

第14章 蛋白质理化性质分析

第15章 系统发育分析



生命健康信息学院

解增言

第14章 蛋白质理化性质分析

- 蛋白质及其功能
- 蛋白质结构
- 蛋白质理化性质
- 蛋白质理化性质分析工具

蛋白质及其功能

➤ 蛋白质是一种由氨基酸串联而成的大分子。

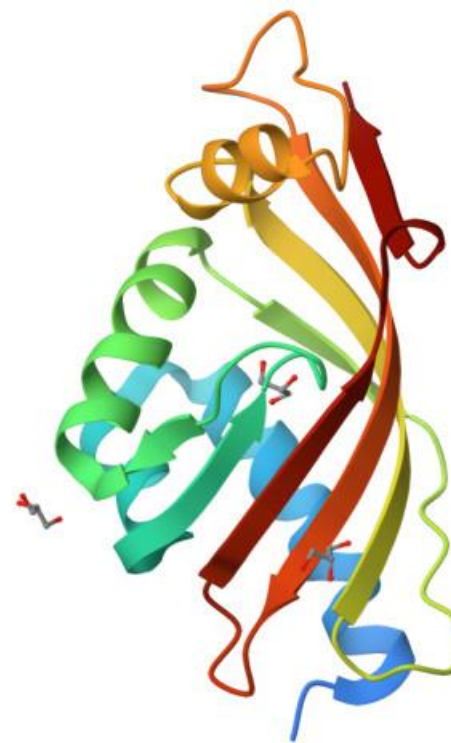
➤ 蛋白质功能：

- **催化：** 酶
- **结构与支持：** 胶原蛋白、角蛋白
- **运输与储存：** 血红蛋白、载脂蛋白、铁蛋白
- **运动：** 肌球蛋白
- **免疫防御：** 抗体
- **信号传递：** 激素、受体
- **调节：** 转录因子

蛋白质结构

- **一级结构**：氨基酸序列
- **二级结构**： α -螺旋、 β -折叠等
- **三级结构**：空间结构
- **四级结构**：亚基数量及排列

>9KW7_1|Chains A, B|MonBI|Streptomyces virginiae (1961)
MNHKVVHHHHHHIEGRHMNEFARAKRALEHSRRINAGDLDAIIDLY
APDAVLEDPVGLPPVTGHDALRAHYEPLLAHLREEAAEPVAGQD
ATHALIQISSVMDYLPVGPLYAERGWLKAPDAPGTARIHRTAMLVIR
MDASGLIRHLKSYWGTSDTLTVLG



9KW7蛋白序列及结构卡通图

一种核转运因子2 (NTF2) 家族蛋白。该家族蛋白的标志是拥有一个保守的NTF2-like 结构域，其内部具有一个疏水性的空腔，用于结合RanGDP（一种小G蛋白），并将其从细胞质运入细胞核。

蛋白质理化性质

- 分子量
- 等电点
- 疏水性/亲水性
- 溶解性
- 稳定性
- 表面可及性
-

蛋白质理化性质分析工具

- infoseq: 序列信息统计
- pepstats: 蛋白质序列理化性质综合统计
- pepwindow: 疏水性图谱可视化

infoseq

```
$ infoseq -only -name -length /home/pub/seq/lde.fa
```

Display basic information about sequences

Name	Length
Musmu	550
Homsa	544
Anoga	558
Caeel	500
Drome	534
Danre	568
Xentr	534

pepstats

```
$ pepstats lde_hs.fa
```

```
Calculate statistics of protein properties
```

```
Pepstats program output file [homsa.pepstats]:
```

```
$ cat homsa.pepstats
```

```
PEPSTATS of Homsa from 1 to 544
```

```
Molecular weight = 61554.93          Residues = 544
```

```
Average Residue Weight  = 113.152    Charge   = -16.0
```

```
Isoelectric Point = 5.1441
```

```
A280 Molar Extinction Coefficients  = 67270 (reduced)    67895 (cystine bridges)
```

```
A280 Extinction Coefficients 1mg/ml = 1.093 (reduced)    1.103 (cystine bridges)
```

```
Probability of expression in inclusion bodies = 0.639
```

