

“实用生物信息技术”课程小组讨论总结报告

组：G2 次：5 组长：边汉青 执笔：刘奇

1. 时间

2026.5.11 上午

2. 方式

线上

3. 主题

TBtools 的课堂练习实例整理

4. 内容

A 癌胚抗原家族成员双序列比对

B 拟南芥和水稻 SBP 转录因子家族成员比对

C 玉米 SBP 转录因子数据库构建和搜索

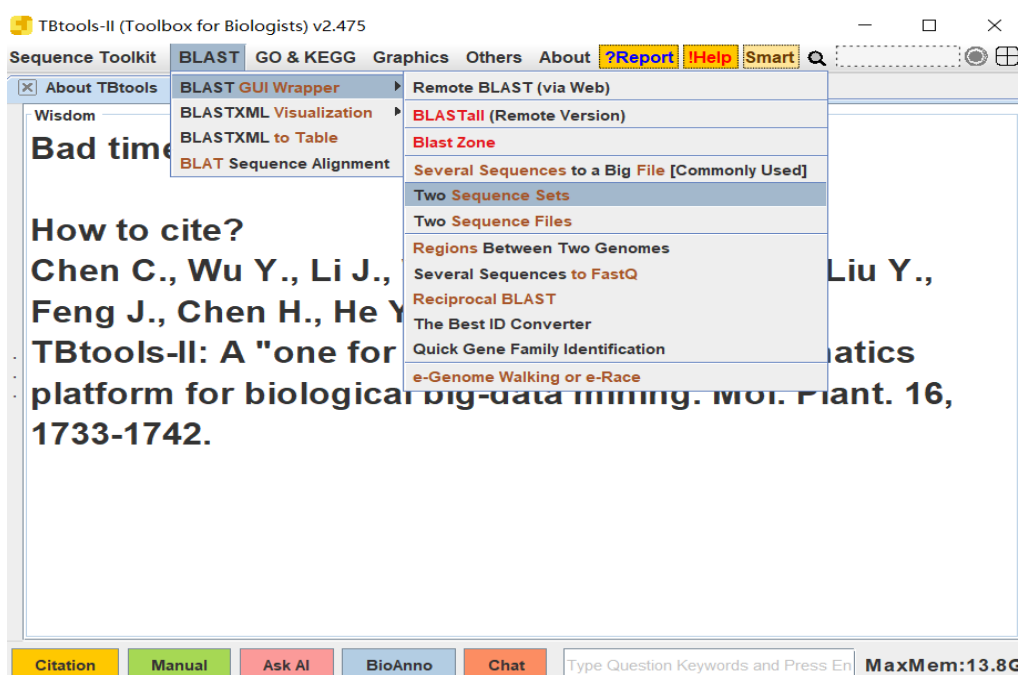
D 玉米 SBP 转录因子数据集中序列提取

E 韦恩图 Venn Diagram 分析

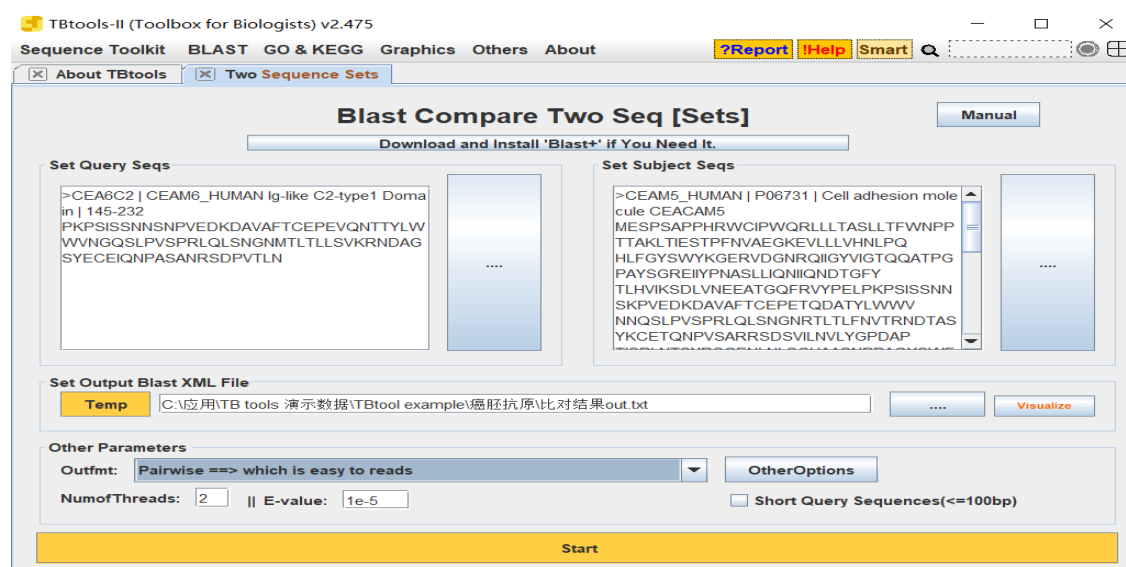
F 热图 Heatmap 分析

A TBtools 双序列对比，以癌胚抗原家族成员双序列比对为例

1) 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 中选择两组序列对比 Two Sequence Sets。



