Fingerprinted BAC Clone Physical Maps

I. Bosdet, S. Barber, S. Chan, S. Chand, R. Chiu, A. Cloutier, C. Fjell, S. Flibotte, D. Fuhrmann†, M. Krzywinski, D. Lee, C. Mathewson, T. Olson, K. Osoegawa*, A. Prabhu, P. Saeedi, H. Shin, M. Tsai, N. Wye, P.J. de Jong*, J. Schein, S. Jones and M. Marra

British Columbia Cancer Agency
Vancouver, British Columbia, Canada

Genome Sciences Centre

www.bcgsc.ca
info@bcgsc.ca

1. Overview

We are constructing high-resolution BAC-based physical maps employing an agarose gel-based fingerprinting methodology (Marra et al., 1997). This technology has been used to produce BAC-based maps of several genomes, including human, mouse and *Cryptococcus neoformans*. We are in the process of constructing a number of physical maps, including those of the rat and bovine genomes, and are fingerprinting currently 15,000 BAC clones per week.

The mouse physical map contains a total of 305,768 whole-clone HindIII fingerprints from two large-insert BAC libraries. Incorporated into the map are data for 16,997 markers provided by other researchers. Manual editing of the resulting contigs, performed at the Washington University Genome Sequence Center, has reduced the number of contigs to 325, spanning an estimated 99% of the mouse genome. This map provides a resource around which sequencing of the mouse genome is being organized.

Physical maps for the fungal pathogen *Cryptococcus neoformans* serotype A (H99) and serotype D (JEC21) were constructed in a similar manner. Markers and BAC-end sequence data have also been added to the two maps. These resources will be used for future comparative genomics studies of the genetics and virulence of this organism, and are assisting the sequencing of JEC21 at The Institute for Genomic Research (TIGR).

The rat physical map, which contains 185,292 clones, is being used to support sequencing of the rat genome. We are selecting, from the assembled contigs, a set of minimally overlapping BAC clones to be sequenced at the Baylor College of Medicine Human Genome Sequencing Centre. To date 15,417 fingerprinted clones have been selected from the FPC database for sequencing, representing approximately 90% of the entire rat genome.

All FPC fingerprint databases and associated data are updated on a weekly basis and are publicly available for download (http://www.bcgsc.ca). FPC databases may also be viewed with our new Java-based program, iCE (Internet Contig Explorer, available at http://ice.bcgsc.ca).

2. Fingerprinting activities

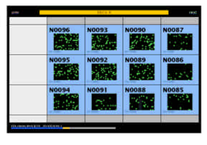
Project	Clones (gels)
Mouse	336,096 (3,501)
Rat	199,872 (2,082)
Bovine (in progress)	231,278 (2,451)
Poplar	48,384 (504)
Cryptococcus neoformans (A, B, D)	9,216 (96)
Leishmania major	6,144 (64)
Ustilago hordei	2,304 (24)
Chlamydia trachomatis (D, E, L2)	1,536 (16)
Salmonella typhymurium	768 (8)
Total:	835,598(8,171)

3. Status of Fingerprint Maps

Organism	No. fingerprints	No. contigs (>2)	Genome Size	Status (coverage)	Collaborator
Mus musculus (C57BL/6)	305,768	7,582 (6,770)	~ 3 GB	Complete (18 X)	Washington U. GSC Whitehead Institute NHGRI
C. neoformans (JEC21)	2,642	20 (20)	~ 16 MB	Complete (16 X)	Jim Kronstad, UBC
C. neoformans (H99)	2,612	20 (19)	~ 16 MB	Complete (14 X)	Jim Kronstad, UBC
Ustilago hordeii	2,032	51 (50)	~ 20 MB	Complete (9 X)	Jim Kronstad, UBC
S. typhimurium	736	7 (7)	~4.9 MB	Complete (15 X)	Washington U. GSC Rod Wing, Clemson
Rattus norvegicus	185,292	11,262 (9,982)	~ 3 GB	Complete (10 X)	Washington U. GSC Baylor HGC NHGRI
Bos taurus holstein and hereford	190,456	12,572 (10,922)	~ 3 GB	In Progress (10 X)	ASRA (R. Church) S. Moore / B. Benkel USDA
Leishmania major	4,561	78 (75)	34 MB	Complete (12X)	Peter Myler SBRI (Seattle)
P. tremuloides	46,024	5,950 (4,338)	~500 MB	Complete (10X) Editing in progress	Carl Douglas, UBC Brian Ellis, UBC
Atlantic salmon	N/A (300,000)	N/A	~3 GB??	Library U/C P. de Jong	W. Davidson, SFU Ben Koop, UVic

4. Human BAC Re-array

We have recently completed assembly of a re-arrayed set of human BAC clones from the RPCI-11 (http://www.chori.org/bacpac) and Caltech D libraries. This array will be valuable for future projects including whole-genome CGH studies. Future plans include chromosome- and arm-specific versions as well as arrays of the mouse and rat genomes.

