FF0019

Automated Finishing System at JCVI

Hoda Khouri, Luke Tallon, Heather Forberger, Brent Bradley and Dan Kosack.
The J. Craig Venter Institute. Rockville, Maryland.

Abstract

In a continuing effort to further automate finishing, the autoCloser was developed and implemented at JCVI. Software tools analyze assembly results, identify finishing targets, design primers, select clones, and choose laboratory reactions to resolve each target. Finishing features targeted by autoCloser include intra-scaffold gaps, low coverage regions, repeat elements, and scaffold ends (physical gaps). Laboratory reactions are grouped by type and clustered to form work orders. A separate LIMS element, called Clover, processes the work orders into laboratory consumable instructions including barcoding, primer orders, clone locations within the template blocks (source plates) and new clone locations in re-arrayed plates (destination plates). A Hamilton® MicroLabSTAR robotic system reads the files generated by Clover, re-arrays the templates, adds the correct primer to each sample, and performs sequencing reactions to generate the finishing reads. The finishing reads are detected by an automated tracking system and incorporated into the targeted contigs and scaffolds using an assembly stitching process. Results are reported automatically to finishing project managers.

Introduction

In this study, we analyze genomes sequenced to 8X coverage with traditional Sanger methods and genomes sequenced to 5X Sanger plus 454 LS pyrosequencing. The reads were assembled into contigs using the Celera assembler and then further grouped into scaffolds. Intrascaffold gaps and low coverage regions were targeted using autoCloser. The resulting finishing reads were incorporated into the project upon reassembly. An average of 60% of targeted gaps and 70% of low coverage areas were resolved after a single iteration of autoCloser.

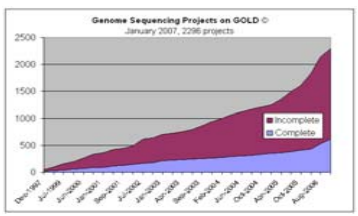
In this poster we report the autoCloser data, compare the genomes studied, and evaluate the cost and the efficiency of the system. AutoCloser proves useful for both complete or partial finishing of both Sanger only and hybrid (Sanger/454) sequenced genomes.

Complete Genomes:

742 Bacteria, 41 Archea, 49 Eukaryotes and 90 Metagenomes have been completely sequenced.

Source : GOLD www.genomesonline.org/gold.cgi

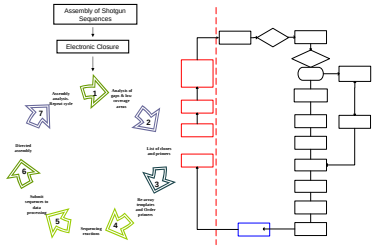
Updated June 3 2007



Genomes analyzed

23 bacterial genomes and 4 fungal genomes where analyzed.
Autocloser was used to close the small (<0.5 Kb) and medium gaps (0.5 – 3 Kb) in all the genomes.
The low coverage features were attempted in the bacterial genomes only.
The average number of gaps/Mb is 22
The average number of features/Mb is 105

AutoClosure Integration



Software – AutoCloser & Clover

Autocloser

- Identifies and categorizes features, by type and size
  - Small, medium or large sequencing gaps
  - Physical gaps
  - Repetitive areas
  - Clone and sequencing coverage
- Creates reaction recipes (clone/primer)
- Utilizes Primer 3 Software to design primers (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)
- Incorporates MUMMER to check for uniqueness of primers

Clover

- Creates all output files for laboratory process:
  - Primer order lists
  - Plate storage location
  - Worklist file for liquid handling equipment
  - Reaction records compatible with the LIMS system

Closure Criteria: Three Tier Approach

For a genome sequenced to a minimum of 8x Sanger-only coverage, we assume the library representation to be random

Tier I – AutoClosure: Given a typical microbial genome, each of two iterations of autoCloser will close an average of 60% of sequence gaps and 58% of low coverage regions. At the completion of AutoClosure a very high-quality draft genome can be expected.

Tier II – Routine FInishing: At the completion of routine closure, a Sanger-only genome can be expected to have 1.5 sequencing gaps per Mbp. In the presence of a reference genome, the order and orientation of all scaffolds along the chromosome will be confirmed and 70% of physical gaps will be closed. Tandem arrays will remain unfinished. Polymorphisms in the sequence will remain unconfirmed.

Tier III – Full 'Gold Standard' Finishing. This tier will yield a continuous consensus of DNA sequence, no ambiguous consensus bases, a minimum of at least 2X sequence coverage and 2X clone coverage over the entire genome, and complete confidence in all repetitive areas.

