

“实用生物信息技术”课程小组讨论总结报告

组：G1 次：R5 组长：薛林蕾 执笔：吴鸿珍，江浩燊，薛林蕾

1. 时间

2026.5.17

2. 方式

线上讨论

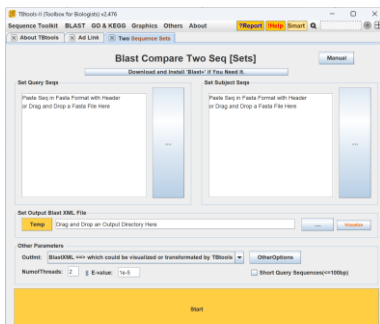
3. 主题

TBtools 应用实例

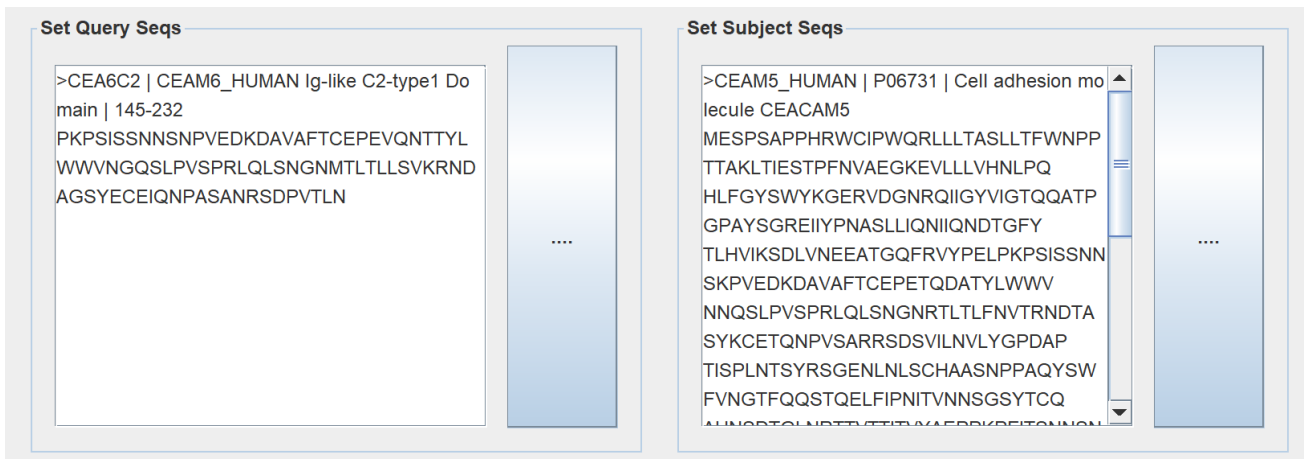
4. 内容

练习 1 — 癌胚抗原家族成员双序列比对

打开 TBtools，在工具选择菜单中选择 BLAST，在 BLAST 图形用户界面封装程序 BLAST GUI Wrapper 中选择两组序列比对 Two Sequence Sets



