CRITICA: Coding Region Identification Tool Invoking Comparative Analysis

Improvements and Applications

Jonathan H. Badger

Department of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.
jhbadger@monod.uwaterloo.ca

Abstract

CRITICA (Coding Region Identification Tool Invoking Comparative Analysis) [1] is a collection of programs for gene prediction which combines traditional gene prediction techniques with a novel comparative analysis. If a pair of ORFS in a nucleotide alignment can be found in which the conceptual translated products are more conserved than would be expected from the observed nucleotide conservation (that is, the alignment is significantly enriched in synonymous substitutions), this is evolutionary evidence that the DNA sequences are protein coding. Regions found by this method are used to generate traditional dicodon frequencies for further analysis. CRITICA is not dependent on (often erroneous) sequence annotations, which many other algorithms base their training set upon, and uses comparative information in a more biologically meaningful way than a simple similarity search.

CRITICA has been used in practice; it was used in the published analyses of the *Archaeoglobus fulgidus*[2] and *Aquifex aeolicus* [3] genomes, and is presently being used by several groups, including the WIT2 metabolic reconstruction project [4] team. This poster describes current progress on CRITICA. CRITICA can now automatically detect frameshifts due to sequencing errors. A frameshift error appears as the sudden loss of coding support in one frame together with the sudden gain of coding support in another frame. Similarly, CRITICA is being extended to handle eukaryotic sequences with introns, as introns appear to be regions without coding support in the midst of coding regions. Improvements in detection of transcriptional and translational start sites have also been made.

References

[1] JH Badger and Olsen, GJ. CRITICA: Coding Region Identification Tool Invoking Comparative Analysis. *Molecular Biology and Evolution* **16:** 512-524. (1999).

[2] H-P Klenk, Clayton RA, Tomb JF *et al.* (48 co-authors). The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390:** 364-370. (1997).

[3] G Deckert, Warren PV, Gaasterland T *et al.* (12 co-authors). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392:** 353-358. (1998).

[4] http://wit.mcs.anl.gov/WIT2/