Next Generation Sequencing, Pipeline Development, and Data Analysis

Authors:
Jordan Taylor, Julian Pierre,
Amit Upadhyay, Bhanu Rekepalli

Overview

- Overview of Bioinformatics
- Next Generation Sequencing
- Data Analysis
- Genome Assembly
- Downstream Analysis
- Results/Comparisons
- Conclusion
- References

Overview of Bioinformatics

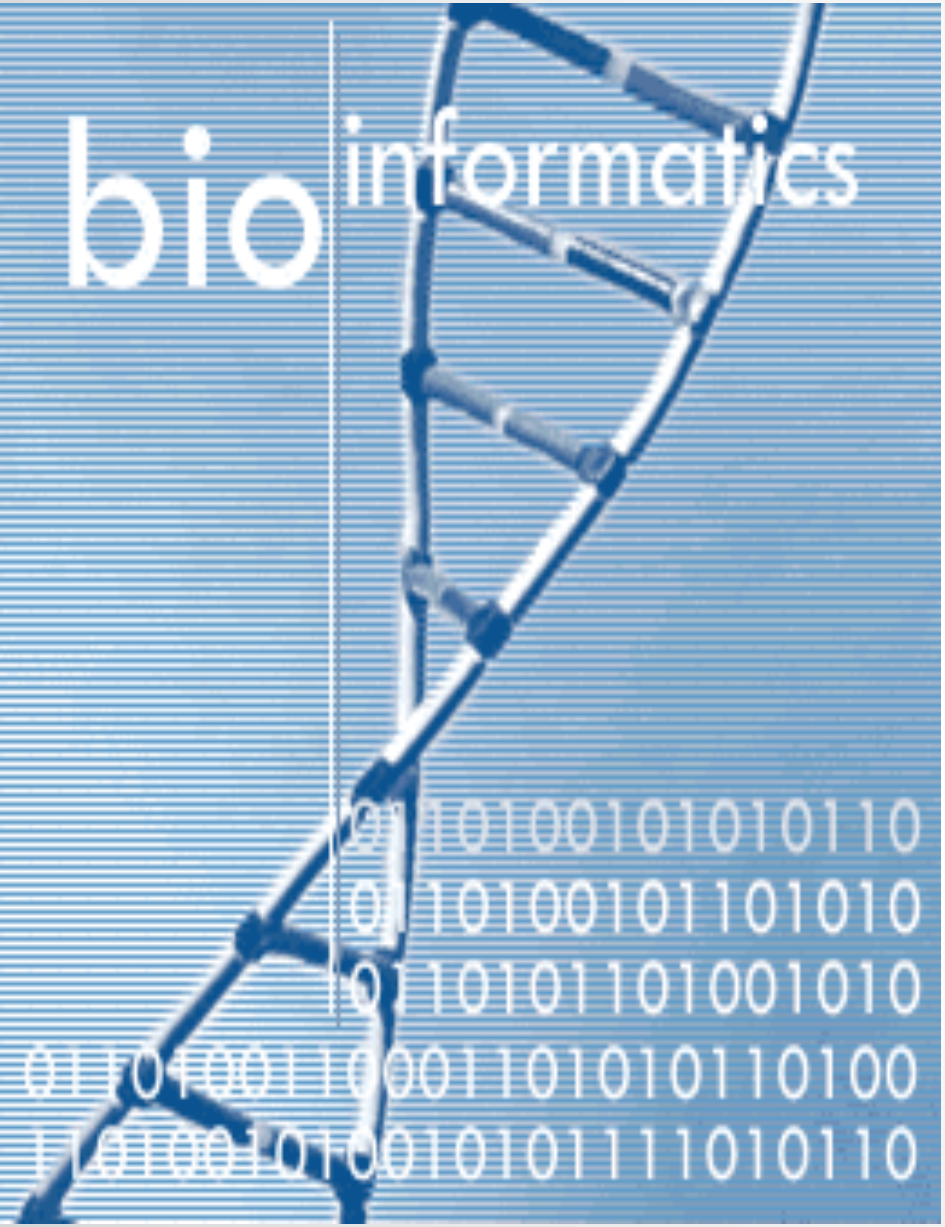


Image taken from Scripps Translational Science Institute

- Encompasses Multiple Scientific Disciplines
- Software Development
- Data Management
- Knowledge Discovery