Hamilton® Microlab STAR Liquid Handling Robot

Hamilton® Microlab STAR

- 8 channel robotic arm with X, Y, and Z independent movement.
- Capable of transferring 1uL to 1000uL.
- Used primarily for re-array of primers and templates.
- Uses picklist generated from Clover to move desired clones from source plates into destination plates.
- Integrated bar code scanner allows plates to be loaded randomly and picked efficiently.
- No sorting of plates is required

Deck Layout

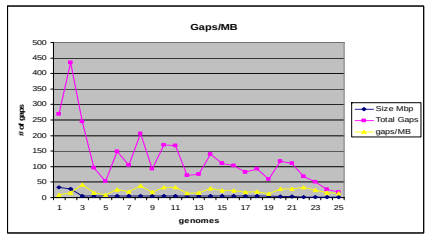
Carriers 3-8: 30 Source Plates
Carrier 2: 5 Destination Plates
Carrier 1: P480 Tips (5 Boxes of 96)

Re-array

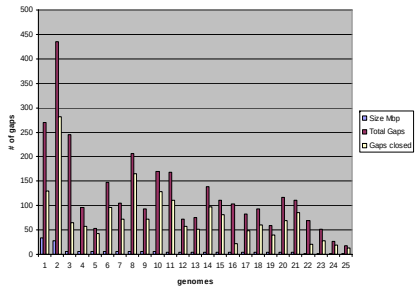
Gap Closure Results

	EUK	PROK	HYBRID	TOTAL
# of Genomes	4	23	3	30
# Mb	112.2	102.1	16.8	231.8
Gaps Attempted	1929	2168	125	4276
Gaps Closed	1079	1503		
% Success	55.9	69		
Attempted Gaps/Mb	17.2	21.2		18.27
Closed Gaps/Mb	9.6	14.7		

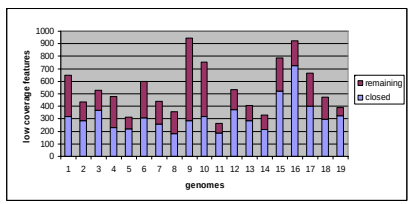
Gap Distribution



Closed gaps after one autoCloser run

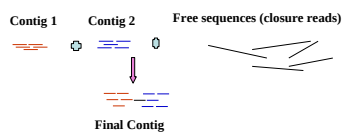


Low coverage features results after one autoCloser run



Contig Stitching

Contig stitching is assembling new reads into existing contigs and gaps without changing the multiple alignments of the existing contigs.



Conclusions

- AutoCloser effectively designs primers for sequencing gaps, low coverage and physical gaps all at once.
- At least 60% of the sequencing gaps are closed with 2 x coverage without any manual intervention.
- The remaining gaps are either too large and need another round of autoCloser or present some difficulties and require manual closure
- Average contig size increases and overall quality improves due to added coverage.
- Clover can batch several small genomes to optimize laboratory operations
- Automation reduces the number of trials and eliminates human error in primer and template rearray.

Recommendations

- Reduce typical genome coverage and utilize autoCloser for more targeted improvement
- Enable autoCloser to recognize reference-based synteny links to target scaffold ends
- Run multiple iterations to finish large gaps.
- Further integrate AutoCloser into LIMS
- Use autoCloser to target islands of interest in metagenomic data
- Automate contig stitching step

References

AutoCloser User Manual (JCVI)

Delcher Al, Phillippy A, Carlton J. Salzberg SL. Fast Algorithms for Large Scale Genome Alignment and Comparison. Nucleic Acids Research (2002), Vol. 30, No. 11:2478-2483.

Fleischmann R.D., et. al. Whole -Genome Sequencing of *Haemophilus influenza* Rd. Science (269): 496-512. 1995.

Goldberg, S et al.., 2006. Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. PNAS Vol (103) 11240-11245.

Hamilton Micro Lab STAR User Manual

Myers W. Eugene, G. Sutton, A. Delcher et. al. A Whole-Genome assembly of Drosophila. 24 March Vol 287 Science(287) 2196-2204. 2000.

Sutton G.G., et. al. TIGR Assembler: A New Tool for Assembling Large Shotgun Sequencing Projects. Genome Science and Technology. 1(1):9-19. 1995

Venter, J.C.; Remington, K.; Heidelberg, J.F.; et. al. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science*, 304, 66-74.

Acknowledgements

The authors would like to thank the JTC sequencing and closure teams as well as:
Martin Shumway, Danny Katzel ,Daniela Puiu,
Karena Sylvester and Miguel Covarrubias
Most of the grants where funded by NIH/NIAID as part of the Microbial Sequencing Center at JCVI
(http://www.genome.gov/dmid/genomes/mscs/)

HT Closure Laboratory Proce