

Julian Pierre<sup>1</sup>, Jordan Taylor<sup>2</sup>, Amit Upadhyay<sup>3</sup>, Bhanu Rekepalli<sup>3</sup>

1 Texas Southern University, 2 Austin Peay State University, 3 University of Tennessee

## ABSTRACT

DNA sequencing is one of the most important platforms for study in biological systems today. The High-Throughput-next generation sequencing technologies delivers fast, inexpensive, and accurate genome information. Next Gen Sequencing can produce over 100 times more data than methods based on Sanger Sequencing. The next gen sequencing technologies offered from Illumina /Solexa, ABI/SOLID, 454/Roche, and Helicos has provided unprecedented opportunity for high-throughput functional genomic research. Next gen sequence technology offer novel and rapid ways for genome-wide characterization and profiling of mRNAs, transcription factor regions, and DNA patterns.

## Next Generation Sequencing

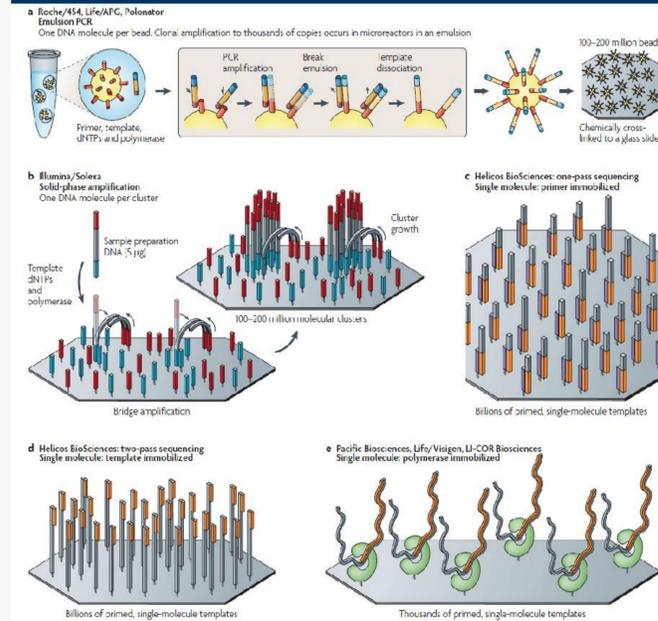


Fig. 1) This is a figure comparing two different methods of Next Generation Sequencing. [2]

- 454/Roche – 454 Life Sciences is a Biotechnology company that is a part of Roche and based in Branford, Connecticut. The center develops ultra-fast high-throughput DNA sequencing methods and tools.
- Illumina/Solexa– Illumina is a company that develops and manufactures integrated systems for the analysis of gene variation. Solexa was founded to develop genome sequencing technology.
- ABI/SOLID - (Sequencing by Oligonucleotide Ligation and Detection) is a next-generation DNA sequencing technology developed by Life Technologies and has been commercially available since 2006. This next generation technology generates hundreds of millions to billions of small sequence reads at one time.
- Helicos - Helicos's technology images the extension of individual DNA molecules using a defined primer and individual fluorescently labeled nucleotides, which contain a "virtual terminator" preventing incorporation of multiple nucleotides per cycle.

## Data Analysis

In Next-Generation Sequencing, data analysis is one of the most expensive processes. While the cost of genome sequencing goes down, the cost of analyzing data is still expensive. In the future, the "\$1,000 genome will come with a \$20,000 analysis price tag."

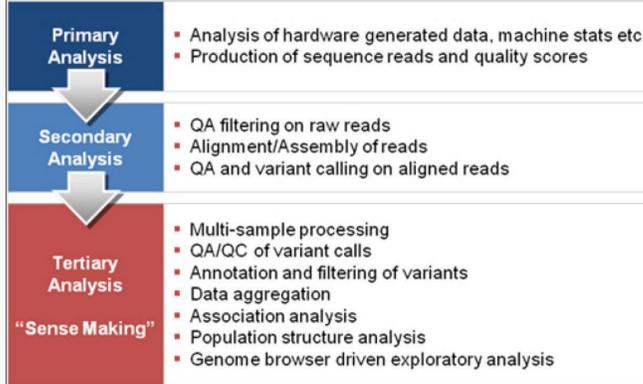


Fig. 2) Taken from A Hitchhiker's Guide to Next-Generation Sequencing, by Gabe Rudy

