

# 天津医科大学理论课教案首页

(共 3 页、第 1 页)

课程名称：分子生物计算

课程内容/章节：遗传密码 / 第 8 章

教师姓名：伊现富

职称：讲师

教学日期：2018 年 12 月 24&26 日 10:00-11:40&15:30-17:10

授课对象：生物医学工程与技术学院 2016 级生信班（本）

听课人数：28

授课方式：理论讲授

学时数：4

教材版本：Perl 语言在生物信息学中的应用——基础篇

教学目的与要求（分掌握、熟悉、了解、自学四个层次）：

- 掌握：散列的初始化及常见操作；正则表达式在密码子中的应用。
- 熟悉：Perl 语言中读取 FASTA 文件的方法；密码子和阅读框的概念。
- 了解：数据结构和关系数据库等基本概念；FASTA 格式等常见的序列格式。
- 自学：在 Perl 语言中使用关系数据库的方法。

授课内容及学时分配：

- (5') 引言与导入：回顾已经学习的知识点，简单介绍即将学习的内容。
- (25') 散列：介绍散列数据类型，讲解散列的初始化和常见操作等基本应用。
- (25') 数据结构和算法：介绍数据结构和算法之间的关系，通过实例分析数据结构和算法的选择，总结比较和排序的方法，介绍数据库和关系数据库的基本概念及其在 Perl 语言中的使用方法。
- (25') 遗传密码：介绍遗传密码的基本概念，讲解实现密码子翻译的不同方法并进行比较。
- (20') DNA 翻译成蛋白质：通过 Perl 程序讲解把 DNA 序列翻译成蛋白质的具体方法。
- (60') 读取 FASTA 文件：介绍常见的序列格式，比较读取文件的不同策略，通过 Perl 程序讲解读取 FASTA 文件并进行格式化输出的具体操作。
- (35') 阅读框：介绍开放阅读框及阅读框的基本概念，通过 Perl 程序讲解实现六种阅读框翻译的具体方法。
- (5') 总结与答疑：总结授课内容中的知识点与技能，解答学生疑问。

教学重点、难点及解决策略：

- 重点：散列的初始化及常见操作；正则表达式在密码子中的应用。
- 难点：散列的初始化及常见操作。
- 解决策略：通过实例演示帮助学生理解、记忆。

专业外语词汇或术语：

散列 (hash)

关系数据库 (relational database)

算法 (algorithm)

密码子 (codon)

数据结构 (data structure)

阅读框 (reading frame)

辅助教学情况：

- 多媒体：散列的结构；折半查找算法；关系数据库的结构；密码子表；常见序列格式；阅读框示意图。
- 板书：Perl 语言中三种数据类型的比较；正则表达式在密码子中的应用。
- 演示：密码子到氨基酸的翻译；读取 FASTA 文件并进行格式化输出。

复习思考题：

- 比较 Perl 语言中的三种数据类型。
- 如何对字符串/数字进行排序？
- 列举散列的常见操作及相关函数。
- 比较两种不同的索引方式。

参考资料：

- Beginning Perl for Bioinformatics, James Tisdall, O'Reilly Media, 2001.
- Perl 语言入门（第六版），Randal L. Schwartz, brian d foy & Tom Phoenix 著，盛春 译，东南大学出版社，2012。
- Mastering Perl for Bioinformatics, James Tisdall, O'Reilly Media, 2003.
- 维基百科等网络资源。

主任签字：

年 月 日

教务处制

## 一、引言与导入 (5 分钟)

### 1. 已经学习

- 查找基序
- 模拟 DNA 突变

- 生成随机序列

- DNA 转录成 RNA

### 2. 即将学习

- 散列

- DNA 翻译成蛋白质

- 处理 FASTA 文件

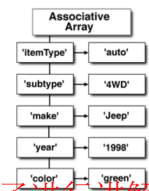
## 二、【重点、难点】散列 (25 分钟)

### 1. 简介：散列/关联数组 (和字典相类比)

### 2. 初始化：胖箭头 (=>)

### 3. 常见操作：keys, values, each, delete, exists

### 4. 三种数据类型：标量变量 vs. 数组 vs. 散列



	%Perez	%Kleiner
Age	21	21
Name	"Jeff"	"Dave"
Eyes	"Br"	"Blue"
Temp	98.6	99.0

\$Perez{'Eyes'} = "Br"; \$Kleiner{'Temp'} = 99.0;

