

MagnaportheDB: a federated solution for integrating physical and genetic map data with BAC end derived sequences for the rice blast fungus *Magnaporthe grisea*

Stanton L. Martin, Barbara P. Blackmon¹, Ravi Rajagopalan, Thomas D. Houfek, Robert G. Sceeles, Sheila O. Denn, Thomas K. Mitchell, Douglas E. Brown, Rod A. Wing¹ and Ralph A. Dean*

Fungal Genomics Laboratory, North Carolina State University, 840 Main Campus Drive, Suite 1200, Raleigh, NC 27606, USA and ¹Clemson University Genomics Institute, Clemson University, Clemson, SC, USA

Received September 4, 2001; Revised and Accepted October 26, 2001

ABSTRACT

We have created a federated database for genome studies of *Magnaporthe grisea*, the causal agent of rice blast disease, by integrating end sequence data from BAC clones, genetic marker data and BAC contig assembly data. A library of 9216 BAC clones providing >25-fold coverage of the entire genome was end sequenced and fingerprinted by *Hind*III digestion. The Image/FPC software package was then used to generate an assembly of 188 contigs covering >95% of the genome. The database contains the results of this assembly integrated with hybridization data of genetic markers to the BAC library. AceDB was used for the core database engine and a MySQL relational database, populated with numerical representations of BAC clones within FPC contigs, was used to create appropriately scaled images. The database is being used to facilitate sequencing efforts. The database also allows researchers mapping known genes or other sequences of interest, rapid and easy access to the fundamental organization of the *M.grisea* genome. This database, MagnaportheDB, can be accessed on the web at http://www.cals.ncsu.edu/fungal_genomics/mgdatabase/int.htm.

INTRODUCTION

The ascomycetous fungus *Magnaporthe grisea* (anamorph *Pyricularia grisea*) causes rice blast, one of the most important diseases of rice, the staple food for half of the world's population (1,2). Once a problem of moderate significance, crop losses associated with this disease have been magnified in recent times with the intensification of rice production. We can expect increasing pressure from this disease as crop production is enhanced to meet the demands of a growing world population. Genetic resistance has been and continues to be the major method of disease control for blast. However, as with many

diseases, *M.grisea* is able to rapidly overcome host resistance (3). Fungal plant pathogens, including *M.grisea*, impact more than food security. In addition to affecting the food supply directly through reduction in yield, they also affect food quality. The wholesale destruction of large fields can affect ecosystem function and have devastating effects on local economies.

Since there are many pathogens worthy of investigation, and given our limited pool of resources, the most reasonable research approach is to fully understand a 'model' pathogenic organism and apply that knowledge to design control strategies for other related pathogens. A candidate 'model' organism should have the following characteristics:

1. It should use an infection strategy that is evolutionarily related to other pathogens.
2. It should have a relatively small genome size.
3. It should be economically important.
4. It should be relatively well studied and amenable to molecular and classical genetic experimentation.

Magnaporthe grisea is ideal as a 'model organism' to represent plant pathogenic filamentous fungi (4,5). The fungus has a small genome size (40 Mb) making it an excellent target for whole genome sequencing (6,7). Decades of intense investigation have resulted in extensive genetic maps and a well established transformation system (8,9). In addition to being representative of many phytopathogenic fungi, *M.grisea* is closely related to other prominent non-pathogenic model fungi, such as *Neurospora crassa* and *Aspergillus nidulans* (10).

The strategy we initially chose for our genome sequencing project was to create a BAC library, sequence the ends, and then use fingerprinting analysis to create a physical framework spanning the entire genome, from which a minimum tiling path of BAC clones would be deduced for sequencing. BAC end sequence and fingerprinting data were obtained from a large insert (~130 kb) *Hind*III BAC library containing 9216 clones prepared from rice-infecting strain 70-15 using the BAC vector pBACwch. This library provides >25-fold coverage of the rice blast genome (11). Additional information can be found via the links shown in Table 1.

