

# 宁波大学 2017 年博士研究生招生考试初试试题(A 卷)

(答案必须写在考点提供的答题纸上)

科目代码: 3824 科目名称: 生物信息学

## 一、名词解释(每小题 3 分,共 15 分)。

1. 比较基因组学
2. Contig
3. 直系同源
4. 蛋白质组
5. 系统生物学

## 二、选择题(每小题 2 分,共 26 分)。

1. ortholog 的含义是\_\_\_\_\_。  
A. 直系同源 B. 旁系同源 C. 直接进化 D. 间接进化
2. UTR 的含义是\_\_\_\_\_。  
A. 编码区 B. 非编码区 C. 低复杂度区域 D. 开放阅读框
3. 基本局部比对搜索工具是\_\_\_\_\_。  
A. Mega B. ClustalW C. BLAST D. GCG
4. 将核酸序列按照 6 条链翻译成蛋白质序列后搜索蛋白质序列数据库使用的程序是\_\_\_\_\_。  
A. blastp B. blastx C. tblastn D. tblastx
5. 常用系统发育分析软件是\_\_\_\_\_。  
A. BLAST B. BOWTIE C. PHYLIP D. FASTA
6. 人类基因组大小大约是\_\_\_\_\_Mb。  
A. 130 B. 300 C. 3000 D. 30000
7. 根据研究发现,人类基因组中真正编码蛋白质的区域仅占 DNA 序列的\_\_\_\_\_。  
A. 1-5% B. 5-10% C. 10-15% D. 15-20%
8. 以下哪一项不属于启动子研究范围? \_\_\_\_\_  
A. CpG 岛预测 B. 转录起始点预测  
C. 糖基化修饰 D. 甲基化检测
9. INDEL 指\_\_\_\_\_。  
A. 点突变 B. 基因组小片段插入 C. 拷贝数变异 D. 序列扩增
10. 常用于差异表达分析的统计方法是\_\_\_\_\_。  
A.  $\chi^2$  检验 B.  $t$  检验 C. logistic 回归 D. cox 回归
11. 目前应用于基因芯片表达数据统计分析的主要方法是\_\_\_\_\_。  
A.  $\chi^2$  检验 B. 相关分析 C. 聚类分析 D. 正态性分布检验
12. 从头预测真核基因的原因有\_\_\_\_\_。  
A. 外显子/内含子边界难以确定  
B. 内含子长度可能只有几个碱基对  
C. 编码区域的 GC 含量并不总是与非编码区相同  
D. 以上三个方面的原因

# 宁波大学 2017 年博士研究生招生考试初试试题(A 卷)

(答案必须写在考点提供的答题纸上)

科目代码: 3824 科目名称: 生物信息学

13. 下列 Fasta 格式正确的是\_\_\_\_\_。

- A. seq1: agcggatccagacgctgcgtttgctggctttgatgaaaacttaactaaactccctta
- B. >seq1 agcggatccagacgctgcgtttgctggctttgatgaaaacttaactaaactccctta
- C. seq1:agcggatccagacgctgcgtttgctggctttgatgaaaacttaactaaactccctta
- D. >>seq1agcggatccagacgctgcgtttgctggctttgatgaaaacttaactaaactccctta

### 三、填空题（每空 1 分，共 10 分）。

1. 初级序列数据库包含: \_\_\_\_\_、\_\_\_\_\_和 DDBJ。
2. 常用的进化分析算法有\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。
3. 预测蛋白质三级结构的三种方法是\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。
4. 人类基因组计划要完成的 4 张图谱为\_\_\_\_\_、\_\_\_\_\_、遗传图谱和基因图谱。

### 四、简答题（每小题 8 分，共 24 分）。

1. 序列的相似性与同源性有什么区别与联系?
2. 基因组注释的内容包括哪些?
3. ChIP-seq 技术有哪些应用?

### 五、综合分析题（25 分）。

1. 某研究者为研究基因 A 在大肠癌中的功能，构造了基因 A 过表达的 HCT116 细胞系，以空载体转染的 HCT116 细胞系为对照，分别选择 3 个生物学重复样本，进行 RNA-seq，得到测序数据。
  - 1) 如何基于 RNA-seq 数据鉴定新转录本? (8 分)
  - 2) 获得基因表达谱后可以作何基础分析? (7 分)
  - 3) 怎样结合其他数据库的资源进行调节网络的分析? (10 分)