2) 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:\应用\TB tools 演示数据\TBtool example\癌胚抗原\比对结果 out.txt; 在输出格式 Outfmt 选择框选择成对比较 Pairwise。



3) 用 notepad 查看结果。发现 CEA5 有三个结构区域与第一个恒定结构域

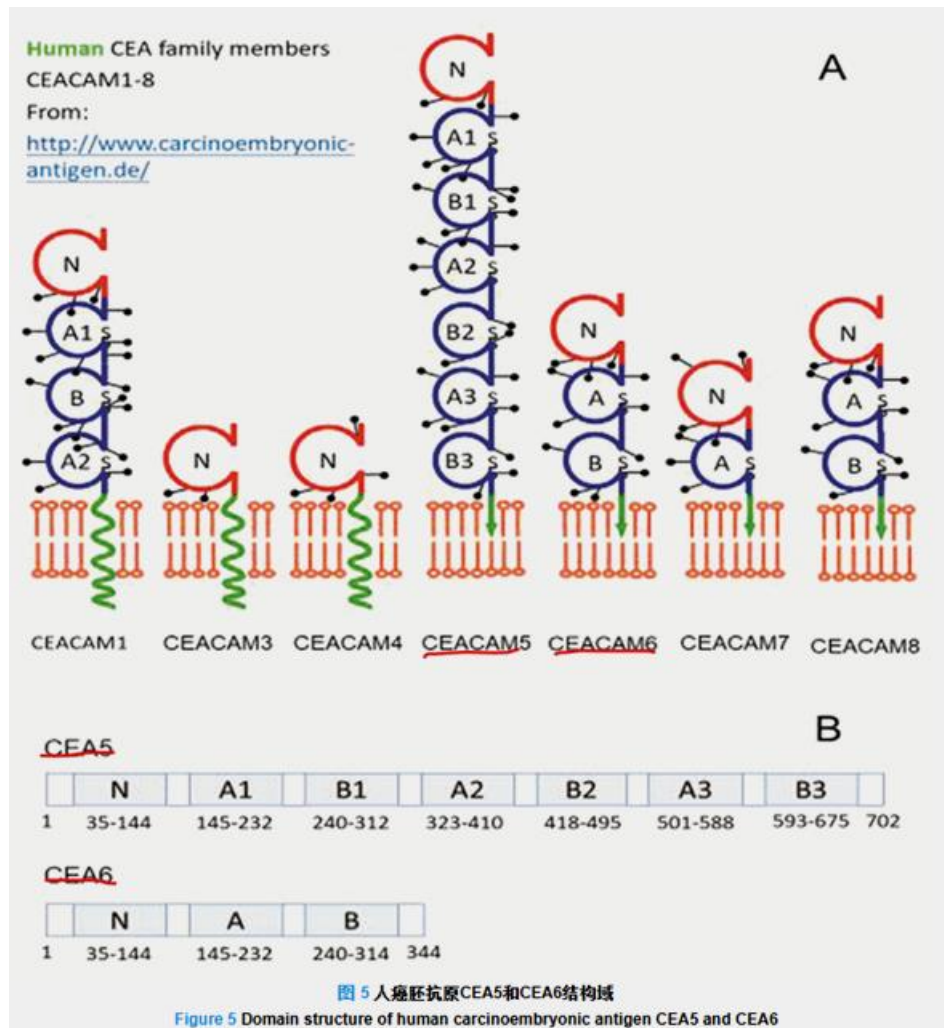
Ig-like C2-type1 序列 CEA6C2 相似。

```

28 Sequences producing significant alignments: (Bits) Value
29
30 CEAM5_HUMAN | P06731 | Cell adhesion molecule CEACAM5 153 2e-49
31
32
33 >CEAM5_HUMAN | P06731 | Cell adhesion molecule CEACAM5
34 Length=702
35
36 Score = 153 bits (386), Expect = 2e-49, Method: Compositional matrix adjust.
37 Identities = 75/88 (85%), Positives = 78/88 (89%), Gaps = 0/88 (0%)
38
39 Query 1 PKPSISSNNSNFVEDKDAVAFTCEFEVQNTTYLWVWVNGQSLFVSPRLQLSNGNMTLTLLS 60
40 PKPSISSNNS PVEDKDAVAFTCEFE QNTTYLWVWVNGQSLFVSPRLQLSNGN TLT +
41 Sbjct 501 PKPSISSNNSKPVEDKDAVAFTCEFEAQNTTYLWVWVNGQSLFVSPRLQLSNGNRTLTLEN 560
42
43 Query 61 VKRNDAGSYECIQNPASANRSDPVTLN 88
44 V RND A +Y C IQN SANRSDPVTL+
45 Sbjct 561 VTRNDARAIVCGIQNSVANSRSDPVTLD 588
46
47
48 Score = 146 bits (368), Expect = 6e-47, Method: Compositional matrix adjust.
49 Identities = 71/88 (81%), Positives = 74/88 (84%), Gaps = 0/88 (0%)
50
51 Query 1 PKPSISSNNSNFVEDKDAVAFTCEFEVQNTTYLWVWVNGQSLFVSPRLQLSNGNMTLTLLS 60
52 PKPSISSNNS PVEDKDAVAFTCEFE Q+ TYLWVWV NGQSLFVSPRLQLSNGN TLT +
53 Sbjct 145 PKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWVWVNGQSLFVSPRLQLSNGNRTLTLEN 204
54
55 Query 61 VKRNDAGSYECIQNPASANRSDPVTLN 88
56 V RND SY+CE QNF SA RSD V LN
57 Sbjct 205 VTRNDTASYKCEQNFVSAARRSDSVILN 232
58
59
60 Score = 143 bits (361), Expect = 6e-46, Method: Compositional matrix adjust.
61 Identities = 70/88 (80%), Positives = 74/88 (84%), Gaps = 0/88 (0%)
62
63 Query 1 PKPSISSNNSNFVEDKDAVAFTCEFEVQNTTYLWVWVNGQSLFVSPRLQLSNGNMTLTLLS 60
64 PKP I+SNNSNFVED+DAVA TCEFE+QNTTYLWVWV NGQSLFVSPRLQLSN N TLTLLS
65 Sbjct 323 PKPFITSNNSNFVEDEDAVALTCEFEIQNTTYLWVWVNGQSLFVSPRLQLSNDNRTLTLLS 382
66
67 Query 61 VKRNDAGSYECIQNPASANRSDPVTLN 88
68 V RND G YEC IQN S + SDPV LN
69 Sbjct 383 VTRNDVGFYECIQNELSVHSDPVTILN 410
70

