Plant cis-acting regulatory DNA elements (PLACE) database: 1999

Kenichi Higo*, Yoshihiro Ugawa, Masao Iwamoto and Tomoko Korenaga

Department of Genetic Resources, National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Ibaraki 305, Japan

ABSTRACT

PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/) is a database of nucleotide sequence motifs found in plant cis-acting regulatory DNA elements. Motifs were extracted from previously published reports on genes in vascular plants. In addition to the motifs originally reported, their variations in other genes or in other plant species in later reports are also compiled. Documents for each motif in the PLACE database contains, in addition to a motif sequence, a brief definition and description of each motif, and relevant literature with PubMed ID numbers and GenBank accession numbers where available. Users can search their query sequences for cis-elements using the Signal Scan program at our web site. The results will be reported in one of the three forms. Clicking the PLACE accession numbers in the result report will open the pertinent motif document. Clicking the PubMed or GenBank accession number in the document will allow users to access to these databases, and to read the abstract of the literature or the annotation in the DNA database. This report summarizes the present status of this database and available tools.

INTRODUCTION

Recent rapid progress in genome research of higher plants, such as *Arabidopsis* or rice, has produced many nucleotide sequences, and will continue to produce an unprecedented amount of DNA sequence data in the next several years. However, many of these sequences do not offer much information as to the functions of the genes they code. Therefore it will be beneficial if biological information can be extracted from the nucleotide sequences by computer analyses. A database of nucleotide sequence motifs found in plant cis-acting regulatory DNA elements (cis-elements) and a tool for homology searches within such a database will be helpful in estimating the mode of gene regulation, regions involved in such regulation, and other pertinent regions in the DNA sequence.

Previous databases created for similar purposes such as TFD (1), or its successor ooTFD (2), and TRANSFAC (3) contain many eukaryotic cis-elements, but only a small number of those from plant genes were compiled. Therefore it is an excellent time to start compiling motifs and their biological information from the point of view of the plant scientist. In response to such a need we have embarked on the construction of the PLACE, a database of nucleotide sequence motifs found in plant cis-acting regulatory DNA elements (4). In this brief report, we describe the present status of the PLACE database, its content and the database analysis tools, which are currently available.

DATABASE ORGANIZATION

The PLACE database is a compilation of motifs found in plant cis-acting regulatory DNA elements, extracted from previously published reports, and also from article reviews on the regulatory regions of various plant genes. The originally reported motifs, as well as their variations in other genes or other plant species in later reports, are included. The PLACE database also includes some motifs in non-plant cis-elements in the hope that this may assist in finding plant homologues.

In the PLACE database, a document for each motif consists of the description of the following items in each line: ID, a unique identifier; AC, a unique accession number; DT, date of update; DE, a brief description of the motif; KW, keywords; OS, common name and/or scientific name of plant species; RA, author name(s) of a relevant report; RT, title of the report; RL, bibliographic information of the report; RD, PubMed ID numbers (of MEDLINE database at NCBI-NIH, USA) and/or GenBank accession number, where available; (repeat RA to RD for up to seven reports); SQ, motif sequence.

A unique identifier (ID) is assigned to each motif. Early entries have IDs identical to the motif names, which appeared in the reports. As many variations of different motifs appeared, a unique ID to the PLACE database was assigned by the following format: [name of cis-element motif (or trans-acting factor binding site)] [plant scientific name (in initials)] and [gene name]. For example, LTRE (low temperature responsive element) found in *Arabidopsis thaliana* lti78 gene was registered as LTREATLTI78 in the PLACE database. Furthermore identical motif sequence(s) with the same function found in other gene(s) and/or in other plant species are described in the DE field of the document for LTREATLTI78. The motif names commonly used, or their synonyms, are also described in the same DE field.

