

Algorithms in Genome Research  
Winter 2013/2014

Exercises

Number 5, Discussion: 2013 December 6

1. **Reads.** Enumerate advantages and defaults of single reads versus paired-end reads.
2. **Sequencing approaches.** Explain and compare Whole Genome Shotgun (WGS) sequencing and hierarchical shotgun sequencing.
3. **N50 length.** Compute the N50 length of 3 genomes respectively composed by contigs of the following lengths:
  - (a) 450k, 300k, 270k, 200k, 180k, 170k, 115k, 100k, 90k, 75k, 50k.
  - (b) 200k, 190k, 180k, . . . , 30k, 20k, 10k.
  - (c) 500k, 500k, 400k, 400k, 100k, 50k, 25k, 10k, 5k.
4. **Lander-Waterman statistics.** Let  $N$  be the number of reads,  $G$  be the genome length,  $L$  be the sequence read length and  $a$  be the coverage. What is the number of expected gaps? The number of reads per contig? The expected size of sequenced region? The mean contig size?
5. **Overlap/layout/consensus approach.** From an overlap graph, propose an algorithm to arrange all reads into an approximate multi-alignment (layout step).