```

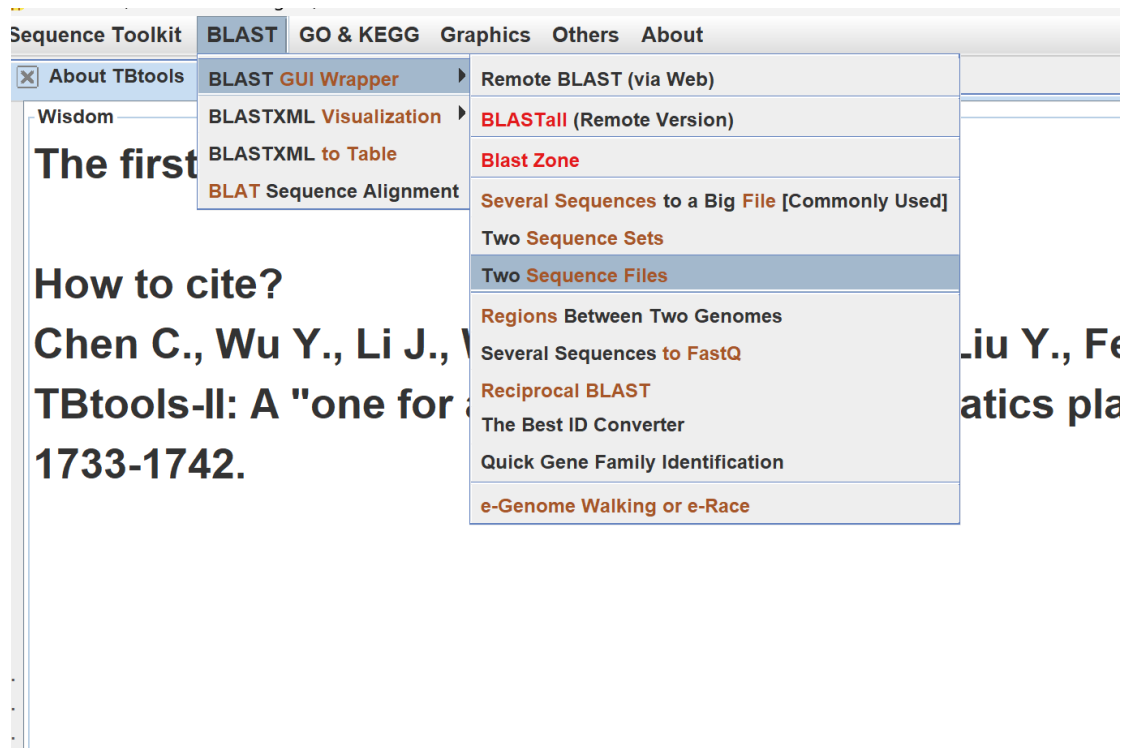
4) 这与文献显示结果相同，即 CEA5 含有三个类似结构重复序列，证明对比结果可信。