Residue	Number	Mole%	DayhoffStat
A = Ala	36	6.618	0.769
B = Asx	0	0.000	0.000
C = Cys	11	2.022	0.697
D = Asp	42	7.721	1.404
.....			
Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	143	26.287
Small	(A+B+C+D+G+N+P+S+T+V)	265	48.713
Aliphatic	(A+I+L+V)	142	26.103
Aromatic	(F+H+W+Y)	67	12.316
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	269	49.449
Polar	(D+E+H+K+N+Q+R+S+T+Z)	275	50.551
Charged	(B+D+E+H+K+R+Z)	167	30.699
Basic	(H+K+R)	81	14.890
Acidic	(B+D+E+Z)	86	15.809

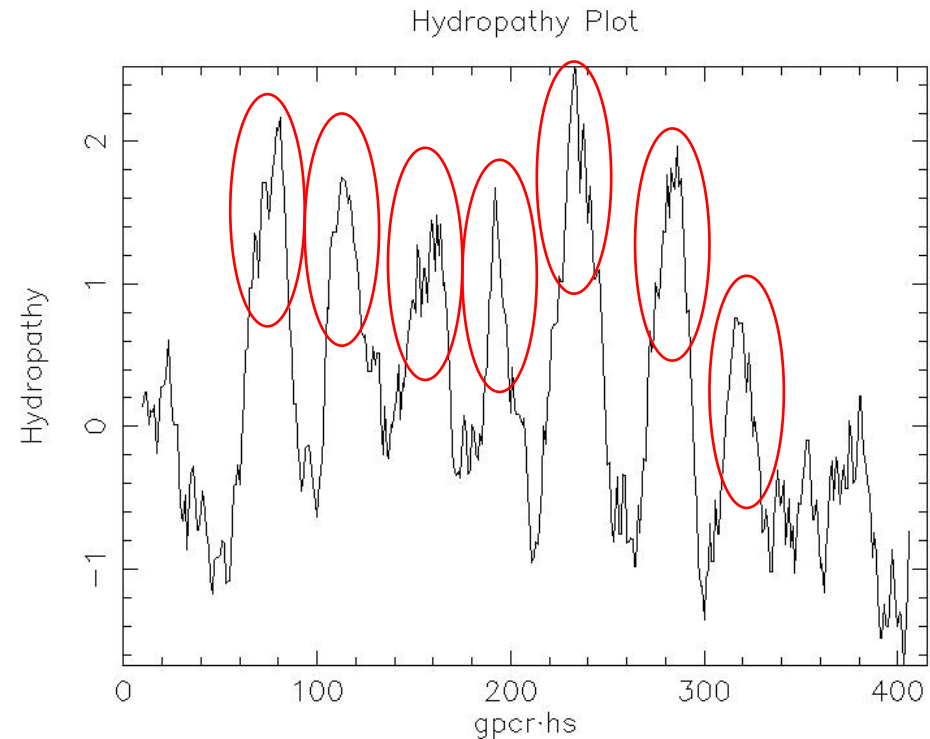
pepwindow

```
$ pepwindow GPCR_hs.fa
```