将 CEA6C2 粘贴到 TBtools 左侧输入框，将 CEA5 粘贴到 TBtools 右侧输入框



在 TBtools 输出文件选择框中输入结果存放文件夹和输出文件名 C:\Users\w1572\Desktop\作业\TBtools\CEA.txt，在输出格式 Outfmt 选择框选择成对比较 Pairwise


```

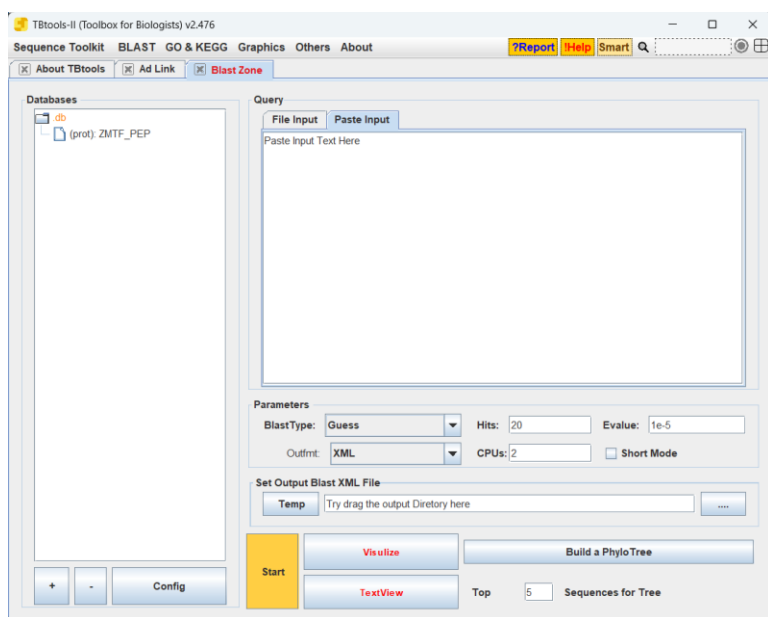
1 # BLASTP 2.12.0+
2 # Query: SPL1_ARATH
3 # Database: 190J_SPL.FAS.TBtoolsDB
4 # Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, value, bit score
5 # 20 hits found
6 SPL1_ARATH SPL6_ORYSJ 42.549 973 446 26 8 879 7 967 0.0 677
7 SPL1_ARATH SPL1_ORYSJ 37.728 933 429 23 20 879 7 860 0.0 560
8 SPL1_ARATH SPL15_ORYSJ 34.506 597 324 17 330 880 564 1139 3.98e-98 323
9 SPL1_ARATH SPL15_ORYSJ 41.150 226 94 7 105 314 186 388 4.26e-38 144
10 SPL1_ARATH SPL9_ORYSJ 25.433 578 335 14 106 661 171 674 2.84e-40 150
11 SPL1_ARATH SPL2_ORYSJ 61.905 105 39 1 105 208 91 195 1.28e-36 134
12 SPL1_ARATH SPL9_ORYSJ 69.136 81 24 1 106 185 185 265 3.02e-34 127
13 SPL1_ARATH SPL16_ORYSJ 69.620 79 24 0 106 184 118 196 6.56e-34 126
14 SPL1_ARATH SPL14_ORYSJ 74.324 74 19 0 106 179 104 177 8.70e-34 125
15 SPL1_ARATH SPL7_ORYSJ 68.421 76 24 0 106 181 108 183 1.44e-33 123
16 SPL1_ARATH SPL18_ORYSJ 64.286 84 30 0 106 189 115 198 5.83e-32 121
17 SPL1_ARATH SPL17_ORYSJ 67.089 79 26 0 106 184 74 152 1.21e-31 118
18 SPL1_ARATH SPL19_ORYSJ 65.333 75 26 0 105 179 92 166 1.38e-31 117
19 SPL1_ARATH SPL3_ORYSJ 59.770 87 35 0 97 183 173 259 6.19e-31 117
20 SPL1_ARATH SPL13_ORYSJ 68.000 75 24 0 106 180 110 184 9.47e-31 111
21 SPL1_ARATH SPL12_ORYSJ 57.471 87 37 0 97 183 171 257 1.06e-30 117
22 SPL1_ARATH SPL12_ORYSJ 36.049 405 184 13 1 348 6 392 1.92e-36 184
23 SPL1_ARATH SPL10_ORYSJ 66.216 74 25 0 106 179 69 141 3.90e-30 110
24 SPL1_ARATH SPL5_ORYSJ 63.095 84 27 1 106 189 207 286 1.18e-28 110
25 SPL1_ARATH SPL11_ORYSJ 59.459 74 30 0 106 179 67 140 5.09e-26 100
26 # BLASTP 2.12.0+
27 # Query: SPL2_ARATH
28 # Database: 190J_SPL.FAS.TBtoolsDB
29 # Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, value, bit score
30 # 19 hits found
31 SPL2_ARATH SPL3_ORYSJ 45.525 257 124 2 1 245 6 258 1.63e-60 195
32 SPL2_ARATH SPL12_ORYSJ 36.049 405 184 13 1 348 6 392 1.92e-36 184
33 SPL2_ARATH SPL4_ORYSJ 59.200 125 39 1 169 281 68 192 5.69e-46 150
34 SPL2_ARATH SPL11_ORYSJ 41.026 195 100 3 167 346 65 259 1.27e-39 136
35 SPL2_ARATH SPL7_ORYSJ 49.219 128 62 1 120 244 56 183 3.58e-36 127
36 SPL2_ARATH SPL16_ORYSJ 69.136 81 25 0 165 245 114 194 5.94e-36 128
37 SPL2_ARATH SPL18_ORYSJ 56.250 112 48 1 165 276 111 221 8.94e-36 128
38 SPL2_ARATH SPL8_ORYSJ 48.837 129 39 3 118 245 160 262 1.16e-34 124
39 SPL2_ARATH SPL17_ORYSJ 47.407 135 68 1 115 246 17 151 3.84e-34 122
40 SPL2_ARATH SPL10_ORYSJ 49.180 122 52 2 118 237 136 249 9.71e-34 122
41 SPL2_ARATH SPL2_ORYSJ 62.353 85 31 1 162 245 85 169 2.71e-33 120
42 SPL2_ARATH SPL14_ORYSJ 65.789 76 26 0 169 244 104 179 7.96e-33 119
43 SPL2_ARATH SPL19_ORYSJ 64.474 76 27 0 169 244 93 168 1.52e-32 117
44 SPL2_ARATH SPL15_ORYSJ 64.557 79 28 0 167 245 185 263 1.30e-31 118
45 SPL2_ARATH SPL6_ORYSJ 63.291 79 29 0 167 245 150 228 1.51e-31 118
46 SPL2_ARATH SPL5_ORYSJ 62.195 82 31 0 165 246 203 284 2.03e-31 115

