	0	0	0	0	0	0	0	0
T.barbarae	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G.setzeri	0	0	1	0	0	0	0	0	0	0	0	0	0	0
G.cherriei	0	0	0	0	0	1	0	0	0	0	0	0	0	0
G.panamensis	0	0	0	0	1	0	0	0	0	0	0	0	0	0
G.costaricensis	0	0	0	0	0	0	1	0	0	0	0	0	0	0
G.chapini	0	0	0	1	0	0	0	0	0	0	0	0	0	0
G.thomomys	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G.actuosi	0	1	0	0	0	0	0	0	0	0	0	0	0	0
G.expansus	0	0	0	0	0	0	0	0	0	0	0	1	0	0
G.perotensis	0	0	0	0	0	0	0	0	0	0	0	0	1	0
G.trichopi	0	0	0	0	0	0	0	0	0	0	0	0	0	1
G.nadleri	0	0	0	0	0	0	0	0	0	0	0	0	0	1
G.texanus	0	0	0	0	0	0	0	0	1	0	0	0	0	0
G.oklahomensis	0	0	0	0	0	0	0	0	0	0	1	0	0	0
G.ewingi	0	0	0	0	0	0	0	1	0	0	0	0	0	0
G.geomydis	0	0	0	0	0	0	0	0	0	1	0	0	0	0

The names of the hosts, corresponding to the columns of the matrix, are:

T.talpoides	T.bottae	O.underwoodi	O.hispidus	O.cavator
O.cherriei	O.heterodus	G.breviceps	G.personatus	G.bursarius_majus
G.bursarius_halli	C.castanops	C.merriami	P.bulleri	Z.trichopus

Matrix B: File B(17x16).txt contains the 16 principal coordinates (as columns) derived from the patristic distance matrix shown below. The data lines, each one containing 16 values, are folded here because they are too long.

```

0.01583 0.09887 -0.06985 0.00276 0.11530 -0.01345 -0.00249 -0.01018 0.00000 0.00284
-0.00633 0.07009 0.00138 -0.00146 -0.00005 -0.00004
0.01349 0.07749 -0.04449 0.00274 0.06277 -0.00794 0.00201 -0.00595 -0.00117
-0.00401 -0.00258 -0.10679 -0.00103 0.00098 0.00002 0.00002
0.08906 -0.04633 0.00101 0.00090 -0.00093 0.03810 -0.02556 -0.02026 -0.00305
0.08802 0.00901 -0.00235 0.06882 -0.00402 -0.00002 -0.00002
0.12890 -0.08998 -0.00635 -0.00038 0.00387 -0.06812 0.03798 0.01243 0.00106
-0.03872 -0.00407 0.00050 0.00301 -0.05672 -0.00009 -0.00007
0.08406 -0.04227 0.00177 0.00111 -0.00092 0.03152 -0.02050 -0.01732 -0.00281
0.06806 0.00649 0.00106 -0.08227 -0.00325 0.00002 0.00002
0.11068 -0.07225 -0.00301 0.00035 0.00251 -0.04701 0.02679 0.00629 0.00005 -0.02551
-0.00333 0.00361 0.00270 0.07200 0.00005 0.00003
0.06435 -0.01834 0.00234 0.00036 -0.00336 0.10401 -0.06714 -0.01272 0.00373
-0.09222 -0.01121 0.00380 0.00301 -0.00234 -0.00003 -0.00002
0.01552 0.08897 0.17785 0.01641 0.00645 -0.03027 -0.00050 -0.05166 -0.00105
-0.00622 -0.00162 0.00422 0.00111 -0.00142 -0.00006 -0.00005
-0.06963 -0.01787 0.00077 -0.01701 -0.00116 0.06131 0.11096 -0.02304 0.08793
0.00003 -0.00174 0.00278 0.00055 -0.00065 -0.00009 -0.00005
-0.08998 -0.03363 -0.00335 0.02213 -0.00002 0.03390 0.06210 -0.01565 -0.12822
-0.00621 -0.00268 0.00293 0.00068 -0.00083 -0.00009 -0.00009
0.01129 0.05051 0.04980 0.00474 -0.00295 0.02710 0.01115 0.14497 -0.00241 0.01417
-0.00450 0.00380 0.00038 -0.00028 -0.00001 -0.00001
0.01417 0.08190 -0.04730 0.00189 -0.07963 -0.01371 0.00149 -0.00673 -0.00137
-0.01339 0.11675 0.00807 0.00047 -0.00052 -0.00003 -0.00002
0.01502 0.08987 -0.05715 0.00176 -0.10695 -0.02073 -0.00121 -0.01484 -0.00030
0.00950 -0.09522 0.00700 0.00078 -0.00091 -0.00004 -0.00003
-0.09137 -0.03285 0.00579 -0.11146 0.00174 -0.02291 -0.03156 0.00265 0.00017
0.00059 0.00012 0.00048 0.00015 -0.00026 0.03411 -0.00002
-0.11014 -0.05075 -0.00693 0.09434 0.00076 -0.02446 -0.03602 0.00470 0.02393
0.00123 0.00042 0.00017 0.00005 -0.00004 -0.00003 -0.02478
-0.09242 -0.03340 0.00574 -0.11342 0.00180 -0.02334 -0.03243 0.00284 0.00022
0.00078 0.00018 0.00027 0.00005 0.00001 -0.03360 -0.00004
-0.10885 -0.04993 -0.00665 0.09277 0.00072 -0.02400 -0.03507 0.00447 0.02328
0.00104 0.00033 0.00037 0.00018 -0.00030 -0.00005 0.02516