Sequence Analysis refers to the process of subjecting a DNA, RNA or peptide sequence to a wide range of analytical methods to:

- Compare sequences to find similarities and infer if they are Homologous
- To identify the features of the sequence such as gene structure, distribution, introns and exons, and regulation of gene expression
- Identify Sequence differences and variations such as mutations

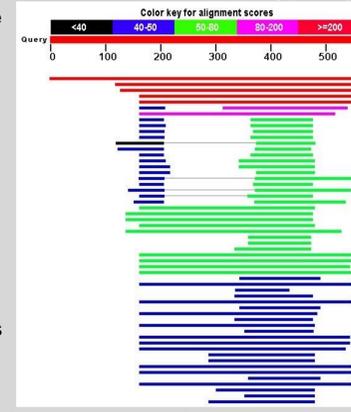


Fig. 3 Taken from bio.davidson.edu/courses. Shows alignment results for yeast.

## Genome Assembly

Galaxy is an open, web-based platform for data intensive biomedical research. It can be used on its own free public server where you can perform, reproduce, and share complete analyses.

An example of how Galaxy reflects its data is shown in Fig 5.

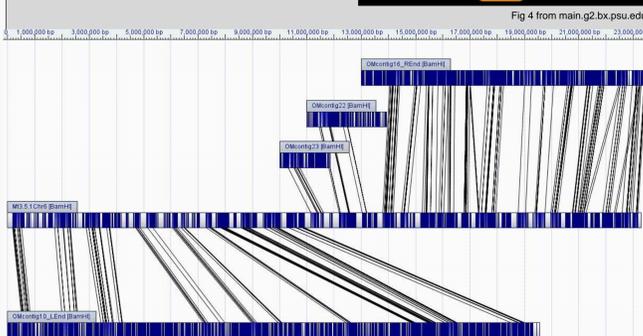


Fig 5 from jvci.org shows the mapping of chr6 of a Human Genome

## Downstream Analysis

Next Gen Sequencing uses a wide array of tools to obtain results based on the genome sequence. The most widely used Tools are BLAST, HMMER, and MUMMER.

- BLAST (Basic Local Alignment Search Tool) is a multi-sequence alignment tool developed by NIH (National Institute of Health). It is used find similar regions in different sequences and then compare their similarities.
- MUMmer is a rapid alignment system used for rapidly aligning entire genomes. It can also align incomplete genomes and can easily handle 1000's of contigs from a shotgun sequencing project.
- HMMER is used for searching sequence databases for homologs of protein sequences, and for making protein sequence alignments. It implements methods using probabilistic models called profile hidden Markov models (HMMs)

This is a small sample from our contig database in a FASTA format (.fna). This file contains all of the contig IDs, the length of the contig sequence, the number of reads, and the sequence itself.