```

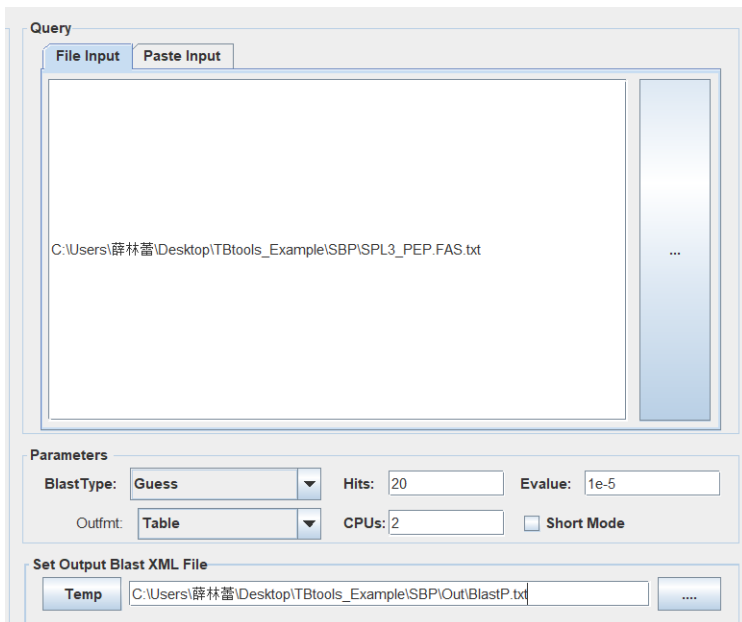
与 NCBI/CNCB 在线 BLAST2Seq 相比, TBtools 两组序列比对是本地、全图形化、免网页操作、可批量、结果直接导出表格、可无缝衔接后续分析, 不需要联网、不需要复制粘贴序列、不需要手动下载结果。打开分析界面, 输入文件, 设置输出结果文件夹和文件, 点击 **start**。

