

生物信息学札记¹¹

樊龙江

浙江大学生物信息学研究所/作物科学研究所
Bioinplant 实验室 (Bioinplant Lab), 华家池校区, 杭州 310029
fanlj@zju.edu.cn

2001 年 6 月

¹¹这份材料是我学习和讲授《生物信息学》课程时的备课笔记，材料大多是根据当时收集的一些外文资料翻译编辑而成。学生在学习过程中经常要求我给他们提供一些中文的讲义或材料，这促使我把我的这份笔记整理并放到网上，供大家参考。要提醒使用者的是，这份材料仅是根据我对生物信息学的一些浮浅的认识整理而成，其中的错误和偏颇只能请读者自鉴了。

本材料可通过 Bioinplant Lab 网址获得：

www.cab.zju.edu.cn/cab/xueyuanxiashubumen/nx/bioinplant.htm

《生物信息学札记》目录

第一章 生物信息学通论	(1)
第一节 生物信息与生物信息学	
一、迅速膨胀的生物信息	
二、生物信息学的概念	
第二节 生物信息学发展简史	
第三节 基因组时代：生物信息学的应用与展望	
第二章 分子数据库	(10)
第一节 初级数据库	
一、DNA 数据库	
二、基因组数据库	
三、蛋白质序列数据库	
四、蛋白质结构数据库	
第二节 初级序列数据的注释	
第三节 数据库信息检索系统	
第四节 数据库的冗余与偏误	
第五节 向数据库发送序列数据及其它	
第三章 序列分析与比较	(24)
第一节 序列组成和单一序列分析	
一、碱基组成	
二、碱基相邻频率	
三、同向重复序列分析	
四、DNA 序列的几何学分析——Z 曲线	
第二节 序列联配	
一、Needleman-Wunsch 算法	
二、Smith-Waterman 算法	
三、Karlin-Altchul 统计方法	
1、二进制值(Bit Score)或标准比值; 2、P 值(P-value);	
3、BLAST 和 FASTA 的数据库搜索策略; 4、空位列线	
(gapped alignment) 的统计问题; 5、边际效应(edge	
effect); 6、替换矩阵的选择; 7、空位罚值(gap	
penalties)	
四、FASTA 算法	
五、替换矩阵	
1、替换矩阵的一般原理; 2、PAM 氨基酸替换矩阵; 3、	
BLOSUM 氨基酸替换矩阵; 4、DNA 替换矩阵	

六、多序列联配	
第三节 数据库搜索引擎——BLAST 和 FASTA 应用.....	(50)
一、数据之海与一叶轻舟	
二、BLAST: 核酸数据库搜索	
1、BLAST 实战操作(1); 2、BLAST 的检索报告; 3、BLAST 选项; 4、BLAST 实战操作(2)	
三、BLAST: 蛋白质数据库搜索	
四、FASTA: 另一种搜索策略	
1、FASTA 选项; 2、FASTA 实战操作及其检索报告	
第四节 限制性酶作图和寡核苷酸设计	
一、限制性酶作图	
1、限制性酶位点寻找; 2、绘制限制酶作图	
二、寡核苷酸设计	
1、引物设计; 2、用于检测相关基因的简并探针	
第四章 基因组测序与分析	(65)
第一节 DNA 测序与序列片段的拼接	
一、DNA 测序的一般方法	
1、DNA 测序的基本原理; 2、双脱氧测序法(Sanger 法);	
3、化学测序法(Maxam-Gilbert 法); 4、荧光自动测序仪	
二、DNA 测序策略	
1、从遗传图谱、物理图谱到基因组序列图谱; 2、鸟枪测序法(shotgun sequencing); 3、引物步查法(primer walking); 4、限制性酶切-亚克隆法(restriction endonuclease digestion and subcloning)	
三、序列片段的拼接方法	
第二节 基因区域的预测	
一、从序列中寻找基因	
1、基因及基因区域预测; 2、发现基因的一般过程; 3、解读序列(making sense of the sequence)	
二、基于编码区特性: 最长 ORF 法等	
三、数据库相似性搜索法	
四、神经网络	
五、隐马尔可夫模型(HMM)	
六、RNA 二级结构预测	
第三节 基因组的比较	
一、比较基因组学	
二、基因组多样性	
三、基因组比较的方法	
第五章 分子进化	(97)
第一节 系统树及其它	

一、系统树	
二、遗传模型和序列距离	
三、分子进化与系统发育分析软件	
第二节 距离矩阵法	
一、平均连接聚类法 (UPGMA 法)	
二、Fitch-Margoliash 算法	
三、邻接法	
第三节 简约法	
第四节 似然法	
一、DNA 序列的似然模型	
二、两条序列的系统树	
三、多条序列的系统树	
四、对系统树 Bootstrap 抽样	
第六章 蛋白质结构与功能预测 (119)
第一节 蛋白质功能预测	
一、根据序列预测功能的一般过程	
二、通过比对数据库相似序列确定功能	
三、序列特性: 疏水性、螺旋等	
四、通过比对模序数据库等确定功能	
第二节 蛋白质结构预测	
一、蛋白质结构及其数据库	
二、二级结构预测	
三、三级结构预测	

附录:

生物信息学主要英文术语及释义 (125)
核苷酸和氨基酸代码 (150)
分子生物学主要数据库和应用工具网址一览表 (152)
序列分析软件目录 (198)
与核苷酸和蛋白质序列相关的特征关键词表 (207)
主要参考文献 (213)