BIOINFORMATICS APPLICATIONS NOTE
Vol. 19 no. 13 2003, page 1726
DOI: 10.1093/bioinformatics/btg225

CRANN: detecting adaptive evolution in protein-coding DNA sequences

C.J. Creevey* and J.O. McInerney

Bioinformatics and Pharmacogenomics Laboratory, Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland

Received on December 16, 2002; Accepted on February 10, 2003

ABSTRACT

Summary: A software program CRANN has been developed in order to detect adaptive evolution in protein-coding DNA sequences.

Availability: CRANN is available from http://bioinf.may.ie/crann/

Contact: chris.creevey@may.ie

Supplementary Information: CRANN has been written in the C programming language. Source code is available on request.

CRANN is a software program written in the C programming language that can be used to investigate adaptive evolution in a number of ways. The program requires a set of protein-coding DNA sequences, aligned so that the first residue of the alignment corresponds to the first position of a codon. The program checks to ensure that the alignment has valid codons and gap characters.

First of all, the program implements some of the most popular methods of measuring synonymous and non-synonymous distances between a pair of sequences (Li *et al*., 1985; Li, 1993). Phylogenetic hypotheses based upon these distances can also be generated using CRANN, which implements the Neighbor-joining algorithm (Saitou and Nei, 1987). In addition, CRANN can also carry out a moving window analysis using either synonymous (d_s) or non-synonymous (d_n) distances. This kind of analysis can be useful for locating regions of the protein that appear to be under varying selective pressures. CRANN can analyse the entire dataset and compile a cumulative result, or subsets of the dataset can be chosen (perhaps clades of sequences).

The most powerful part of this program is to be found in its ability to detect adaptive evolution along evolutionary lineages. There are two methods implemented in the software—the method described by Messier and Stewart (1997) and the method of Creevey and McInerney (2002). In both cases, hypothetical ancestral sequences are reconstructed at the internal nodes of a given phylogenetic tree. This tree can either be constructed as mentioned previously, using the neighbor-joining algorithm based on either synonymous or non-synonymous pairwise distances or the tree can be read from a file, provided it is in standard New Hampshire format.

The first output file is a tree in nested parentheses format (New Hampshire format) with those lineages identified where the ratio of the rate of non-synonymous invariable to non-synonymous variable changes is significantly different to the synonymous variable to synonymous invariable changes. The second kind of output is a file that details for each internal branch the counts of silent and replacement variable and invariable substitutions. The software also produces a file that details the site-by-site substitution rate for each internal node of the tree. In this way, residues that are potentially under positive selection can be identified.

Identification of positive selection and the residues involved is very important for studies of protein function and CRANN will assist in these analyses. CRANN will also permit comparison of methods for detecting adaptive evolution as it can carry out three different kinds of analyses.

REFERENCES

Creevey,C.J. and McInerney,J.O. (2002) An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene*, **300**, 43–51.

Li,W.H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.*, **36**, 96–99.

Li,W.H., Wu,C.I. and Luo,C.C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.*, **2**, 150–174.

Messier,W. and Stewart,C.B. (1997) Episodic adaptive evolution of primate lysozymes. *Nature*, **385**, 151–154.

Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.

*To whom correspondence should be addressed.