练习 3 — 玉米 SBP 转录因子数据库构建和搜索

打开 TBtools, 在工具选择菜单中选择 BLAST, 在 BLAST 图形用户界面封装程序 BLAST GUI Wrapper 中选择 BLAST Zone 点击 BLAST Zone 用户界面左侧数据库 Database 窗口下方添加数据库按钮"+", 在弹出窗口中选择数据集 ZMTF_PEP.FAS, 在数据库名称输入框中输入 ZMTF_PEP, 点击 OK, 数据库构建完毕



将 SPL_PEP 拖入右侧搜索序列输入框, 将输出格式 Outfmt 改为表格 Table, 在输出文件选择框中设定比对结果存放文件夹和文件名 TBtool_Example\SBP\Out\BlastP.txt



点击开始 Start 按钮，用 FireFox 浏览器查看比对结果，查看 764 个玉米转录因子数据集中 4 个拟南芥 SPL3_PEP 相似序列 [BlastP.txt](#)

```
# BLASTP 2.12.0+
# Query: AT2G33810.1 Arabidopsis SBP family transcription factor, protein sequence, 131 AA
# Database: ZMTF_PEP
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q.
start, q. end, s. start, s. end, evaluate, bit score
# 4 hits found
AT2G33810.1 PTZm00608.1 45.082 122 67 0 10 131 127 248 9.03e-34 115
AT2G33810.1 PTZm00605.1 45.082 122 67 0 10 131 100 221 1.32e-33 114
AT2G33810.1 PTZm00606.1 56.757 37 16 0 54 90 180 216 1.36e-12 57.0
AT2G33810.1 PTZm00607.1 30.000 90 63 0 10 99 127 216 9.62e-12 55.1
# BLAST processed 1 queries
```

结果显示，在 764 个玉米蛋白中，找到了四个在进化上和拟南芥 SPL3 蛋白最像的玉米蛋白。与 NCBI/CNCB BLAST 比较，TBtools 自建本地数据库可以用于处理大量数据，可以建立一个专属的、可重复使用的自己的数据库：比如，建立一个只包含你研究领域所有已知转录因子蛋白的数据库，随时用来比对新的测序结果。而且数据库查询是离线操作，不受制于网络状态。

练习 4 — 玉米 SBP 转录因子数据集中序列提取（基于课外练习 3 结果，从 764 个预测所得玉米转录因子数据集中提取拟南芥 SPL3_ARATH 相似序列）

打开 TBtools，在工具选择菜单中选择序列处理工具包 Sequence Toolkit，在下拉菜单中选择 FASTA 格式序列提取 Fasta Extract

在序列输入框中输入 764 个玉米转录因子 ZMTF_PEP.FAS

在输出框中选择 TBtools_Example\SBP\Out\ZMTF_SPL.txt

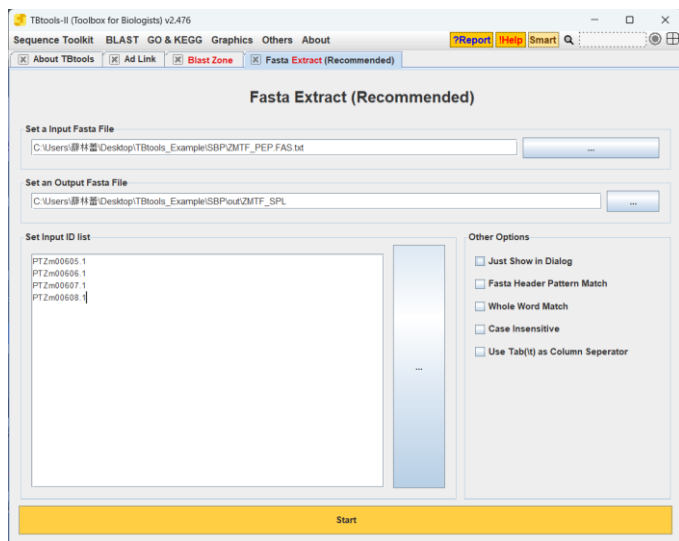
在输入数据标识符 Input ID List 中逐行输入课外练习 3 结果中 4 个转录因子标识符