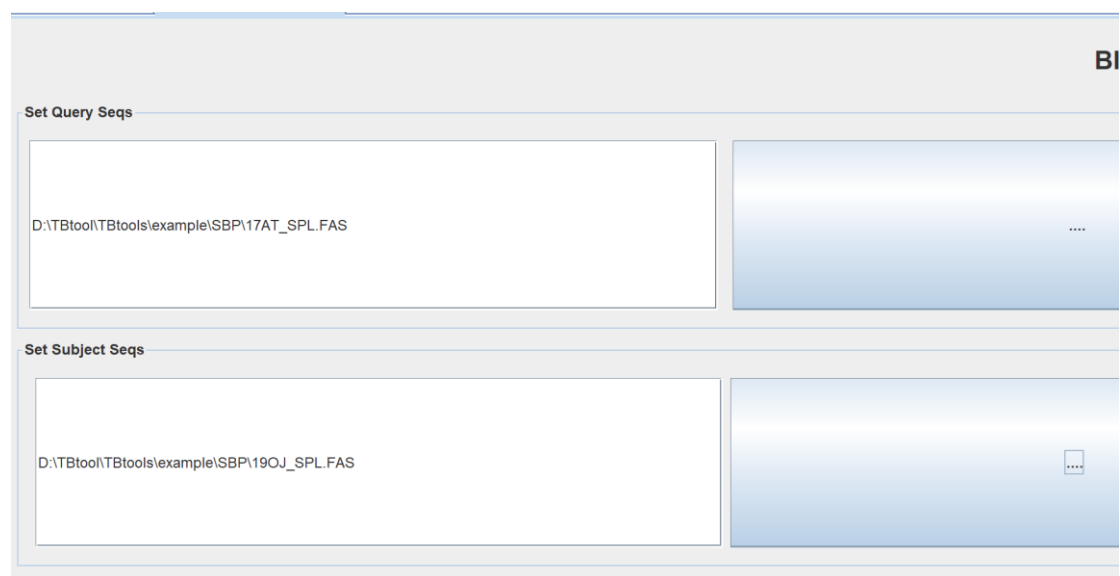
B 拟南芥和水稻 SBP 转录因子家族成员比对

研究目的 — 找出拟南芥 SPL7_ARATH 与粳稻中的直系同源基因

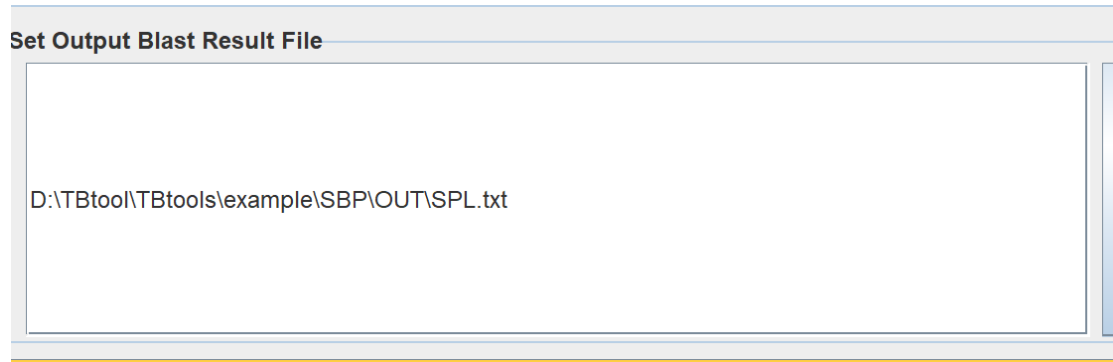
1) TBtools, 在工具选择菜单中选择 BLAST, 在 BLAST 图形用户界面封装程序 BLAST GUI Wrapper 中选择双序列比对 Two Sequence Files



