RED-T: utilizing the Ratios of Evolutionary Distances for determination of alternative phylogenetic events

Kamyar Farahi¹, William B. Whitman¹,∗ and Eileen T. Kraemer²

¹Department of Microbiology and ²Computer Science Department, The University of Georgia, Athens, GA 30602, USA

Received on November 11, 2002; revised on March 24, 2003; accepted on March 28, 2003

ABSTRACT

Summary: RED-T is a Java application for phylogenetic analysis based on a unique method, RED, that utilizes the ratios of evolutionary distances (Ed) to distinguish between alternative evolutionary histories. RED-T allows the user to examine if any given experimental gene shares the same evolutionary history as the designated control gene(s). Moreover, the tool detects any differences in evolutionary history and allows the user to examine comparisons of Ed for a likely explanation. Lateral gene transfer, which may have a significant influence in organismal evolution is one mechanism that could explain the findings of these RED-T analyses.

Availability: The application is available online at http://www.arches.uga.edu/~whitman/RED.

Contact: whitman@arches.uga.edu

Deciphering organismal evolution and detecting phylogenetic events have captured biologists’ attention for generations. With the growth of genomic data and development of advanced methods of analysis, evolutionary biology has taken a great leap forward in understanding the evolutionary history of prokaryotes. An important discovery has been the realization that lateral gene transfer (LGT) occurs at high rates and that LGT may have made major contributions to the formation of the modern lineages of prokaryotes (Lawrence and Ochman, 2002; Snel et al., 2002).

Methods to detect LGT depend upon searching for incongruities in phylogenetic trees or differences in the DNA composition of LGT genes (Lawrence and Ochman, 2002). We sought to develop an alternative strategy to detect LGTs. The RED method relies on a matrix of evolutionary distance (Ed) that is imported from standard phylogenetic programs. It assumes that if the rates of evolution are constant and the evolutionary history of two genes are the same, then plots of the Ed for one gene against another gene should be linear. In addition to LGT, any perturbation of the evolutionary clock, such as changes in gene function or evolutionary rate, could in theory be detected. Therefore, it is possible to use empirical tests rather than tree building to identify non-vertical evolution. For the application of this method, an interactive research tool entitled RED-T—Ratios of Evolutionary Distances for determination of alternative phylogenetic events—was developed. RED-T is a Java application capable of generating scatter plots from distance matrices to analyze evolutionary relationships among various levels of taxa. In addition, it is fully capable of importing new gene data for comparison with the control set we developed (RED control) or allowing the user to develop a new control.

With the assistance of the application wizard, the user can import Phylib formatted distance matrices (Felsenstein, 1989) or columnar formats (see RED help files). RED-T allows the user to rename and map each sequence ID that is imported to the appropriate organism and taxonomy provided by the tool’s catalog. This resource is in accordance with EMBL classifications (Stoesser et al., 2002) and fully adjustable by advanced users. In addition, imported sequence IDs from the WIT/ERGO database are automatically mapped and can then be edited by the user (Overbeek et al., 2000). One other advantage of the mapping step is to allow the user to analyze multiple paralogous genes of an organism or examine unknown sequences.

The imported data can be used as the experimental gene for comparison with the tool’s default control. This control was developed by examining a large number of genes to find those which produced linear RED plots for the organisms studied. For this reason, they are assumed to possess a constant evolutionary history. The user can also develop an original control based on RED’s protocol, which is simulated by the tool’s wizard. This allows the analysis of genes with different evolutionary histories such as rapid gene evolution in Mycoplasma or even transposon phylogeny.

RED-T provides the user with extensive analysis tools to evaluate evolutionary relationships of a gene, ranging from intra-domain and inter-domain comparisons (e.g. archaeal domain versus groups of proteobacteria) to specific relationships of two organisms. In addition, the tool calculates
Fig. 1. Screen shot of RED-T application. The main window is separated into three resizable partitions: the main partition contains the plot generator as illustrated here in highlight mode and a bar displaying quantitative variables that reflect values for the selected portion of the plot. The second partition contains a text box for taking notes of analysis. The last partition contains five tabbed windows to provide different levels of taxonomic selection. In addition, the floating window displays an imported phylogenetic tree. Illustrated here is the analysis of the isoleucyl-tRNA synthetase gene, in highlight mode. This mode allows the user to visualize the relationship between the selected comparisons to the entire data set. In addition, the m-line allows for further interpretation of the mean of $K$ values. High $K$ (ratio of $E_d$ of the experimental and the control) values resulted from comparisons between bacterial clades B1 and B2 (outlined box), suggesting that one of these clades obtained its gene from another domain. It is clear that the eukaryote and bacterial clade B2 (arrows in plot and tree window) comparisons had much lower $K$ values than the eukaryote and bacterial clade B1 comparisons, thus bacterial clade B2 was the likely recipient of the eukaryotic isoleucyl-tRNA synthetase gene.

quantitative variables for the selected comparisons, provides highlight mode and zoom settings to evaluate the user’s selection with other comparisons, and a journal interface for analysis notes (Fig. 1). Any content of the analyses can be printed for further study. Also, phylogenetic trees can be imported for the user to compare and contrast with the RED-T plot. This import feature also allows the user to import other graphical references, such as tables, graphs or diagrams to be saved under the same analyses folder.

In order to evaluate the RED method, RED-T was applied to a data set of ribosomal proteins believed to have been vertically inherited (Hansmann and Martin, 2000). This data set was used to standardize the variability in the absence of LGT (Farahi et al., submitted for publication). Then, 20 aminoacyl-tRNA synthetases were examined for alternative phylogenetic events, such as LGT. Although these synthetases are members of a functionally conserved family, LGT has been proposed for the evolution of many members of this ancient family (Woese et al., 2000). Although some of the proposed LGT events were confirmed in these analyses, many were shown to be better explained by other evolutionary models (Farahi et al., submitted for publication). These and other genes are available as samples in the RED-T package, version 2.1. The advantages of this approach are: the variability inherent in comparing protein sequences is transparent, the direction of LGT and the relative rates of evolution are readily identified, and it is possible to detect other types of evolutionary events.
The RED-T application is available as a Java™Archive (JAR) file. This compression format allows the user to launch the full application on any operating environment provided the system has JRE (Java™Runtime Environment) version 1.4 or above. In addition, the complete source code and documentation are available for downloading by registered users. For additional information regarding current and upcoming features, please visit: http://www.arches.uga.edu/~whitman/RED

REFERENCES


