

武汉大学生命科学学院  
2018—2019 学年度第一学期期末考试  
《生物信息学》(A) 试卷

专业\_\_\_\_\_ 学号\_\_\_\_\_ 姓名\_\_\_\_\_ 得分\_\_\_\_\_

一、填空：(14×1=14 分)

1. 计算机网络是计算机技术与\_\_\_\_\_技术相结合的产物，包括计算机软/硬件、网络系统结构以及\_\_\_\_\_等内容。
2. 数据库系统的体系结构有三个数据抽象级别：\_\_\_\_\_、\_\_\_\_\_和内模式，它们相互独立而形成两级数据独立性。
3. 序列相似性分析可以用来进行\_\_\_\_\_推测、\_\_\_\_\_推测和\_\_\_\_\_推测。
4. 目前，最重要的公共核苷酸序列数据库是\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_，它们每天进行数据交换，同时更新序列资料，所以无论是投递还是查询都基本等价。
5. 真核生物的基因预测十分复杂，主要原因是\_\_\_\_\_占比可能较低。现在一般的预测方法是基于基因结构分级建模的从头预测，另外就是借助\_\_\_\_\_水平的证据。
6. 蛋白质组表达分析主要涉及大规模蛋白质的分离和鉴定，其中分离方面的支撑技术是\_\_\_\_\_，鉴定方面的支撑技术是\_\_\_\_\_。

二、单项选择 (8×2=16 分)

1. ( ) 以下哪一个数据库是数据源库，而不是次级数据库：  
A. RefSeq B. PDB C. Interpro D. SCOP
2. ( ) UCSC 是著名的\_\_\_\_\_数据库。  
A. 基因信息 B. 基因组信息 C. 蛋白质信息 D. RNA 信息
3. ( ) 以下\_\_\_\_\_是一种 RNA 远程结构，不属于狭义的 RNA 二级结构。  
A. Bulge loop B. Kissing hairpin loop  
C. Internal loop D. Multibranched loop
4. ( ) ExPASy 是著名的\_\_\_\_\_服务器，现在其内容逐渐向整个生物信息学领域扩展。  
A. 基因信息 B. 基因组信息 C. 蛋白质信息 D. RNA 信息
5. ( ) 蛋白质同源建模要求的目标序列和模板序列的相似性的底限是\_\_\_\_\_。  
A. 50% B. 40% C. 35% D. 25%
6. ( ) Python 中，range(3,6)对应的序列是：  
A. 4,5 B. 3,4,5 C. 3,4,5,6 D. 4,5,6
7. ( ) Python 中，(1,'3',5,'u',4,'tt') 是一个：  
A. 数组 B. 列表 C. 元组 D. 字典
8. ( ) Python 中，b=[14, 23, 7, 43, 66]，那么 b[2:4]指的是：  
A. [23, 7] B. [23, 7, 43] C. [7, 43, 66] D. [7, 43]



二、判断 (正确--V; 错误--X, 10×2=20分)

1. ( ) 简约法的原理是试图用最少的改变 (次数) 来解释进化带来的差异。
2. ( ) 对于同样数目的分类单位, 可能构造的无根树的数量多于有根树。
3. ( ) 研究亲缘关系远的物种进化关系应该用进化速率快的序列。
4. ( ) KEGG 是著名的描述代谢通路的数据库。
5. ( ) 第二代测序的最大特点是单分子测序, 而且其读段明显长于第一代测序。
6. ( ) 算法效率的最重要特性主要体现在“问题规模增大时, 算法运行时间的增长速度”。
7. ( ) RNA 据其碱基配对情况而形成二级结构, 而其三级结构是指其三维空间结构。
8. ( ) FASTA 格式是描述蛋白质结构信息的著名格式, 很多分析软件都能识别这种格式。
9. ( ) RNA-seq 分析的靶标并不局限于 mRNA, 是真正的、全面的转录组分析技术。
10. ( ) Perl 的座右铭是“*There is more than one way to do it*”, 其编程灵活, 但难以胜任大型程序。

四、英译汉: (6分)

Bioinformatics is the combination of biology and information technology. It is the branch of science that deals with the computer-based analysis of large biological data sets. Bioinformatics incorporates the development of databases to store and search data, and of statistical tools and algorithms to analyze and determine relationships between biological data sets, such as macromolecular sequences, structures, expression profiles, biochemical pathways and evolution history.

五、问答: (44分)

1. 请从序列、结构和功能这三个水平解释蛋白质域 (domain)。 (6分)
2. 请解释单系类群、并系类群和复系类群的概念, 并分别举例说明。 (9分)
3. 试描述 Blast 的算法过程。其精髓是追求速度, 表现在哪些步骤中? (9分)
4. 试描述肽质指纹分析的算法思想。 (8分)
5. 为什么说“大分子的序列决定结构, 继而决定功能”这一自然原理对生命现象之存在和演化有着重要意义 (提示: 从“生物信息”的概念来分析)。 (12分)

