COMPASS (Compatibility Analysis Site Stripping) Manual

COMPASS is a UNIX program for identifying and removing fast evolving sites from morphological or molecular data matrices using a number of compatibility-based methods.

Installation:
The source code and supplementary information on compatibility methods can be downloaded from http://www.ncl.ac.uk/microbial_eukaryotes/downloads.html

The program can be extracted from the tgz file using the following command:
Tar -xvzf COMPASS.tgz

This will create a folder called COMPASS. Inside you will find:
1) This manual
2) A folder called src containing the program source file
3) A folder called example containing example input and output files
4) The Makefile for the program

To install COMPASS:
Enter the COMPASS folder by typing "cd COMPASS"
Type "Make"

Citing COMPASS:
COMPASS is currently unpublished, but has been submitted to Bioinformatics.

Output Files:
COMPASS creates a number of output files. By default a log file is created which describes the analysis and gives compatibility values for each site in the alignment. If a boildown is carried out, the log file also records the order of characters removed and the compatibility value at the time of removal. The program also creates .out and .boil files which are checkpoint files for the program. NEXUS charset and character mask files can also be output (see below).

Running COMPASS:
To run the program type ./COMPASS from the directory containing the program.

An example input file and output files from an LQP boildown of that data can be found in the examples folder in the COMPASS folder.

Usage:

-i	Input file in nexus format	
-t	Compatibility type	A = Count of pairwise Incompatibilities
		C = CCSR (coefficient of character-state randomness)
		L = LQP (LeQuesne Probability)
		N = Normal Deviate
-p	Number of permutations	0 - 10,000
-z	Boildown stop value (default = 0.05)	

-b	Run boildown	0 = No
	(default = 0)	1 = Fast boildown
		2 = Slow boildown
-f	Fuzzy analysis	0 = No
	(default = 0)	1 = Yes
-s	Start character (optional)	
-e	End character (optional)	
-r	Starting seed (optional, default = random)	
-m	Create mask file (optional)	
-c	Create charset file (optional)	
-x	LQP cutoff for charset and mask (optional, default = 1)	
-I	Launch interactive mode	

-p This option defines the number of random permutation to run in CCSR, NDEV and LQP analyses. If 0 is chosen a simple count of observed incompatibilities will be output.

-z allows you to choose to stop the boildown analysis at a particular LQP/CSSR etc. value. If specify a value of zero, the boildown will continue until a minimal value is reached (e.g. in LQP when no permutations in any remaining character are 'better' than the actual value)

-b If you want to run a boildown, it is strongly advised to always use option 1 (fast boildown). This method is MUCH faster than the traditional method (option 2).

-f Fuzzy compatibility is described in the compatibility methods pdf included with the COMPASS package, but is unpublished.

-s and -e allow you to run a subset of the dataset. This allows large datasets to be run across multiple machines or processors. To do this you run LQP analyses on multiple machines with a number of subsets of characters. E.g.:
-s 1 -e 50
-s 51 -e 100 etc.
However YOU MUST USE THE SAME NUMBER OF PERMUTATIONS AND THE SAME RANDOM SEED.
To join the analyses back together, you run the same parameters on the whole data with using the -d option followed by a list of the '.out' files created by your previous analyses. The multiple analyses will be joined and boildowns can be run.

-m and -c produce additional output files. -m creates a file with masks showing which characters have been removed while -c creates a list of charsets of the same.