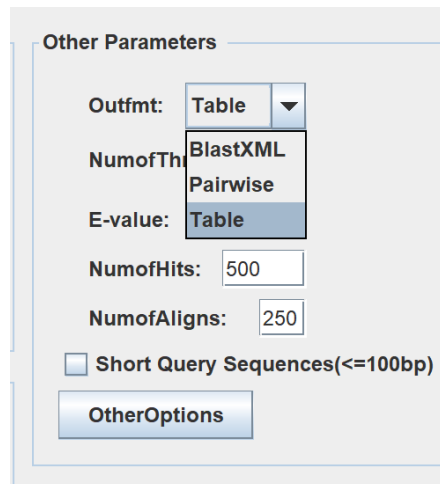
2) 点击 TBtools BLAST 用户界面搜索序列文件选择框 Set Query Seqs 选择菜单, 选择 17AT_SPL.FAS; 点击 TBtools BLAST 用户界面目标序列文件选择框 Set Subject Seqs 选择菜单, 选择 19OJ_SPL.FAS



3) 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\SPL.txt



4) 在其它参数 Other Parameter 选择菜单中设置输出格式 Outfmt 为表格方式 Table, 点击开始 Start 按钮, 用 Windows 系统写字板程序 WordPad 查看比对结果, 找出水稻中拟南芥 SPL7_ARATH 的直系同源基因是 SPL9_ORYSJ

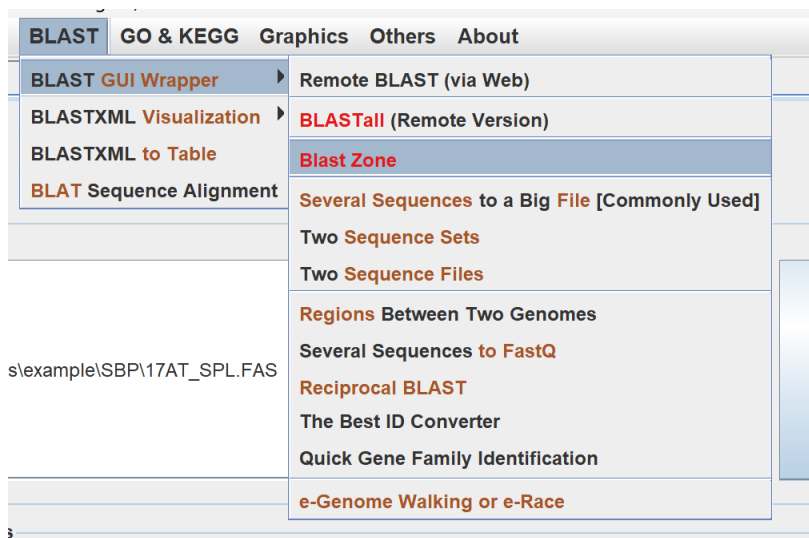


```
# Database: 159J_SPL7no_ID1001SBP
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s.
end, evalue, bit score
# 21 模糊位 查询序列 目标序列 相似度 比对长度 错配数 空位数 开始和结束的位置
# 21 模糊位 模糊位 比对分数
SPL7_ARATH SPL9_ORYSJ 37.170834 436 19 32 797 23 836 1.64e-169 502
SPL7_ARATH SPL1_ORYSJ 29.530447 266 11 138 550 107 538 2.55e-50 181
SPL7_ARATH SPL15_ORYSJ 28.632234 163 2 320 553 599 828 4.11e-32 124
SPL7_ARATH SPL15_ORYSJ 49.36779 39 1 138 215 187 265 2.15e-22 93.2
SPL7_ARATH SPL6_ORYSJ 29.675246 156 4 307 548 432 664 1.20e-26 106
SPL7_ARATH SPL6_ORYSJ 47.664107 53 2 125 228 138 244 9.88e-25 100
SPL7_ARATH SPL14_ORYSJ 57.33375 32 0 137 211 103 177 2.34e-25 100
SPL7_ARATH SPL2_ORYSJ 42.453106 53 3 127 224 77 182 1.73e-23 94.4
SPL7_ARATH SPL16_ORYSJ 33.728169 93 4 127 288 102 258 3.61e-23 94.0
SPL7_ARATH SPL7_ORYSJ 50.64977 38 0 135 211 105 181 5.42e-23 92.0
SPL7_ARATH SPL10_ORYSJ 54.28670 32 0 137 206 180 249 6.95e-23 92.8
SPL7_ARATH SPL18_ORYSJ 47.25391 43 2 138 224 115 204 1.01e-22 92.8
SPL7_ARATH SPL8_ORYSJ 40.594101 59 1 137 236 184 284 1.36e-22 91.7
SPL7_ARATH SPL19_ORYSJ 50.00076 37 1 138 213 93 167 2.39e-22 90.1
SPL7_ARATH SPL4_ORYSJ 43.67887 49 0 137 223 67 153 2.95e-22 87.8
SPL7_ARATH SPL5_ORYSJ 53.08681 37 1 137 217 206 285 4.20e-22 90.5
SPL7_ARATH SPL17_ORYSJ 52.56478 36 1 137 214 73 149 7.71e-22 89.0
SPL7_ARATH SPL3_ORYSJ 50.64977 37 1 137 213 181 256 5.91e-21 87.0
SPL7_ARATH SPL12_ORYSJ 48.68476 38 1 138 213 180 254 9.36e-21 86.7
SPL7_ARATH SPL13_ORYSJ 46.15478 42 0 138 215 110 187 5.08e-20 80.1
SPL7_ARATH SPL11_ORYSJ 44.30479 44 0 137 215 66 144 2.13e-19 80.9
# BLASTP 2.12.0+
# Query: SPL8_ARATH
```