## Parsed Results (BLAST)

```
contig0001: NODE_12272, length_271, cov_3.26203, NODE_16584, length_310, cov_2.91206,
NODE_17485, length_351, cov_3.25999
contig0004: NODE_14622, length_16, cov_2.110448
contig0007: NODE_19461, length_30, cov_3.18708
contig0008: NODE_19670, length_40, cov_4.22584
contig0011: NODE_19110, length_393, cov_4.85254, NODE_20220, length_368, cov_5.717315
contig0016: NODE_28920, length_343, cov_4.56256
contig0027: NODE_4678, length_1029, cov_3.18587
contig0061: NODE_28910, length_422, cov_2.18247, NODE_56103, length_158, cov_2.32277
contig0069: NODE_17049, length_35, cov_4.01684
contig0080: NODE_14005, length_342, cov_5.52646
contig0084: NODE_15275, length_348, cov_3.54885, NODE_17724, length_772, cov_4.47403
contig0087: NODE_15740, length_10, cov_2.017472, NODE_36411, length_70, cov_4.011761, NODE_34412, length_34, cov_2.102458
contig0088: NODE_13846, length_198, cov_2.44248, NODE_24884, length_181, cov_2.81487
contig0093: NODE_14622, length_16, cov_2.110448
contig0094: NODE_14622, length_16, cov_2.110448
contig0095: NODE_14622, length_16, cov_2.110448
contig0096: NODE_14622, length_16, cov_2.110448
contig0097: NODE_14622, length_16, cov_2.110448
contig0098: NODE_14622, length_16, cov_2.110448
contig0099: NODE_14622, length_16, cov_2.110448
contig0100: NODE_14622, length_16, cov_2.110448
contig0101: NODE_14622, length_16, cov_2.110448
contig0102: NODE_14622, length_16, cov_2.110448
contig0103: NODE_14622, length_16, cov_2.110448
contig0104: NODE_14622, length_16, cov_2.110448
contig0105: NODE_14622, length_16, cov_2.110448
contig0106: NODE_14622, length_16, cov_2.110448
contig0107: NODE_14622, length_16, cov_2.110448
contig0108: NODE_14622, length_16, cov_2.110448
contig0109: NODE_14622, length_16, cov_2.110448
contig0110: NODE_14622, length_16, cov_2.110448
contig0111: NODE_14622, length_16, cov_2.110448
contig0112: NODE_14622, length_16, cov_2.110448
contig0113: NODE_14622, length_16, cov_2.110448
contig0114: NODE_14622, length_16, cov_2.110448
contig0115: NODE_14622, length_16, cov_2.110448
contig0116: NODE_14622, length_16, cov_2.110448
contig0117: NODE_14622, length_16, cov_2.110448
contig0118: NODE_14622, length_16, cov_2.110448
contig0119: NODE_14622, length_16, cov_2.110448
contig0120: NODE_14622, length_16, cov_2.110448
contig0121: NODE_14622, length_16, cov_2.110448
contig0122: NODE_14622, length_16, cov_2.110448
contig0123: NODE_14622, length_16, cov_2.110448
contig0124: NODE_14622, length_16, cov_2.110448
contig0125: NODE_14622, length_16, cov_2.110448
contig0126: NODE_14622, length_16, cov_2.110448
contig0127: NODE_14622, length_16, cov_2.110448
contig0128: NODE_14622, length_16, cov_2.110448
contig0129: NODE_14622, length_16, cov_2.110448
contig0130: NODE_14622, length_16, cov_2.110448
contig0131: NODE_14622, length_16, cov_2.110448
contig0132: NODE_14622, length_16, cov_2.110448
contig0133: NODE_14622, length_16, cov_2.110448
contig0134: NODE_14622, length_16, cov_2.110448
contig0135: NODE_14622, length_16, cov_2.110448
contig0136: NODE_14622, length_16, cov_2.110448
contig0137: NODE_14622, length_16, cov_2.110448
contig0138: NODE_14622, length_16, cov_2.110448
contig0139: NODE_14622, length_16, cov_2.110448
contig0140: NODE_14622, length_16, cov_2.110448
contig0141: NODE_14622, length_16, cov_2.110448
contig0142: NODE_14622, length_16, cov_2.110448
contig0143: NODE_14622, length_16, cov_2.110448
contig0144: NODE_14622, length_16, cov_2.110448
contig0145: NODE_14622, length_16, cov_2.110448
contig0146: NODE_14622, length_16, cov_2.110448
contig0147: NODE_14622, length_16, cov_2.110448
contig0148: NODE_14622, length_16, cov_2.110448
contig0149: NODE_14622, length_16, cov_2.110448
contig0150: NODE_14622, length_16, cov_2.110448
contig0151: NODE_14622, length_16, cov_2.110448