```

The (17 x 17) patristic distance matrix for the lice, from which these principal coordinates were computed, is the following. Again, data lines are folded because they are too long.

```

0.00000 0.18786 0.25607 0.28227 0.24860 0.26145 0.25854 0.28273 0.26606
0.27265 0.24718 0.24101 0.24827 0.26012 0.26704 0.26176 0.26530
0.18786 0.00000 0.23384 0.26004 0.22637 0.23922 0.23631 0.26050 0.24383
0.25042 0.22495 0.21878 0.22604 0.23789 0.24481 0.23953 0.24307
0.25607 0.23384 0.00000 0.20815 0.15285 0.18733 0.21177 0.27423 0.25755
0.26415 0.23868 0.24671 0.25398 0.25161 0.25854 0.25325 0.25679
0.28227 0.26004 0.20815 0.00000 0.20068 0.13426 0.23797 0.30043 0.28375
0.29035 0.26488 0.27292 0.28018 0.27781 0.28474 0.27946 0.28299
0.24860 0.22637 0.15285 0.20068 0.00000 0.17986 0.20431 0.26676 0.25009
0.25668 0.23121 0.23925 0.24651 0.24414 0.25107 0.24579 0.24933
0.26145 0.23922 0.18733 0.13426 0.17986 0.00000 0.21716 0.27961 0.26294
0.26953 0.24406 0.25210 0.25936 0.25699 0.26392 0.25864 0.26218

```

0.25854	0.23631	0.21177	0.23797	0.20431	0.21716	0.00000	0.27670	0.26002
0.26662	0.24115	0.24919	0.25645	0.25408	0.26101	0.25573	0.25926	
0.28273	0.26050	0.27423	0.30043	0.26676	0.27961	0.27670	0.00000	0.28422
0.29082	0.24624	0.27338	0.28065	0.27828	0.28521	0.27992	0.28346	
0.26606	0.24383	0.25755	0.28375	0.25009	0.26294	0.26002	0.28422	0.00000
0.22840	0.24867	0.25671	0.26397	0.21586	0.22279	0.21751	0.22105	
0.27265	0.25042	0.26415	0.29035	0.25668	0.26953	0.26662	0.29082	0.22840
0.00000	0.25527	0.26330	0.27057	0.21905	0.20786	0.22069	0.20611	
0.24718	0.22495	0.23868	0.26488	0.23121	0.24406	0.24115	0.24624	0.24867
0.25527	0.00000	0.23783	0.24510	0.24273	0.24966	0.24437	0.24791	
0.24101	0.21878	0.24671	0.27292	0.23925	0.25210	0.24919	0.27338	0.25671
0.26330	0.23783	0.00000	0.21561	0.25076	0.25769	0.25241	0.25595	
0.24827	0.22604	0.25398	0.28018	0.24651	0.25936	0.25645	0.28065	0.26397
0.27057	0.24510	0.21561	0.00000	0.25803	0.26496	0.25967	0.26321	
0.26012	0.23789	0.25161	0.27781	0.24414	0.25699	0.25408	0.27828	0.21586
0.21905	0.24273	0.25076	0.25803	0.00000	0.21344	0.06775	0.21169	
0.26704	0.24481	0.25854	0.28474	0.25107	0.26392	0.26101	0.28521	0.22279
0.20786	0.24966	0.25769	0.26496	0.21344	0.00000	0.21508	0.05001	
0.26176	0.23953	0.25325	0.27946	0.24579	0.25864	0.25573	0.27992	0.21751
0.22069	0.24437	0.25241	0.25967	0.06775	0.21508	0.00000	0.21334	
0.26530	0.24307	0.25679	0.28299	0.24933	0.26218	0.25926	0.28346	0.22105
0.20611	0.24791	0.25595	0.26321	0.21169	0.05001	0.21334	0.00000	

