研究生课程教学大纲

课程中文名称：实用生物信息技术

课程英文名称：Applied Bioinformatics Course (ABC)

课程编号：B02213

一、课程概述

本课程旨在利用丰富的网络生物信息资源和分析工具，解决自己正在进行或即将开始的研究课题中的实际问题。本课程是一门以上机操作为主的实验课，采用腾讯会议线上教学，并通过电子邮件和微信交流等网络教学手段，进行课外讨论和辅导。选修本课程的同学以便于课外一起讨论为原则结合成小组，除了课堂教学外，还进行课外小组讨论、完成必要的课外练习和期末总结。

本课程与分子育种、基因工程、比较基因组学与分子进化、基因编辑技术、代谢组学等分子生物学相关课程相辅相成。

二、课程学分、学时

学分：2.5

总学时： 48 ，其中讲授 32 学时，实验实习 16 学时。

三、先修课程

普通生物学或细胞生物学、生物化学、分子生物学或分子遗传学。

四、课程目标

通过本课程学习，学会网络文档查阅、数据库查询、数据库相似性搜索，核酸和蛋白质序列分析、蛋白质结构分析和预测、分子系统发育树构建等常用生物信息技术。

1. 熟练使用 CNCB、NCBI 和 EBI 等国内外著名生物信息中心数据资源、分析工具和网络教程。熟练下载安装和使用 TBtools、mEMBOSS 等国内外常用生物信息软件。
2. 学会 PubMed 文献高级检索，了解存储系统 MyNCBI 和医学主题词 MeSH 等特色工具基本用法。
3. 熟练使用 UniProt 蛋白质序列数据库、RefSeq 核酸和蛋白质参考序列数据库、ENSEMBL 基因组数据库；通过北京基因组研究所国家生物信息中心数据库浏览和搜索功能，查找课题相关物种、相关基因和蛋白质专用数据库。
4. 了解序列比对的基本原理和常用算法，理解和熟练进行双序列整体或局部比对，掌握计分矩阵和空位罚分选择和设置方法，并对比对结果的生物学意义进行分析。
5. 了解 BLAST 数据库相似性搜索的基本原理，全面系统掌握 NCBI 网站提供的各种蛋白质和核酸序列 BLAST 搜索程序；学会如何选择不同数据库和不同物种；理解种子词长 WORD、期望值 E、计分矩阵和空位罚分等参数对搜索结果的影响；熟悉搜索结果中得分、覆盖率、相似性等含义；熟练运用结果筛选、格式转换等工具。

6. 深入了解相似性和同源性、直系同源和旁系同源、基因树和物种树等基本概念；通过 TimeTree 物种演化网站，了解课题相关物种分类学地位、起源时间、分歧年代和系统发生关系；熟练掌握 MEGA 等常用软件构建系统发生树的方法；了解不同程序、不同参数的选择和对结果的影响；学会所构建的系统发生树的不同展示方式，利用 iTOL 网站进行修饰；搞清所构建的系统发生树的生物学意义。
7. 熟悉蛋白质结构数据库 PDB 的查询、浏览、结构显示和比较等基本功能，了解 SCOP 和 CATH 等蛋白质结构分类数据库的用途；熟练掌握 Swiss-PDBViewer 和 PyMOL 等蛋白质结构显示和分析软件的使用方法，对血红蛋白、胰岛素、免疫球蛋白、核小体、绿色荧光蛋白等典型蛋白质及课题相关蛋白的结构功能关系进行深入分析。
8. 通过蛋白质结构预测网站 CASP 了解蛋白质结构预测主要方法、优缺点和适用范围；通过 UniProt 序列条目中相关链接了解课题相关蛋白质或其它生物中的同源蛋白的结构；熟练掌握利用 SwissModel 进行同源模建；对构建的模型及其模板进行深入分析，探索序列、结构、功能关系。
9. 通过斑头雁血红蛋白 Hemoglobin、豌豆内膜蛋白 PPF1、癌胚抗原 CEA、植物特异转录因子 SBP 家族、抗菌肽 AFP 和蜘蛛毒素 HWT 等富含半胱氨酸的多肽，以及金属硫蛋白 MT 等典型蛋白质分析，了解课题相关蛋白质及其编码基因的基本分析方法。