Image taken from Applied Biosystems

Roche 454 SEQUENCING

Image taken from 10th Cyanobacterial Molecular Biology Workshop

illumina®

Image taken from Mobile Health 360

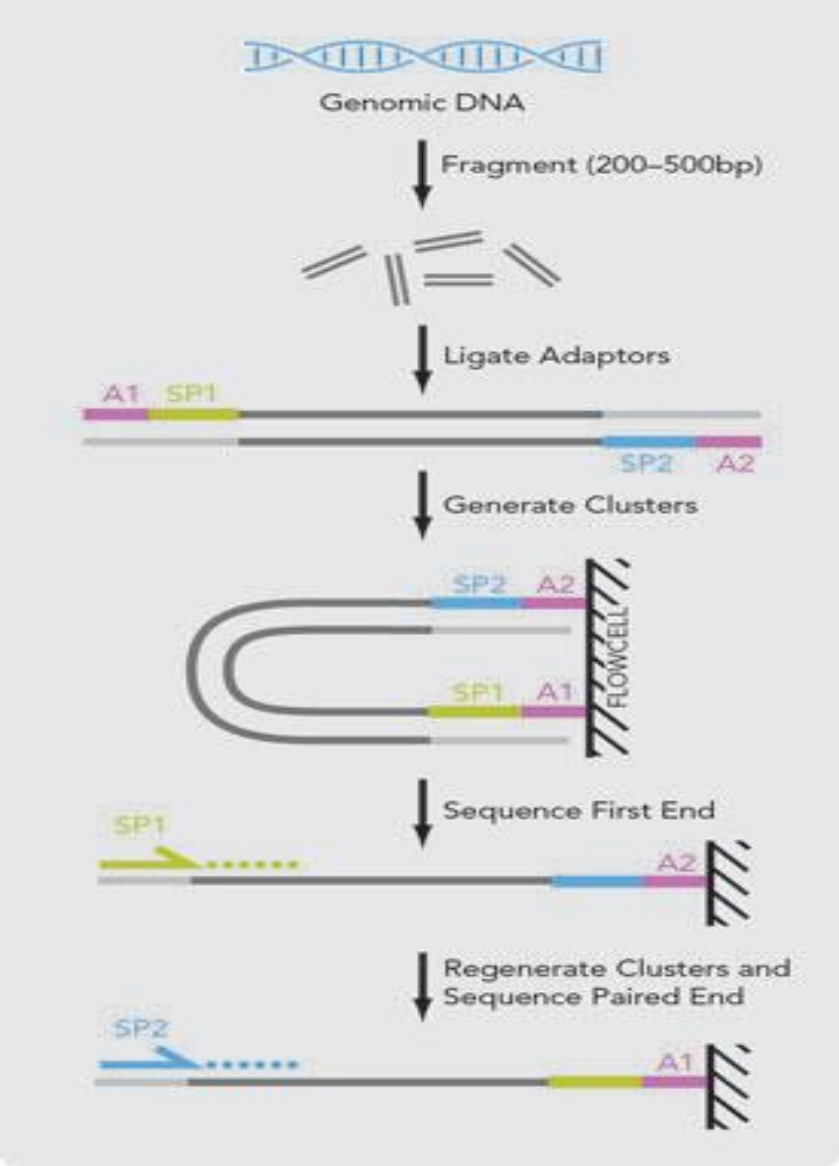
Helicos BioSciences Corporation

Image taken from Xconomy

AB Applied Biosystems

Image taken from Industrial Exhibition 2004

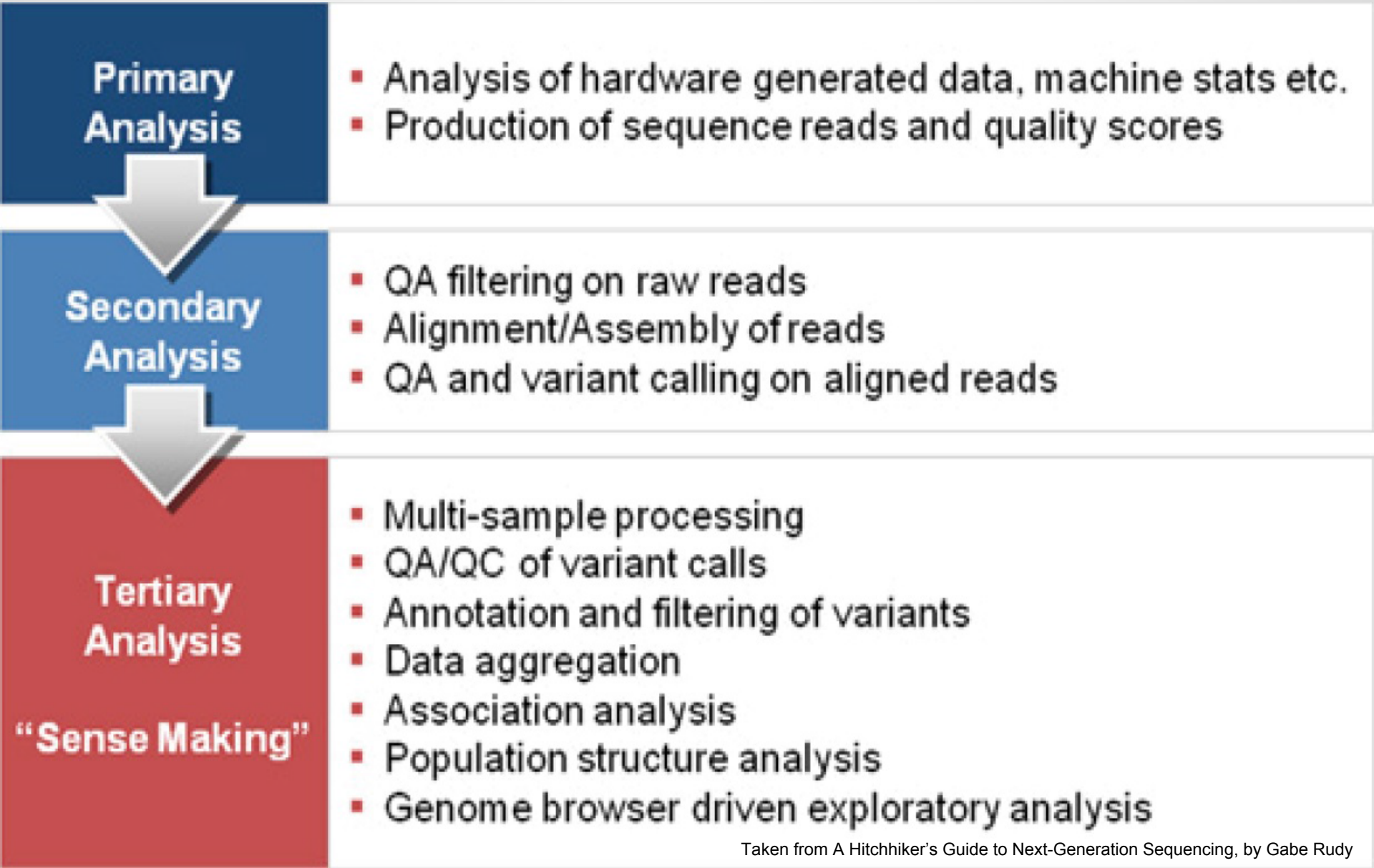
Assembly



- Single End Sequence Assembly
- Paired End Sequencing Assembly
- Which is Preferred?

Image taken from Illumina

Data Analysis



Taken from A Hitchhiker's Guide to Next-Generation Sequencing, by Gabe Rudy

Downstream Analysis

- BLAST (Basic Local Alignment Search Tool)

contig09789 : NODE_52575_length_348_cov_3.548851, NODE_97374_length_772_cov_4.497409, NODE_204370_length_106_cov_2.075472, NODE_284311_length_72_cov_4.611111, NODE_514121_length_244_cov_2.102459

contig09788 : NODE_138945_length_196_cov_2.464286, NODE_246064_length_161_cov_2.614907

contig09783 : NODE_14623_length_1135_cov_3.317181, NODE_23403_length_416_cov_3.625000, NODE_57393_length_270_cov_2.614815, NODE_160727_length_508_cov_3.454724, NODE_198812_length_374_cov_3.762032, NODE_306930_length_186_cov_3.284946, NODE_311476_length_111_cov_7.243243, NODE_346931_length_216_cov_2.680556, NODE_511004_length_133_cov_2.015038, NODE_537359_length_317_cov_3.987382, NODE_547876_length_181_cov_2.331492

- MUMmer (Maximum Unique Matches)