contig0152: NODE_14622, length_16, cov_2.110448
contig0153: NODE_14622, length_16, cov_2.110448
contig0154: NODE_14622, length_16, cov_2.110448
contig0155: NODE_14622, length_16, cov_2.110448
contig0156: NODE_14622, length_16, cov_2.110448
contig0157: NODE_14622, length_16, cov_2.110448
contig0158: NODE_14622, length_16, cov_2.110448
contig0159: NODE_14622, length_16, cov_2.110448
contig0160: NODE_14622, length_16, cov_2.110448
contig0161: NODE_14622, length_16, cov_2.110448
contig0162: NODE_14622, length_16, cov_2.110448
contig0163: NODE_14622, length_16, cov_2.110448
contig0164: NODE_14622, length_16, cov_2.110448
contig0165: NODE_14622, length_16, cov_2.110448
contig0166: NODE_14622, length_16, cov_2.110448
contig0167: NODE_14622, length_16, cov_2.110448
contig0168: NODE_14622, length_16, cov_2.110448
contig0169: NODE_14622, length_16, cov_2.110448
contig0170: NODE_14622, length_16, cov_2.110448
contig0171: NODE_14622, length_16, cov_2.110448
contig0172: NODE_14622, length_16, cov_2.110448
contig0173: NODE_14622, length_16, cov_2.110448
contig0174: NODE_14622, length_16, cov_2.110448
contig0175: NODE_14622, length_16, cov_2.110448
contig0176: NODE_14622, length_16, cov_2.110448
contig0177: NODE_14622, length_16, cov_2.110448
contig0178: NODE_14622, length_16, cov_2.110448
contig0179: NODE_14622, length_16, cov_2.110448
contig0180: NODE_14622, length_16, cov_2.110448
contig0181: NODE_14622, length_16, cov_2.110448
contig0182: NODE_14622, length_16, cov_2.110448
contig0183: NODE_14622, length_16, cov_2.110448
contig0184: NODE_14622, length_16, cov_2.110448
contig0185: NODE_14622, length_16, cov_2.110448
contig0186: NODE_14622, length_16, cov_2.110448
contig0187: NODE_14622, length_16, cov_2.110448
contig0188: NODE_14622, length_16, cov_2.110448
contig0189: NODE_14622, length_16, cov_2.110448
contig0190: NODE_14622, length_16, cov_2.110448
contig0191: NODE_14622, length_16, cov_2.110448
contig0192: NODE_14622, length_16, cov_2.110448
contig0193: NODE_14622, length_16, cov_2.110448
contig0194: NODE_14622, length_16, cov_2.110448
contig0195: NODE_14622, length_16, cov_2.110448
contig0196: NODE_14622, length_16, cov_2.110448
contig0197: NODE_14622, length_16, cov_2.110448
contig0198: NODE_14622, length_16, cov_2.110448
contig0199: NODE_14622, length_16, cov_2.110448
contig0200: NODE_14622, length_16, cov_2.110448
contig0201: NODE_14622, length_16, cov_2.110448
contig0202: NODE_14622, length_16, cov_2.110448
contig0203: NODE_14622, length_16, cov_2.110448
contig0204: NODE_14622, length_16, cov_2.110448
contig0205: NODE_14622, length_16, cov_2.110448
contig0206: NODE_14622, length_16, cov_2.110448
contig0207: NODE_14622, length_16, cov_2.110448
contig0208: NODE_14622, length_16, cov_2.110448
contig0209: NODE_14622, length_16, cov_2.110448
contig0210: NODE_14622, length_16, cov_2.110448
contig0211: NODE_14622, length_16, cov_2.110448
contig0212: NODE_14622, length_16, cov_2.110448
contig0213: NODE_14622, length_16, cov_2.110448
contig0214: NODE_14622, length_16, cov_2.110448
contig0215: NODE_14622, length_16, cov_2.110448
contig0216: NODE_14622, length_16, cov_2.110448
contig0217: NODE_14622, length_16, cov_2.110448
contig0218: NODE_14622, length_16, cov_2.110448
contig0219: NODE_14622, length_16, cov_2.110448
contig0220: NODE_14622, length_16, cov_2.110448
contig0221: NODE_14622, length_16, cov_2.110448
contig0222: NODE_14622, length_16, cov_2.110448
contig0223: NODE_14622, length_16, cov_2.110448
contig0224: NODE_14622, length_16, cov_2.110448
contig0225: NODE_14622, length_16, cov_2.110448
contig0226: NODE_14622, length_16, cov_2.110448
contig0227: NODE_14622, length_16, cov_2.110448
contig0228: NODE_14622, length_16, cov_2.110448
contig0229: NODE_14622, length_16, cov_2.110448
contig0230: NODE_14622, length_16, cov_2.110448