五、适用对象

适用于从事分子生物学、遗传学和基因组学等相关课题研究的硕士和博士研究生。

六、授课方式

本课程采用课堂讲授和上机实习同步进行的方式，充分利用腾讯会议、电子邮件、微信等基于网络的先进技术，以及 DeepSeek 等人工智能软件平台，特别是任课教师本人创建、维护、更新的本课程教学专用网站（abc.gao-lab.org 或 abc.cbi.pku.edu.cn）。

七、课程内容

第一章 网络文档

第一节 分子月报

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。
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。
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。
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。

第二节 蛋白质分子精选

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. 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 Peruts 对血红蛋白三维结构测定所做的贡献及其背后的故事。
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。
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。

第三节 网络教程

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。
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. 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. 熟悉伯克利大学 Evolution 101 演化生物学网络教程。
5. 通过 TimeTree 网站熟练掌握查找物种之间的亲缘关系。
6. 通过 TWiW 网站熟悉病毒基本知识。

第四节 PubMed 文献

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。
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。
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. 运用相似文献和引用文献等功能，扩充检索范围。
5. 熟练掌握文献保存、共享、导出、发送等基本功能.
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。
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。

第二章 序列比对

第一节 序列比对基本概念

序列相似性（Similarity）和序列同源性（Homology）

1. 直系同源（Ortholog）和旁系同源（Paralog）
2. 动态规划算法（Dynamic Programing）和启发式算法（Heuristic Programing）
3. 计分矩阵（Scoring Matrix）和空位罚分（Gap Penalty）
4. PAM (Point Accept Mutation) 计分矩阵和 BLOSUM (Block Substitution)计分矩阵
5. 全局比对（Global Alignment）和局部比对（Local Alignment
6. 双序列比对（Pairwise Sequence Alignment）和多序列比对（Multiple Sequence Alignment）
7. 序列比对点阵图方法（Dot Plot）
8. 多序列比对序列图标（Sequence Logo）

5. 如何选择搜索字长 WORD

第三节 NCBI Blast 平台其它工具

1. NCBI Blast 系统用户界面和常用程序
2. 模式生物蛋白质序列数据库搜索程序 SmartBlast
3. 测序载体接头序列检测程序 VecScreen
4. 引物设计程序 Primer-Blast
5. 免疫球蛋白数据库搜索专用程序 IgBlast

第四节 Blast 使用经验点滴

1. 如何选择数据库
2. 如何选择物种
3. 如何选择不同程序
4. 如何改变输出格式
5. 如何分析输出结果
6. 如何下载输出结果
7. 如何保存搜索策略
8. 如何搜索短序列
9. 如何屏蔽重复序列

第五节 Linux 系统本地 Blast 简介

1. Linux 系统简介
2. Linux 系统常用命令
3. SBP 转录因子本地 Blast 运行实例
4. SBP 转录因子搜索结果分析

第五章 系统发生树构建

第一节 分子演化和系统发生树基本概念

1. 物种分化和分子演化
2. 分支图和系统树
3. 物种树和基因树
4. 有根数和无根树
5. 二叉树与多岐树
6. 外部节点和内部节点
7. 内部节点和根节点
8. 系统发生树稳定性检验

第二节 常用系统发生树构建方法

1. 距离法
2. 最大简约法
3. 最大似然法
4. 贝叶斯推断
5. 组分矢量 CVTree 方法

2. 蛋白质结构预测常用网站
3. 同源模建基本原理和步骤
4. 癌胚抗原（Carcinoembryonic Antigen, CEA）结构预测实例
5. 癌胚抗原 CEA21 结构预测

第六节 富含半胱氨酸多肽

1. 研究背景
2. 低相似度多肽序列比对和手工调整
3. 结构预测结果分析

课程重点

本课程是一门上机操作实验课，结合具体实例，介绍常用生物信息技术和分析思路，为课题研究提供参考。本课程的特色可以概括为三个“题”：结合课题、带着问题、多做例题，对应英文三个 P：Project, Problem, Practice。

课程难点

本课程重点为掌握生物信息数据库资源和软件工具的实际使用，选修本课程的学生必须具有较好的分子生物学和生物化学基础，具备课外自由上机的能力和上网条件，必须保证每周 6 学时以上的课外上机条件。