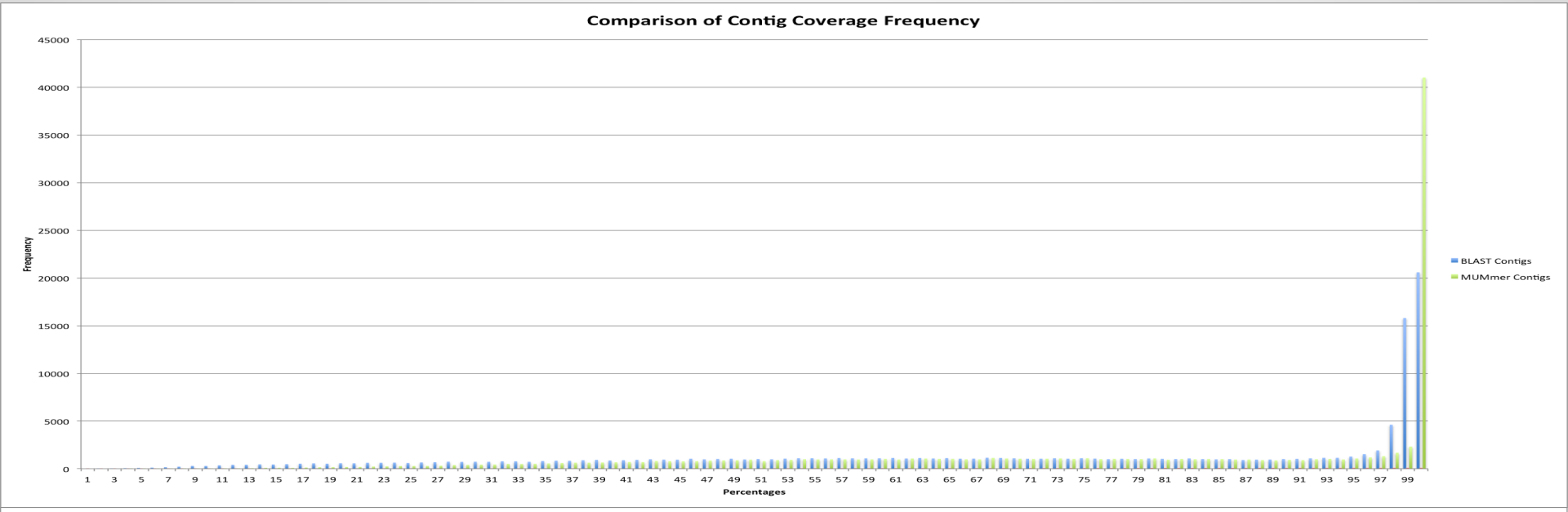
contig122151 : NODE_145039_length_491_cov_3.128309

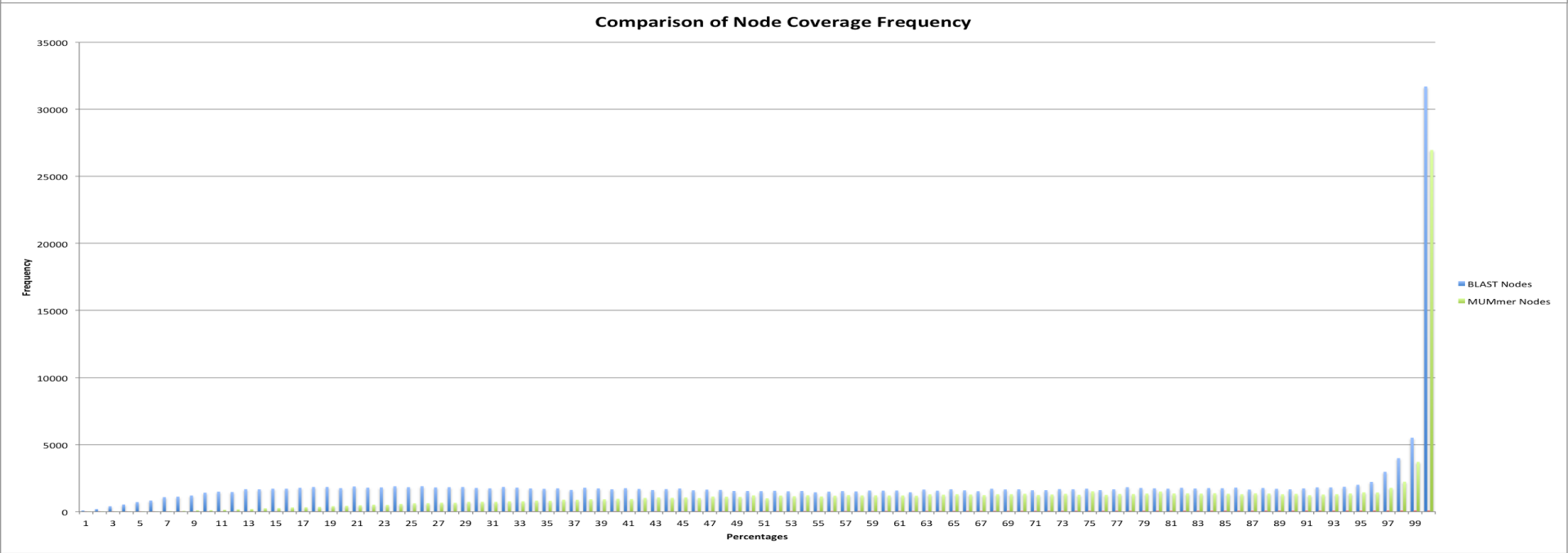
contig09789 : NODE_79919_length_122_cov_3.483607, NODE_204370_length_106_cov_2.075472, NODE_173926_length_330_cov_3.427273, NODE_97374_length_772_cov_4.497409, NODE_97374_length_772_cov_4.497409, NODE_97374_length_772_cov_4.497409