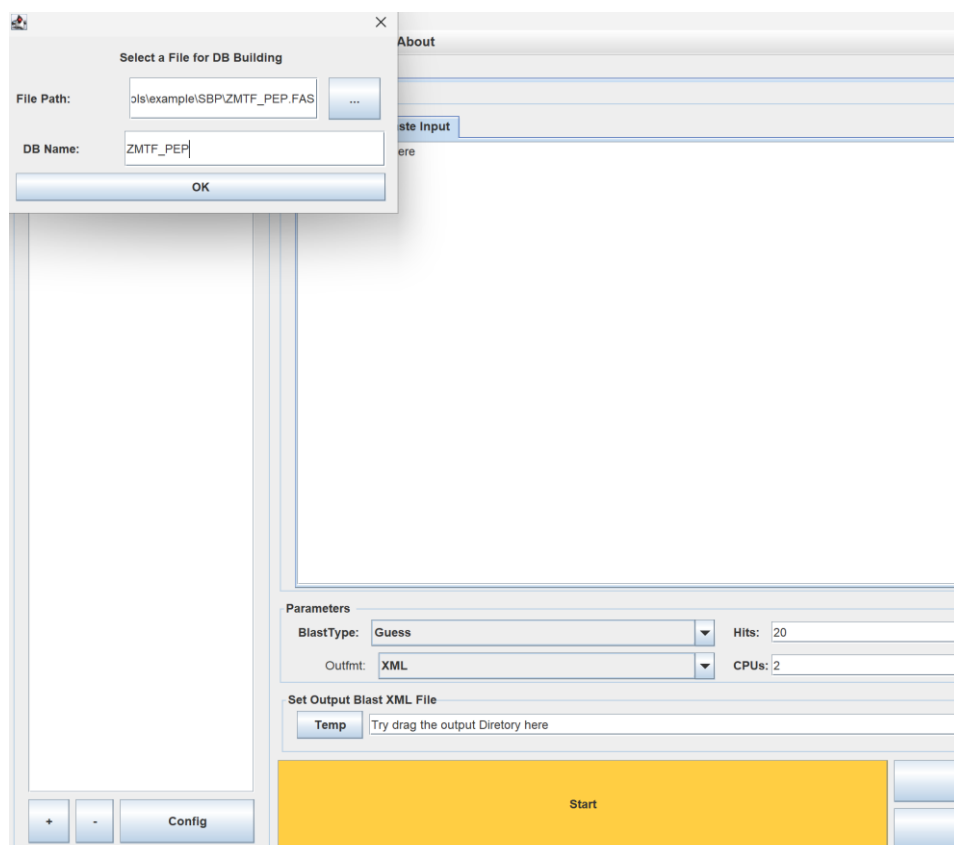
C 玉米 SBP 转录因子数据库构建和搜索

构建预测所得玉米转录因子数据库，搜索拟南芥 SPL3_ARATH 相似序列

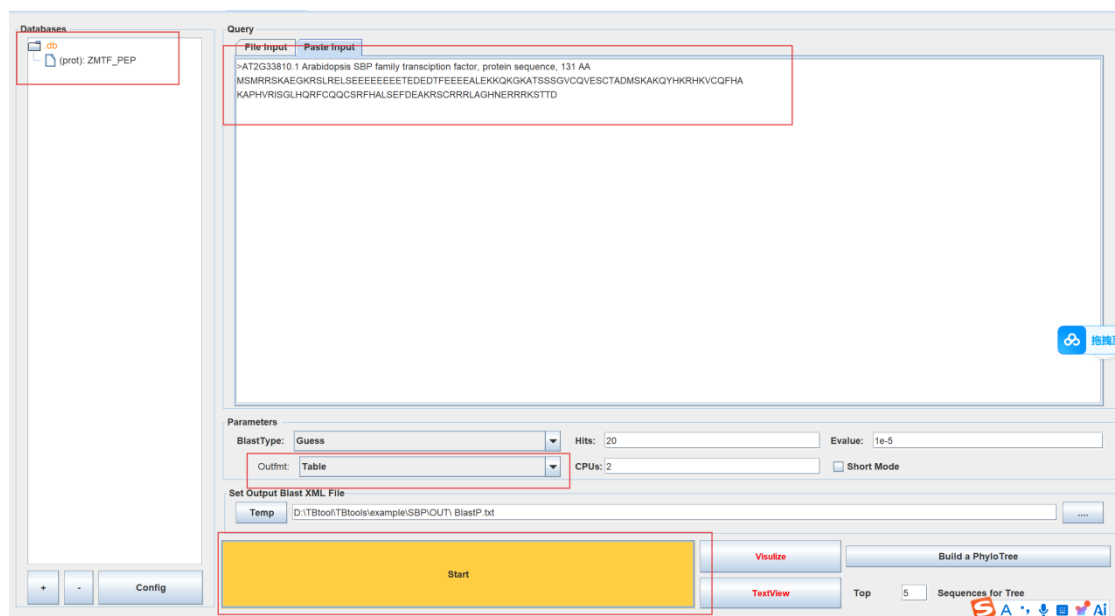
1) 打开 TBtools，在工具选择菜单中选择 BLAST，在 BLAST 图形用户界面封装程序 BLAST GUI Wrapper 中选择 BLAST Zone



