

天津医科大学实验课教案首页

(共 2 页、第 1 页)

课程名称：生物信息学		实验名称：原核生物和真核生物的基因识别	
教师姓名：伊现富	职称：讲师	教学日期：2013 年 9 月 11 日 14 时 -16 时	
授课对象：生物医学工程学院 2010 级生信班（本）			实验人数：23
实验类型（验证型、综合型、设计型、创新型）：验证型			实验分组：一人一机
学时数：2		教材版本：生物信息学实验讲义（自编教材）	

实验目的与要求：

- 了解隐马尔科夫模型在基因识别中的应用。
- 掌握原核基因和真核基因的结构特征。
- 掌握 GeneMarkS 和 GENSCAN 的使用方法。

实验内容及学时分配：

- (5') 基因结构：回顾原核生物和真核生物基因的结构特点。
- (5') 基因识别的方法：回顾基因识别的三大类方法。
- (5') 基因识别的工具：介绍 GeneMarkS 和 GENSCAN。
- (85') 实验操作：对大肠杆菌基因组序列进行基因识别，对人类 CD9 基因进行结构分析。

主要仪器和实验材料：

- 实验材料：大肠杆菌基因组，人类 CD9 基因。
- 主要仪器：联网的计算机。
- 分析工具：GeneMarkS，GENSCAN。

实验重点、难点及解决策略：

- 难点：FASTA 格式与序列的区别；解决策略：通过实例进行讲解。
- 重点：GeneMarkS 和 GENSCAN 的使用；解决策略：通过练习熟练掌握。

思考题：

- 原核基因和真核基因的结构有何异同？
- GeneMarkS 和 GENSCAN 的适用范围有何差别？
- GeneMarkS 和 GENSCAN 对输入格式的要求有何差别？

参考资料：

- NCBI
- GeneMarkS
- GENSCAN

主任签字：

年 月 日

教务处制

一、基因结构 (5 分钟)

基因结构的复杂性直接影响着基因预测的策略及最终的准确度。

1. 原核基因：连续基因
2. 真核基因：不连续性

二、基因识别的方法 (5 分钟)

1. 间接识别法
2. 从头预测法：基因预测
3. 比较基因组学的方法

三、基因识别的工具 (5 分钟)

分析工具有各自的适用范围。

1. GeneMarkS：迭代隐马尔科夫模型，适用于原核生物的基因预测
2. GENSCAN：广义隐马尔科夫模型，脊椎动物基因预测软件

四、实验操作 (85 分钟)

基因结构的复杂性直接影响着基因预测的准确度。

1. 大肠杆菌基因组序列的基因识别

- 获取序列：NCBI 中的 Nucleotide 数据库，U00096，FASTA 格式和 GenBank 格式 (自学 GenBank 格式)
- 截取序列：EMBOSS, extractseq, 1-10000bp
- 基因预测：GeneMarkS, FASTA 格式
- 结果分析：和 GenBank 格式中的信息进行比较

2. 人类 CD9 基因的结构分析

- 获取序列：NCBI 中的 Nucleotide 数据库，AY422198，FASTA 格式和 GenBank 格式
- 基因预测：GENSCAN，纯序列 (注意不是 FASTA 格式)
- 结果分析：和 GenBank 格式中的信息进行比较