

中国科学院北京基因组研究所

2008 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：2008 年 1 月 20 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色（共 47 分）：

- 1) Tree of Life (3 分)，
- 2) dbSNP (6 分)，
- 3) PKUCBI (5 分)，
- 4) FlyBase (6 分)，
- 5) PDB (5 分)，
- 6) BLOCKS (5 分)，
- 7) COG (6 分)，
- 8) SCI (6 分)，
- 9) KEGG (5 分)，

二、简要说明以下算法的用途，并写出公式（第 3 小题除外）（共 21 分）：

- 1) PSSM (6 分)，
- 2) Needleman-Wunsch (7 分)，
- 3) Simulated annealing (3 分)，
- 4) Simple pseudocounts (5 分)，

三、简要说明以下软件的用途（共 27 分）：

- 1) Entrez (5 分)，
- 2) PHYLIP (6 分)，
- 3) Primer3 (5 分)，
- 4) GenScan (5 分)，
- 5) Phred-Phrap-Consed (6 分)，

四、请说明真核生物基因的基本结构和主要原件。（10 分）

五、简要介绍带仿射空位分值的联配算法，画出与之对应的有限状态自动机图示，并用它描述联配  $VLSPAD-K$  的形成过程。（25 分）

六、简要介绍  $HL-AESK$  CLUSTALW 算法的过程，并列出至少 2 个其中所使用的启发式机制。（20 分）