2) 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



3) 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



4) 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, evalue, 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
```

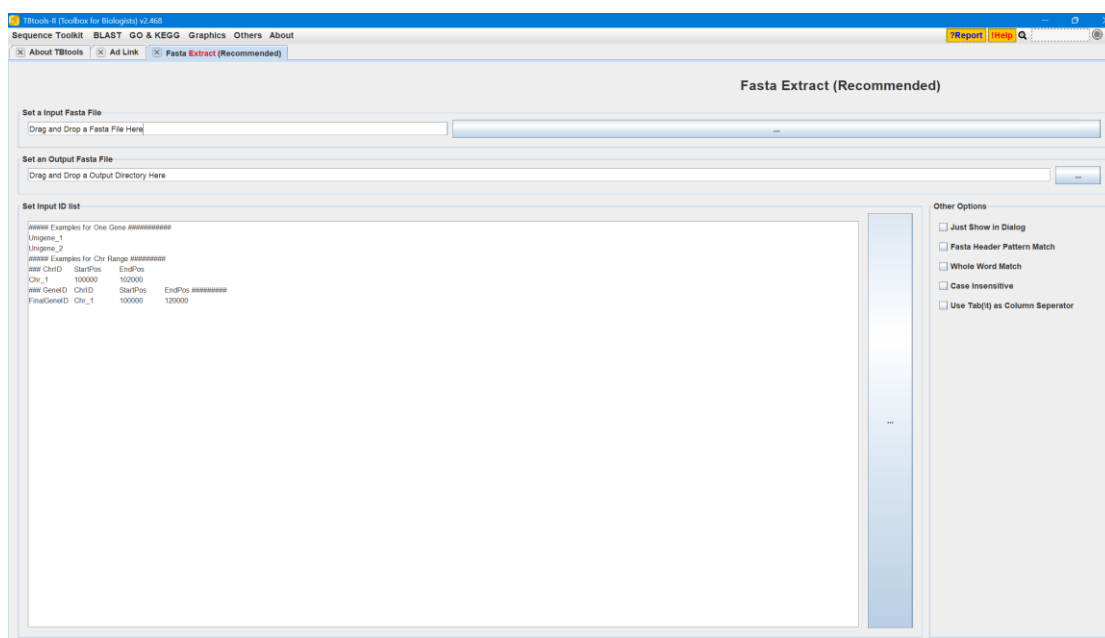
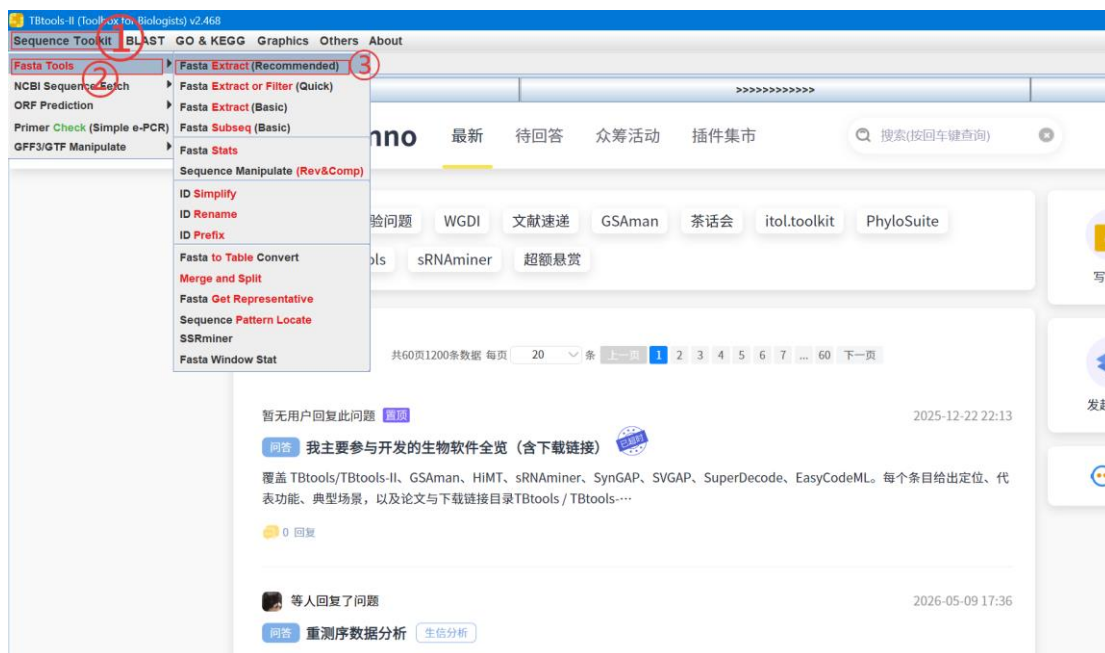