八、思政元素有机融入课程情况

序号	章节位置	课程思政目标	课程思政内容概述
1	第一章 第一节/第二节	培养学生从事科学研究必须具有无私奉献精神	通过介绍分子月报（Molecule of the Month）、蛋白质分子精选（Protein Spotlight）等科普网站专栏作家自 2000 年起，每月撰写一篇蛋白质结构或蛋白质序列功能科普文章： https://pdb101.rcsb.org/motm/ https://www.proteinspotlight.org/ 鼓励同学从事科学研究必须有持之以恒的奉献精神。
2	第三章 第一节	鼓励学生在研究生期间发挥主观能动性	介绍蛋白质序列和功能数据库创始人 Amos Bairock 二十世纪八十年代就读研究生期间创建了蛋白质序列注释数据库 Swiss-Prot，该数据库九十年代与 TrEMBL 合并，成为国际上最著名的蛋白质序列和功能数据库 UniProt（https://www.uniprot.org/），为全世界从事生命科学研究的人员提供无偿服务。
3	第三章 第四节	国家生物信息中心创建	1999 年 6 月，郝柏林院士写了“建议尽快组建国家级的生物医学信息中心”的院士建议。2019 年，中国科学院北京基因组研究所加挂“国家生物信息中心”牌子，成为国际上主要生物信息中心之一。https://www.cncb.ac.cn/
4	第五章 第二节	生物信息领域创新性研究	郝柏林院士二十世纪五十年代就读于苏联，曾任中国科学院理论物理研究所所长，九十年代起从事数学、物理在生命科学领域中的应用，提出了不用序列比对的组分矢量（Composition Vector）

			方法，并用于细菌基因组分类，取得了创新性成果。曾多次应邀到农科院研究生院作学术报告，鼓励同学在各自研究领域中勇于探索。2018 年去世后，基因组研究所建立了郝柏林院士纪念网站：https://ngdc.cncb.ac.cn/education/biography/haobl/
5	第八章 第 4 节	生命科学领域创新性研究	蛋白质序列-结构-功能关系是生命科学领域重要研究方向之一，斑头雁每年春秋两季飞跃高海拔、低氧分压的喜马拉雅山。北京大学生命科学学院顾孝诚教授及其研究团队测定了斑头雁及其近缘物种灰雁的血红蛋白分子结构，揭示了斑头雁适应长途迁徙的分子机制。顾孝诚教授曾任 CUSBEA 项目中方负责人，二十世纪八十年代选送了大批优秀留学生。2012 年去世后，她的同事和学生自发撰写了许多悼念文章：https://ngdc.cncb.ac.cn/education/biography/guxc/

九、考核要求

本课程不进行笔试，以课堂表现、课外交流、平时练习、期末总结等为主要考核指标。

期末（30 分）：

期末总结报告 WORD 文档（15 分），期末小组交流汇报 PPT 文档（15 分）

平时（70 分）：

考勤（10 分）。全勤 10 分，无故缺勤 3 次或 3 次以上者本课程成绩不及格；无故缺勤超过总学时三分之一者总成绩为零分。

课堂表现（20 分），包括课上演示、提问、解答问题等。

课外练习（20 分），共 5 次练习，每次 4 分。

小组讨论（20 分），每周至少参加 3 学时小组讨论，轮流撰写小组讨论总结报告，平均每人 4 次，每次 5 分。

十、课程资源

常用网站：

由主讲教师构建、维护、更新的本课程专用教学网站：

https://abc.gao-lab.org/ 或 http://abc.cbi.pku.edu.cn/

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，第 1-19 页.
5. 罗静初，序列数据库搜索系统 BLAST 简介，《生物信息学》，2015 年（知网首发）。

6. 罗静初，顾孝诚教授与北京大学生物信息中心，《中国科学: 生命科学》，2022 年第 52 卷，第 10 期，第 1555-1560 页.
7. 罗静初，EMBOSS 和 EMBnet，《生物信息学》，2021 年，第 19 卷，第 4 期，第 223-231 页.

十一、编写成员名单与编写日期

编写成员：罗静初 教授（北京大学生命科学学院）

编写日期：2025 年 5 月