

# 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：序列和字符串 / 第 4 章

教师姓名：伊现富

职称：讲师

教学日期：2016 年 11 月 16&23 日 15:30-17:30

授课对象：生物医学工程与技术学院 2013 级生信班（本）

听课人数：30

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：Perl 语言的基础知识；标量和数组的使用；字符串的常见操作。
- 熟悉：文件的读取；标量上下文和列表上下文。
- 了解：Perl 语言中双引号和单引号的区别。
- 自学：常见函数的高级用法。

授课内容及学时分配：

- (5') 引言与导入：回顾分子生物学的基础知识，介绍本章将要学习的内容。
- (10') 序列数据的表征：回顾核酸和蛋白质序列的表征方法，介绍序列和字符串之间的关系。
- (35') 存储 DNA 序列：通过存储 DNA 序列的 Perl 程序讲解 Perl 语言的基础知识。
- (20') 拼接 DNA 片段：通过拼接 DNA 片段的 Perl 程序讲解字符串的拼接操作。
- (20') DNA 转录成 RNA：通过 DNA 转录成 RNA 的 Perl 程序讲解字符串的替换操作。
- (10') 使用 Perl 文档：总结查找 Perl 文档的方法。
- (25') 序列反向互补：通过获取 DNA 反向互补序列的 Perl 程序讲解字符串的翻译反转操作。
- (25') 从文件读取数据：通过读取文件数据的 Perl 程序讲解读取文件的基本方法和策略。
- (30') 数组：通过读取文件中多行数据的 Perl 程序讲解数组的基本使用与常见操作。
- (15') 上下文：通过实例介绍上下文的概念，比较标量上下文和列表上下文。
- (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：Perl 语言的基础知识；字符串的常见操作；文件的读取；数组的基本使用。
- 难点：字符串的常见操作；数组的基本使用。
- 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：

字符串 (string)

赋值操作符 (assignment operator)

语句 (statement)

字符串内插 (string interpolation)

标量变量 (scalar variable)

标量上下文 (scalar context)

数组 (array)

列表上下文 (list context)

辅助教学情况：

- 多媒体：序列数据的表征；Perl 程序实例；数组的常见操作。
- 板书：常见的字符串操作。
- 演示：Perl 程序实例。

复习思考题：

- 列举拼接 DNA 片段的不同方法。
- 在 Perl 语言中如何从文件中读取数据？
- 比较 Perl 语言中双引号和单引号的区别。
- 列举常见的数组操作。
- 列举字符串的常见操作。
- 举例说明 Perl 语言中的上下文。

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. 回顾知识点

- Perl 语言：变量，字符串操作符，……
- 分子生物学：中心法则，DNA、RNA 和蛋白质

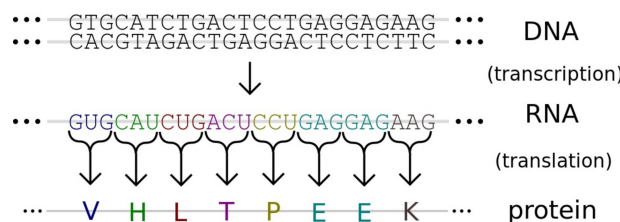
### 2. 本章内容

- Perl 语言：标量变量，数组变量，字符串操作，从文件中读取数据
- Perl 语言在分子生物学中的应用：拼接 DNA 片段，DNA 转录成 RNA，获取 DNA 的反向互补序列，从文件中读取序列信息，计算序列的统计信息

## 二、序列数据的表征 (10 分钟)

### 1. 序列表征

- DNA/RNA: A、C、G、T/U、M、R、W、S、Y、K、V、H、D、B、N
- 蛋白质: A、B、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、W、X、Y、Z



### 2. 序列与字符串

- (生物学) DNA/RNA/蛋白质序列  $\Rightarrow$  (计算机科学) 字符串

## 三、存储 DNA 序列 (35 分钟)

### 1. Perl 程序 4.1: 使用 Perl 程序存储 DNA 序列并将其打印出来

### 2. 【重点】程序 4.1 的解释说明 (结合程序逐一讲解)

- 编辑运行: ASCII 或纯文本格式; 可执行权限
- 控制流: 顺序执行; 条件/循环流程控制 (和 RPG 中的主线支线相类比)
- 注释: 以 # 起始, 空行和注释会被解释器忽略
- 命令解释: #!/usr/bin/perl
- 语句: 以 ; 结尾 (和中文/英文中的句号相类比)
- 变量: 命名规范; 标量变量 (以 \$ 起始)
- 字符串: 单引号 vs. 双引号
- 赋值: = vs. ==
- 打印输出: print, STDOUT
- 退出: exit; vs. 自动退出

