

RTG Investigator™ – High Throughput Metagenomics

Richard Littin, Ph.D. and Brian Hilbush, Ph.D.

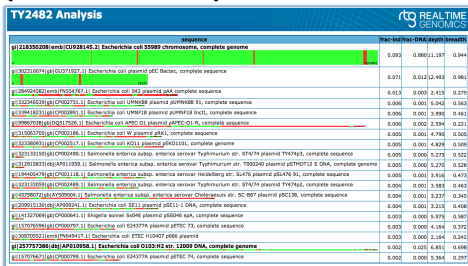


RTG Investigator

Real Time Genomics has built RTG Investigator, a software toolkit with metagenomics and variance detection features bundled into a single package. A novel hash-based indexing search engine provides high-throughput mapping and gapped alignment of shotgun reads to both genome references and protein databases. The engine has been adapted to perform large-scale k-mer phylogenetic analysis on genome sequence and metagenomic samples. Multiple outputs are produced to provide a rich set of alternative views and afford unique insights into phylogenetic relationships.

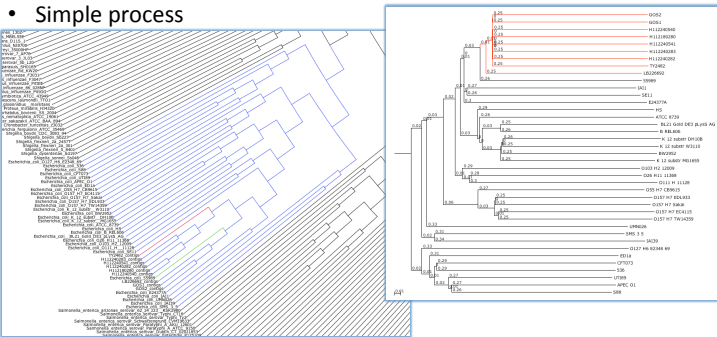
E. coli Novel Species Composition

- RTG map -> species
- RTG map -> coverage
- Ion Torrent reads
- DB of 1,750 species
- 3,200 sequences
- Fast turn around
- < 35 minutes
- Confirmed results



Novel Species Phylogeny

- RTG similarity (phylogenetic tree generation)
- Multiple semi-assembled samples/isolates
- Bacteria sequence database
- < 3 minutes (30 E. coli sequences + 9 isolates)
- < 20 minutes (1,300 bacteria sequences + 9 isolates)
- Simple process



All processing performed on one dual CPU, quad-core Xeon E5520, 48GB RAM computer system running Linux (Centos 5.6).

Acknowledgements

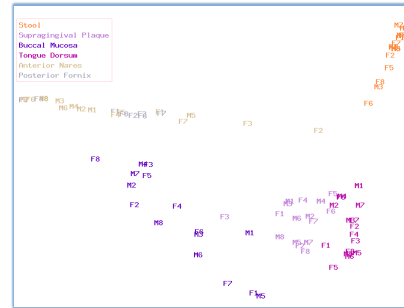
- Metagenomic sample analysis
- E. coli read data



HMP Metagenomic Sample Analysis

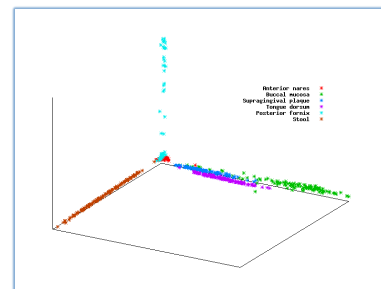
Tools are required to differentiate metagenomic samples from various human body sites. The samples contain a high proportion of unknown bacteria, which causes loss of information when mapping against known species. RTG Investigator provides multi-sample k-mer overlap analysis with RTG similarity.

- RTG similarity -> SVD
- 88 samples
- 5/6 body sites
- 8 male/female
- < 15 minutes
- Confirmed hypotheses
- Robust w.r.t. k-mer size
- New insights
- Simple process



More Samples

- RTG similarity -> SVD
- 628 metagenomic samples
- 6 body sites, 3 oral
- 54 to 138 samples per site
- 1M reads per sample used
- Clustering by site
- Separation within oral cavity
- 3D gives better visualization



Applications

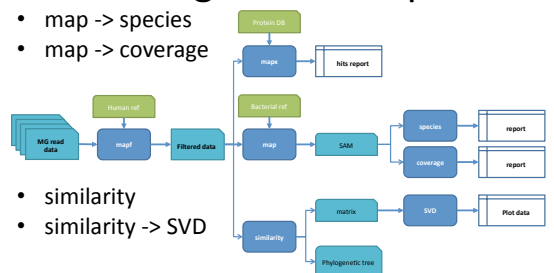
- Novel species discovery
- Sample QA
- Biodiversity analysis
- Population shift analysis
- Contaminant detection

References

- Phylogenetic trees rendered with Archaeopteryx <http://www.phylosoft.org/archaeopteryx/>
- Rohde H, Qin J, Cui Y, et al. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med* 2011. DOI: 10.1056/NEJMoa1107643.
- Smet A, Van Nieuwerburgh F, Vandekerckhove TT, et al. Complete nucleotide sequence of CTX-M-15-plasmids from clinical *Escherichia coli* isolates: insertional events of transposons and insertion sequences. *PLoS ONE* 2010;5(6):e11202.

RTG Investigator	
map	Sensitive sequence alignment <ul style="list-style-type: none"> • Illumina, Ion Torrent, 454 • Complete Genomics
coverage	Depth, breadth statistics <ul style="list-style-type: none"> • Coverage histogram • Gap analysis
mapx	Translated nucleotide search <ul style="list-style-type: none"> • Illumina style reads • 100x to 1000x on BLASTX
species	Species frequency estimator <ul style="list-style-type: none"> • Bayesian technique • Abundance statistics
similarity	NxN sample comparisons <ul style="list-style-type: none"> • Similarity matrix • Phylogenetic tree

RTG Metagenomics Pipelines



All tools in a single, easy to install, Java executable.



RTG Investigator
 Free for individual use
 web: www.realttimegenomics.com
 email: info@realttimegenomics.com