## 三、数据结构和算法 (25 分钟) (通过基因表达的例子进行讲解)

### 1. 简介：不同的算法通常需要不同的数据结构

### 2. 排序：cmp (字符串), <=> (数字), sort, 折半查找

### 3. 数据库：关系数据库 (表格, 行与列 (与 Excel 表格进行比较)), SQL

### 4. DBM: 采用散列结构, 与散列结合使用; dbmopen/dbmclose, tie/untie

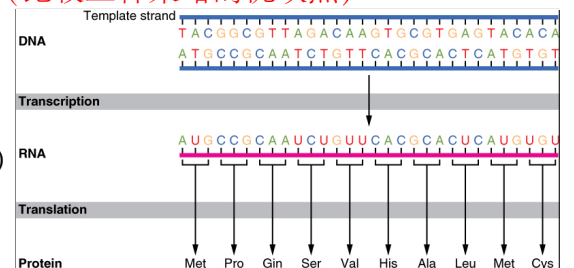
## 四、遗传密码 (25 分钟)

### 1. 简介：密码子, 三个核苷酸翻译成一个氨基酸

### 2. 【重点】策略：模式匹配, 正则表达式, 散列 (比较三种策略的优缺点)

### 3. 知识点

- 文件句柄：STDIN, STDOUT, STDERR
- 正则表达式：元字符, 选择, 修饰符
- 函数：uc (转换为大写), exists (检查键)



## 五、DNA 翻译成蛋白质 (20 分钟)

### 1. Perl 程序 8.1: 把 DNA 序列翻译成蛋白质序列

### 2. for (my \$i = 0; \$i < (length(\$dna) - 2); \$i += 3) {} (结合密码子的属性进行理解)

### 3. \$codon = substr (\$dna, \$i, 3); (结合密码子的属性进行理解)

### 4. 把主程序变成子程序 (比较主程序和子程序的利弊)

## 六、读取 FASTA 文件 (60 分钟)

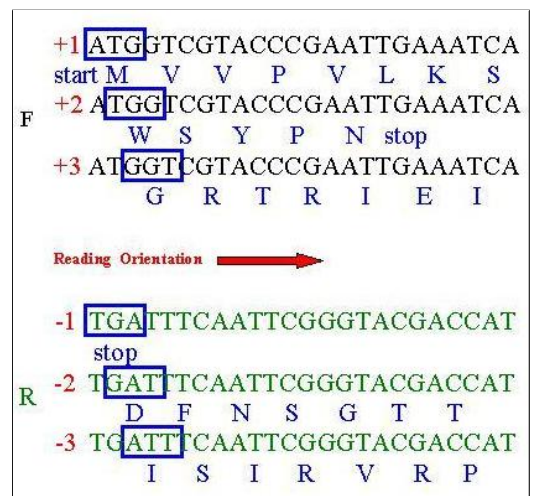
### 1. 序列格式：纯序列, FASTA, GenBank, EMBL, GCG, .....

### 2. 读取文件：一次性读入所有行, 一次读入一行 (比较两种策略的优缺点)

### 3. 读取文件失败：退出程序, 询问用户, 使用默认文件, ..... (比较各种方法的利弊)

### 4. Perl 程序 8.2: 读取 FASTA 文件并从中提取序列数据 (体验子程序的优势)

### 5. Perl 程序 8.3: 读取 FASTA 文件, 把其中的序列翻译成蛋白质并进行格式化输出



## 七、阅读框 (35 分钟)

### 1. 简介：ORF, +1, +2, +3, -1, -2, -3

### 2. 子程序库：充分利用已有的子程序, 添加缺少的子程序

### 3. Perl 程序 8.4: 在六种框架下翻译 DNA 序列 (组合需要的子程序即可)

### 4. 索引方式：从 0 起始 vs. 从 1 起始 (注意面向程序和面向用户的选择)

## 八、总结与答疑 (5 分钟)

### 1. 知识点

- 散列：初始化，键值，常见操作及相关函数
- 数据类型：标量变量，数组，散列
- 正则表达式：元字符，字符集，修饰符，择一匹配
- 其他：比较排序，文件句柄，循环，索引，……

### 2. 技能

- 熟练使用 Perl 语言中的散列
- 能编写把 DNA 翻译成蛋白质相关的 Perl 程序