

# 实用生物信息技术

## Applied Bioinformatics Course

中国农业科学院研究生院

2019年9月8日 13:30-16:30 一教

罗静初

北京大学生命科学学院

[abc.cbi@pku.edu.cn](mailto:abc.cbi@pku.edu.cn)

<http://abc.cbi.pku.edu.cn/>

# 课程网站和邮箱

---

- 课程网站

<http://abc.cbi.pku.edu.cn/> (北大)

<http://abc.ic4r.org/> (中科院基因组所)

- 课程邮箱

[abc.cbi@pku.edu.cn](mailto:abc.cbi@pku.edu.cn)

- 邮箱请用实名, 不要自动回复
- 邮件主题: CAAS19MS

# 课程安排

---

- 时间: 2019/9/15-2020/1/5
  - 1班: 周日上午8:00-11:00
  - 2班: 周日下午1:30-4:30
- 地点
  - 17教(待定)

# 教学方式

---

- 教师

讲解：介绍课题背景、研究目的、分析思路

演示：实际上网演示操作步骤、分析结果

- 同学

课上：上机练习、分组讨论

课下：分组讨论、自行练习

# 本课程不涉及

---

- 高通量组学数据分析
- 统计分析软件MetLab, SAS, R使用
- Linux操作系统及其在生物信息学中的应用
- 计算机语言编程Perl, Python, Java, PHP, C
- 数据库管理系统MySQL, Oracle和数据库构建
- 生物信息学算法和软件开发
- 文献管理系统Endnote使用

# 教学内容：常用方法

---

- 网络文档和文献高级检索
- 序列比对
- 数据库高级检索和应用
- 数据库相似性搜索
- 系统发生树构建和演化分析
- 蛋白组结构分析和预测

# 教学内容：实例

---

1. 血红蛋白序列、结构、功能和演化分析
2. 癌胚抗原家族比较分析和三维空间结构预测
3. 豌豆内膜蛋白及拟南芥中同源蛋白比较分析
4. 植物特异转录因子家族系统发生分析及功能预测

# 生物信息数据库

---

- 国际蛋白质序列数据库UniProt

<https://www.uniprot.org/>

- NCBI核酸序列数据库

<https://www.ncbi.nlm.nih.gov/nucleotide/>

- 国际蛋白质结构数据库PDB

<http://www.rcsb.org/>

# 分析工具

---

- 双序列比对（在线分析）

<https://www.ebi.ac.uk/Tools/emboss/> (EBI)

<https://blast.ncbi.nlm.nih.gov/> (NCBI)

- 数据库相似性搜索Blast（在线分析）

<https://blast.ncbi.nlm.nih.gov/> (NCBI)

- 系统发生树构建MEGA（下载安装）

<https://www.megasoftware.net/>

- 蛋白质结构分析Swiss-PDBViewer（下载解压）

<https://spdbv.vital-it.ch/>

# 分析工具

---

- 荷兰瓦格宁根大学EMBOSS Explorer分析平台  
<http://www.bioinformatics.nl/emboss-explorer/>
- 瑞士蛋白质序列特征分析工具ProtScale  
<https://web.expasy.org/protscale/>
- 丹麦理工大学序列分析平台CBS  
<http://www.cbs.dtu.dk/services/>
- 德国慕尼黑理工大学蛋白质功能预测平台  
<https://www.predictprotein.org/>
- 英国帝国理工大学蛋白质结构预测平台  
<http://www.sbg.bio.ic.ac.uk/phyre2/>

# 血红蛋白分析实例

---

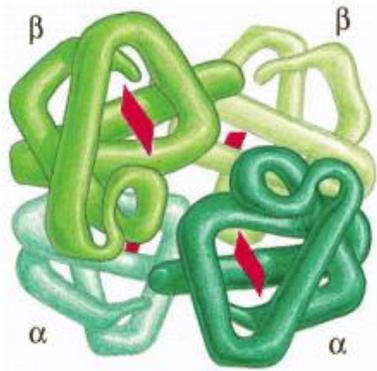
- 人、小鼠、大鼠血红蛋白alpha亚基及其编码序列比对
- 利用高级检索从UniProt数据库中检索血红蛋白9个成员和珠蛋白家族12个成员
- 利用双序列比对查看alpha珠蛋白和beta珠蛋白、alpha珠蛋白和脑红蛋白间的序列相似性
- 以alpha珠蛋白为检测序列，利用NCBI的Blast数据库相似性搜索平台搜索脑红蛋白
- 构建人的12个珠蛋白系统发生树；构建人、小鼠、大鼠37个珠蛋白编码区序列系统发生树
- 斑头雁和灰雁血红蛋白序列和结构比较分析

