

“实用生物信息技术”课程教学大纲

2023年3月1日

第一讲 网络文档

1. 分子月报

- 1.1. 了解蛋白质结构数据库 PDB 科普短文专栏 Molecule of the Month 主要内容、写作风格，及其与 PDB 数据库的关系。
- 1.2. 阅读血红蛋白 Hemoglobin 科普短文，了解血红蛋白亚基组成和四级结构、构象变化和协同作用、镰刀状贫血症的分子机制等基本知识。
- 1.3. 阅读胰岛素、免疫球蛋白、绿色荧光蛋白、转录因子等经典蛋白质的科普文章，探索蛋白质结构和功能关系。
- 1.4. 搜索课题相关蛋白质，阅读科普文章，并通过与 PDB 数据库的链接以及所提供的文献，深入了解该蛋白质的研究背景。

2. 蛋白质分子精选

- 2.1. 了解蛋白质序列数据库 UniProt 科普短文专栏 Protein Spotlight 主要内容、写作风格，及其与 UniProt 数据库的关系。
- 2.2. 阅读血红蛋白 Hemoglobin 科普短文，了解血红蛋白研究背景及其在生命科学研究中的重要意义，了解 Max Perutz 对血红蛋白三维结构测定所做的贡献及其背后的故事。
- 2.3. 阅读胰岛素、绿色荧光蛋白、肥胖症相关蛋白等经典蛋白质的科普文章，探索蛋白质结构和功能关系。
- 2.4. 搜索课题相关蛋白质，阅读科普文章，并通过与 UniProt 数据库的链接以及所提供的注释信息，深入了解该蛋白质的研究背景。

3. 网络教程

- 3.1. 熟悉欧洲生物信息学研究所 EBI 网络教程的主要内容、教程类别和学习方式。
- 3.2. 以蛋白质数据库 UniProt 和蛋白质结构数据库为例，说明如何利用网络教程学会数据库和软件工具的使用。
- 3.3. 查找基因组测序、蛋白质结构预测等研究方向相关教程，进行自学和小组讨论。

4. PubMed 文献

- 4.1. 了解 NCBI 网站生物医学文献摘要数据库 PubMed 的主要内容和特点。
- 4.2. 熟练使用高级检索功能，学会快速、高效查找作者、单位和研究方向相关文献。
- 4.3. 熟练掌握结果筛选、展示方式更改和统计等功能，对文献检索结果进行后处理。
- 4.4. 运用相似文献和引用文献等功能，扩充检索范围。
- 4.5. 熟练掌握文献保存、共享、导出、发送等基本功能。
- 4.6. 利用 MyNCBI 工具，掌握保存检索策略、定制自动发送等高级功能。
- 4.7. 了解医学主题词 MeSH 的含义和用途，熟练运用 MeSH 进行精准检索。

第二讲 序列比对

1. 序列比对基本概念

- 1.1. 序列相似性 (Similarity) 和序列同源性 (Homology)
- 1.2. 直系同源 (Ortholog) 和旁系同源 (Paralog)
- 1.3. 动态规划算法 (Dynamic Programming) 和启发式算法 (Heuristic Programming)
- 1.4. 计分矩阵 (Scoring Matrix) 和空位罚分 (Gap Penalty)
- 1.5. PAM (Point Accept Mutation) 计分矩阵和 BLOSUM (Block Substitution) 计分矩阵
- 1.6. 全局比对 (Global Alignment) 和局部比对 (Local Alignment)
- 1.7. 双序列比对 (Pairwise Sequence Alignment) 和多序列比对 (Multiple Sequence Alignment)
- 1.8. 序列比对点阵图方法 (Dot Plot)
- 1.9. 多序列比对序列图标 (Sequence Logo)

2. 序列比对常用工具和网站

- 2.1. 欧洲生物信息学研究所 (EBI) 在线序列比对工具
- 2.2. 美国国家生物信息中心 (NCBI) 在线序列比对工具
- 2.3. 荷兰瓦格宁根大学生物信息中心 (WUR) 序列分析在线服务 (EMBOSS Explorer)
- 2.4. 北京大学生物信息中心 (CBI) 生物信息网上实验室 (WebLab)

3. 常用程序

- 3.1. 双序列全局比对程序 Needle
- 3.2. 双序列局部比对程序 Water
- 3.3. EBI 多序列比对程序 ClustalW
- 3.4. NCBI 多序列比对程序 COBALT
- 3.5. BLAST 查找重复序列
- 3.6. 点阵图 Dotlet 寻找重复序列
- 3.7. 序列图标 WebLogo 寻找保守位点