Draw a hydropathy plot for a protein sequence

Graph type [x11]: png

Created pepwindow.1.png

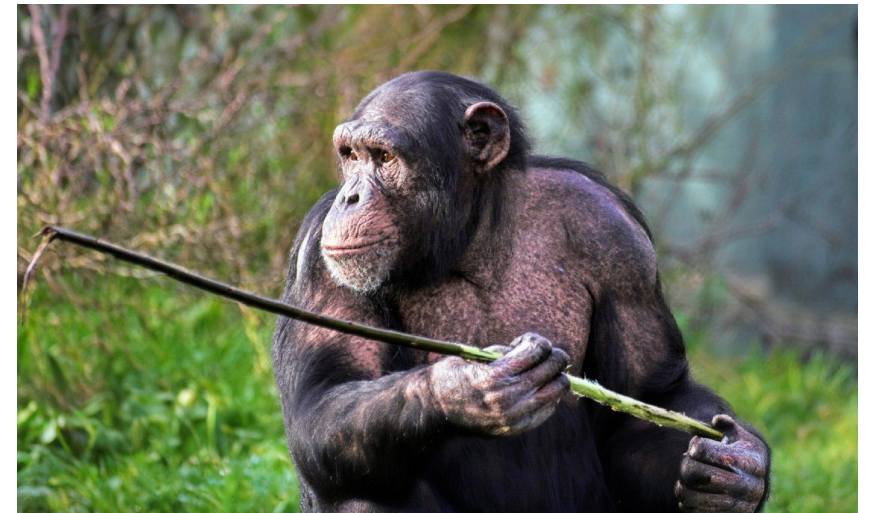


第15章 系统发育分析

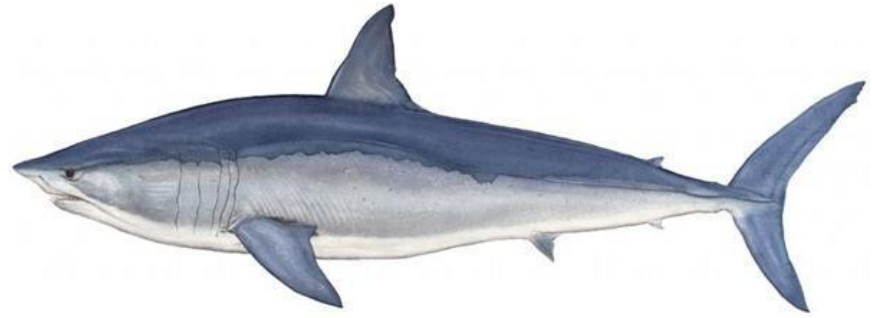
- 系统发育与分子进化
- 系统发育树构建方法及工具
- 系统发育树构建流程
- 系统发育树实例

系统发育与分子进化

- **系统发育**：利用形态、生理生化或分子等数据推断物种之间进化关系
- **分子系统发育**：利用分子数据（主要是蛋白质和核酸序列）推断物种之间进化关系
- **分子进化**：研究分子在一个物种基因组内或不同物种基因组间的序列及功能演化

