

中国科学院北京基因组研究所
2009 年招收攻读博士学位研究生入学统一考试专业试题
科目名称：生物信息学

考试时间：2009 年 3 月 22 日上午 8:30-11:30

考生须知：

1. 本试卷满分为 100 分，全部考试时间总计 180 分钟。
2. 所有答案必须写在答题纸上，写在试题纸上或草稿纸上一律无效。

一、基本概念（共 40 分）

- a) 简要说明下列数据库的主要内容及特色。
 - i. UCSC Genome (4 分)
 - ii. PDB (4 分)
 - iii. RISE (4 分)
 - iv. UniGene (4 分)
- b) 简要说明下列算法的原理及用途。
 - i. UPGMA (4 分)
 - ii. Smith-Waterman, 并列公式 (6 分)
 - iii. Ab initio Folding (6 分)
- c) 简要介绍下列软件的用途。
 - i. Phred-Phrap-Consed (4 分)
 - ii. ROBETTA (4 分)

二、问答题（共 60 分）

- a) 简要介绍隐马尔可夫模型，并阐述该模型在序列比对问题上的应用。（20 分）
- b) 请阐述“分子钟”模型，简要介绍构建进化树的各种算法，并比较其优缺点。（20 分）
- c) 简述聚类分析算法及其在转录组学分析中的应用。（20 分）

