

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均无效。

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) AARDSDb

二、简要说明以下算法的用途。（共 2 小题，每题 5 分，共 10 分）

- a) support vector machines
- b) Smith-Waterman

三、简要说明以下软件的用途。（共 5 小题，每题 10 分，共 50 分）

- a) MEGA
- b) Tophat
- c) Trinity
- d) OrfFinder
- e) GeneMark

四、简述神经网络算法，并阐述该算法在蛋白质二级结构预测问题上的应用？（30 分）

五、简述第一代和第二代测序所产生数据的特点及其适用的研究范围。（30 分）