# 癌胚抗原分析实例

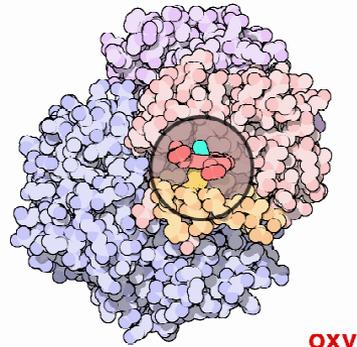
---

- 癌胚抗原CEA家族研究背景和网络资源
- 癌胚抗原家族不同成员的结构域信息
- 癌胚抗原结构预测的意义
- 蛋白质结构预测的常用方法和主要平台
- 利用Phyre2网站预测CEA21结构及结果分析
- 利用PredictProtein网站预测CEA21功能和突变敏感性分析

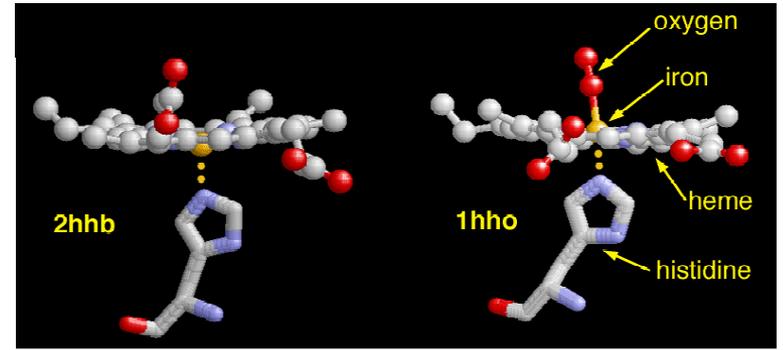
# 斑头雁血红蛋白序列/结构/功能分析



血红蛋白四个亚基



血色素携带氧分子



脱氧和含氧两种状态下的血色素分子

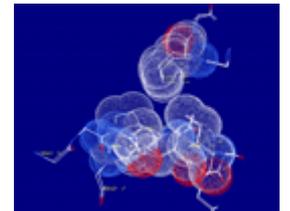
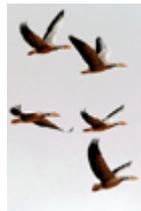
```

1  VLSAADKTNVKGVFSKIGGHAEYGAETLERMFTAYPQTKTYFPHFDLQH  50
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1  VLSAADKTNVKGVFSKISGHAEYGAETLERMFTAYPQTKTYFPHFDLQH  50

51  GSAQIKAHGKKVAAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKF  100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
51  GSAQIKAHGKKVAAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKF  100

101 LGHCFLVVVAIHHPALTPEVHASLDKFLCAVGTVLTAKYR  141
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
101 LGHCFLVVVAIHHPALTAEVHASLDKFLCAVGTVLTAKYR  141
    
```

	α119	β55
斑头雁	Ala	Leu
安第斯雁	Pro	Ser
灰雁	Pro	Leu



# 豌豆内膜蛋白分析实例

---

- 利用ORF分析工具提取Y12618编码区及所码氨基酸序列
- 利用密码子统计工具分析Y12618编码区密码子偏好
- 利用限制性内切酶分析工具分析Y12618酶切位点
- UniProt数据库中豌豆内膜蛋白PPF\_PEA注释信息
- UniProt数据库中拟南芥中PPF1\_PEA的同源蛋白查找
- TAIR和AraPort数据库中ALB3注释信息
- 利用ProtScale网站分析PPF1\_PEA序列特征
- 利用CBS网站分析PPF1\_PEA定位和跨膜螺旋
- 利用Phyre2网站预测PPF1\_PEA结构
- 利用PredictProtein网站预测PPF1\_PEA功能

# SBP转录因子家族分析实例

---

- 植物转录因子数据库PlantTFDB
- 植物特异转录因子家族SBP
- 如何查找数据库中SBP转录因子信息
- 序列全局比对和局部比对
- 系统发生数构建和分析
- 保守结构域预测和分析
- 序列图标构建和应用

# 拟南芥和水稻SBP转录因子分析