PTZm00605.1

PTZm00606.1

PTZm00607.1

PTZm00608.1



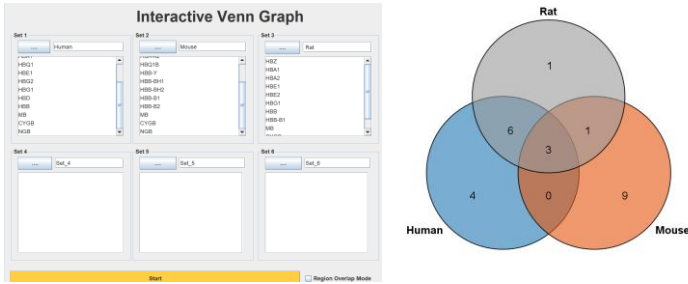
点击开始 Start 按钮，用 WordPad 查看结果 ZMTF_SPL.txt

```
PTZm00605.1 |PUT-155a-Zea_mays-3434
FLEGLRSPVLMGSGFMNWNQKDPMVWVWDEHLVPSVSNVAVTRHGSANSSGGTLTNSSELG
HGSSKVKNTGTNGRVDSDGNSPSSMIAFNQGEPLISLKGKRAYFENACGGQDAKVASAAS
DVTSAAASVVKTKVVSQNAKRWYCVQVERCKVLDLSSAKDYNRKHVCVHHSKATKVVVAGL
ERRFCQCSRHFGLAEFDQNKRCRRRLMHHNARRRKPQADTISFNSTMFYDTRQRTNL
FFSQPLVYGQVRSNAGSSWDNLGGLKFMETKHPVHPTKTASPDELHFSALQITSAAAHTG
HHHLDLDFMFAFKGTSTKVLNQVGEAWXX
>PTZm00606.1 |PUT-155a-Zea_mays-44562
GAAAQITDGLMGSGFMNWNQKASVLDWVWENLPPVAVGASGSENPRMAAQPALHSACGTI
SSSEMGVSSKSSVASADSSPKAKGKSMELNFAADPKV/PDKNTDSGKVDARTSPSSP
VIAISSAEPVLSLKGRTYFEDVCGGQSVKLSPDSATAVPPASPKTKVTPNKKSYC
QVEGCKVDLSSAKEYHRKHRCVCEPHSKAPKVVVAGL
>PTZm00607.1 |PUT-155a-Zea_mays-9077737
FLEGLRSPVLMGSGFMNWNQKDPMVWVWDEHLVPSVSNVAVTRHGSANSSGGTLTNSSELG
HGSSKSSIASIDSPSGVGNLEFNFAAVERHVKNVTGTNGRVDSDGNSPSSMIAFNQGEPL
LISLKGKRAYFENACGGQDAKVASAASDVTSAAASVVKTKVVSQNAKRWYCVQEGCKVDL
SSAKDYNRKHVCVHHSKATKVVVAGLGVVNSVADTRQRTNLFFSQPLVYGQVRSNAGS
SWDNLGGLKFMETKHPVHPTKTASPDELHFSALQITSAAAHTGHHHLDLDFMFAFKGTST
KXXX
>PTZm00608.1 |PUT-155a-Zea_mays-9197737
FLEGLRSPVLMGSGFMNWNQKDPMVWVWDEHLVPSVSNVAVTRHGSANSSGGTLTNSSELG
HGSSKSSIASIDSPSGVGNLEFNFAAVERHVKNVTGTNGRVDSDGNSPSSMIAFNQGEPL
LISLKGKRAYFENACGGQDAKVASAASDVTSAAASVVKTKVVSQNAKRWYCVQEGCKVDL
SSAKDYNRKHVCVHHSKATKVVVAGLERFCQCSRHFGLAEFDQNKRCRRRLMHHNARRR
RRRKPQADTISFNSTMFYDTRQRTNLFFSQPLVYGQVRSNAGSSWDNLGGLKFMETKHPV
VHPTKTASPDELHFSALQITSAAAHTGHHHLDLDFMFAFKGTSTKVLNQVGEAWXX
```

