Introduction to bioinformatics 2015

A

A1.
What are the two branches of computer uses in biology? Give a (detailed) narrow definition and a broad definition for bioinformatics.

A2.
Give a generalized description of structures as entities and relationships - examples for entities and relationships in various systems!

A3.
Describe the core data-types used in bioinformatics! Name the underlying models, the principal form of description and give examples of extended and simplified descriptions (annotations)!

A4.
Give a short overview of data representation types (structured, unstructured, mixed), and what is granularity!

A5.
What are the substitution matrices? Describe PAM and Blosum matrices!

A6.
What types of alignments exist according to what we compare with what? Give examples for the usage of each method!

A7.
Describe pairwise and multiple alignments! Give the relationship between pairwise and multiple alignment (difference in complexity and its consequences, construction of pairwise from multiple alignment and multiple from pairwise alignment)!

A8.
Protein groups are heterogeneous from a number of points of view. Name a few group properties that make protein classification difficult! Explain the role of within-group and between-group similarities!

A9.
Describe how similarity search can help to understand the protein universe. What are the class annotated databases (COG, SBASE), what are the non-annotated cluster databases, how are they created and how can they help to get to know an unknown protein.

A10.
What is the maximum likelihood algorithm and for which kind of data it should be used for? What is the Bayesian interference algorithm and for which kind of data it should be used for? How can you determine the root of a phylogenetic tree? Describe consensus tree generating methods and data re-sampling!

A11.
Describe the main steps of preparing DNA for NGS data analysis! Analysis of mutation (examples for medical applications briefly)

A12.
Describe the main difference between genomics and functional genomics and the main difference between microarray and NGS based genomics methods! What are RNA-seq (applications: genomic mapping, transcriptome mapping, de Novo assembly, gene expression analysis), ChIP-seq and GRO-seq techniques?

A13.
Give a short overview about the future of NGS (challenges, recommendations)!

Introduction to bioinformatics 2015

B

B1.
Describe data comparison - score, common alignment patterns!

B2.
Give a short overview of the statistics of alignments (e-value, p-value, z-score, linearization, Karlin-Altschul statistics)!

B3.
Explain the Dot Plot method! Write down the algorithm, describe the parameters and list some usages!

B4.
What is the difference between global and local alignment? Explain the main algorithms (Needleman&Wunsch and Smith&Waterman), and gap-handling!

B5.
What is the main idea of dynamic programming - example: multiple alignment of 3 sequences (additive approach)? Describe the principle of the ClustalW algorithm (iterative approach)!

B6.
Describe the Hidden Markov Model! (What is the connection between protein profiles and Hidden Markov Model?)

B7.
Why do we need databases in biology, what are the main tasks of databases? What is their technical and logical structure? What are the main data-types (give examples to them) and data description formats (syntax, semantics)?

B8.
What is similarity searching in a database, what are the main steps, the importance of it, what kind of heuristics do you know?

B9.
Describe BLAST algorithms in details! What kind of heuristics does BLAST use? How can the BLAST specificity be improved? List 5 different BLAST programs (query type)!

B10.
Describe the phylogenetic method Parsimony (character set, homology and homoplasy, character fit)! Define the exhausted tree generation!

B11.
Describe the distance based phylogenetics! List some nucleotide substitution models! What is the connection between them? Describe the Neighbor-joining algorithms at tree generation!

B12.
Explain the basics of next generation sequencing (definitions, file formats). Compare traditional sequencing and NGS approaches. Describe assembling strategies, hierarchical strategies for long reads, short read strategies (De Brujin graph), scaffolding, difficulties and used heuristics and Lander-Waterman assumptions!

B13.
Define short read alignment problem, alignment strategies (Blast, Burrows Wheeler aligners, spaced seed alignment) and name some software for short-read mapping!