

<<DNA和蛋白质序列数据分析工具>>

图书基本信息

书名：<<DNA和蛋白质序列数据分析工具>>

13位ISBN编号：9787030270528

10位ISBN编号：7030270525

出版时间：2010-4

出版时间：科学

作者：薛庆中

页数：275

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<DNA和蛋白质序列数据分析工具>>

前言

当今生物基因组DNA测序数据总量正在以指数倍的速度增长。如何对数据库的海量数据进行科学的搜集、管理、挖掘、注释已成为基因组学和蛋白组学研究的热点。为普及和提高我国科学工作者基因组科学知识，学习并掌握序列数据分析的实用操作技能，及时了解该领域的最新进展，自2003年以来，浙江大学和中国科学院基因组研究所紧密协作已举办了24期基因组科学培训班。

培训学员来自全国各地29个省市，人次多达1800余人，每次培训班中都不仅常见到较多教授和副教授们的身影，年轻的研究生更是踊跃参加。他们的专业背景虽然各自不同，涵盖理学、工学、医学、农学等不同门类和学科，但渴求知识、不断进取的态度却是一样的。

基因组科学培训班由杨焕明、于军、郑树、林标扬、胡松年、薛庆中、徐宇虹等教授担当主讲教师。

他们不仅对基因组科学的基本概念加以正确诠释，对当今的最新进展进行全面介绍，并能结合自己的科研工作，分别讲解他们在医学、农学、微生物等领域具体的应用实例。

学员们反映，通过这些生动趣味的讲座加深了他们对DNA数据挖掘的理解，有助于开阔研究视野和工作思路，同时激发了学习这门前沿科学的兴趣和热情。

培训班主要学习数据库搜索和实用工具的操作，采用“跟我学”的教学方式，指导教师边讲边示范，学员们每人备有电脑，跟着大屏幕一步步操作；辅导教员随时在旁帮助解难，使学员们在较短时间内尽快初步掌握基本操作程序。

为满足培训的需要，我们先后编写了《基因组数据分析手册》（胡松年和薛庆中，2003）和《EST数据分析手册》（胡松年，2005），得到良好的反映和发行量。

教学内容的不断更新是培训班久盛不衰的保证。

近期培训内容中我们又新增芯片数据、蛋白质数据和系统生物学网络结构显示与分析等内容。

为此，在前两本书的基础上，我们新编写了《DNA和蛋白质序列数据分析工具》一书。

全书分9章。

第1章，阐述序列比较的核心方法，即运用BLST和ClustalX等工具做序列比对。

第2章，重点介绍核苷酸序列分析工具，主要包括：基因可读框的识别，CpG岛、转录终止信号和启动子区域的预测分析，用mRNA序列预测基因等。

第3章，介绍电子克隆的概念和具体操作方法。

第4章，用MEGA4做分子进化遗传分析，绘制系统进化树，为研究基因进化打好基础。

第5章，对蛋白质基本理化性质、二级结构、结构域和三维空间结构、预测目标蛋白的生物学功能等工具做逐一介绍。

第6章，通过GeneOntology和KEGG两个数据库，挖掘基因和蛋白质的功能并做代谢途径分析。

<<DNA和蛋白质序列数据分析工具>>

内容概要

在众多生物基因组测序项目完成之际，我们面临的重大挑战是如何对DNA和蛋白质数据进行分析科学的分析和注释。

《DNA和蛋白质序列数据分析工具(第2版)》分三个层次解读基因数据库和网络工具：基因组学层面重点介绍序列比对工具BLAST和ClustalX的使用、真核生物基因结构的预测、电子克隆及分子进化遗传分析工具(MEGA4)的使用；蛋白质组学层面介绍了蛋白质结构与功能预测、序列模体的识别和解析、蛋白质谱数据分析、基因芯片数据处理和分析，以及应用GO注释基因功能和通过KEGG分析代谢途径；系统生物学层面从网络结构分析阐述了蛋白质与蛋白质的相互作用；此外，还增添了使用Bioperl模块进行数据分析和Windows操作系统下Bioperl程序包的安装等内容。