练习 5 — 韦恩图 Venn Diagram 分析

打开 TBtools，选择图形 Graphics 中韦恩图分析 Venn

在 Set1 输入窗口中输入人的珠蛋白家族基因名，在 Set2 输入窗口中输入小鼠珠蛋白家族基因名，在 Set3 输入窗口中输入大鼠珠蛋白基因家族名，点击开始按钮 Start，生成韦恩图 Globin.svg。

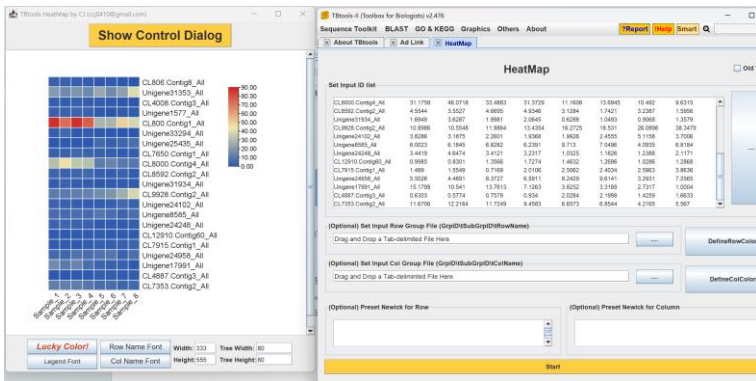


三个物种均有肌红蛋白 MB, 胞红蛋白 CYGB 和脑红蛋白 NGB, 人和大鼠另有 6 个同名基因, 小鼠和大鼠另有 1 个同名基因。

练习 6 — 热图 Heatmap 分析

打开 TBtools, 选择图形 Graphics 热图分析 Heatmap。点击开始按钮 Start, 生成初始热图: HeatMap0。

在图形窗口中打开会话窗口 Show Control Dialog, 设置并调节参数, 生成新的热图。在图形窗口中点击 Save Graph 按钮, 选择 SVG 格式, 保存新的热图: Heatmap1



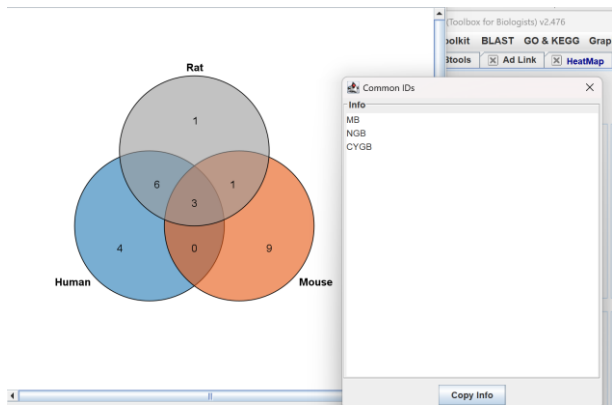
5. 问题

1、Two Sequence Files 的作用是什么

判断不同物种 (拟南芥 ARATH 和水稻 ORYSJ) 的蛋白之间是否存在同源关系: 结果里的每一行, 都代表一个 query 序列 (如 SPL1_ARATH) 和数据库中 subject 序列 (如 SPL6_ORYSJ) 的比对结果。可以通过 % identity (一致性)、evalue (期望值) 和 bit score (比特得分) 判断同源性高低: E 值越小、bit score 越高, 说明序列间的同源性越可靠, 随机匹配的可能性越低。比如 SPL1_ARATH 和 SPL6_ORYSJ 的 E 值为 0, 一致性 42.549%, 说明二者同源性极高, 是可靠的同源蛋白。

2、怎么查看 venn 的具体数据?

双击 venn 图数字位置, 会显示该数字包含的具体基因



3、怎么调整热图颜色？

点击 show control dialog ,然后点击色块调整颜色。