*To whom correspondence should be addressed. Tel: +1 919 513 0020; Fax: +1 919 513 0024; Email: ralph_dean@ncsu.edu

Table 1. Links to further information

Description of rice blast project	http://www.cals.ncsu.edu/fungal_genomics/fromMagna/project_riceblast.html
BAC end sequencing and physical mapping	http://www.cals.ncsu.edu/fungal_genomics/fromMagna/BAC_ENDS.html
Status of BACS in the minimum tiling path which have been sequenced	http://www.cals.ncsu.edu/fungal_genomics/fromMagna/current_bacs.html
Blast against BAC end sequences	http://fungi.cals.ncsu.edu/tdsblastscript.html
Database Description: technical details and Scriptable access instructions	http://www.cals.ncsu.edu/fungal_genomics/mgdatabase/magnaporthedb.htm
CVS Repository: downloadable scripts and tools used in the FGL Bioinformatics Lab	http://fungi.cals.ncsu.edu/cgi-bin/cvsweb.cgi/

Table 2. MagnaporthDB summary statistics

Number of fingerprinted BAC clones	7338						
Number of <i>M. grisea</i> BAC end sequences	17684						
Number of TBLASTX hits with significant (E-value < 1×10^{-10}) homology to a BAC end sequence	1609						
Number of contigs from the fingerprint assembly	188						
Number of markers in the database	188						
Number of anchored contigs by chromosome	I	II	III	IV	V	VI	VII
	13	11	12	14	6	14	8

DATABASE DESCRIPTION

We have created an integrated physical and genetic map that interfaces with a searchable database containing end sequence data from the BAC clones, genetic marker data and FPC contig assembly data that represent the majority of the rice blast genome. The database contains:

1. A *Hind*III digestion profile 'fingerprint' of each BAC clone and FPC contig assemblies.
2. Nucleotide sequence data and analysis of BAC ends.
3. Genetic marker data generated previously and assembled into a genetic map at the University of Wisconsin (9) (I. Yap and S. McCouch; <http://ascus.plbr.cornell.edu/blastdb/>).
4. Hybridization of genetic markers to the BAC library, which physically anchors FPC contigs onto one of the seven chromosomes in the *M. grisea* genome.
5. Sequence information for GenBank sequences with significant similarity to *M. grisea* BAC end sequences.

We are using this database as a foundation for generating annotated sequence data of chromosome 7 (4.2 Mb) and to 'fill in the gaps' in the parts of the genome which will remain to be sequenced following whole genome shotgun sequencing. Table 2 contains summary statistics of the data contained in the database. The combined physical and genetic map shown in Figure 1 is an image map whose links provide a 'gateway' to the database used to view *M. grisea* BAC end sequence data, FPC assembled BAC contigs and marker data. The database engine is AceDB, originally developed for use with the *Caenorhabditis elegans* sequencing project (12). Its native graphical libraries have made it a de facto standard for database engines for medium-sized sequencing projects. Figure 2 shows

a graphical representation of an example of an FPC contig presented by the database.

DATABASE INFRASTRUCTURE

The Image/FPC package, available from the Sanger Centre was used to analyze the gels generated in the BAC fingerprinting project and to construct contigs (13). A MySQL database was used to tabulate FPC data. A perl script reads the database. A program written in C and known as 'The Fly' (<http://www.w3perl.com/fly/>) was used to actually generate graphics in gif format to represent the physical and genetic maps of the chromosomes. The graphical maps of the contigs were generated by the native graphics capability in AceDB. AceBrowser (14) was used to display the AceDB graphics over the web. The Sitedefs.pm module in AceBrowser was modified to interface with both MySQL and AceDB. The HTML web pages are served via a Dell PowerEdge 3500 running Redhat Linux 7.0 and the Apache webserver.

AVAILABILITY

The database can be accessed via the International Rice Blast Genome Consortium homepage: <http://www.riceblast.org>. Curator: Stan Martin, stan_1_martin@ncsu.edu.

