

Databases and ontologies

DictyMOLD-a *Dictyostelium discoideum* genome browser database

Marius Felder¹, Karol Szafranski¹, Rüdiger Lehmann¹,
Ludwig Eichinger², Angelika A. Noegel², Matthias Platzer¹ and
Gernot Glöckner^{1,*}

¹IMB Jena, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena, Germany and

²Center for Biochemistry, Medical Faculty, University of Cologne, Joseph-Stelzmann-Strasse 52, 50931 Köln, Germany

Received on January 14, 2004; revised on February 23, 2004; accepted on January 5, 2005

Advance Access publication January 28, 2005

ABSTRACT

Summary: With the *Dictyostelium* Genome Project nearing completion, we initiated the construction of a data repository for all *Dictyostelium discoideum* genomic data. Up to now this database, called DictyMOLD (Dicty Map Of Linked Data), incorporates the recently completed *D. discoideum* chromosomes 1 and 2 sequences together with related annotations. To visualise maps, sequences and annotations and to provide access for the scientific community a perl-based browser was developed.

Availability: The DictyMOLD database is freely accessible via <http://genome.imb-jena.de/dictyostelium/>

Contact: gernot@imb-jena.de

INTRODUCTION

The lower eukaryote *Dictyostelium discoideum* is a well suited model for the investigation of cell functions related to growth, motility, differentiation and development (Loomis, 1996). An international effort to sequence the genome of this organism was initiated in 1998 (Kay and Williams, 1999). The current version of the sequence assembly (May 2002 freeze) covers ~6.8 Mb of chromosome 2 (C2) in 66 ordered and oriented contigs (Glöckner *et al.*, 2002). Another 0.9 Mb of C2 sequence is available but not mapped yet. The chromosome 1 (C1) assembly (July 2003 freeze) covers with 4.7 Mb the whole unique part of the chromosome in 4 ordered and oriented contigs (G. Glöckner, unpublished data).

Using the gene prediction program GeneID (Parra *et al.*, 2000) in the gene prediction process for C1 and C2 yielded, in total, 4599 predicted genes >40 amino acids (C1: 1770 genes; C2: 2829 genes). BLAST searches (Altschul *et al.*, 1990) were used to determine the chromosomal location of *D. discoideum* repetitive elements (Glöckner *et al.*, 2001), mapping markers (Williams and Firtel, 2000) and ESTs (Morio *et al.*, 1998).

The amino acid sequences of all predicted genes were compared to the full sets of proteins from completed model organism genomes as well as against SWISS-PROT and TrEMBL. They were also

checked for the presence of InterPro domains using the InterPro database (Apweiler *et al.*, 2000). The gene ontology terminology (<http://www.geneontology.org/>) (The Gene Ontology Consortium, 2000) was applied to determine potential roles for the proteins in the cell.

To facilitate storage and exploration of the data, we developed the DictyMOLD genome browser together with its underlying database.

THE DictyMOLD INTERFACE

To enter the chromosome browser different possibilities are offered. Two of the entry points can be accessed via <http://genome.imb-jena.de/cgi-bin/dicty/analysis.cgi>. From there further links lead to a search interface and the genome browser. The search page provides an easy way to query the DictyMOLD database with search strings (e.g. gene identifiers, InterPro numbers). The users can also enter the chromosome browser via a second link, through a specially developed DictyMOLD interface. The web interface allows easy access to selected regions by clicking on the appropriate sequence range of the chromosomes that are schematically represented on this page. Alternatively the search function can be used to search for gene identifiers, accession numbers, HAPPY markers, ESTs and direct access to the corresponding chromosomal location is provided.

The user can navigate through the chromosome and zoom between a 100 kb (Fig. 1) and a 10 kb range view using navigation buttons. The graphical representation of each annotated gene incorporates a direct link to the data sheet containing all available information about the gene product. The third entry point to the DictyMOLD database is provided via a BLAST Server.

The current version of DictyMOLD provides a basic set of annotations that will be expanded in the future. Furthermore, the database will soon be extended to chromosome 3.

Additional *D. discoideum* databases are available under: <http://dictybase.org>, <http://dictygenome.bcm.tmc.edu/>, http://www.sanger.ac.uk/Projects/D_discoideum/, <http://www.uni-koeln.de/dictyostelium/> and <http://glamdring.ucsd.edu/others/dsmith/dictydb.html>.

*To whom correspondence should be addressed.