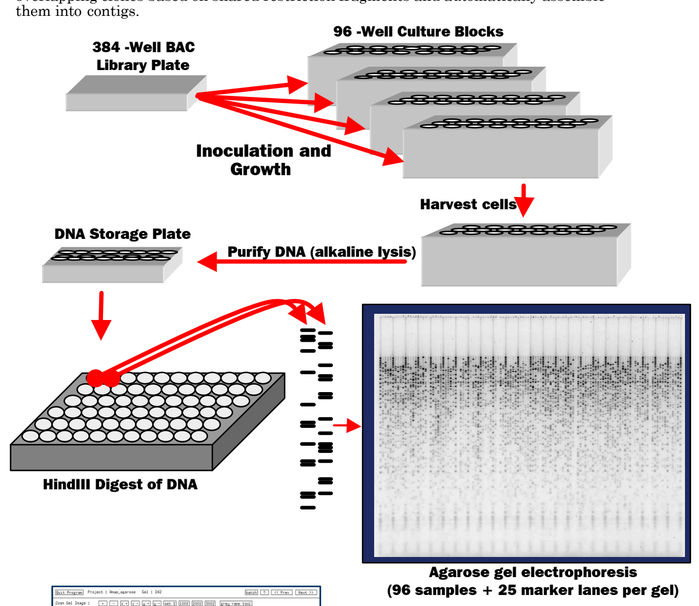


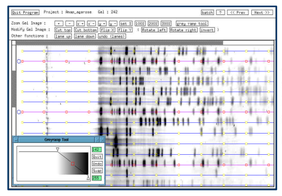
Almost 29,035 BAC clones, selected from the human physical map, were re-arrayed (from the >300,000 total clones in the libraries) using a Genetix QPIX II robot. For more details on this array see poster by M. Krzywinski et al.

Dr. Michael Smith
1932-2000
Founding Director

5. Fingerprint Data Generation and Map Construction

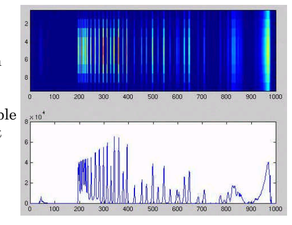
Large-insert BAC clones are grown, purified and digested with a restriction enzyme in 96-well format. The resulting restriction fragments are resolved using agarose gel electrophoresis and stained with SYBR-Green dye. Gel images are collected on a fluorimager and restriction fragment identification is performed automatically using the Bandleader software package. The gel and restriction fragment data is then exported to FPC (http://www.genome.clemson.edu/fpc/), which is used to computationally identify overlapping clones based on shared restriction fragments and automatically assemble them into contigs.

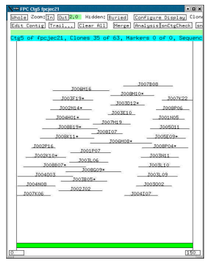




An **automated lane tracking** application is used to analyze the gel images for identification of marker and fingerprint lanes, a process which has previously been performed manually.

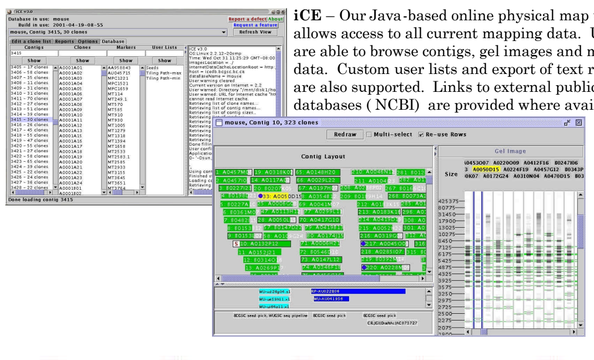
Bandleader – Detection of restriction fragments is accomplished using a customized image analysis package written in MATLAB. Bandleader is able to detect marker lanes and fingerprint data with a high degree of sensitivity and specificity (see analysis in **6.**).