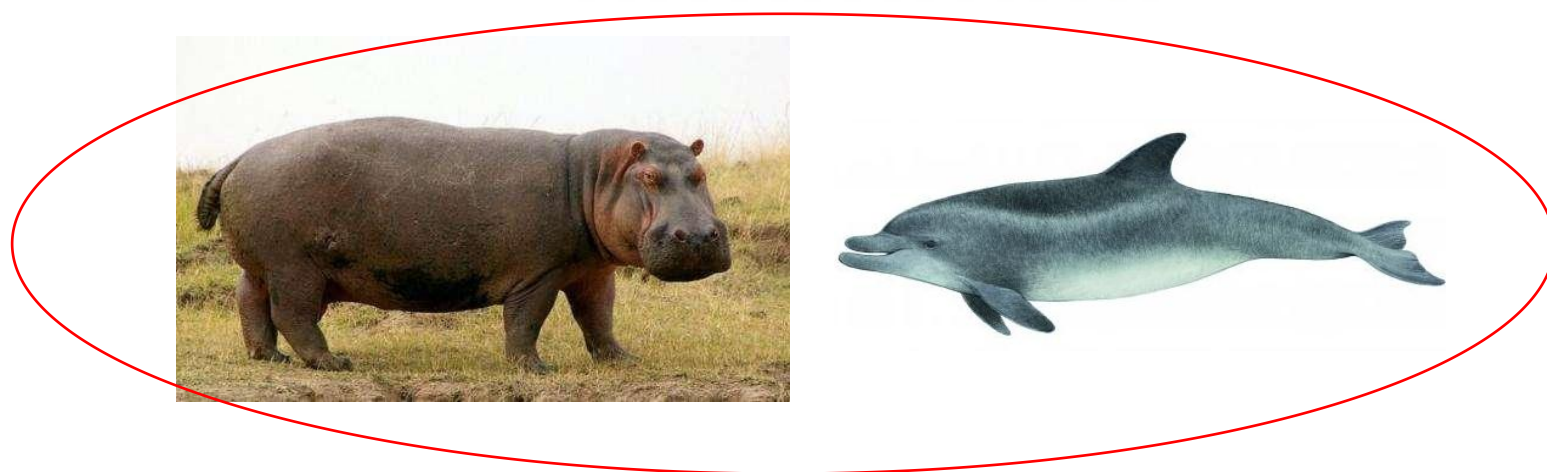
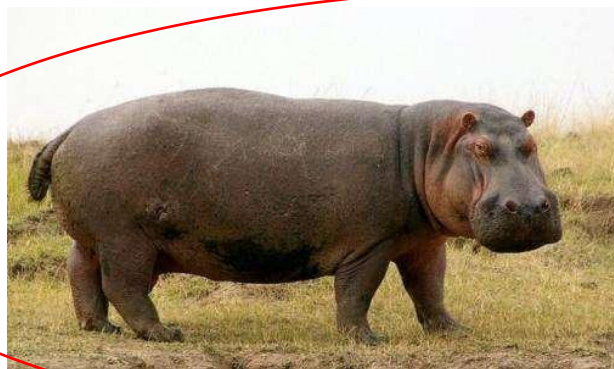
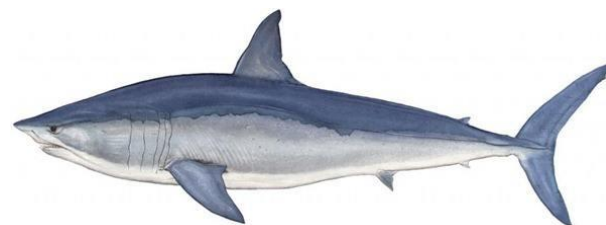






	**** *:*:*:*:::****:***** *****:***:***:*****.* *****:*
Dorphen	-MTNIRKTHPLMKILNDAFIDLPTPSNISSWWNFGSLLGLCLIMQILTGLFLAMHYTPDTSTAFSSVAH
Hippo	-MTNIRKSHPLMKIINDAFVDLPAPSNISSWWNFGSLLGVCLILQILTGLFLAMHYTPDTLTAFSSVTH
Shark	MALNIRKIHPLLLKIMNQTLIDLAPSNISIWVNFGSLLGLCLVIQIVTGLFLAMHYTADITMAFSSVTH