Matrix C: File C(14x15).txt contains the 14 principal coordinates (as rows) derived from the patristic distance matrix shown below. Data lines are folded here because they are too long.

```

0.06099 0.05965 -0.10935 -0.08248 -0.11000 -0.09119 -0.08470 0.07360 0.04480
0.06477 0.06997 0.03239 0.03015 0.01199 0.02941
0.09796 0.09423 -0.00836 -0.00376 -0.00850 -0.00439 -0.00307 -0.08384 -0.02342
-0.06603 -0.07596 0.02913 0.02682 0.01220 0.01698
-0.06765 -0.06433 -0.01433 -0.00716 -0.01444 -0.01092 -0.01004 -0.02664 -0.00745
-0.02238 -0.02521 0.10061 0.08580 0.02520 0.05894
-0.00709 -0.00612 -0.00504 -0.00075 -0.00515 -0.00170 -0.00081 -0.00999 0.00321
-0.00514 -0.00716 -0.05822 -0.04634 0.01831 0.13202
-0.00839 -0.00709 -0.00822 0.00156 -0.00831 -0.00512 -0.00475 -0.00508 0.00144
-0.00587 -0.00653 -0.01737 -0.01362 0.11879 -0.03142
0.07757 -0.08045 -0.00006 0.00003 -0.00007 0.00017 0.00020 -0.00023 0.00010 0.00045
0.00047 0.00017 0.00025 0.00091 0.00050
0.00131 0.00055 -0.00248 0.00258 -0.00262 0.00230 0.00269 0.07773 0.00718 -0.03724
-0.05279 -0.00092 0.00129 -0.00113 0.00156
0.00102 0.00098 0.03240 -0.08337 0.03319 0.00093 0.00101 0.00332 0.00123 -0.00067
-0.00166 0.00115 0.00135 0.00681 0.00232
-0.00476 -0.00476 0.00620 0.00456 0.00662 -0.01319 -0.01197 -0.01851 0.07665
-0.01034 -0.01144 -0.00129 -0.00550 -0.00614 -0.00613
-0.00214 -0.00189 -0.02581 -0.02013 -0.02677 0.04649 0.03288 -0.00597 0.01542
-0.00346 0.00038 0.00945 -0.01371 -0.00188 -0.00287
-0.00149 -0.00137 -0.00558 -0.00432 -0.00566 0.00961 0.00558 -0.00302 0.00526
-0.00227 0.00087 -0.04517 0.05173 -0.00183 -0.00235
-0.00002 -0.00002 0.04293 -0.00011 -0.04216 -0.00031 -0.00027 -0.00003 0.00006
-0.00006 0.00004 -0.00005 0.00004 -0.00002 -0.00002
0.00060 0.00055 0.00076 0.00051 0.00087 -0.00212 0.00061 0.00308 0.00001 -0.03693
0.03047 0.00096 -0.00049 0.00038 0.00075
0.00025 0.00023 0.00113 0.00087 0.00116 0.01662 -0.02099 0.00044 -0.00048 -0.00065
0.00063 0.00036 -0.00012 0.00021 0.00034