4. 序列比对应用实例：人、小鼠和大鼠血红蛋白及其编码区序列比对

- 4.1. 研究背景
- 4.2. 血红蛋白 alpha 亚基氨基酸序列比对
- 4.3. Alpha 珠蛋白编码区核苷酸序列比对
- 4.4. 结果和讨论

第三讲 数据库检索和应用

1. 蛋白质序列和功能数据库 UniProt

- 1.1. UniProt 数据库简介
- 1.2. UniProt 数据库中序列条目注释信息
- 1.3. UniProt 数据库中序列位点和结构域注释信息
- 1.4. UniProt 数据库序列条目与其它数据库的链接
- 1.5. UniRef 参考数据集
- 1.6. UniParc 归档库
- 1.7. UniProt 蛋白组
- 1.8. UniProt 数据库高级检索应用实例

1.9. UniProt 数据库系统实用程序

1.10. UniProt 数据库用户界面

1.11. UniProt 数据库统计报表

1.12. UniProt 数据库帮助文档

2. 参考序列数据库 RefSeq

2.1. RefSeq 数据库特点简介

2.2. RefSeq 数据库与 GenBank 数据库的区别

2.3. RefSeq 数据库中序列条目登录号规则和含义

2.4. RefSeq 数据库高级检索应用实例

2.5. RefSeq 数据库序列条目与 NCBI 其它数据库的链接

3. 基因组数据库

3.1. 常用基因组数据库和基因组浏览器

3.2. Ensembl 基因组数据库简介

3.3. Ensembl 基因组数据库用户界面

3.4. Ensembl 基因组数据库应用实例

3.5. 主要模式生物基因组数据库 MGI, RGD, FlyBase, WormBase, ZFIN, SGD

3.6. 植物基因组资源网站 Phytozome

3.7. 微生物基因组数据库 PATRIC 和 GenoList

3.8. 病毒分类和生物信息资源数据库 ViroZone

4. 专业数据库

4.1. 蛋白质家族和结构域数据库 PFam

4.2. 基因和蛋白质表达数据库 HPA

4.3. 蛋白质相互作用网络数据库 STRING

4.4. 代谢通路数据库 KEGG

4.5. 反应过程数据库 REACTOME

4.6. 系统发生树数据库 TreeFam

4.7. 基因本体数据库 GO

4.8. 蛋白质结构分类数据库 SCOP

4.9. 单核苷酸多态性数据库 dbSNP

4.10. microRNA 数据库 miRBase

4.11. 拟南芥资源网站 TAIR

4.12. 基因组研究所国家生物信息中心生物信息数据库资源导航 Database Common

4.13. NAR 期刊数据库专辑

4.14. Database 生物信息数据库专刊

第四讲 序列相似性数据库搜索

1. Blast 算法简介和参数设置

1.1. BLAST 序列相似性数据库搜索的基本思路和主要步骤

- 1.2. 如何选择种子序列 Seed 长度 Word Size
- 1.3. 如何选择计分矩阵 Scoring Matrix
- 1.4. 如何则空位罚分 Gap Penalty
- 1.5. 如何设置期望值 Expected Value (E)
- 1.6. 如何分析搜索结果
- 1.7. 如何保存搜索策略

2. 常用程序及应用实例

- 2.1. 常用 BLAST 程序 BLASTP, BLASTN, BLASTX, TBLASTN
- 2.2. 蛋白质序列数据库搜索专用程序位点特异性迭代搜索程序 PSI-Blast、序列模体搜索程序 PHI-Blast、结构域数据库搜索程序 DELTA-Blast
- 2.3. 如何选择蛋白质数据库：非冗余数据库 NR、参考序列数据库 RefSeq_Protein、人工审阅蛋白质序列数据库 Swiss-Prot
- 2.4. 如何选择核酸序列数据库：非冗余数据库 NR、参考序列 NRA 数据库 RefSeq_NRA、参考序列基因组数据库 RefSeq_Genome
- 2.5. 用户界面和结果分析
- 2.6. 血红蛋白蛋白质搜索实例
- 2.7. 植物特异转录因子 SBP 家族搜索实例

3. 专用 Blast 程序

- 3.1. 模式生物蛋白质序列数据库搜索程序 SmartBlast
- 3.2. 引物设计程序 Primer-Blast
- 3.3. 保守结构域搜索程序 CD-Search
- 3.4. 免疫球蛋白数据库搜索专用程序 IgBlast
- 3.5. 测序载体接头检测程序 VecScreen

4. Linux 系统本地 Blast 简介

- 4.1 Linux 系统简介
- 4.2 Linux 系统常用命令
- 4.3 SBP 转录因子本地 Blast 运行实例
- 4.4 SBP 转录因子搜索结果分析