FUTURE DEVELOPMENT

The database will be expanded to include additional sequencing information and annotation. Every effort is being made to ensure that the annotation taxonomy is compliant with the Gene Ontology Consortium (15) schema to ensure universal accessibility. New modules are being developed that

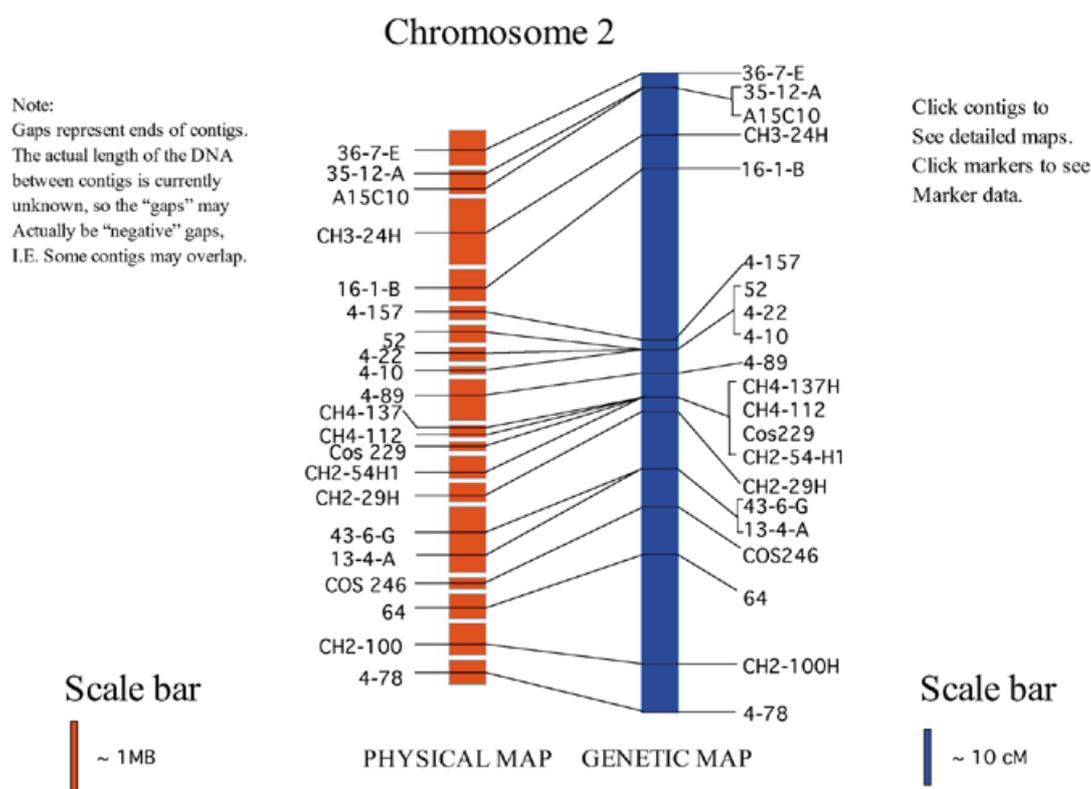


Figure 1. Chromosome 2 of *M. grisea*. Integrated physical and genetic map. All integrated *M. grisea* chromosome maps serve as gateways to the federated database.