contig0231: NODE_14622, length_16, cov_2.110448
contig0232: NODE_14622, length_16, cov_2.110448
contig0233: NODE_14622, length_16, cov_2.110448
contig0234: NODE_14622, length_16, cov_2.110448
contig0235: NODE_14622, length_16, cov_2.110448
contig0236: NODE_14622, length_16, cov_2.110448
contig0237: NODE_14622, length_16, cov_2.110448
contig0238: NODE_14622, length_16, cov_2.110448
contig0239: NODE_14622, length_16, cov_2.110448
contig0240: NODE_14622, length_16, cov_2.110448
contig0241: NODE_14622, length_16, cov_2.110448
contig0242: NODE_14622, length_16, cov_2.110448
contig0243: NODE_14622, length_16, cov_2.110448
contig0244: NODE_14622, length_16, cov_2.110448
contig0245: NODE_14622, length_16, cov_2.110448
contig0246: NODE_14622, length_16, cov_2.110448
contig0247: NODE_14622, length_16, cov_2.110448
contig0248: NODE_14622, length_16, cov_2.110448
contig0249: NODE_14622, length_16, cov_2.110448
contig0250: NODE_14622, length_16, cov_2.110448
contig0251: NODE_14622, length_16, cov_2.110448
contig0252: NODE_14622, length_16, cov_2.110448
contig0253: NODE_14622, length_16, cov_2.110448
contig0254: NODE_14622, length_16, cov_2.110448
contig0255: NODE_14622, length_16, cov_2.110448
contig0256: NODE_14622, length_16, cov_2.110448
contig0257: NODE_14622, length_16, cov_2.110448
contig0258: NODE_14622, length_16, cov_2.110448
contig0259: NODE_14622, length_16, cov_2.110448
contig0260: NODE_14622, length_16, cov_2.110448
contig0261: NODE_14622, length_16, cov_2.110448
contig0262: NODE_14622, length_16, cov_2.110448
contig0263: NODE_14622, length_16, cov_2.110448
contig0264: NODE_14622, length_16, cov_2.110448
contig0265: NODE_14622, length_16, cov_2.110448
contig0266: NODE_14622, length_16, cov_2.110448
contig0267: NODE_14622, length_16, cov_2.110448
contig0268: NODE_14622, length_16, cov_2.110448
contig0269: NODE_14622, length_16, cov_2.110448
contig0270: NODE_14622, length_16, cov_2.110448
contig0271: NODE_14622, length_16, cov_2.110448
contig0272: NODE_14622, length_16, cov_2.110448
contig0273: NODE_14622, length_16, cov_2.110448
contig0274: NODE_14622, length_16, cov_2.110448
contig0275: NODE_14622, length_16, cov_2.110448
contig0276: NODE_14622, length_16, cov_2.110448
contig0277: NODE_14622, length_16, cov_2.110448
contig0278: NODE_14622, length_16, cov_2.110448
contig0279: NODE_14622, length_16, cov_2.110448
contig0280: NODE_14622, length_16, cov_2.110448
contig0281: NODE_14622, length_16, cov_2.110448
contig0282: NODE_14622, length_16, cov_2.110448
contig0283: NODE_14622, length_16, cov_2.110448
contig0284: NODE_14622, length_16, cov_2.110448
contig0285: NODE_14622, length_16, cov_2.110448
contig0286: NODE_14622, length_16, cov_2.110448
contig0287: NODE_14622, length_16, cov_2.110448
contig0288: NODE_14622, length_16, cov_2.110448
contig0289: NODE_14622, length_16, cov_2.110448
contig0290: NODE_14622, length_16, cov_2.110448
contig0291: NODE_14622, length_16, cov_2.110448
contig0292: NODE_14622, length_16, cov_2.110448
contig0293: NODE_14622, length_16, cov_2.110448
contig0294: NODE_14622, length_16, cov_2.110448
contig0295: NODE_14622, length_16, cov_2.110448
contig0296: NODE_14622, length_16, cov_2.110448
contig0297: NODE_14622, length_16, cov_2.110448
contig0298: NODE_14622, length_16, cov_2.110448
contig0299: NODE_14622, length_16, cov_2.110448
contig0300: NODE_14622, length_16, cov_2.110448
contig0301: NODE_14622, length_16, cov_2.110448
contig0302: NODE_14622, length_16, cov_2.110448
contig0303: NODE_14622, length_16, cov_2.110448
contig0304: NODE_14622, length_16, cov_2.110448
contig0305: NODE_14622, length_16, cov_2.110448
contig0306: NODE_14622, length_16, cov_2.110448
contig0307: NODE_14622, length_16, cov_2.110448
contig0308: NODE_14622, length_16, cov_2.110448
