

中国科学院北京基因组研究所

2013 年招收攻读硕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

科目代码：887

考试时间：2013 年 1 月 6 日下午 2: 00-5: 00

考生须知：

1. 本试卷满分为 150 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、简要说明以下数据库存储的主要内容及其特点。（共 6 小题，每题 5 分，共 30 分）

- a) DDBJ
- b) SWISS-PROT
- c) EcoliWiki
- d) OMIM
- e) UTRdb
- f) AARSDB

二、简要说明以下算法的原理。（共 4 小题，每题 5 分，共 20 分）

- a) Smith-Waterman
- b) Needleman-Wunsch
- c) 基于序列的序列相似性度量（举一个例子，不用打分，只要写出原理即可）
- d) BLAST
- e) TopHat
- f) STAR
- g) cuffdiff
- h) cuffdiff
- i) cuffdiff

三、假设有一个长度为 1000 的序列，其 GC 含量为 40%，请计算该序列的熵。（5 分）

四、假设有一个长度为 1000 的序列，其 GC 含量为 40%，请计算该序列的 GC 含量。（5 分）