`framefinder` : An error-tolerant approach for prediction of the peptide implied by assembled EST consensi.

Guy St.C. Slater [1] gslater@hgmp.mrc.ac.uk
Martin J. Bishop [1] mbishop@hgmp.mrc.ac.uk
Winston Hide [2] winhide@sanbi.ac.za

[1] *Human Genome Mapping Project RC,*
Wellcome Trust Genome Campus,
Hinxton, Cambridge, CB10 1SB. UK.

[2] *South African National Bioinformatics Institute,*
Private Bag X17, University of the Western Cape,
Belleville, South Africa.

Abstract

We have developed a method for the prediction of the extent and reading frame of coding sequence from EST derived consensus sequences.

This method is based on the use of log-odds hexamer frequency statistics and multiple dynamic programming based frameshift models, but also employs an ambiguity tolerant translation model and context-dependent frameshift penalties to increase tolerance of errors which commonly occur in EST data.

We present here details of the algorithms used, and some results from its use with EST consensus datasets.