Introduction to Bioinformatics using the eBioKit Platform

Commission for Biotechnology & Genetic Engineering, National

Center for Research, December 1 - 6, 2014

Assembly with velvet

Andreas Gisel, IITA – Ibadan, Nigeria

Reads:	reads1.fq reads2.fq readpairs.fq
Ref	ref1.fa
Scripts:	n50.pl fastalen.pl

1. in your home create a directory ‘results’ (mkdir results)
2. copy the data above into the directory ‘results (cp path results)
3. enter into the directory results (cd results)
4. check the content of the current directory (ls -l)

Now we can start with the assembly:
Write a short protocol of the results!

A) Velvet assembly using single reads

1. we test the results of different k-mer length (19, 21, 23, 25, 27)
2. cmd: velveth kmerSXX XX –fastq –short reads1.fq reads2.fq
3. cmd velvetg kmerSXX -exp_cov 15 -min_contig_lgth 1000 -unused_reads yes

4. with the perl script we calculate the contig lengs
5. cmd: perl fastalen.pl kmerSXX/contigs.fa > kmerSXX_len.txt
6. with the perl script we calculate the N50
7. cmd: perl n50.pl kmerSXX_len.txt
8. what is the conclusion????

B) Velvet assembly using paired-end reads

1. We run velvet with the optimal kmer from the previous test
2. cmd: velveth kmerPXX XX -fastq –short Paired readpairs.fq
3. cmd: velvetg kmerPXX -exp_cov 15 -ins_length 250 -min_contig_lgth 1000 -unused_reads yes
4. with the perl script we calculate the contig lengs
5. cmd: perl fastalen.pl kmerPXX/contigs.fa > kmerPXX_len.txt
6. with the perl script we calculate the N50
7. cmd: perl n50.pl kmerPXX_len.txt
8. what is the next conclusion????

C) Clean up the account

1. rm –r kmerXX

kmer23 single
503 sequences, sum_length: 4461101
min_length: 1023, max_length: 60572, N50: 13984

kmer23 paired
109 sequences, sum_length: 4527173
min_length: 1036, max_length: 174071, N50: 87045