

实验六 DNA序列的基本信息及特征分析

一、实验目的

人类CD9抗原基因属于4跨膜蛋白超家族，与癌细胞黏附、迁移及增生有关，具有肿瘤转移抑制作用。应用集成工具EMBOSS的网页版EMBOSS Explorer，对人类CD9抗原基因（序列号为AY422198）进行核酸序列组分分析，同时对大肠杆菌基因组序列（U00096）进行ORF的分析，从中识别真实的ORF。

1. 学习和掌握应用NCBI查询核酸序列信息。
2. 学习和掌握EMBOSS的基本使用方法。
3. 学习和掌握核酸序列ORF的性质。

二、实验内容

1. 人类CD9基因的序列组分分析

1. 获取序列。打开NCBI主页 (<http://www.ncbi.nlm.nih.gov/>)，在核酸 (Nucleotide) 数据库查询人类CD9抗原基因 (AY422198)，将其序列以FASTA格式保存到本地计算机上。
2. 打开EMBOSS。有<http://emboss.bioinformatics.nl/>和<http://genome.csdb.cn/emboss/>多个EMBOSS Explorer可供选用，任选其一打开链接即可使用。
3. 碱基组成分析。找到程序compseq，在Input section项目中，使用upload上传CD9的序列；在Required section项目中，把Word size to consider修改为“1”；其他参数默认即可。最后，点击Run compseq获得ATGC四种碱基的数目及百分比含量。
4. GC含量分析。找到程序geecee，以上传文件的方式提交DNA序列，之后点击Run geecee得到CD9基因的GC含量。
5. 序列转换。找到程序revseq，如前所述上传CD9基因的DNA序列，点击Run revseq后，得到其反向互补序列。通过调整Advanced section项目中的参数，可以仅获得反向序列或互补序列。
6. 比较AY422198及其互补序列的组成成分。

2. 大肠杆菌基因组序列的ORF分析

1. 获取序列。登录NCBI主页 (<http://www.ncbi.nlm.nih.gov/>)，在核酸 (Nucleotide) 数据库查询大肠杆菌基因组序列 (U00096)，将其以FASTA格式保存到本地计算机上。
2. 截取序列。找到EMBOSS中的程序extractseq，在Required section项目中，通过设定Regions to extract为1-3000截取大肠杆菌基因组序列的前3000bp。
3. ORF预测。找到程序getorf，在序列输入框中粘贴上一步截取的前3000bp序列片段，Code to use中选择“Bacterial”，Type of output中选择“Nucleic sequences between START and STOP codons”，其他默认即可。之后点击Run getorf得到ORF的预测结果。
4. 和NCBI的ORF Finder预测结果进行比较分析。