```

The (15 x 15) patristic distance matrix for the gophers, from which these principal coordinates were computed, is the following. Again, data lines are folded because they are too long.

```

0.00000 0.15811 0.22964 0.21973 0.23025 0.21382 0.20633 0.21684 0.17844
0.19454 0.20484 0.21121 0.20018 0.20277 0.22228
0.15811 0.00000 0.22716 0.21725 0.22777 0.21134 0.20385 0.21437 0.17596
0.19206 0.20237 0.20873 0.19770 0.20029 0.21980
0.22964 0.22716 0.00000 0.12738 0.08511 0.09666 0.08917 0.22248 0.18408
0.20018 0.21048 0.20811 0.19709 0.19138 0.21919
0.21973 0.21725 0.12738 0.00000 0.12799 0.11156 0.10408 0.21257 0.17417
0.19026 0.20057 0.19820 0.18717 0.18147 0.20927
0.23025 0.22777 0.08511 0.12799 0.00000 0.09727 0.08978 0.22309 0.18469
0.20079 0.21109 0.20872 0.19770 0.19199 0.21980
0.21382 0.21134 0.09666 0.11156 0.09727 0.00000 0.04087 0.20666 0.16826
0.18435 0.19466 0.19229 0.18126 0.17556 0.20337
0.20633 0.20385 0.08917 0.10408 0.08978 0.04087 0.00000 0.19918 0.16077
0.17687 0.18717 0.18480 0.17378 0.16807 0.19588
0.21684 0.21437 0.22248 0.21257 0.22309 0.20666 0.19918 0.00000 0.14014
0.12388 0.13418 0.20405 0.19302 0.19561 0.21512
0.17844 0.17596 0.18408 0.17417 0.18469 0.16826 0.16077 0.14014 0.00000
0.11783 0.12814 0.16565 0.15462 0.15721 0.17672

```

0.19454	0.19206	0.20018	0.19026	0.20079	0.18435	0.17687	0.12388	0.11783
0.00000	0.07036	0.18174	0.17071	0.17330	0.19281			
0.20484	0.20237	0.21048	0.20057	0.21109	0.19466	0.18717	0.13418	0.12814
0.07036	0.00000	0.19205	0.18102	0.18361	0.20312			
0.21121	0.20873	0.20811	0.19820	0.20872	0.19229	0.18480	0.20405	0.16565
0.18174	0.19205	0.00000	0.10167	0.18124	0.20075			
0.20018	0.19770	0.19709	0.18717	0.19770	0.18126	0.17378	0.19302	0.15462
0.17071	0.18102	0.10167	0.00000	0.17021	0.18972			
0.20277	0.20029	0.19138	0.18147	0.19199	0.17556	0.16807	0.19561	0.15721
0.17330	0.18361	0.18124	0.17021	0.00000	0.19232			
0.22228	0.21980	0.21919	0.20927	0.21980	0.20337	0.19588	0.21512	0.17672
0.19281	0.20312	0.20075	0.18972	0.19232	0.00000			

The program dialogue window contains the following:

Français: tapez (1)

English: type (2)

2

ParaFit: A test of host-parasite coevolution

Pierre Legendre

Département de sciences biologiques

Université de Montréal.

© Pierre Legendre, 2001

Type n1, n2, n3, n4.

A(n1,n2) B(n1,n4)

C(n3,n2) D(n3,n4)

17 15 14 16

Name of input data file A(n1,n2) ?

A(17X15)

Name of input data file B(n1,n4) ?

B(17x16)

Name of input data file C(n3,n2) ?

C(14x15)

The input data matrices have been read.

How many permutations? (e.g., 999, 9999, ...)

999

Computation time: 35.63 sec.

Results are found in file 'Host-parasite.out'

End of the program.

Pay great attention when typing the values of n1, n2, n3 and n4. These dimensions determine the shape of the three input matrices as well as that of matrix **D**, as can be seen in the block diagram representing matrices **A**, **B**, **C** and **D**. In the Macintosh version of the program, a dialogue box appears when the name of an input data file is requested by the program.

Output file

The output file ("Host-parasite.out") contains the following information:

ParaFit: A test of host-parasite coevolution

Pierre Legendre
Département de sciences biologiques
Université de Montréal.
© Pierre Legendre, 2001

Matrix A = A(17X15)
Matrix B = B(17x16)
Matrix C = C(14x15)

Number of permutations: 999

Global test of cospeciation: ParaFitGlobal = 0.01390 Prob = 0.00100

Test of individual host-parasite links:

