

# 分子生物计算

## (*Perl* 语言编程)

天津医科大学  
生物医学工程与技术学院

2016-2017 学年上学期 (秋)  
2014 级生信班

# 条条大路通罗马

伊现富 (Yi Xianfu)

天津医科大学 (TIJMU)  
生物医学工程与技术学院

2017 年 1 月



```
1 # The Perl 5 language interpreter
2 man perl
3 # Look up Perl documentation in Pod format
4 man perldoc
5 perldoc perldoc
6
7 # Perl built-in function
8 perldoc -f BuiltinFunction
9 perldoc -f substr
10 # FAQ
11 perldoc -q FAQKeyword
12 perldoc -q random
13 # Perl predefined variable
14 perldoc -v PerlVariable
15 perldoc -v '$_'
```



# 启用警告

```
1 # Method1
2 # 使用命令行选项-w
3 perl -w script.pl
4
5 # Method2
6 # 在命令解释行使用-w
7 #!/usr/bin/perl -w
8
9 # Method3
10 # 使用use
11 use warnings;
```



# 字符串拼接

```
1 # Method1
2 $DNA3 = "$DNA1$DNA2"; print "$DNA3\n";
3
4 # Method2
5 print "$DNA1$DNA2\n";
6
7 # Method3
8 print $DNA1, $DNA2, "\n";
9
10 # Method4
11 $DNA3 = $DNA1 . $DNA2; print "$DNA3\n";
12
13 # Method5
14 $DNA3 = join "", $DNA1, $DNA2; print "$DNA3\n";
```



# 文件读取

```
1 # Method1 (只读取一行)
2 $protein = <PROTEINFILE>;
3
4 # Method2 (读取所有行, 一次性保存进数组)
5 @proteins = <PROTEINFILE>;
6
7 # Method3 (依次读取每一行)
8 while (<PROTEINFILE>) {
9     ...actions...
10 }
11
12 # Method4 (读取所有行, 一次性保存进标量)
13 $protein = do { local $/; <PROTEINFILE>; };
```



# 获取数组元素个数

```
1 # Methods1
2 $num = scalar @bases;
3
4 # Methods2
5 $num = @bases;
6
7 #Methods3
8 $num = $#bases + 1;
```



# 获取数组的第一个元素

```
1 # Method1
2 $first_gene = $genes[0];
3 #($first_gene) = $genes[0];
4
5 # Method2
6 ($first_gene) = @genes;
7
8 # Method3
9 $first_gene = shift @genes;
10 # 注意: shift会影响原始的@genes数组
```





# 获取数组的最后一个元素

```
1 # Method1
2 $last_gene = $genes[-1];
3
4 # Method2
5 $last_gene = $genes[$#genes];
6
7 # Method3
8 $last_gene = pop @genes;
9 # 注意: pop会影响原始的@genes数组
```



# 变量值 +1

```
1 # Methods1
2 $count++;
3
4 # Methods2
5 ++$count;
6
7 # Methods3
8 $count += 1;
9
10 # Methods4
11 $count = $count + 1;
```



# 计数核苷酸

```
1 # Methods1 (使用数组)
2 @DNA = split( ' ', $DNA );
3 foreach $base (@DNA) { ...count... }
4 # Methods2 (操作字符串)
5 for ( $pos=0 ; $pos < length $DNA ; ++$pos )
6 {
7     $base = substr( $DNA, $pos, 1 ); ..count..
8 }
9 # Methods3 (模式匹配)
10 while ( $DNA =~ /a/ig ) { $a++ }
11 while ( $DNA =~ /[^acgt]/ig ) { $e++ }
12 # Methods4 (使用tr函数)
13 $a = ($DNA =~ tr/Aa/Aa/);
14 $basecount = ($DNA =~ tr/ACGTacgt/ACGTacgt/);
15 $nonbase = (length $DNA) - $basecount;
```





TEX

LATEX

X<sub>Y</sub>TEX

Beamer

