

PlantCARE, a plant *cis*-acting regulatory element database

Stephane Rombauts, Patrice Déhais, Marc Van Montagu and Pierre Rouzé^{1,*}

Laboratorium voor Genetica, Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), Universiteit Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium and ¹Laboratoire Associé de l'Institut National de la Recherche Agronomique (France), Universiteit Gent, B-9000 Gent, Belgium

Received August 31, 1998; Accepted September 28, 1998

ABSTRACT

PlantCARE is a database of plant *cis*-acting regulatory elements, enhancers and repressors. Besides the transcription motifs found on a sequence, it also offers a link to the EMBL entry that contains the full gene sequence as well as a description of the conditions in which a motif becomes functional. The information on these sites is given by matrices, consensus and individual site sequences on particular genes, depending on the available information. PlantCARE is a relational database available via the web at the URL: <http://sphinx.rug.ac.be:8080/PlantCARE/>

INTRODUCTION

The genome sequencing projects have generated and will continue to generate enormous amounts of sequence data. One of the intensively studied and sequenced organisms is the higher plant *Arabidopsis thaliana*. Given the amount of genome sequences, strategies are necessary to discover the location and structure of the genes and to elucidate the expression of those genes and the elements regulating it.

Because experimental information is lacking, *in silico* gene discovery through gene prediction programs is attempted, but gene modeling remained of poor quality due to the lack of descriptions of gene borders and promoters. Knowledge on promoters will soon become available on a large scale for plants as well as for other organisms with the analysis of co-expressed genes on micro-arrays (1–3). This information on promoters will allow the allocation of a cellular function to a gene or will offer biotechnology the possibility to control more specifically the expression of genes. Therefore, the need for managing this knowledge emerged and urged us to build an adequate database.

Earlier databases, such as TFD (4–7) or TRANSFAC (8–10), with the more general purpose of storing any known promoter element and offering search capabilities, have only a few cases of established plant *cis*-acting regulatory elements. This under-

representation of plant data makes these systems not efficiently applicable to plant promoter sequences.

We thus began to collect as much information as possible about plant promoters with their *cis*-acting regulatory elements and to organize them in a local relational database. As we aimed to store every known plant element and gather all the described plant promoters, we designed our database in such a way as to avoid redundancy and to have a clear view on the different sites. We stored every promoter element as we found them in the literature with the degree of confidence that we could find from their description according to the procedures followed, such as experimental procedures or homology searches. At present, we have collected 319 *cis*-acting regulatory elements: 108 from monocots, 206 from dicots and five from other higher plants (such as conifers), describing more than 70 individual promoters from higher plant genes.

DATABASE ORGANISATION

The central 'node' of our database is the name of a site within a plant species. The site behind the name can then be represented by a matrix, a consensus, individual site sequences on particular promoters, or by all of these together. Thanks to the organization of our data we could respond to two needs: (i) collecting what is known about *cis*-acting regulatory elements within plants and (ii) gathering described plant promoters for future *in silico* analysis. Linked to the name of a site, we also provide information on the factor binding to each specific element as far as this information is given in the literature.

Because knowledge of plant promoters is still limited and the number of elements with experimental evidence scarce in the plant kingdom, we plan to collect all the existing knowledge on promoters and their elements by creating three categories for the *cis*-acting regulatory elements complying to the degree of investigation that was made for each particular element. This classification will enable us to include in our listing, besides experimental sites, putative and 'likely to be' elements, and to leave the possibility of investigating such sites to those who encounter them, leading to a confirmation or rejection of the

*To whom correspondence should be addressed at: Laboratorium voor Genetica, Universiteit Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium. Tel: +32 9 264 51 89; Fax: +32 9 264 53 49; Email: pirou@gengenp.rug.ac.be

element. This view allows a different organization of the data compared with PLACE (11).

DATABASE USAGE

Accessing the data from PlantCARE

PlantCARE is a relational database accessible through the internet via the URL: <http://sphinx.rug.ac.be:8080/PlantCARE/>. For instance, the database can be queried on site names, motifs, functions, species, cell type, genes, factors and bibliographic references. These queries will result in a listing of entries with links to other information within the database or beyond through accession numbers from other databases such as EMBL, GenBank and TRANSFAC.

Homology search against the PlantCARE database

We made a web page enabling the user to submit a promoter sequence to search for sites matching elements from the database. This web page proposes a field where a user's sequence can be pasted. The resulting report will show besides the site name, motif and location, a link to a page with more detailed information on the particular site.

Future prospects

Our database will be updated regularly according to the availability of new data. We intend to enhance our search tools to enable searches for promoters and their elements in a more modular way. New fields such as links to PubMed will be added.

As microarray analysis of whole genomes are becoming available, the promoter and coexpression information that will be generated will have to be stored and made accessible for the community with programs that will allow one to search for them on promoter sequences; we intend to develop our database towards this goal.

CITATION OF PlantCARE DATABASE

Users are asked to cite this article when publishing results which have been obtained with the PlantCARE database as described above.

REFERENCES

- 1 Schena, M. (1996) *Bioessays*, **18**, 427–431.
- 2 Shalon, D., Smith, S.J. and Brown, P.O. (1996) *Genome Res.*, **6**, 639–645.
- 3 Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. (1997) *Proc. Natl Acad. Sci USA*, **94**, 13057–13062.
- 4 Ghosh, D. (1990) *Nucleic Acids Res.*, **18**, 1749–1756.
- 5 Ghosh, D. (1991) *Trends Biochem. Sci.*, **16**, 445–447.
- 6 Ghosh, D. (1993) *Nucleic Acids Res.*, **21**, 3117–3118.
- 7 Ghosh, D. (1998) *Nucleic Acids Res.*, **26**, 360–361.
- 8 Wingender, E., Dietze, P., Karas, H. and Knüppel, R. (1996) *Nucleic Acids Res.*, **24**, 238–241.
- 9 Wingender, E., Kel, A.E., Kel, O.V., Karas, H., Heinemeyer, T., Dietze, P., Romaschenko, A.G. and Kolchanov, N.A. (1998) *Nucleic Acids Res.*, **25**, 265–268.
- 10 Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L. and Kolchanov, N.A. (1998) *Nucleic Acids Res.*, **26**, 362–367.
- 11 Higo, K., Ugawa, Y., Iwamoto, M. and Higo, H. (1998) *Nucleic Acids Res.*, **26**, 358–359.