A unique accession number (AC) is assigned to a motif sequence. The format is: S (for site) followed by a six-digit
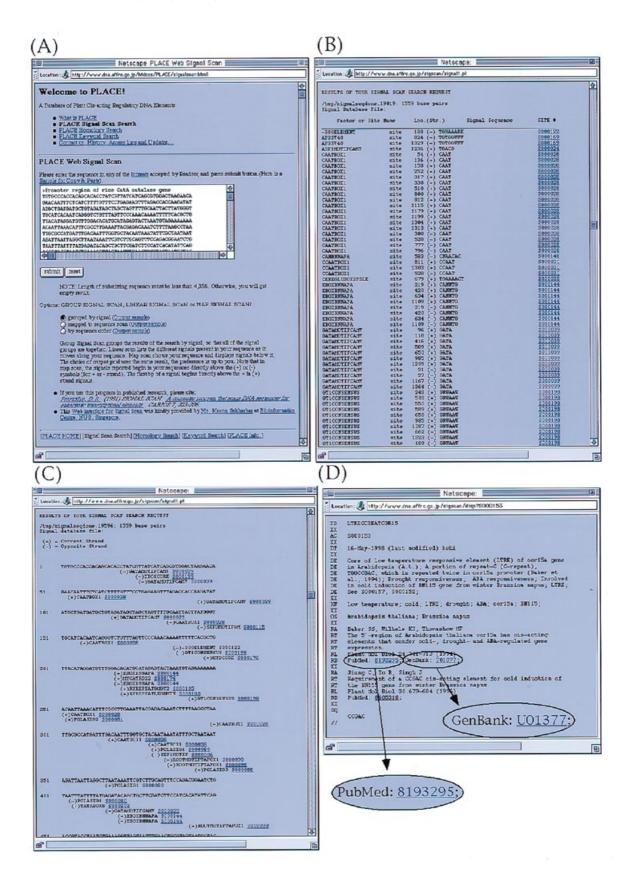
---

*To whom correspondence should be addressed. Tel: +81 298 38 7425; Fax: +81 298 38 7408; Email: kenhigo@abr.affrc.go.jp

Figure 1. A result of sample PLACE query using Signal Scan program. (**A**) A sample entry to the Signal Scan analysis page in the PLACE database. (**B**) A result report in 'grouped-by-signal' format of Signal Scan. (**C**) A result report in 'mapped-to-sequence'. (**D**) The document of the PLACE accession number S000153.

number. This number is strictly associated with a particular registered sequence. When later reports indicate a portion of the registered sequence to be a true motif, a new accession number will be assigned to the shorter motif, and the DE field revised accordingly.

Up to seven references will be included in a document. Additional references will be incorporated after deleting others. The first report on the motif and review articles will normally be retained. Deleted reference(s) may be cited in DE field when space is available.

We added PubMed ID and GenBank accession numbers to the majority of all references. Not all literature is compiled in the PubMed database, and GenBank accession numbers are not listed with every publication. The symbols used, in addition to A, G, C or T, for the sequences of cis-acting element motifs are according to the recommendation by IUPAC-IUB.

HOW TO MAKE USE OF PLACE

Access to the PLACE database

The database is maintained in a server located at the MAFF (Ministry of Agriculture, Forestry and Fisheries, Japan) DNA Bank at the NIAR (National Institute of Agrobiological Resources) in collaboration with colleagues at the DISC (DNA Information and Stock Center), NIAR.

Presently, the PLACE database is accessible via the WWW at: http://www.dna.affrc.go.jp/htdocs/PLACE/ ; it is also available by anonymous ftp from the ftp server: ftp://ftp.dna.affrc.go.jp/pub/dna_place/

At the web site of the PLACE database, users can select either from keyword search, Signal Scan search, or homology search by FASTA.

Keyword search of PLACE database using WAIS

Information on cis-elements can be obtained by keyword search through WAIS (Wide Area Information Service) at the WWW page. Almost any query word can be used as the 'keyword', for example: motif name, name of inducer or plant hormone involved, type of stress, names of tissues or organs in which the gene is expressed, name of author of the report, motif nucleotide sequence, plant species etc.

The resulting report will then show the list of motifs which matches the query word by their accession numbers (AC), and the first line of the description (DE). By clicking the PLACE accession number on the list, a document of the motif will be shown. As the PLACE database is now linked to the PubMed/MEDLINE database Web site at NCBI-NIH, USA, clicking the PubMed ID number in the document will allow users to read the abstract of the paper and additional information. Clicking the GenBank accession number(s) in the PLACE document will show the sequence and annotation of the pertinent sequence in the DDBJ/EMBL/GenBank nucleotide sequence databases.