				F1 = ParaFitLink1		F2 = ParaFitLink2	
Parasite 1	Host 2	F1 =	0.00105	Prob1 = 0.04000	F2 =	0.09387	Prob2 = 0.00500
Parasite 2	Host 1	F1 =	0.00093	Prob1 = 0.02000	F2 =	0.08323	Prob2 = 0.00200
Parasite 3	Host 3	F1 =	0.00172	Prob1 = 0.00100	F2 =	0.15353	Prob2 = 0.00100
Parasite 4	Host 6	F1 =	0.00193	Prob1 = 0.00800	F2 =	0.17212	Prob2 = 0.00100
Parasite 5	Host 5	F1 =	0.00163	Prob1 = 0.00100	F2 =	0.14602	Prob2 = 0.00100
Parasite 6	Host 7	F1 =	0.00158	Prob1 = 0.01700	F2 =	0.14094	Prob2 = 0.00200
Parasite 7	Host 4	F1 =	0.00104	Prob1 = 0.01600	F2 =	0.09306	Prob2 = 0.00100
Parasite 8	Host 1	F1 =	0.00116	Prob1 = 0.08800	F2 =	0.10378	Prob2 = 0.00100
Parasite 9	Host 2	F1 =	0.00067	Prob1 = 0.22100	F2 =	0.05998	Prob2 = 0.04400
Parasite 10	Host 12	F1 =	0.00072	Prob1 = 0.24200	F2 =	0.06450	Prob2 = 0.04600
Parasite 11	Host 13	F1 =	0.00045	Prob1 = 0.46400	F2 =	0.03986	Prob2 = 0.13700
Parasite 12	Host 15	F1 =	0.00085	Prob1 = 0.07500	F2 =	0.07591	Prob2 = 0.00400
Parasite 13	Host 14	F1 =	0.00064	Prob1 = 0.31300	F2 =	0.05705	Prob2 = 0.05100
Parasite 14	Host 9	F1 =	0.00074	Prob1 = 0.18000	F2 =	0.06631	Prob2 = 0.06300
Parasite 15	Host 11	F1 =	0.00154	Prob1 = 0.02300	F2 =	0.13782	Prob2 = 0.00400
Parasite 16	Host 8	F1 =	0.00129	Prob1 = 0.04000	F2 =	0.11500	Prob2 = 0.00700
Parasite 17	Host 10	F1 =	0.00139	Prob1 = 0.03400	F2 =	0.12453	Prob2 = 0.00700

There are 17 host-parasite links in matrix A

Number of hosts per parasite:

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Number of parasites per host:

2 2 1 1 1 1 1 1 1 1 1 1 1 1

Computation time: 35.63 sec.

[Using a PPC at 400 MHz]

The global test indicates the presence of a highly significant ($P = 0.002$) overall coevolutionary structure. Results of the tests on individual links indicate, however, that we are in a mixed situation where portions of the two trees are coevolutionary whereas other portions are not. The simulations reported in the main paper (third type of power simulations) show that in these conditions, type I error on the random links is inflated; *ParaFitLink1* is the recommended statistic in that situation. So, we have to be cautious

not to use a liberal significance level. All links that were not significant at $\alpha = 0.05$ were discarded (Fig. 1). While we recognize that this level is arbitrary, using a more conservative level (e.g., $\alpha = 0.03$ or 0.02) would have eliminated a few more links but would not have changed the overall coevolutionary picture. The tree was pruned of all non-significant links and of all hosts and parasites that became unconnected. This resulted in two identical phylogenetic trees displaying perfect coevolution for a subset of the gophers and lice (Fig. 2).