书中提及的各种方法均有充实的例证并附上相关数据和图表，供读者理解和参考；书后还附有中英文的专业术语和词汇。

《DNA和蛋白质序列数据分析工具(第2版)》可作为对生物信息学专业感兴趣的本科生、研究生和研究人员学习、研究的重要工具手册。

书籍目录

第二版前言 第一版前言 第1章 序列比对工具BLAST和ClustalX的使用 1.1 序列比对BLAST 1.1.1 Basic BLAST 1.1.2 网上blastx比对 1.1.3 网上PSI.Blast比对 1.1.4 Specialized BLAST 1.1.5 网上Blast2比对 1.2 本地运行BLAST(Windows系统) 1.2.1 BLAST程序下载 1.2.2 BLAST程序安装 1.2.3 进入DOS命令行提示符状态 1.2.4 搜索数据库的格式化 1.2.5 BLAST搜索程序运行 1.2.6 本地化BLAST搜索结果查看 1.3 多序列比对(ClustalX) 1.3.1 ClustalX的使用 1.3.2 数据的输入 1.3.3 数据的输出 主要参考文献 第2章 真核生物基因结构的预测 2.1 基因可读框的识别 2.2 CpG岛、转录终止信号和启动子区域的预测 2.2.1 CpG岛的预测 2.2.2 转录终止信号的预测 2.2.3 启动子区域的预测 2.3 基因密码子偏好性计算: CodonW的使用 2.4 采用mRNA序列预测基因: Spidey的使用 2.5 ASTD数据库简介 主要参考文献 第3章 电子克隆 3.1 利用UniGene数据库进行电子延伸 3.1.1 目标序列的检索 3.1.2 UniGene数据库检索 3.2 从数据库中获取CDNA全长序列 3.3 本地序列拼接 3.3.1 CAP3序列拼接程序 3.3.2 Velvet序列拼接程序 3.4 基因的电子表达谱分析 3.5 核酸序列的电子基因定位分析 主要参考文献 第4章 分子进化遗传分析工具(MEGA4)的使用 4.1 序列数据的获取和比对 4.1.1 数据库直接检索 4.1.2 多序列比对 4.2 进化距离的估计 4.3 分子钟假说的检验 4.4 系统进化树构建 4.4.1 系统进化树构建方法选择 4.4.2 进化树的树形选择 4.4.3 进化树的拓扑结构调整 4.4.4 进化树树枝形态的优化 4.4.5 进化树的保存 主要参考文献 第5章 蛋白质结构与功能预测 5.1 蛋白质一级结构分析 5.1.1 ProtParam: 蛋白质序列理化参数检索 5.1.2 ProtScale: 蛋白质亲疏水性分析 5.1.3 COILS: 卷曲螺旋预测 5.2 蛋白质二级结构预测 5.2.1 PredictProtein: 蛋白质结构预测 5.2.2 PSIPRED: 不同蛋白质结构预测方法 5.3 InterProScan: 模式和序列谱研究 5.3.1 InterPro' Scan简介 5.3.2 PROSITE: 蛋白质结构域、家族和功能位点数据库 5.3.3 Pfam: 蛋白质家族比对和HMM数据库 5.3.4 BLOCKS: 模块搜索数据库 5.3.5 SMART: 简单模块构架搜索工具 5.3.6 TMHMM.跨膜区结构预测服务器 5.4 蛋白质三级结构预测 5.4.1 Swiss.ModelWorkspace: 同源建模的网络综合服务器 5.4.2 Phyre(Successor of 3D.PSSM): 线串法预测蛋白质折叠 5.4.3 HMMSTR / Rosetta: 从头预测蛋白质结构 5.4.4 Swiss.PdbViewer: 分子建模和可视化工具 主要参考文献 第6章 序列模体的识别和解析 6.1 MEME程序包 6.2 通过MEME识别DNA或蛋白质序列组中模体 6.3 通过MAST搜索序列中的已知模体 6.4 通过GLAM2识别有空位的模体 6.5 通过GLAM2scAN搜索序列中的已知模体 6.6 应用TOMTOM与数据库中的已知模体进行搜索比对 6.7 应用GOM0鉴定模体的功能 主要参考文献 第7章 蛋白质谱数据分析 7.1 生物质谱技术介绍 7.1.1 质谱技术的基本原理 7.1.2 x!Tandem软件 7.1.3 Mascot软件 7.1.4 Sequest软件 7.2 蛋白质组学数据统计分析软件 7.2.1 TPP简介 7.2.2 TPP的安装与配置 7.2.3 样本数据准备 7.2.4 将RAW文件转换成mzXML文件 7.2.5 由out数据文件夹生成pepXML文件 7.2.6 运行PeptideProphet 7.2.7 PeptideProphet处理后的结果分析 7.2.8 运行ProteinProphet 7.2.9 数据的过滤筛选和将结果保存成Excel文件 主要参考文献 第8章 基因芯片数据处理和分析 8.1 芯片数据的获取和处理 8.1.1 ExpressCOnVener 8.1.2 MIDAS 8.2 芯片数据聚类分析和差异表达基因筛选 8.2.1 MeV 8.2.2 Cluster 8.2.3 TreeView 8.3 芯片数据的可视化 8.3.1 GenMAPP的概念 8.3.2 GenMAPP的安装 8.3.3 GenMAPP的使用 8.4 芯片数据的检索和提交 8.4.1 GEO检索 8.4.2 Platform信息 8.4.3 Series信息 8.4.4 Samples信息 8.4.5 芯片数据的提交 主要参考文献 第9章 应用GO注释基因功能和通过KEGG分析代谢途径 9.1 GeneOntology数据库 9.1.1 简介 9.1.2 用关键词检索GO数据库 9.1.3 用序列检索GO数据库 9.2 KEGG数据库 9.2.1 简介 9.2.2 根据代谢途径名称检索 9.2.3 根据基因名称检索 9.2.4 根据序列检索 9.2.5 利用KAAS工具作批量注释 9.2.6 基因芯片数据的代谢途径分析 主要参考文献 第10章 系统生物学网络结构分析 10.1 Cytoscape软件简介..... 第11章 使用Bioperl模块作数据分析 第12章 Winodws环境下Bioperl程序包的安装英汉对照词汇彩图