Searching motif in query sequence using Signal Scan program

The query sequence can be searched for the presence of motifs identical to, or similar to the previously reported cis-element motifs in the PLACE database using the homology search tools. Presently the Web version of the SIGNAL SCAN program (5) is
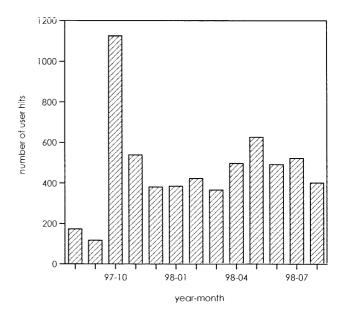

Figure 2. Visits to the Signal Scan analysis (visits per month). Users were counted when the Signal Scan program was used with the PLACE database.

available at our Web site. The query sequence can be entered by copying and pasting onto the window. Due to the limitations in the program, sequences longer than 4355 bases cannot be entered. Longer query sequences must be split into shorter fragments (Fig. 1A).

The appearance of results can be chosen from: 'grouped-by-signal' in which located motifs are listed in alphabetical order of IDs (Fig. 1B); 'mapped-to-sequence' in which motif IDs are shown below the query sequence (Fig. 1C); or 'by sequence order', in which motifs are listed in order of appearance from the 5′ end of the query sequence. The resulting report will then list the motif (site) name, location, motif sequence, and the PLACE accession number. Information on the motif can be obtained by clicking the accession number (Fig. 1D). As described above, clicking the PubMed ID number or the GenBank accession number(s) will allow users to read the abstract of the paper or the sequence and annotation of the pertinent sequence in the DDBJ/EMBL/GenBank nucleotide sequence databases, respectively.

Users of this database should be aware of the fact that these report(s) simply show potential motifs in the query sequence, and that it does not imply that these motifs are actually effective in the regulation of pertinent gene(s).

Notes on PLACE Release 3.1, maintenance policy and future prospects

The number of entries in the current release (Rel. 3.1, as of August 25, 1998) is 208, which has almost doubled since the original release of version 1.0 (114 entries), which was made available to the public on the Internet in 1997. The PLACE database web site has been visited, on average, 1000 times per month. The counter on the Signal Scan analysis recorded an average of 400 visits per month worldwide (Fig. 2); 50% of visits originated from outside of Japan. Records are not kept of any data submitted for a motif search against the PLACE database, except for a simple statistical

analysis of user identification. These statistics are also open to the public via access logs.

Description of a motif (DE line) is updated when a new finding on the motif is reported in a recent publication, as the interpretation of the function(s) of each motif sometimes changes and adjustment of the DE line is required. Due to rather limited resources available for maintenance, update of the database will be made twice a year for the time being. However, more frequent updates are planned as resources become available. Also planned (as an inclusion to the DE field) is information on the position of the motif and other context in each gene. Error corrections of the database will be made when detected. Users are requested to inform us of any error, misinterpretation of reports, and/or missing motifs. Suggestions for improvements to the PLACE database are welcome (http://www.dna.affrc.go.jp/htdocs/PLACE/info.html).

Reference to PLACE database

Users are asked to cite this article when publishing results, which have been obtained through this database.

REFERENCES

1 Ghosh,D. (1993) *Nucleic Acids Res.*, **21**, 3117–3118.
2 Ghosh,D. (1998) *Nucleic Acids Res.*, **26**, 360–361.
3 Heinemeyer,T., Wingender,E., Reuter,I., Hermmjakob,H., Kel,A.E., Kel,O.V., Ignatieva,E.V., Ananko,E.A., Podkolodnaya,O.A., Kolpanov,F.A., Podkolodny,N.L. and Kolchanov,N.A. (1998) *Nucleic Acids Res.*, **26**, 362–367.
4 Higo,K., Ugawa,Y., Iwamoto,M. and Higo,H. (1998) *Nucleic Acids Res.*, **26**, 358–359.
5 Prestridge,D.S. (1991) *Comput. Appl. Biosci.*, **7**, 203–206.