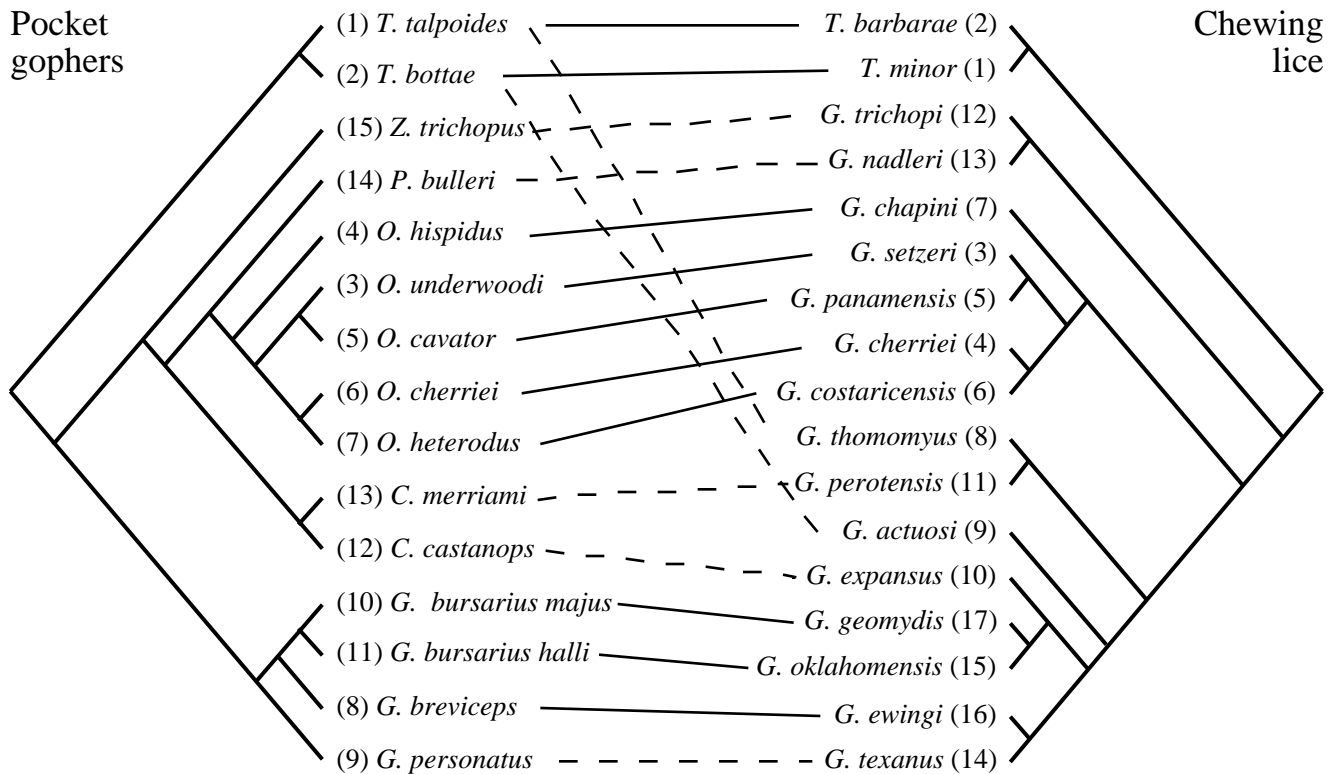


Fig. 1. Pocket gophers and chewing lice phylogenetic trees and host-parasite links. Significant H-P links are represented by full lines, non-significant links by dashed lines. Modified from Fig. 8 of Legendre et al. (2002).

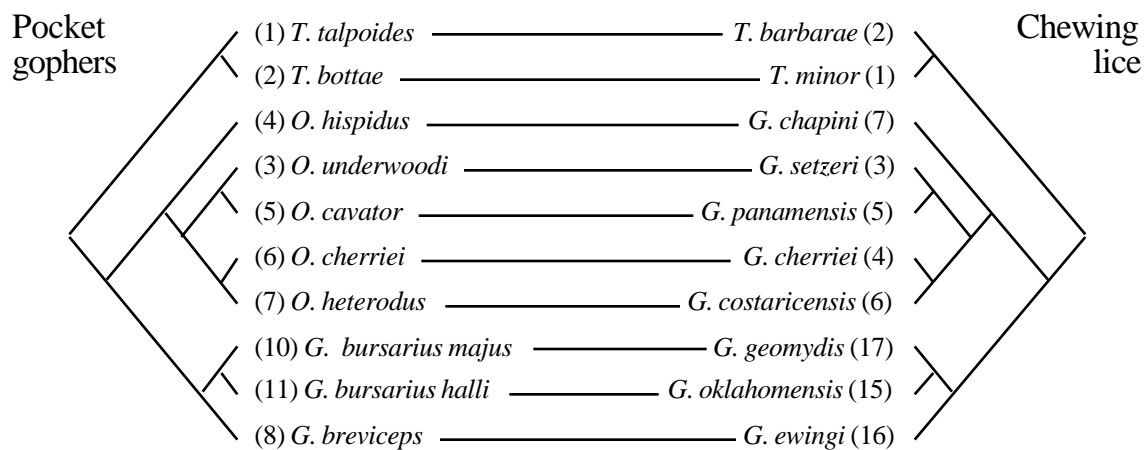


Fig. 2. Pruned trees: the trees are now identical and display perfect coevolution for a subset of the animals. Modified from Fig. 9 of Legendre et al. (2002).