contig0309: NODE_14622, length_16, cov_2.110448
contig0310: NODE_14622, length_16, cov_2.110448
contig0311: NODE_14622, length_16, cov_2.110448
contig0312: NODE_14622, length_16, cov_2.110448
contig0313: NODE_14622, length_16, cov_2.110448
contig0314: NODE_14622, length_16, cov_2.110448
contig0315: NODE_14622, length_16, cov_2.110448
contig0316: NODE_14622, length_16, cov_2.110448
contig0317: NODE_14622, length_16, cov_2.110448
contig0318: NODE_14622, length_16, cov_2.110448
contig0319: NODE_14622, length_16, cov_2.110448
contig0320: NODE_14622, length_16, cov_2.110448
contig0321: NODE_14622, length_16, cov_2.110448
contig0322: NODE_14622, length_16, cov_2.110448
contig0323: NODE_14622, length_16, cov_2.110448
contig0324: NODE_14622, length_16, cov_2.110448
contig0325: NODE_14622, length_16, cov_2.110448
contig0326: NODE_14622, length_16, cov_2.110448
contig0327: NODE_14622, length_16, cov_2.110448
contig0328: NODE_14622, length_16, cov_2.110448
contig0329: NODE_14622, length_16, cov_2.110448
contig0330: NODE_14622, length_16, cov_2.110448
contig0331: NODE_14622, length_16, cov_2.110448
contig0332: NODE_14622, length_16, cov_2.110448
contig0333: NODE_14622, length_16, cov_2.110448
contig0334: NODE_14622, length_16, cov_2.110448
contig0335: NODE_14622, length_16, cov_2.110448
contig0336: NODE_14622, length_16, cov_2.110448
contig0337: NODE_14622, length_16, cov_2.110448
contig0338: NODE_14622, length_16, cov_2.110448
contig0339: NODE_14622, length_16, cov_2.110448
contig0340: NODE_14622, length_16, cov_2.110448
contig0341: NODE_14622, length_16, cov_2.110448
contig0342: NODE_14622, length_16, cov_2.110448
contig0343: NODE_14622, length_16, cov_2.110448
contig0344: NODE_14622, length_16, cov_2.110448
contig0345: NODE_14622, length_16, cov_2.110448
contig0346: NODE_14622, length_16, cov_2.110448
contig0347: NODE_14622, length_16, cov_2.110448
contig0348: NODE_14622, length_16, cov_2.110448
contig0349: NODE_14622, length_16, cov_2.110448
contig0350: NODE_14622, length_16, cov_2.110448
contig0351: NODE_14622, length_16, cov_2.110448
contig0352: NODE_14622, length_16, cov_2.110448
contig0353: NODE_14622, length_16, cov_2.110448
contig0354: NODE_14622, length_16, cov_2.110448
contig0355: NODE_14622, length_16, cov_2.110448
contig0356: NODE_14622, length_16, cov_2.110448
contig0357: NODE_14622, length_16, cov_2.110448
contig0358: NODE_14622, length_16, cov_2.110448
contig0359: NODE_14622, length_16, cov_2.110448
contig0360: NODE_14622, length_16, cov_2.110448
contig0361: NODE_14622, length_16, cov_2.110448
contig0362: NODE_14622, length_16, cov_2.110448
contig0363: NODE_14622, length_16, cov_2.110448
contig0364: NODE_14622, length_16, cov_2.110448
contig0365: NODE_14622, length_16, cov_2.110448
contig0366: NODE_14622, length_16, cov_2.110448
contig0367: NODE_14622, length_16, cov_2.110448
contig0368: NODE_14622, length_16, cov_2.110448
contig0369: NODE_14622, length_16, cov_2.110448
contig0370: NODE_14622, length_16, cov_2.110448
contig0371: NODE_14622, length_16, cov_2.110448
contig0372: NODE_14622, length_16, cov_2.110448
contig0373: NODE_14622, length_16, cov_2.110448
contig0374: NODE_14622, length_16, cov_2.110448
contig0375: NODE_14622, length_16, cov_2.110448
contig0376: NODE_14622, length_16, cov_2.110448
contig0377: NODE_14622, length_16, cov_2.110448
contig0378: NODE_14622, length_16, cov_2.110448
contig0379: NODE_14622, length_16, cov_2.110448
contig0380: NODE_14622, length_16, cov_2.110448
contig0381: NODE_14622, length_16, cov_2.110448
contig0382: NODE_14622, length_16, cov_2.110448
contig0383: NODE_14622, length_16, cov_2.110448
contig0384: NODE_14622, length_16, cov_2.110448
contig0385: NODE_14622, length_16, cov_2.110448
contig0386: NODE_14622, length_16, cov_2.110448
contig0387: NODE_14622, length_16, cov_2.
```