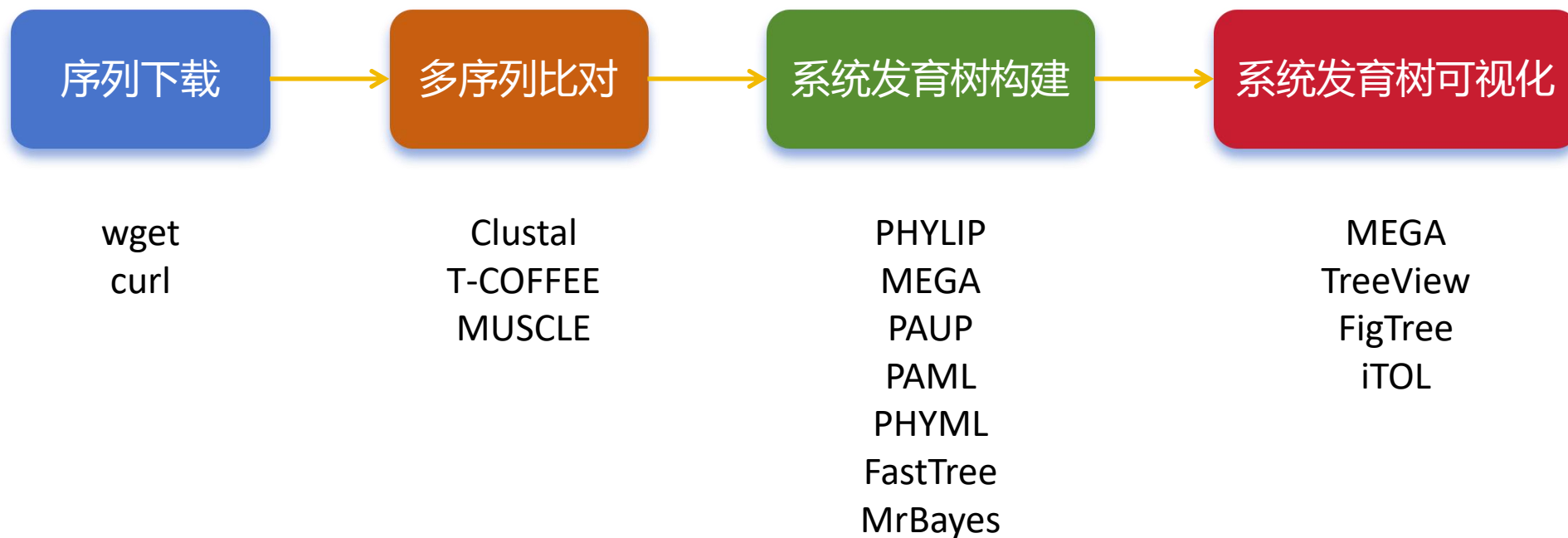
宽吻海豚、河马和大白鲨的细胞色素b蛋白分子序列比对结果



系统发育树构建方法及工具

- 距离法
- 最大似然法
- 最大简约法
- 贝叶斯方法

系统发育树构建流程

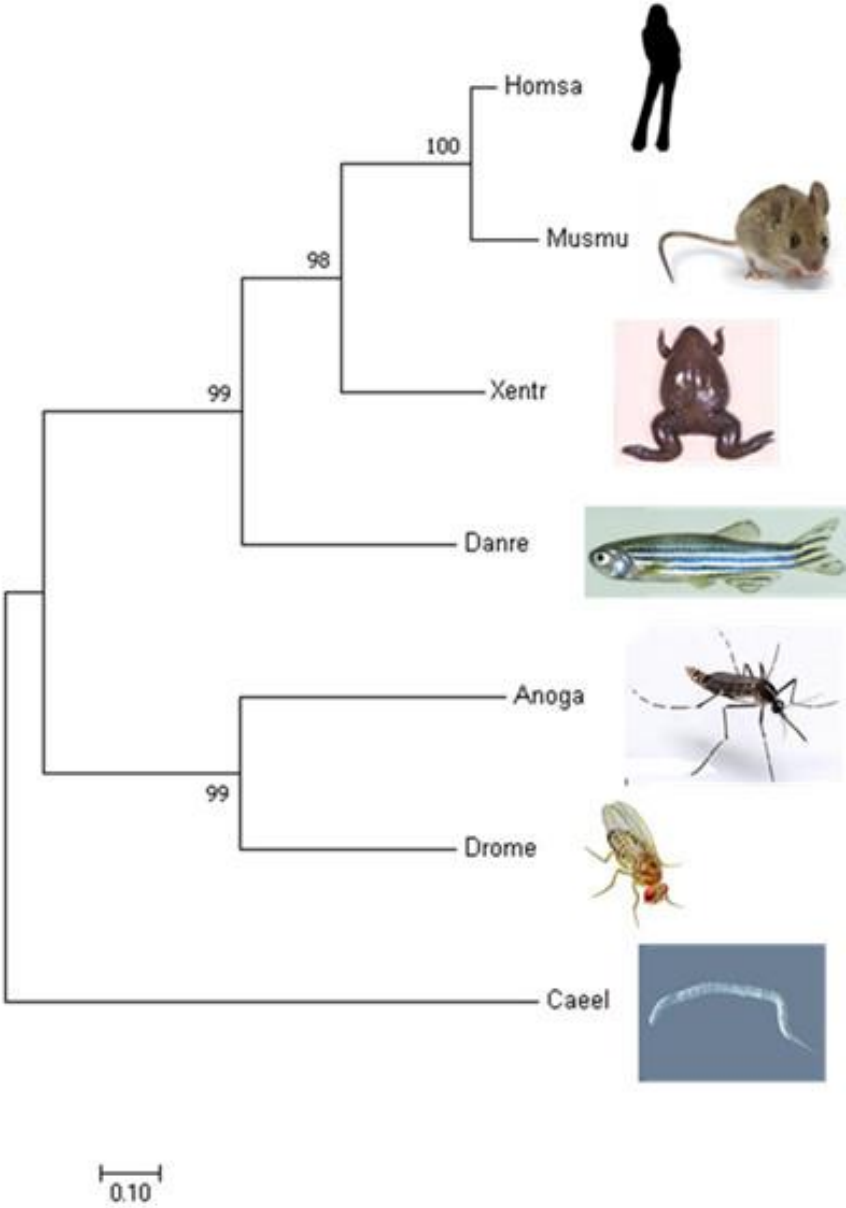


系统发育树实例

物种	拉丁学名	缩写
秀丽隐杆线虫	<i>Caenorhabditis elegans</i>	<i>Caeel</i>
黑腹果蝇	<i>Drosophila melanogaster</i>	<i>Drome</i>
冈比亚按蚊	<i>Anopheles gambiae</i>	<i>Anoga</i>
斑马鱼	<i>Danio rerio</i>	<i>Danre</i>
热带爪蟾	<i>Xenopus tropicalis</i>	<i>Xentr</i>
小鼠	<i>Mus musculus</i>	<i>Musmu</i>
人	<i>Homo sapiens</i>	<i>Homsa</i>

>Homsa debranching RNA lariats 1

MRVAVAGCCHGELDKIYETLALAERRGPGPVDLLCCGDFQAVRNEADLRCMAVPPKYRHMQTFYRYYS
G
EKKAPVLTFLFIGGNHEASNHLQELPYGGWVAPNIYYLGLAGVVKYRGVRIIGGISGIFKSHDYRKGFEC
P
