

BACMan: BAC Data Management for High Throughput Physical Mapping of Genomes

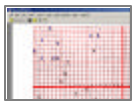
James C. Estill¹, John E. Bowers¹, Barry S. Marler¹, Andrew H. Paterson¹

1. University of Georgia Center for Applied Genetic Technologies, 111 Riverbend Road, Athens, GA 30602

Data Entry

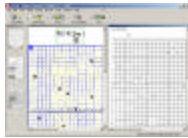
Films with few hits are hand-scored and the data are manually entered, while films with large number of hits (>25) are manually scored on a transparency overlay. These data are then entered into the database using a manual data input method or text recognition software. All data are traceable by the films individual Bar-Code ID.

Manual Bar-Code Data Entry
An MS Access interface where a Bar-Code scanner is used to jump to the data entry page of a particular film which has been hand scored.



Manual BACGrid Data Entry
The MS Word VBA application BACGrid is used to manually digitize data from hand-scored overlays.

Text Recognition Data Entry
Hand scored overlays are scanned and read by the ABBY FineReader ICR software then sent to the MSeXcel VBA application ABBY Eater.



Abstract

BAC Data Management (BACMan) is a RBMS for large insert library hybridization data that provides for high throughput screening of probes, increased quality assurance and interoperability with existing physical mapping software. BACMan increases throughput by allowing for the design and deconvolution of pooled probe multiplex hybridization experiments while incorporating high throughput scoring routines that utilize optical character recognition of hand-scored templates. Quality control incorporates a bar-coding system in which all films are traceable to individual experiment and probe. Further quality assurance is provided by a system in which deconvoluted hybridization data may be error checked with machine generated error checking templates that flag unmatched or overmatched hits and that can be compared to the original scored film. BACMan is fully interoperable with FPC, and is integrated with a sequence data management application currently under development in our lab. Online context sensitive help has been developed as an HTML help file, and an intuitive graphical user interface insures the application has a low learning curve. BACMan is currently deployed under MS Access, but work is underway to allow for deployment as a stand-alone Windows application using the Access run-time environment.

Interoperability

Web Database Browsing

Previous versions of BACMan made web browsing available through active server pages. We are currently converting access tables to MySQL tables, and future web browsing will be via PHP formatted interfaces.



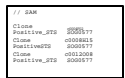
SeqMan

Probe sequence data are linked to a sequence data analysis tool under development in our lab. This tool allows for batch submission to GenBank, automated BLAST against local databases, and export of FASTA formatted data.



FPC

Data can be exported from BACMan into the format necessary for use in FPC.



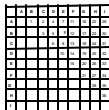
Data Analysis

Multiplex Design and Deconvolution

BACMan can design and deconvolute pooled probe multiplexed experiments in a number of different formats including the grid, triangle, and three dimension multiplex formats.



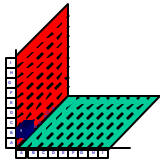
Grid Format
12 Filters: A-L
36 Probes: 1-36



Triangle Format
9 Filters: A-I
36 Probes: 1-36

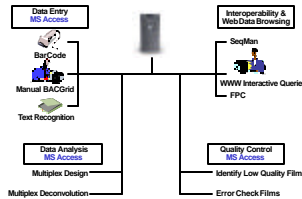
Three Dimension Format

The use of 3D multiplex design can be used to provide a more robust design by adding redundancy. Thus only two of the three expected hits need to be positively matched to identify the probe.



BACMAN

BAC Data Management



Quality Control

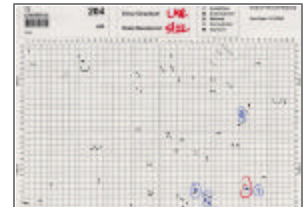
Low Quality Films are Flagged

Films that contain a high proportion of unmatched, undermatched or overmatched hits are flagged as low quality films.



Error Checking

The data may be compared to the original film using an error checking sheet shown below. The data can then be manually changed in the database to reflect any changes made.



DEFINITIONS OF HIT QUALITY

- *Unmatched - No matching hits occurred within the multiplex experiment
- *Undermatched - Only a subset of the expected number of hits occurred in the experiment (2 of 3)
- *Matched - All of the expected matching hits occurred in the experiment (3 of 3)
- *Overmatched - Too many matching hits occurred in the experiment
- *Expected - A hit that was expected based on the third component of an undermatched pair

User Interface

The BACMan (BAC Data Management) user interface is a series of intuitive forms. Help files were developed in HTML Help Workshop and are directly available from within the application.



Funding for this project has been provided by the National Science Foundation. We would also like to thank the UGA Department of Plant Biology for a travel award to J. Estill.

See a demo of this database Monday at 2:00 pm in the California Room

<http://www.plantgenome.uga.edu>

Software

ABBY Fine Reader 5.0 - <http://www.abbyy.com>
 ABBY Eater - <http://www.plantgenome.uga.edu/>
 BACMan - <http://www.plantgenome.uga.edu/>
 BACGrid - <http://www.plantgenome.uga.edu/>
 FPC - <http://www.genome.arizona.edu/software/fpc>
 MS Access 2002 - <http://www.microsoft.com/Access>
 Visual Studio.Net - <http://www.microsoft.com/vstudio>