

期末考试, 闭卷

地点: 二主楼 A203

时间: 6月23号星期一 14:00-15:40, 大家提前15分钟到场

携带学生证

第1章 生物信息学简介

★ 重点掌握

1. 生物信息学的跨学科定义与研究对象
2. “中心法则”
3. 研究的主要数据类型及数据来源
4. 蛋白质的一级/二级/三级结构
5. 本学期生物信息学课程讲解的几个方向, 了解这些方向的历史发展脉络、研究现状、可能的未来发展方向等

● 掌握

- 生物信息学细分方向示例 (基因组、单细胞、结构信息学等)
- 数据库基本概念与一级/二级分类框架 (衔接第4章)
- 分子生物学基础 (密码子、ORF 详解等)
- 通识性的理论、发明及其作者 (如遗传学的开创人是孟德尔等), 其他章节也有同样要求, 不再赘述

○ 了解

- 历史时间线与教材/参考书列表

第2章 序列比对

★ 重点掌握

1. Gap 概念与线性 / 仿射罚分模型、公式及示例
2. Needleman-Wunsch 全局比对算法: 矩阵递推、初始化、Traceback
3. Smith-Waterman 局部比对算法及与 NW 的两点核心差异
4. 等价最优对齐与路径偏好策略, 序列相似度 (sequence identity) 定义及计算
5. 熟练掌握 Gap Open=Gap Extension 罚分的全局/局部动态规划算法

● 掌握

- 对齐得分公式
- 仿射 Gap 动态规划三矩阵推导与复杂度

○ 了解

- 打分矩阵 (PAM、BLOSUM)
- 在线服务器 (NW-align、EMBOSS Water) 的使用界面说明

第3章 多重序列比对 & 系统发育

★ 重点掌握

1. Neighbor-Joining (NJ) 算法目标, 特点等
2. 掌握 NJ 的框架
3. 熟练掌握枝长的推导及计算方法 (已知叶子节点间距离矩阵情况下)
4. 传统定义的 MSA 和结构预测 MSA 的概念演变及区别
5. 构造多重序列比对的意义

● 掌握

- 星形比对与核心序列选择策略
- UPGMA
- 权重分配、SP-Score、NP-完全性质的工程取舍

○ 了解

- 渐进式 (ClustalW) 流程: 距离矩阵→指导树→逐步对齐
- 进化树应用 (Tree of Life、病原溯源示例)

第 4 章 大规模比对 / 数据库检索

★ 重点掌握

1. 大规模比对的定义及意义
2. BLAST 原理、启发式流程及 PSI-BLAST 扩展
3. 核心公共数据库体系:
 - INSDC 三大核酸库 (GenBank/EMBL/DBJ)
 - UniProtKB/Swiss-Prot
 (二级数据库典型)
4. 一级 / 二级数据库定义与举例; PDB 全球镜像现状

● 掌握

- FASTA 算法和 lookup table
- PSI-BLAST 中 profile 迭代与 P-value、E-value 过滤思想
- 数据库检索在同源建模中的应用场景

○ 了解

- Profile-profile 动态规划公式推导与对数似然矩阵细节
- Profile-profile 比对及 Henikoff & Henikoff 序列权重
- Profile-seq 比对

第 5 章 蛋白质结构与功能预测

★ 重点掌握

1. 蛋白质结构四级层次: 一级→二级 (α -螺旋/ β -折叠/Loops)—三级整体折叠—四级装配;
2. 2024 年诺贝尔化学奖方向
3. 蛋白质结构预测发展的历史脉络, 有哪些代表方法, 方法的基本原理, 优缺点等
4. Contact 定义及共进化
5. 掌握 3-5 种预测方法体系: 模板比较建模、Threading、I-TASSER、D-I-TASSER、ab initio、Rosetta、分子动态模拟 MD、QUARK、AlphaFold2 深度学习端到端流程,

● 掌握

- CASP 比赛评测定位

○ 了解

- 二级结构预测
- PDB 文件坐标基本格式与解析要点

参考选择题示例

1. 氨基酸多肽链是通过哪种化学键连接的? A. 氢键 B. 离子键 C. 肽键 (脱水缩合) D. 范德华力 答案: C
2. 蛋白质四级结构指: A. 氨基酸序列 B. α -螺旋和 β -折叠 C. 多个亚基的空间组合 D. 原子间的氢键 答案: C
3. Watson 和 Crick 提出 DNA 双螺旋结构基于谁的 X 射线衍射实验? A. Maurice Wilkins 和 Rosalind Franklin B. Margaret Dayhoff C. Masatoshi Nei D. Sebastian Bassi 答案: A
4. DNA 序列由哪 4 种核酸组成? A. A, C, G, T B. A, U, G, C C. G, C, T, U D. A, T, U, G 答案: A
5. 替换突变示例中, ACT→AT 属于哪种类型? A. 插入 B. 删除 C. 替换 D. 重排 答案: B
6. 在全局比对中, 若 Gap open = Gap extension, 则罚分公式为: A. $w(k)=gk$ B. $w(k)=g+e(k-1)$ C. $w(k)=0$ D. $w(k)=\log k$ 答案: A
7. 结构预测领域常用的 MSA 特点是:

- A. 有中心序列, 只有同源序列可插 Gap
- B. 所有序列等价
- C. 中心序列可插 Gap
- D. 不允许 Gap 答案: A

8. UPGMA 算法的“UPGMA”全称是:

- A. Unweighted Pair Group Method with Arithmetic mean
- B. Unified Protein Grouping Method Algorithm
- C. Universal Phylogeny Generation Model
- D. Ubiquitous Pairwise Genomic Matching 答案: A

9. Neighbor-Joining 方法构建的是:

- A. 有根树
- B. 无根树
- C. 全局比对
- D. 局部比对 答案: B

10. E-value 与 P-value 的区别是:

- A. E-value 指期望数, P-value 指概率
- B. 二者相同
- C. P-value 指期望数
- D. E-value 仅用于蛋白质 答案: A

11. QUARK 方法基于:

- A. 深度学习
- B. 片段组装
- C. UPGMA
- D. X 射线实验 答案: B

12. Contact map 预测常用的信息来源是:

- A. 氨基酸序列共进化信号
- B. 质谱数据
- C. 进化树
- D. UPGMA 距离 答案: A