

中国科学院北京基因组研究所

2008 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：2008 年 3 月 23 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、名词解释（共 17 分）：

- 1) Phylogeny (2 分)，
- 2) GO (3 分)，
- 3) Haplotype (3 分)，
- 4) NCBI (3 分)，
- 5) EMBL (3 分)，
- 6) PIR (3 分)。

二、简要回答以下问题（共 21 分）：

- 1) 从来源、矩阵序号、矩阵元素数值和常用缺省设置四方面比较 PAM 和 BLOSUM 系列打分矩阵（8 分），
- 2) 简述 ClustalW 的功能（2 分），
- 3) 举例构建系统发育树的方法：至少两种聚类法，至少两种提出优化要求的方法（6 分），
- 4) 简述 Phred, Phrap 和 Consed 的功能（5 分）。

三、请画出分子生物学中心法则的简图。（10 分）

四、描述当训练序列的路径未知时，如何使用 Baum-Welch 算法以及与之相关的前向、后向算法进行隐马模型的参数估计。请给出算法的伪代码。（25 分）

五、请写出轮廓隐马模型（profile HMM）中 Viterbi 方程的通式。（12 分）

六、简述如何用极值分布的方法评估局部无空位联配分值的显著性。（15 分）