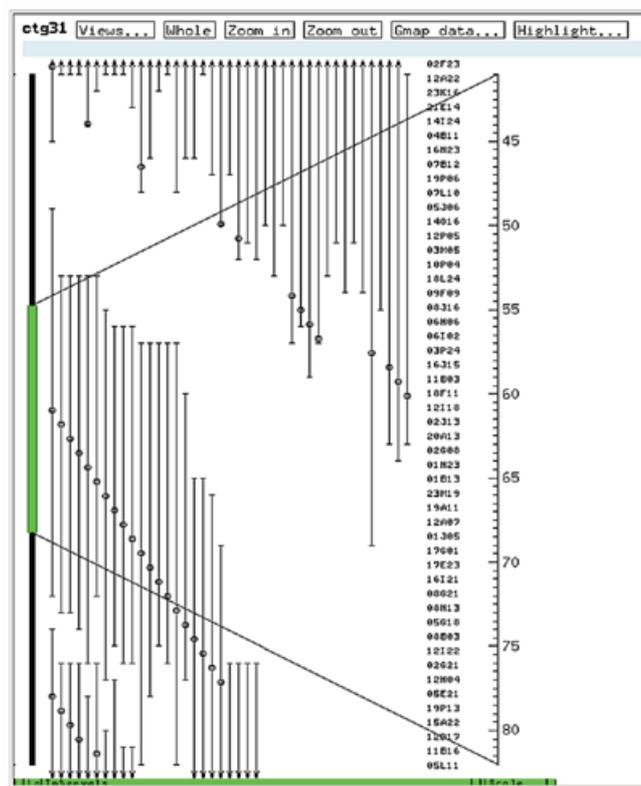


Figure 2. FPC contig 31 as displayed by AceDB. Clones are listed in the same order as they occur in the contig. More information about each clone can be obtained by clicking on the clone name.

will include functional data derived from microarrays, as well as data on mutant phenotypes.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the National Science Foundation, the USDA and Syngenta Corporation.

REFERENCES

- Ou, S.H. (1985) *Rice Diseases*, 2nd edn. Commonwealth Mycological Institute, Kew, UK, pp. 1–380.
- Zeigler, R.S., Leong, S.A. and Teng, P.S. (1994) *Rice Blast Disease*. CAB International, Wallingford, pp. 1–626.
- Bonman, J.M. and MacKill, D.J. (1988) Durable resistance to rice blast disease. *Oryza*, **25**, 103–110.
- Valent, B. (1990) Rice blast as a model system for plant pathology. *Phytopathology*, **80**, 33–36.
- Dean, R.A. (1997) Signal pathways and appressorium morphogenesis. *Annu. Rev. Phytopathol.*, **35**, 211–234.
- Orbach, M.J., Chumley, F.G. and Valent, B. (1996) Electrophoretic karyotypes of *Magnaporthe grisea* pathogens of diverse grasses. *Mol. Plant Microbe Interact.*, **9**, 261–271.
- Zhu, H., Blackmon, B., Sasinowski, M. and Dean, R.A. (1999) Physical map and organization of chromosome 7 in the rice blast fungus, *Magnaporthe grisea*. *Genome Res.*, **9**, 739–750.
- Leung, H.U., Lehtinen, R., Karjalainen, R., Skinner, D., Tooley, P., Leong, S.A. and Ellingboe, A. (1990) Transformation of the rice blast fungus *Magnaporthe grisea* to hygromycin B resistance. *Curr. Genet.*, **17**, 409–411.
- Nitta, N., Farman, M.L. and Leong, S.A. (1997) Genome organization of *Magnaporthe grisea*: integration of genetic maps, clustering of transposable elements and identification of genome duplications and rearrangements. *Theor. Appl. Gen.*, **95**, 20–32.

10. Taylor, J.W., Bowman, B.H., Berbee, M.L. and White, T.J. (1993) Fungal model organisms: phylogenetics of *Saccharomyces*, *Aspergillus*, and *Neurospora*. *Syst. Biol.*, **42**, 440–457.
11. Zhu, H., Choi, S., Johnston, A.K., Wing, R.A. and Dean, R.A. (1997) A large insert (130 kbp) bacterial artificial chromosome library of the rice blast fungus *Magnaporthe grisea*: genome analysis, contig assembly and gene cloning. *Fungal Genet. Biol.*, **21**, 337–347.
12. Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J. and Spieth, J. (2001) WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **29**, 82–86.
13. Soderlund, C., Longden I. and Mott, R. (1997) FPC: A system for building contigs from restriction fingerprinted clones. *Comp. Appl. Biosci.*, **13**, 523–535.
14. Stein, L. and Thierry-Mieg, J. (1998) Scriptable access to the *Caenorhabditis elegans* genome sequence and other ACEDB databases. *Genome Res.*, **8**, 1308–1315.
15. Blake, J. (2001) The gene ontology consortium creating the gene ontology resource: design and implementation. *Genome Res.*, **11**, 1425–1433.