FPC – Fingerprints are assembled into contigs using a distributed version of FPC which allows multiple computers to process the data simultaneously. Assembly of a mammalian-sized map can be completed in approximately 3 hours using a 30-processor compute cluster.

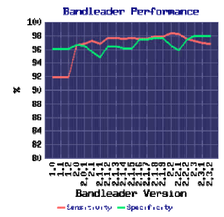
Correct **clone ordering** within contigs is valuable for many applications of fingerprint maps. An algorithm to automatically order clones and merge contigs is currently under development. This will reduce the manual editing of contigs required to produce a complete and well-ordered map, decreasing both the cost and time required for map construction.

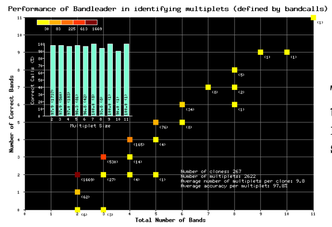


iCE – Our Java-based online physical map viewer allows access to all current mapping data. Users are able to browse contigs, gel images and marker data. Custom user lists and export of text reports are also supported. Links to external public databases (NCBI) are provided where available.

6. Bandleader Development

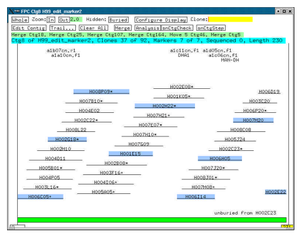
The Bandleader software package provides highly accurate and robust restriction fragment detection. Continued development of Bandleader has resulted in an increase in the accuracy and sensitivity of band detection.





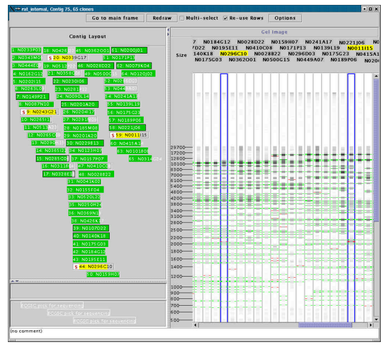
Testing of Bandleader is performed on the fingerprints of 139 human and 128 mouse BACs for which complete sequence is available.

7. Sequencing Seed Selection



Many of our fingerprint maps are used to assist in genomic sequencing activities. A tiling set of minimally-redundant clones is selected from the contigs in the map for sequencing purposes. Clones may also be selected from the map for gap closure. This strategy has been employed for the *Cryptococcus* and poplar genomes.

The rat physical map is also being used for sequencing seed clone selection. During the fingerprinting and map assembly process, high-quality non-overlapping clones were selected in an automated fashion from the unordered contigs of the map. On a weekly basis these clones were provided to the Human Genome Sequencing Center at the Baylor College of Medicine for library construction and shotgun sequencing. In a ten-month period, 15,417 clones were selected and over 11,000 have entered the pipeline. This sequence will be combined with whole-genome shotgun sequence data and used to assist in the whole-genome assembly. (http://www.nih.gov/science/models/rat)
A similar clone-selection strategy was employed for the mouse genome, resulting in 12,527 clones being selected.



8. Future Projects

Genomics Research on Atlantic Salmon Project (GRASP)

The Genome Sciences Centre has received funding from Genome Canada to fingerprint a 300,000 clone BAC library of the *Salmo salar* (Atlantic Salmon) genome. In addition, numerous EST libraries from various tissues and developmental stages are being sequenced (see poster by D. Smailus et al).

GRASP offers the possibility of identifying large numbers of genes and proteins related to disease resistance, reproduction, growth, tolerance to physical factors, product quality, and nutrition. This knowledge will accelerate investment in salmonid fish health, vaccine development, and an array of fisheries and aquaculture related industries.

Acknowledgements

National Institutes of Health
National Human Genome Research Institute
British Columbia Cancer Agency and the British Columbia Cancer Foundation
Genome British Columbia / Genome Canada
Natural Sciences and Engineering Research Council of Canada
Alberta Science and Research Authority
United States Department of Agriculture

***BACPAC Resources, Children's Hospital Oakland Research Institute**
†Department of Electrical Engineering, Washington University St. Louis