MAMMOT—a set of tools for the design, management and visualization of genomic tiling arrays

E. Ryder1,2,*, R. Jackson2, A. Ferguson-Smith2 and S. Russell1
1Department of Genetics, University of Cambridge CB2 3EH, UK and 2Department of Anatomy, University of Cambridge CB2 3DY, UK

Received on December 22, 2005; revised and accepted on January 27, 2006
Advance Access publication February 1, 2006
Associate Editor: Martin Bishop

ABSTRACT
Summary: The MAMMOT software suite is a collection of Perl and PHP scripts for designing, annotating and visualizing genome tiling arrays to, for example, facilitate studies into the epigenetics of gene regulation. The web design allows rapid experimental data entry from multiple users, and results can easily be shared between groups and individuals.
Availability: http://www.mammot.org.uk/
Contact: e.ryder@gen.cam.ac.uk

INTRODUCTION
With the growing prevalence and use of microarrays, a vast amount of data are being generated about genome-wide gene expression in a variety of species. While expression profiling is very important, it can be hard to elucidate from such data how individual genes are controlled in vivo. One method of investigating gene regulation and epigenetic control is to produce a tiling path microarray, using PCR product or oligonucleotide probes, and hybridize labelled samples enriched by chromatin-immunoprecipitation (ChIP) or other experimental strategies to look for potential binding sites, modified chromatin or other useful features (see review by Mockler and Ecker, 2005 for other applications and types of DNA tiling arrays).

To assist in such efforts we have produced MAMMOT, a set of scripts and visualization tools for the production and annotation of genomic DNA tiling arrays. While in no way a direct replacement for genome browsers, such as those available from Ensembl (http://www.ensembl.org) or UCSC (http://genome.ucsc.edu/), the MAMMOT suite does provide dedicated tools for the specific task of managing a genome tiling array project and processing data. Although the visualization software was originally designed for use with arrays based on PCR products, other array types (e.g. those using oligonucleotide probes) can also be easily imported and used.

In addition to MAMMOT, other tools for the design and optimization of genomic tiling arrays have recently been published by Bertone et al. (2006), who used a dynamic programming and heuristic search approach to circumvent repetitive sequences in large-scale oligonucleotide and amplicon-based arrays. This is in contrast to MAMMOT, which, while it uses a much simpler algorithm for amplicon-based tile design, can include repetitive sequences if desired (e.g. in epigenetic studies).

There are also several tools available for visualizing and analysing data produced by Affymetrix array technologies. These include the Integrated Genome Browser (Affymetrix) and ChipViewer (H. Chen and J. Ecker, unpublished data), which can also perform a variety of statistical tests to facilitate identification of novel transcripts, evolutionary conserved non-coding regions and Single Feature Polymorphisms in Arabidopsis thaliana. Another application, ARTADE (Toyoda and Shinozaki, 2005), uses a threshold-based likelihood and a bi-directional Markov model to estimate precise splice points and the exon/intron structure of previously undefined genes from oligonucleotide array expression data.

MAMMOT has so far been successfully used for the design and subsequent production of tiling pathways across a 120 Kb region of the X chromosome of Drosophila melanogaster and a 1.7 Mb imprinted domain from Mus musculus chromosome 12.

TILING DESIGN
The main tiling script takes a FASTA file of sequence data [either raw or masked with Ns in RepeatMasker (A. F. A. Smit and P. Green, unpublished data)], splitting it into jobs based on a predefined amount of repeat DNA (typically 150 bp). If there are no repeats the whole sequence length is considered one job. For each job, the script analyses the length of the sequence and splits the sequence into fragments based on a defined tile size. Sequences are then sent to Primer 3 (Rozen and Skaletsky, 2000) for oligo design to ensure that the primers are matched pairs for PCR. The extent of tile overlap can be defined when running the script and the amount of any resultant gaps can be traded off against overlap size, depending on the experimenter’s needs. After one round of tile design, another two rounds are performed using any gap data available from previous rounds.

All output files are tab delimited and can be easily exported into MySQL for use in the visualization program or for other analysis. Any masked sequence data used can be transformed back into raw or masked with Ns in RepeatMasker output files. Since it may be desirable to exclude some regions from the final path (e.g. due to a high repeat content), a PHP script can rank the tiles according to

*To whom correspondence should be addressed.
primer sequence, amplicon length, repeat content and proximity to genomic features, giving a final ‘usefulness’ score.

THE MAMMOT WEBSITE AND BROWSER

The MAMMOT website is divided into a number of sections covering all aspects of chip and experiment management. Data are stored in a MySQL database, and the user interface is written in PHP and Javascript.

The chip administration section allows the creation, appending and deletion of microarray chips, where data can either be imported directly from the tiling script output file or from external sources (e.g. an oligonucleotide array design). Further annotation can also be added, including PCR amplification results (based on a simple one-digit code; entry system currently set up for the ABgene Electro-Fast® Wide system) and custom feature tracks for the tile viewer.

The experiment and results pages allow the entry and management of ChIP enrichment microarray data. Given the nature of the web interface, experimental data can be easily added by multiple users from different locations to facilitate intra- and inter-lab collaborations. Experiments can be combined with others if required and exported in .BED format for use in the UCSC genome browser.

A batch viewer showing tile annotation (gene hits, GO annotation, etc.) is also included for rapid analysis of areas of interest indicated by microarray results.

The MAMMOT viewer (Fig. 1) consists of a search and viewing pane, and a tile information pane. The viewing pane shows an overview of the selected region, including the tiling pathway (colour-coded to repeat content; a thin line under the tile indicates the PCR status), gene models and a feature track showing repeats.

Moving the mouse over the gene model or repeat provides further information (repeat name/class, GO annotation, etc.) as a floating box. Custom feature tracks, experimental data and a GC content plot can be toggled on and off as required by the end user, as can a filter for failed PCR reactions. Microarray data from ChIP experiments or other sources can also be added and viewed as stacked bar charts or combined in a single chart.

Clicking on a tile displays data about the amplicon in the information pane including primers, amplification conditions, repeat regions, Blastn matches, tile sequence and any experimental result.

ACKNOWLEDGEMENTS

R.J. is funded by an MRC studentship. Aspects of this work were supported by the EU Epigenome Network of Excellence (A.F.-S.) and a BBSRC Investigating Gene Function award (S.R.).

Conflict of Interest: none declared.

REFERENCES