November 1, 2016

Exam in Biomedical Informatics

(lecturer Alexander Gultyaev, exam is also reviewed by lecturer René Olsthoorn)

Questions

All answers require some (short) argumentation. For each of the questions, the maximum number of points contributing to the exam grade is given. The exam grade contributes 80% to the final grade, 20% are contributed by computer assignments made during the course.

1 (0.5). What kind of genome assemblies yield larger sequence contigs: those based on next generation sequencing or those using Sanger sequencing data ?

2 (1.0). What are the advantages and deficiencies of assisted genome assembly algorithms that use alignments of sequencing reads to another genome that is related to the target one?

3 (1.0). Below the sequences near exon-intron and intron-exon junctions of a gene are given, accompanied by the annotation of mRNA sequence:

```
mRNA    join(1..418,8040..8110,37314..37542,39528..39584, 42541..46669)
    37501 atgacgtttt ccaccaccaa ggaaacagcc acctagccaa cggtattttg aaagcgtttg
    37561 tctggagtta gaaagttctc ...
                      (this part is removed for simplicity)
    39481 gagaacaccc gatcgcagag ggtaattttc tggagtttgt tttgcaggga tagctgggag
    39541 tatggccacc ctgctccacg ...
```

RNA-Seq experiments yielded a transcriptome assembly with the following fragments, apparently mapped to these genome regions.

```
(1) ggaaacagccacctagccaacggtattttgaaagcgtttgtctggagtta
(2) ggaaacagccacctagccaacgggatagctgggagtatggccaccctgctccacg
(3) ttttctggagtttgttttgcagggatagctgggagtatggccaccctgctccacg
```

What can be concluded from this information ? What kind of data is also possible to obtain from RNA-Seq in order to quantify transcription of these sequences ?

4 (0.5). What is the main strategy of collinear alignments of large genomic sequences ? Describe briefly the main steps.

5 (1.0). Studies on genes underlying Mendelian disorders (diseases determined by mutation in single genes) usually require sequencing exomes or genomes of both relatives and people that are unrelated to each other. What is the reason behind such strategy? Try to describe briefly its main principles, and how the databases on human variation (e.g. 1000 Genomes project) can be used.

6 (1.0). What is the main difference between amplicon and shotgun sequencing in metagenomics ? How can these approaches improve each other (name just main principles) ?

7 (1.0). In so-called Monte Carlo algorithms for protein folding the calculation is based on stochastic simulation with a goal to identify low free energy conformations. At every iteration a random change is introduced into the previously folded conformation. The new structure is always accepted if its energy is lower than that of the previous one, and with certain probability it can also be accepted even when its energy is higher. What would happen with the algorithm behavior if the generated structures with higher energies would be never accepted ?

8 (1.0). What are the computational strategies to improve predictions of RNA secondary structure using experimental structure probing data ?

9 (1.0). Assume you need to design an RNA fragment that would fold into the stem-loop structure shown below in "bracket view":

```
<<<<....<<<<.....>>>>....>>>>
```

Which of the following sequences is the most suitable for this design ?

```
<<<<....<<<<.....>>>>....>>>>
CCCCAAAAGGGGAAAAAGGGGAAAACCCC (a)
GGGGAAAACCCCAAAAAGGGGAAAACCCC (b)
CCCCAAAAGGGGAAAAACCCCAAAAGGGG (c)
CGCGAAAAGGGGAAAAACCCCAAAAGCGC (d)
CCCCAAAACGCGAAAAACGCGAAAAGGGG (e)
```

10 (2.0). In a cubic lattice HP model for 3D protein structure, the polypeptide chain is considered to contain two types of monomers: hydrophobic (H) and polar ones (P). The folding free energy is calculated as E = - h, where *h* is the number of H-H contacts in the lattice between hydrophobic monomers that are not adjacent to each other in the polymer chain.
(a) How many different 3D conformations with the lowest free energy of cubic lattice folding are possible in the following fragment of 7 monomers: HPHHPHP ?
(b) If the modeling would require taking structure compactness into account, how many different lowest free energy 3D conformations of this fragment could be located within one cubic unit of the lattice ?