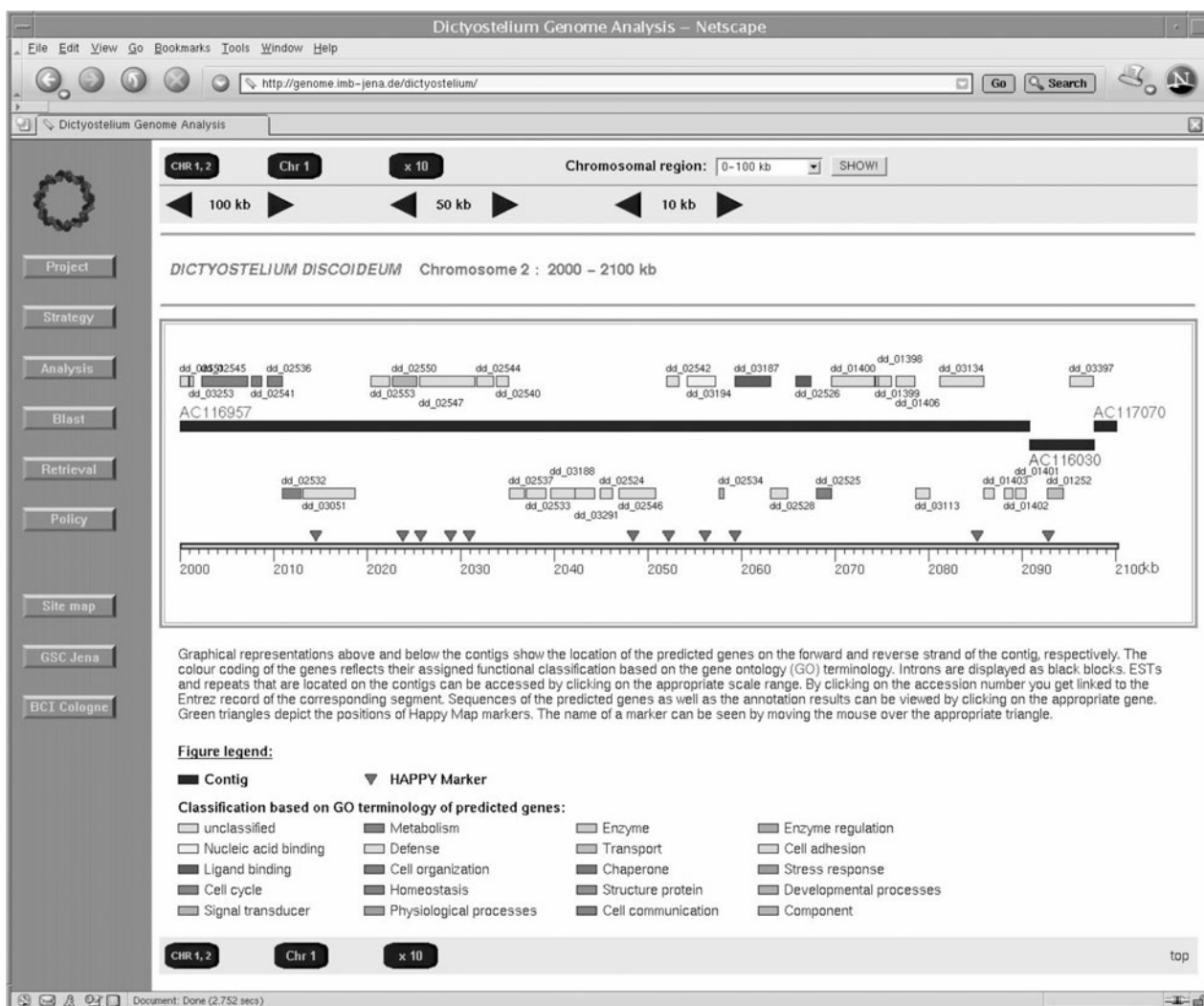


Fig. 1. Screenshot of the genome browser displaying a 100 kb segment of C2.

ACKNOWLEDGEMENTS

We thank the members of the Japanese cDNA project and The *Dictyostelium* Genome Sequencing Consortium for sharing their data. The *D. discoideum* genome project is funded by the Deutsche Forschungsgemeinschaft (DFG), the National Institutes of Health (NIH), the Medical Research Council (MRC) and the European Union (EU). M.F. was supported by the Deutsche Forschungsgemeinschaft (grant PL173/5-2).

REFERENCES

- 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.
- Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Birney,E., Biswas,M., Bucher,P., Cerutti,L., Corpet,F., Croning,M.D. *et al.* (2000) InterPro—an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics*, **16**, 1145–1150.
- Glöckner,G., Eichinger,L., Szafranski,K., Pachebat,J.A., Bankier,A.T., Dear,P.H., Lehmann,R., Baumgart,C., Parra,G., Abril,J.F. *et al.* (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature*, **418**, 79–85.
- Glöckner,G., Szafranski,K., Winckler,T., Dingermann,T., Quail,M.A., Cox,E., Eichinger,L., Noegel,A.A. and Rosenthal,A. (2001) The complex repeats of *Dictyostelium discoideum*. *Genome Res.*, **11**, 585–594.
- Kay,R.R. and Williams,J.G. (1999) The *Dictyostelium* genome project: an invitation to species hopping. *Trends Genet.*, **15**, 294–297.
- Loomis,W.F. (1996) Genetic networks that regulate development in *Dictyostelium* cells. *Microbiol. Rev.*, **60**, 135–150.
- Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T. *et al.* (1998) The *Dictyostelium* developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development. *DNA Res.*, **5**, 335–340.
- Parra,G., Blanco,E. and Guigo,R. (2000) GeneID in *Drosophila*. *Genome Res.*, **10**, 511–515.
- The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genet.*, **25**, 25–29.
- Williams,J.G. and Firtel,R.A. (2000) HAPPY days for the *Dictyostelium* genome project. *Genome Res.*, **10**, 1658–1669.