

《Linux 生物信息基础》小组讨论总结报告

第 4 组，第 11 次讨论

组长：陈奕晗

执笔：高培翔

1 时间

2021 年 6 月 17 日，15:00 ~ 18:00

2 地点

泊星地咖啡厅

3 人员

陈奕晗、邹济平、朱瑾煜、高培翔

4 方式

线下讨论

5 主题

5.1 综合课题进度及报告计划

5.2 MEGA

5.3 Swiss-PDBViewer

6 内容

6.1 综合课题进度及报告计划

我们的综合课题-植物MADS-box转录因子数据库已基本成型，可以实现浏览、高级检索、BLAST、下载序列功能，并配套了帮助、关于页面。

我们小组报告计划：

- 1.介绍项目的动机与前期数据收集过程。
- 2.展示网站的现有功能及用法，介绍各个页面的详情。

3.介绍项目从后台数据到前端实现过程，展示网站的结构。

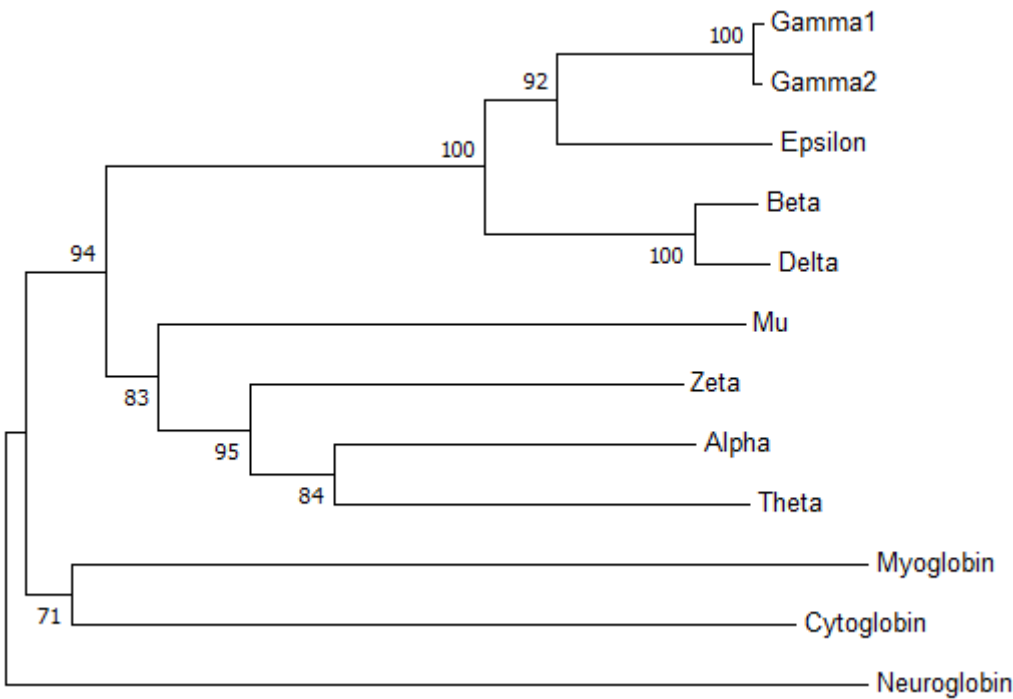
6.2 MEGA

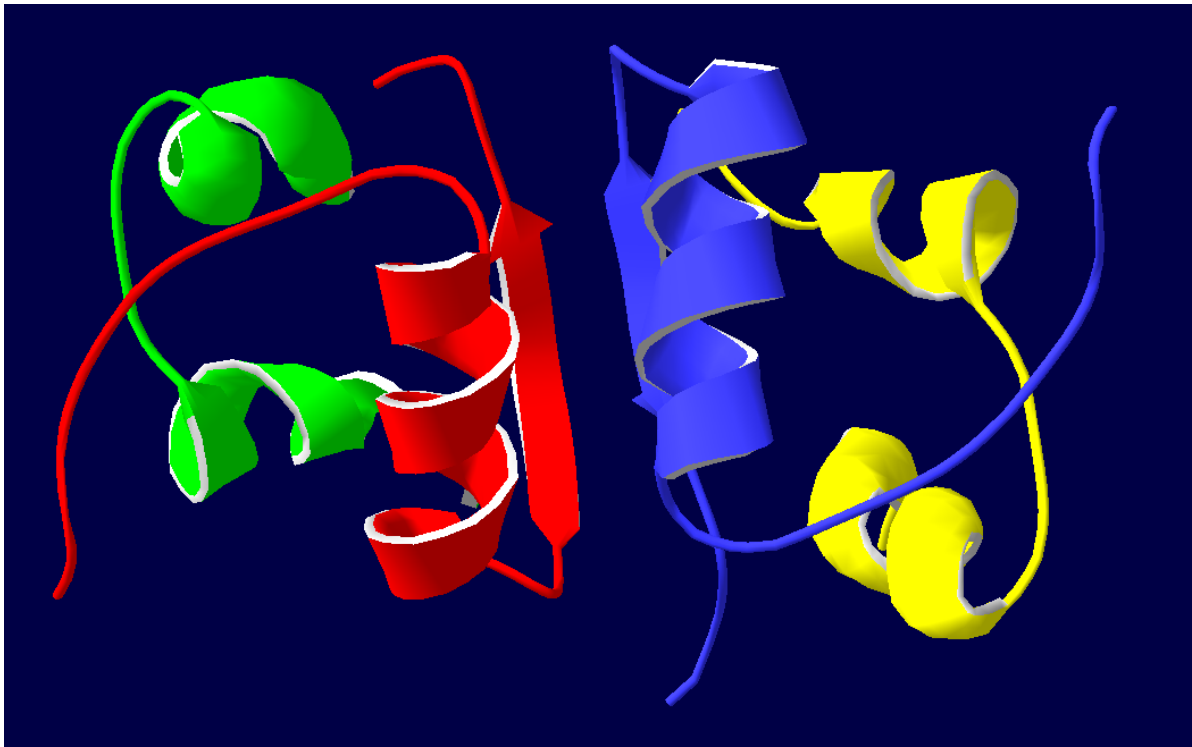
MEGA (Molecular Evolutionary Genetics Analysis) 是一款用于分析核酸序列/蛋白质序列的系统发生关系的软件，有图形界面与命令行界面两种模式。使用MEGA首先需要对多条输入序列执行多序列比对，MEGA内置ClustalW与MUSCLE两种比对工具。获得多序列比对结果后即可进行系统发生分析，结果将以树形图的形式展示。

建树的主要方法有最大似然法(maximum likelihood)，邻接法(neighbour-joining),最小演化法(minimum evolution)。最大似然法：建树时需要设置选取一个特定的替代模型来分析给定的一组序列数据，使得获得的每一个拓扑结构的似然率都为最大值，然后再挑出其中似然率最大的拓扑结构作为最优树。邻接法：计算每两个样本间的距离，建立一个使得系统树总距离达到最小的拓扑结构作为输出结果。最小演化法：计算所有可能拓扑结构的分支总长，选取总长度最小的树作为输出。

建树时需要选取检验方法以确定结果可靠性，一般选用bootstrap method。距离计算时有多种方式可选用，一般用p-distance。

下图是使用人类的12种globin的蛋白序列作为输入，使用邻接法，设定检验方法为500次bootstrap，替换模型采用p-distance所得到的系统发生树。





7 收获

1. 综合项目基本成型;
2. 学习了MEGA的基本用法;
3. 学习了Swiss-PDBViewer的基本用法。