章节摘录

和原核生物相比，真核生物不仅在细胞结构上存在明显差异，并且在基因结构上也更为复杂。模式生物全基因组测序计划的完成，使得基因结构和功能的预测已成为可能。

本章将重点描述真核生物基因可读框、CpG岛、转录终止信号、启动子、密码子偏好等结构，选择介绍一些常用核苷酸序列分析工具，帮助读者了解基因结构预测的方法。

2.1 基因可读框的识别 可读框 (open reading frame, ORF) 指的是从5'端翻译起始密码子 (ATG) 到终止密码子 (TAA、TAG或TGA) 的蛋白质编码碱基序列。

真核生物的可读框除外显子外，还含有内含子，其长度变化范围非常大，因此真核生物基因预测远比原核生物困难。

基因预测软件GENSCAN由斯坦福大学开发 (Burge and Karlin, 1998)，它是针对基因组DNA序列预测可读框及基因结构信息的开放式在线资源，尤其适用于脊椎动物、拟南芥和玉米等真核生物。

进入GENSCAN页面 (图2.1)，选择物种，上传或直接粘贴序列，运行后便可获得提交序列中所包含的基因数目、外显子数目和类型，预测单元的长度、方向、位置及相位、编码区打分值、可信概率、总得分值等预测结果。

例如，提交一个人类cosmid序列 (GenBank号：AC002390)，预测结果表明该克隆存在两个基因，其中第1个基因的起始外显子从532碱基处开始，到657碱基处结束；接着有9个中间外显子；终止外显子在51783碱基处结束，其后还有polyA信号。

在第2个基因起始外显子的前端有启动子区域，它始于59901碱基处 (图2.2)。

由于提交的序列长度只有70311个碱基，所以这里预测的第2个基因的结构并不完整。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>