contig09788 : NODE_469389_length_516_cov_2.474806, NODE_259071_length_143_cov_2.237762, NODE_373623_length_457_cov_2.916849, NODE_194630_length_141_cov_2.801419, NODE_138945_length_196_cov_2.464286, NODE_138945_length_196_cov_2.464286, NODE_138945_length_196_cov_2.464286

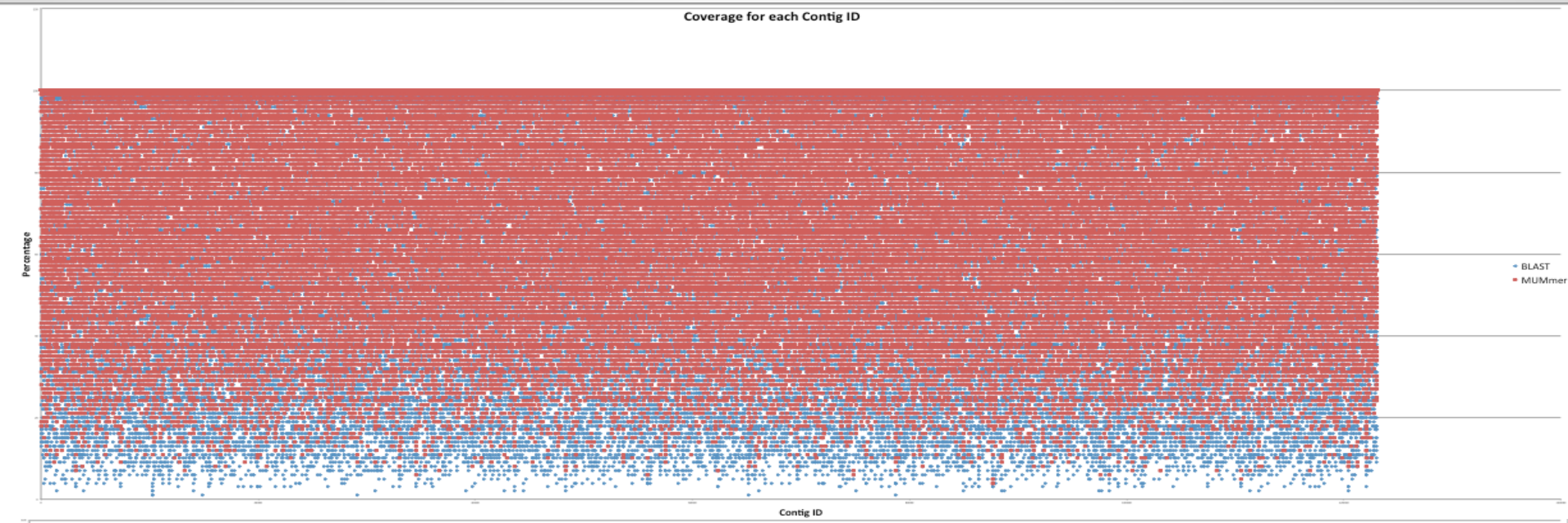
contig68189 : NODE_120143_length_513_cov_4.403509

