Intelligent alternatives for searching the PRINTS database

Philip Scordis, William Wright and Teresa K Attwood.

School of Biological Sciences, Stopford Building, The University of Manchester,
Oxford Road, Manchester, M13 9PT, UK.

scordis@bioinf.man.ac.uk, wright@bioinf.man.ac.uk, attwood@bioinf.man.ac.uk

By reconciling an unknown gene or protein with a known family, we can infer a wealth of previously compiled information pertinent to that family and its members. Here, we introduce a pattern-recognition method (FingerPRINTScan (Scordis et al., 1999)) that classifies sequences using familial definitions from the PRINTS database (Attwood et al., 1999). We also provide a new perspective on the use of BLAST (Altschul et al., 1990) with PRINTS.

FingerPRINTScan is a tool for searching the PRINTS fingerprint database. Fingerprints consist of groups of motifs excised from areas of conservation in multiple sequence alignments. As PRINTS continues to expand, it becomes increasingly important to provide tools for fast, efficient and discriminatory searching of the resource. We have developed a procedure that characterises sequences using signatures in PRINTS, allowing progress to be made with the identification of distant evolutionary relationships. The approach makes use of the contextual information inherent in a multiple-motif method, and has the power to identify hitherto unidentified relationships in mass genome data. This analysis tool combines a simple, user friendly interface with the capacity to provide a biologically-relevant result, directed ‘intelligently’ at the user in a hierarchical manner. The package contains three modules: the main tool (FPScan), which provides the scanning functionality; and two subsidiary tools – GRAPHScan, which provides a visual representation of the resultant sequence-fingerprint match, and MULScan, which facilitates bulk data submission.

Complementing the fingerPRINTScan suite, an implementation of BLAST for searching the fingerprint database has been developed (Wright et al., 1999). The interface allows submission of either protein or DNA queries, and returns the familiar form of output, but modified by means of direct links both to the matched fingerprints and to the GRAPHScan visualisation software. The server thus couples the rapidity of BLAST searching with the sensitivity of fingerprint diagnoses, providing alternative perspectives on a given query; we show that BLAST recognises generic similarity, and lacks the diagnostic sensitivity of family-specific discriminators.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. J. Mol Biol., 215, 403-410.

Attwood, T.K., Flower, D.R., Lewis, A. P., Mabey, J. E., Morgan, S. R., Scordis, P., Selley, J. and Wright, W. (1999). PRINTS prepares for the new millennium. Nucleic Acids Research, 27(1), 220-225.

Scordis, P., Flower, D. R., and Attwood, T. K. (1999). FingerPRINTScan: Intelligent searching of the PRINTS motif database. Bioinformatics, submitted.

Wright, W., Scordis, P., and Attwood, T. K. (1999). BLAST PRINTS – alternative perspectives on sequence analysis. Bioinformatics, 15, in press.