第五讲 系统发生分析和系统发生树构建

1. 分子演化和系统发生基本概念

- 1.1. 物种分化和分子演化
- 1.2. 分支图和系统树
- 1.3. 物种树和基因树
- 1.4. 有根数和无根树
- 1.5. 二叉树与多歧树
- 1.6. 外部节点和内部节点
- 1.7. 内部节点和根节点

1.8. 系统发生树稳定性检验

2. 常用系统发生分析软件

2.1. MEGA

2.2. DAMBE

2.3. Phylip

2.4. PAML

3. 常用系统发生树构建方法

3.1. 距离法

3.2. 最大简约法

3.3. 最大似然法

3.4. 贝叶斯推断

4. 系统发生树构建和分析实例

4.1. 人珠蛋白家族 12 个成员基因树

4.2. 人、小鼠、大鼠珠蛋白家族 37 个成员物种和基因树

4.3. 拟南芥和水稻 SBP 转录因子家族

4.4. 典型系统发生树实例

第六讲 蛋白质结构比较和分析

1. 蛋白质结构基本概念

1.1. 蛋白质结构层次

1.2. 氨基酸种类和性质

1.3. 多肽链构象特征

1.4. 蛋白质分子内部作用力

1.5. 蛋白质中的水分子、金属和有机小分子

2. 蛋白质结构数据库

2.1. 蛋白质结构数据库 PDB 简介

2.2. 蛋白质结构数据库统计报表

2.3. 蛋白质结构数据库高级检索

2.4. 蛋白质结构在线显示

2.5. 蛋白质序列和结构分析

3. 蛋白质结构分析软件 ChimeraX

3.1. 用户界面

3.2. 基本步骤

3.3. 高级操作

3.4. 帮助文档

3.5. 分析实例

4. 蛋白质结构显示分析实例

4.1. 猪胰岛素 (Insulin, 4INS)

- 4.2. 人免疫球蛋白 (Immunoglobulin, IIGT)
- 4.3. 绿色荧光蛋白 (Green Fluorescent Protein, IGFL)
- 4.4. 非洲爪蟾锌脂蛋白 (Zinc Finger Protein, 1TF6)
- 4.5. 非洲爪蟾核小体 (Nucleosome, 1AOI)

第七讲 分析实例

1. 斑头雁和灰雁血红蛋白结构分析

- 1.1 研究背景
- 1.2 序列、结构比较
- 1.3 结果讨论

2. 癌胚抗原蛋白质结构预测

- 2.1 蛋白质结构预测常用方法
- 2.2 蛋白质结构预测常用网站
- 2.3 同源模建基本原理和步骤
- 2.4 癌胚抗原 (Carcinoembryonic Antigen, CEA) 结构预测实例
- 2.5 癌胚抗原 CEA21 结构预测

3. 植物转录因子家族 SBP 分析实例

- 3.1 SBP 转录因子家族研究背景
- 3.2 植物转录因子数据库 PlantTFDB
- 3.3 保守结构域预测网站 MEME
- 3.4 结构域搜索网站 SMART
- 3.5 序列谱分析工具 HMMER

4. 豌豆内膜蛋白 PPF1 分析实例

- 4.1 研究背景
- 4.2 读码框分析工具 PlotORF/ShowORF/GetORF
- 4.3 蛋白质一级结构分析工具 ProtScale
- 4.4 跨膜螺旋预测程序 TMHMM
- 4.5 蛋白质螺旋轮显示程序 PepWheel
- 4.6 亚细胞定位预测程序 TargetP
- 4.7 叶绿体定位程序 ChloroP
- 4.8 蛋白质功能预测网站 PredictProtein
- 4.9 密码子分析程序 CUSP

- 1.1 限制性内切酶分析程序 ReMap

5. 富含半胱氨酸多肽

- 5.1 研究背景
- 5.2 低相似度多肽序列比对和手工调整
- 5.3 结构预测结果分析

参考文献

1. 罗静初, 实用生物信息技术课程教学实例, 《生物技术通报》, 2015年, 第31卷, 第11期, 第102-111页.
2. 罗静初, UniProt 蛋白质数据库简介, 《生物信息学》, 2019年, 第17卷, 第3期, 第131-144页.
3. 罗静初, EMBOSS 软件包序列分析程序应用实例, 《生物信息学》, 2021年, 第19卷, 第1期, 第1-25页.
4. 罗静初, 双序列比对基础和应用实例, 《生物信息学》, 2023年, 第21卷, 第1期。
5. Luo J. Teaching the ABCs of bioinformatics: a brief introduction to the Applied Bioinformatics Course. *Brief Bioinform.* 2014 Nov;15(6):1004-13.
6. 曹志伟编译. 生物学信息学: 序列和基因组分析. 科学出版社, 2006.