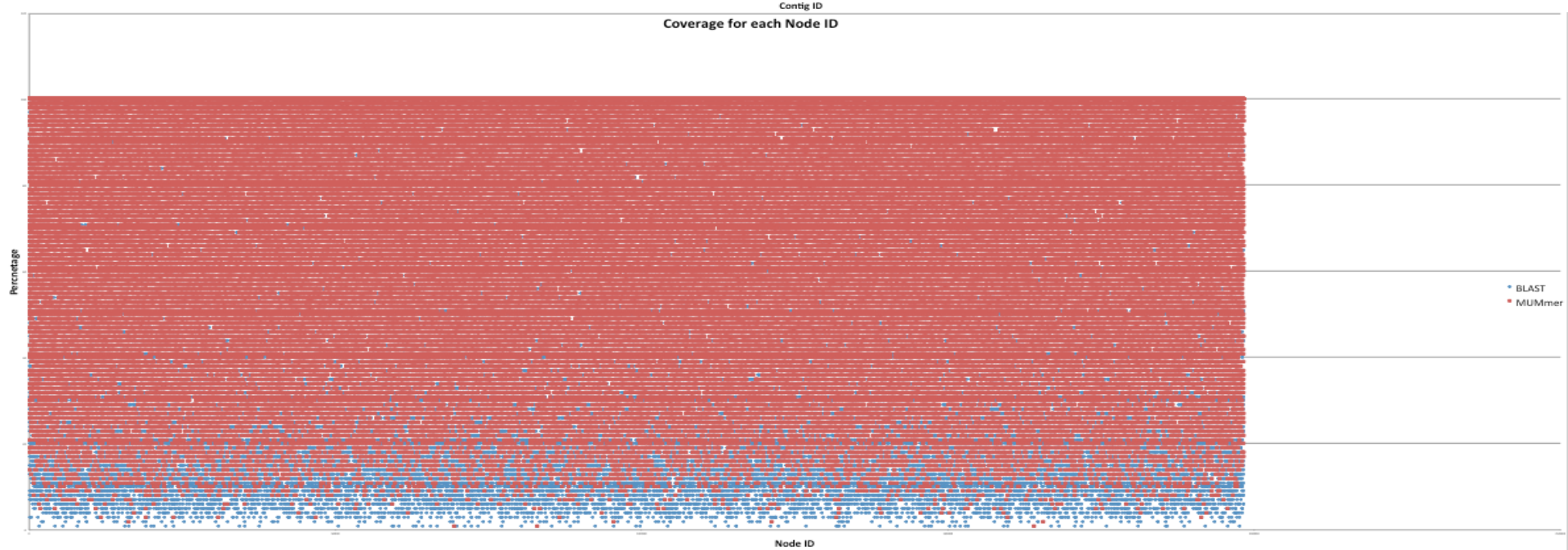
Results/Comparisons



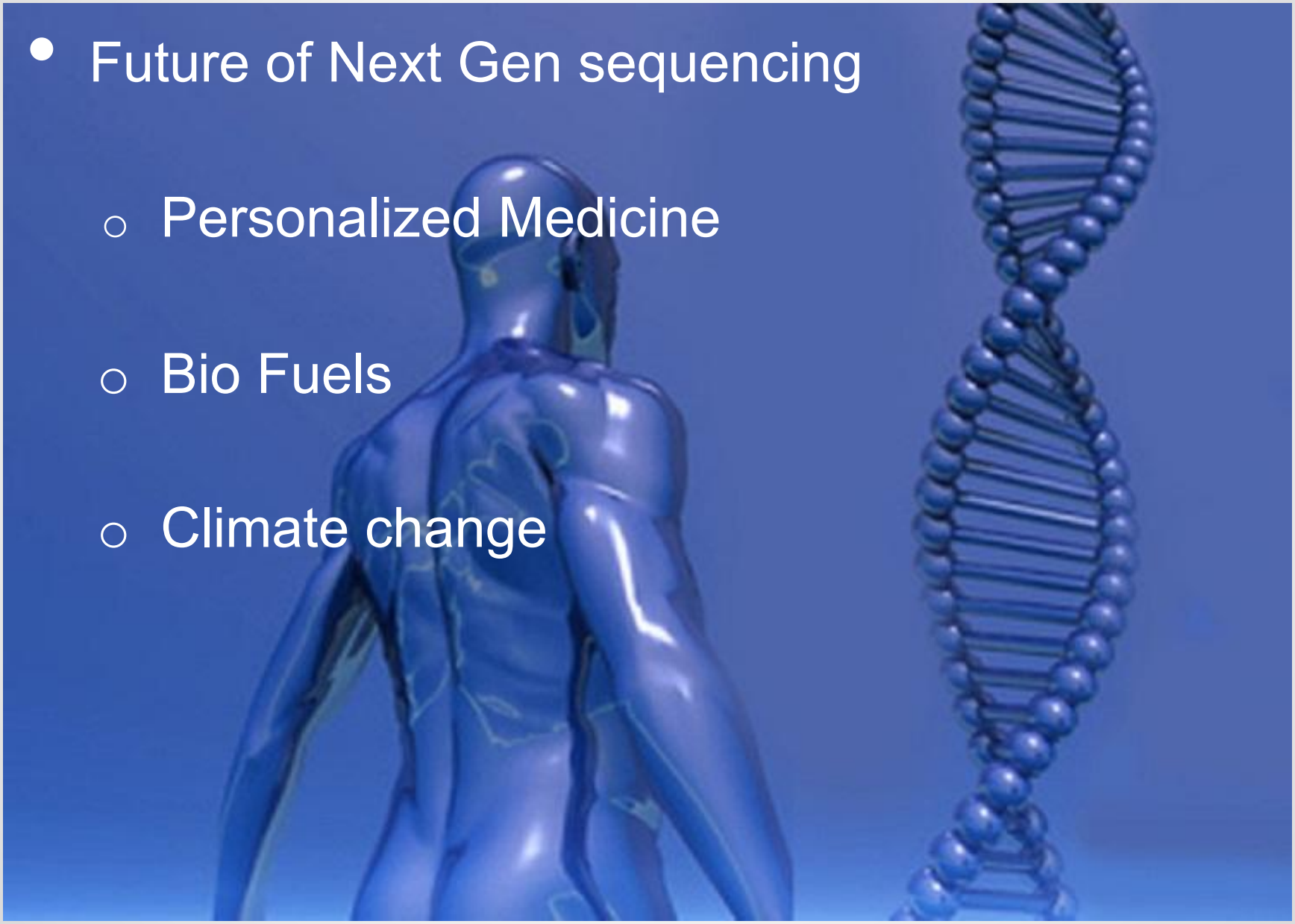


Results/Comparisons





Conclusion



- Future of Next Gen sequencing
  - Personalized Medicine
  - Bio Fuels
  - Climate change

References

- Zhang W, Chen J, Yang Y, Tang Y, et al. (2011) A Practical Comparison *of De Novo* Genome Assembly Software Tools for Next-Generation Sequencing Technologies. PLoS ONE 6(3): e17915. doi:10.1371/journal.pone.0017915
- Goecks J, Nekrutenko A, Taylor J, and The Galaxy Team, Galaxy: a comprehensive for supporting accessible, reproducible, and a transparent computational research in the life sciences. *Genome Biol.* 2012 Aug 25;11(8):R86
- Misra N., Panda P. K., Parida B. K., Mishra B. K., Phylogenomic Study of Lipid Genes Involved in Microalgal Biofuel Production - Candidate Gene Mining and Metabolic Pathway Analyses, *Evolutionary Bioinformatics* 2012:8 545-564, doi: 10.4137/EBO.S10159
- http://blast.ncbi.nlm.nih.gov/Blast.cgi
- http://mummer.sourceforge.net/
- Gabe Rudy, (2010) *A Hitchhiker's Guide to Next-Generation Sequencing,* :1-9, Golden Helix
- John D. McPherson, (2009) *Next-Generation Gap,* 6:1-4, Nature Methods Supplement
- Michael L. Metzker, *(2010) Sequencing Technologies, - the next generation,* 11:1-5, Nature Reviews
- http://www.sanger.ac.uk/resources/software/artemis/
- Rutherford K, Parkhill J, Crook J, et al., *Bioinformatics (Oxford, England)* 2000;16;10;944-5
- http://environment.nationalgeographic.com/environment/global-warming/biofuel-profile/
- Radakovits R. et al. Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana. Nat. Commun. 3:686 doi: 10.1038/ncomms1688 (2012)

Thank you!

Acknowledgements:

- CSURE (Computational Science for Undergraduate Research Experiences)
- University of Tennessee

Questions?