IndelCoder

Michael S. Rosenberg

Copyright © 2005-2010

This program takes an input nexus file (DNA only) and exports a new nexus file with gaps coded using the simple indel coding scheme of (Simmons and Ochoterena 2000). The program was developed for use in :

> Ogden, T.H., and M.S. Rosenberg. 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation. *Molecular Phylogenetics and Evolution* 42(3):817-826.

The program is a command-line program and can be run under Windows operating systems (from a command line) or Linux (pre-compiled executables for both are included). Macintosh executables can be created if there is an interest.

No special installation is necessary. The archive contains the compiled executable (indelcoder_win.exe, indelcoder_linux), the source code (indelcoder.dpr, Delphi-based object-oriented pascal) and files with the command used to compile this code under different operating systems using the Free Pascal Compiler (fpc...). There is also a sample input nexus file called simpletest.nxs (see below).

To use the program:

Simply type:

```
indelcoder inputfile outputfile
```

(changing "indelcoder" to match the name of the executable for your operating system), where `inputfile` and `outputfile` are the names of the input nexus file and the desired output file. The program should be able to accept a variety of nexus styles, but is fairly simplistic in what it is looking for and specialized notation found in the input file may not make it into the output file. Please examine your output files carefully to ensure that all necessary designations have transferred correctly.

Example:

The sample input data fie (simpletest.nxs) contains the following:

```
#NEXUS

Begin data;
      Dimensions ntax=5 nchar=22;
      Format datatype=dna gap=-;
      Matrix
A -AAAAATTTTTGGGGGCCCCC-
B -AAAA-TTTT-GGGG-CCCC--
C -----AT---TG---GC---CA
D --AAAATTTTTGGGGGCCCCCA
E AA----TTTTT------CCC--
      ;
End;
```

After running the indelcoder on this file, the output fill will contain:

```
#NEXUS
BEGIN DATA;
  DIMENSIONS NTAX=5 NCHAR=30;
  FORMAT MISSING=? DATATYPE=DNA GAP=- EQUATE="0=A 1=C";
  OPTIONS GAPMODE=MISSING;
MATRIX
A      ?AAAAATTTTTGGGGGCCCCC?00000000
B      ?AAAA-TTTT-GGGG-CCCC??11100000
C      ?????AT---TG---GC---CA00011100
D      ??AAAATTTTTGGGGGCCCCCA00000000
E      AA----TTTTT------CCC??-0-0-011
;
END;

BEGIN SETS;
  CHARSET Original=1-22;
  CHARSET InDelChar=23-30;
END;

[ Indel Character         Sites
  ---------------     ---------------
          23               6 - 6
          24              11 - 11
          25              16 - 16
          26               8 - 10
          27              13 - 15
          28              18 - 20
          29               3 - 6
          30              12 - 17
]
```

The new file has the following changes/additions: (1) External gaps have been changed to question marks; (2) a series of 0/1 codes have been appended to the sequences to indicate each internal gap character; (3) character sets have automatically been created to separate the original characters from the indel characters; and (4) a comment at the end of the file contains a list of each new gap character and the sites which it encompasses within the DNA sequences.

Technical Problems

For technical problems, contact me at msr@asu.edu. If possible, include a copy of the problematic data file.

References

Ogden, T. H., and M. S. Rosenberg. 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation. *Molecular Phylogenetics and Evolution* 42:817-826.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49:369-381.

Release History

Version 1.0.1.1 (March 23, 2010)

- Initial release