

实用生物信息学

Practical Bioinformatics

(以下内容标题为黑体 4 号，具体内容为五号宋体)

课程编号：4018-1

课程属性：必修课

讲课课时：30

实验实习课时：14

讨论课时：4

学分：2.5 (共 48 课时)

教学对象：硕士生、博士生

知识背景：分子生物学、遗传学、基因组学

教学目的和要求：

本课程旨在介绍在生物信息学中最常用的编程语言 (Unix shell, R 和 Python) 中的最基础和最核心的部分，讲解常用的生物信息学分析软件包的基本原理和使用方法，希望学生在掌握本课程的知识技能后，能够自己解决正在进行或即将开始的研究课题中的实际问题。本课程不是以课堂讲授为主的理论课，而是一门将老师讲解、上机操作和项目制学习相结合的课程。除课堂教学外，还将研究领域相似的学生划分成不同的小组，整个课程期间小组内同学互相学习和讨论、完成必要的课外练习。通过本课程学习，除了掌握常用编程语言的核心用法外，学生们还将掌握连接远程服务器，记录并备份编程语言，对 Illumina 数据进行数据比对，基因差异表达分析，ChIP-Seq 数据分析，关联图谱分析，群体遗传结构分析，基因功能区域划分和最基础的机器学习原理等。课程的最后，小组内同学在领域内选择已经发表的有关处理大数据的文章，讨论基于已经发表的数据可以设计一些新的分析方法，解决一些新的问题，最后以 PPT 展示出来。对于优秀的成果，可以后期组织学生撰写论文。

本课程面向生命科学领域从事分子生物学、遗传学和基因组学等相关课题研究的硕士和博士研究生。选修本课程的学生，必须具有较好的生物学基础知识，必须具有较好的英语阅

读理解能力，必须保证足够课外时间利用计算机完成课堂作业和进行小组讨论。特别欢迎已经进入实验室开始或即将开始分子生物学、基因组学、遗传学、微生物学和免疫学 等相关领域研究的学生选修本课程，鼓励博士生、博士后和其他科研人员选修，带着课题研究中遇到的实际问题，边学边用。

教学方法与教学设备要求：

本课程采用课堂讲授和上机实习同步进行的方式，必须具备较好网络环境和 PPT 放映系统。

课程重点与难点：

本课程重点为掌握最常用的编程语言（Unix shell, R 和 Python）中的最基本和最核心的元素，熟悉常见生物信息学软件工具的实际使用，选修本课程的学生必须具备课外自由上机的能力和上网条件，必须保证每周 至少 4 学时的课外上机条件。

考试方式：

本课程以开卷考试为主，采用项目制的考试方式。将同学根据其细分领域分成小组，在领域内选择已经发表的有关处理大数据的文章，讨论基于已经发表的数据可以设计一些新的分析方法，解决一些实际问题，以 PPT 展示和撰写论文为主。

课程内容：

第一章：基本的生物信息学工具

第一次：为什么学习生物信息学？—— yuwen and li wang

第二次：连接远程服务器& git 命令操作指南 —— yi zou & li wang

第三次：基本的 Unix 操作命令—— xuezhu

第二章：R 编程语言

第一次：R 语言相关软件的介绍和安装（R, Rmarkdown, Rstudio）读取和存储数据—— yang fu

第二次：基本的统计分析（chi-square, t.test, permutation, hypergeometric test）--liu yuwen

第三次：R 的数据可视化 box plot, point plot, linear regression etc —— li wang

第三章：Python 编程语言

第一次：Python 基本用法；Jupyter 笔记本；读写文件；模块和程序包 (Cheng Li)

第二次：pandas 软件包—— Weigang

第三次：Week 10 (bioinformatic pipeline with unix and python) (Weigang) & SOS pipeline (Yang fu)

第四章：多组学分析实用工具介绍：

第一次：基因组数据分析： phylogenetic tree；群体遗传结构的分析 STRUCTURE, PCA, ADMIXTURE； GWAS ----- Cheng li & li Wang

第二次：**转录组数据分析**：序列比对： Bowtie, B；基因表达定量；基因差异表达： EdgeR, DeSeq2, ----- Xuezhu & Cheng li

第三次：表观调控组学：ChIP-Seq Peak 分析:贝叶斯模型；软件应用 Macs2 —— Chao Wang

第四次：基因功能区域划分：马可夫链；软件应用 ChroHMM, Phastcons ----- Yuwen Liu

第五次：机器学习: CNN 模型；软件应用 DeepSEA ----- Wenlong Ma

教材或参考书：

1. Vince Buffalo. Bioinformatics Data Skills. O'Reilly, 2015.
2. Steven H. D. Haddock, Casey W. Dunn. Practical Computing for Biologists. Sinauer Associates, 2011

撰写人：王丽，刘毓文（中国农业科学院农业基因组研究所）

撰写日期： 2020 年 4 月