PYNSSIRSIYHVRNIEVYKLKQLKQPIDIFLSHDWPRSIYHYGNKKQLLTKSFFRQEVENNTLGSPAA
S
SELLEHLKPTYWFSAPHLHVKAALMQHQAQDKGQTARATKFLALDKCLPHRDFLQILEIEHDPSAPDYLE
Y
YDIEWLTILRATDDLINVTGRLWNMPENGLHARWDYSATEEGMKEVLEKLNHDLKVPCNFSVTAACYDP
S
SKPQTQMQLIHRINPQTTEFCAQLGIIDINVRLQKSKEEHHVCGEYEEQDDVESNDSGEDQSEYNTDTS
A
LSSINPDEIMLDEEDEDSDIVSAHSGMNTSPVEPSDQASEFSASFSDVRILPGSMIVSSDDTV DSTIDRE
G
KPGGTVESNGEDLTKVPLKRLSDEHEPEQRKKIKRRNQAIYAAVDDDDDDAA



用Shell脚本构建系统发育树

```
#!/bin/bash
```

```
#从Fasta格式序列开始构建系统发育树
```

```
infile=$1
```

```
test -f pips.aln.fa.tmp && rm -f pips.aln.fa.tmp
```

```
clustalo -i infile -o pips.aln.fa.tmp --outfmt=fasta
```

```
FastTree pips.aln.fa 2>/dev/null
```

```
rm -f pips.aln.fa.tmp
```

单元测试4

15分钟