The Eukaryotic Promoter Database

Rouaida Cavin Périer[1,2]
`Rouaida.Perier@isrec.unil.ch`
Viviane Praz[1,2]
`Viviane.Praz@isrec.unil.ch`
Thomas Junier[1,2]
`Thomas.Junier@isrec.unil.ch`
Claude Bonnard[1,2]
`Claude.Bonnard@isrec.unil.ch`
Christian Iseli[1,3]
`Christian.Iseli@licr.org`
Philipp Bucher[1,2]
`Philipp.Bucher@isrec.unil.ch`

[1] Swiss Institute of Bioinformatics, Chemin des Boveresses 155,
1066 Epalinges, Switzerland
[2] Swiss Institute for Experimental Cancer Research
[3] Office of Information Technology, Ludwig Institute for Cancer Research

Keywords: promoter, transcription, database, gene expression

1 Introduction

The Eukaryotic Promoter Database (EPD) is an annotated non-redundant collection of eukaryotic POL II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide sequence entries. The annotation part of an entry includes description of the initiation site mapping data, exhaustive cross-references to the EMBL nucleotide sequence database, SWISS PROT, TRANSFAC, and other databases, as well as bibliographic references. EPD is one of the first databases which exploits the recently introduced sequence version (SV) numbers to guarantee the accuracy of sequence feature positions relative to EMBL entries. Its contents are structured in a way that facilitates dynamic extraction of biologically meaningful promoter subsets for comparative sequence analysis. Web-based interfaces have been developed that enable the user to view EPD entries in different formats, to select and extract promoter sequences according to a variety of criteria, and to navigate to related databases exploiting newly introduced cross-references. The WWW-based user interfaces can be accessed at `http://www.epd.isb-sib.ch`; the flat file release is available by anonymous ftp from ftp.epd.isb-sib.ch.

2 Method and Results

The information in EPD promoters is collected mostly from the journal literature. Cross-links to other databases, except EMBL and TRANSFAC, are established manually. Perl scripts are used to identify candidate EMBL and TRANSFAC entries potentially corresponding to EPD promoters. These entries are, however, also manually checked before being integrated in EPD. Recently, we have started to systematically relate EPD promoter entries from human and selected model organisms to public gene expression data sets based on SAGE or microarray technology, including those available at:

- `http://genome-www.stanford.edu/`
- `http://idefix.upr420.vjf.cnrs.fr`
- `http://rsb.info.nih.gov/mol-physiol/homepage.html`
- `http://www.ncbi.nlm.nih.gov/SAGE/`

We are in the process of developing tools that can use these links to select EPD promoters according to gene expression criteria, and to automatically extract the corresponding DNA sequences for motif search analysis.

References

[1] Bucher, P., Regulatory elements and expression profiles, *Curr. Opin. Struct. Biol.*, 9(3):400–407, 1999.

[2] Junier T. and Bucher, P., SEView: A java applet for browsing molecular sequence data, *In Silico Biol.*, 1:13–20, 1998.

[3] Périer, R. C., Praz V., Junier, T., Bonnard C., Bucher, P., The Eukaryotic Promoter Database (EPD), *Nucleic Acids Res.*, 28:302–303, 2000.