## 四、【重点、难点】拼接 DNA 片段 (20 分钟) (结合程序讲解字符串的拼接)

### 1. Perl 程序 4.2: 使用 Perl 程序把两个 DNA 片段拼接起来 (应用: 外显子剪接)

### 2. 同样的拼接, 不同的方法 (方法不止一种, 注意开拓思路)

- (1) `$DNA3 = "$DNA1$DNA2"; print "$DNA3\n";`
- (2) `$DNA3 = $DNA1 . $DNA2; print "$DNA3\n";`
- (3) `print $DNA1, $DNA2, "\n";`
- (4) `print "$DNA1$DNA2\n";`

## 五、【重点、难点】DNA 转录成 RNA (20 分钟) (结合程序讲解字符串的替换)

### 1. 问题转化: (生物学) DNA 转录成 RNA $\Rightarrow$ (编程) 把 DNA 序列中的所有 T 替换成 U

### 2. Perl 程序 4.3: 使用 Perl 程序把 DNA 转录成 RNA

### 3. 绑定操作符和替换: `$RNA =~ s/T/U/g;`

### 4. 方法不止一种

- (1) `$RNA = $DNA; $RNA =~ s/T/U/g;`
- (2) `($RNA = $DNA) =~ s/T/U/g;`

## 六、使用 Perl 文档 (10 分钟)

### 1. 官网: <http://www.perl.com>

### 2. 手册: `man perldoc`

## 七、【重点、难点】序列反向互补 (25 分钟) (和手工反向互补序列的思路进行比较; 结合程序讲解字符串的反转与翻译)

1. Perl 程序 4.4: 使用 Perl 程序获取 DNA 序列的反向互补序列
2. 两种方法: 一错 vs. 一对 (出错很正常, 关键是找到症结解决问题); 一繁 vs. 一简
3. 两个函数
  - reverse: 反转字符串、数组元素等的顺序
  - tr: 字符集翻译; `$revcom =~ tr/ACGT/TGCA/;`

## 八、从文件读取数据 (25 分钟)

1. 创建数据: 文件和文件夹的组织与命名 (养成好习惯, 终生会收益)
2. Perl 程序 4.5/4.6: 使用 Perl 程序从文件中读取蛋白质序列数据 (一行/多行)
3. 【重点】读取文件 (三步走: 打开-读取-关闭)
  - (1) 关联文件和文件句柄: `open (PROTEINFILE, $proteinfilename);` (文件句柄通常用大写)
  - (2) 通过文件句柄读取文件数据: `$protein = <PROTEINFILE>;`
  - (3) 解关联文件和文件句柄: `close PROTEINFILE;` (有开有关、有始有终)

## 九、数组 (30 分钟)

1. Perl 程序 4.7: 通过数组, 使用 Perl 程序从文件中读取全部的蛋白质序列数据 (和 4.5/4.6 进行比较)
2. 【重点、难点】数组操作 (演示数组操作实例)
  - (1) 标量 vs. 数组: 单数 vs. 复数; `$` vs. `@`
  - (2) 初始化和输出: `print @bases;` vs. `print "@bases";`
  - (3) 元素访问: 使用索引, 从 0 开始 (编程初期常见错误: 从 1 开始进行索引)
  - (4) 头尾操作: `pop`, `shift`, `unshift`, `push`
  - (5) 其他操作: 反转 (`reverse`), 获取元素个数 (`scalar`), 插入元素 (`splice`)

## 十、上下文 (15 分钟)

1. Perl 程序 4.8: 使用 Perl 程序演示标量上下文和列表上下文
2. 上下文 (和自然语言中的语境相类比)
  - 标量上下文: `$a = @bases;` (vs. `$a = scalar @bases;`)
  - 列表上下文: `($a) = @bases;`

## 十一、总结与答疑 (5 分钟)

1. 知识点
  - Perl 语言基础: 命令解释、注释、语句、赋值、……
  - Perl 中的变量: 标量 vs. 数组
  - 字符串操作: 拼接, 替换, 翻译, 反转
  - 数组: 初始化, 索引, 常见操作
  - 上下文: 标量上下文 vs. 列表上下文
2. 技能
  - 编写 Perl 程序处理 DNA 序列: 存储 DNA 序列, 拼接 DNA 片段, 把 DNA 转录成 RNA, 获取 DNA 的反向互补序列
  - 编写 Perl 程序从文件中读取蛋白质序列数据
  - 掌握 Perl 语言中数组的常见操作
  - 熟练运用标量上下文和列表上下文