D 玉米 SBP 转录因子数据集中序列提取

1. 研究背景

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

2. 操作步骤

1) 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;



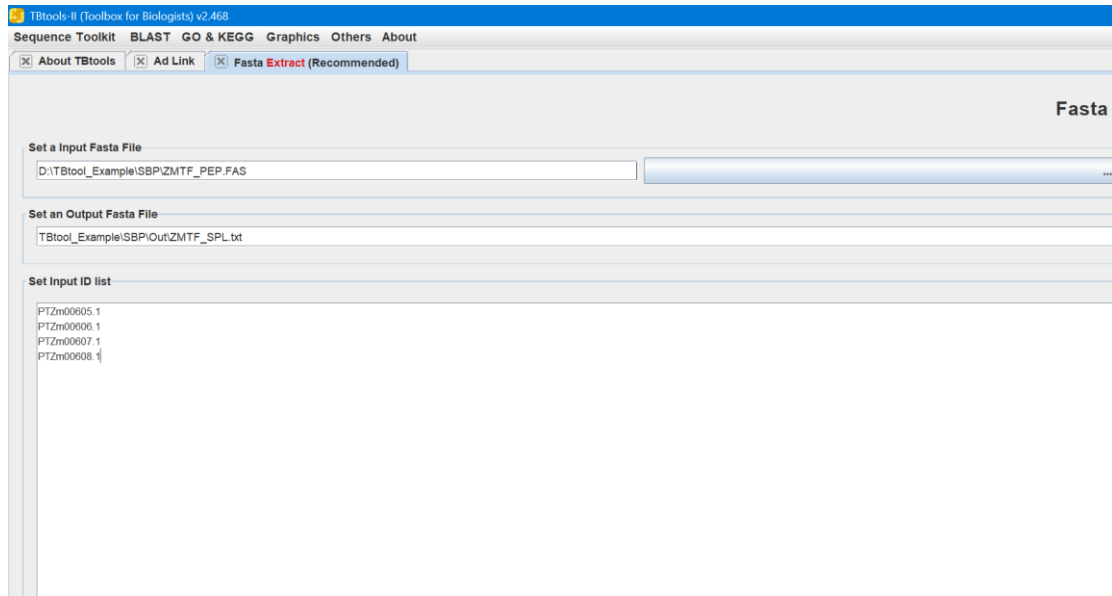
2) 在序列输入框中输入 764 个玉米转录因子 ZMTF_PEP.FAS; 在输出框中选择 TBtool_Example\SBP\Out\ZMTF_SPL.txt; 在输入数据标识符 Input ID List 中逐行输入课外练习 3 结果中 4 个转录因子标识符:

PTZm00605.1

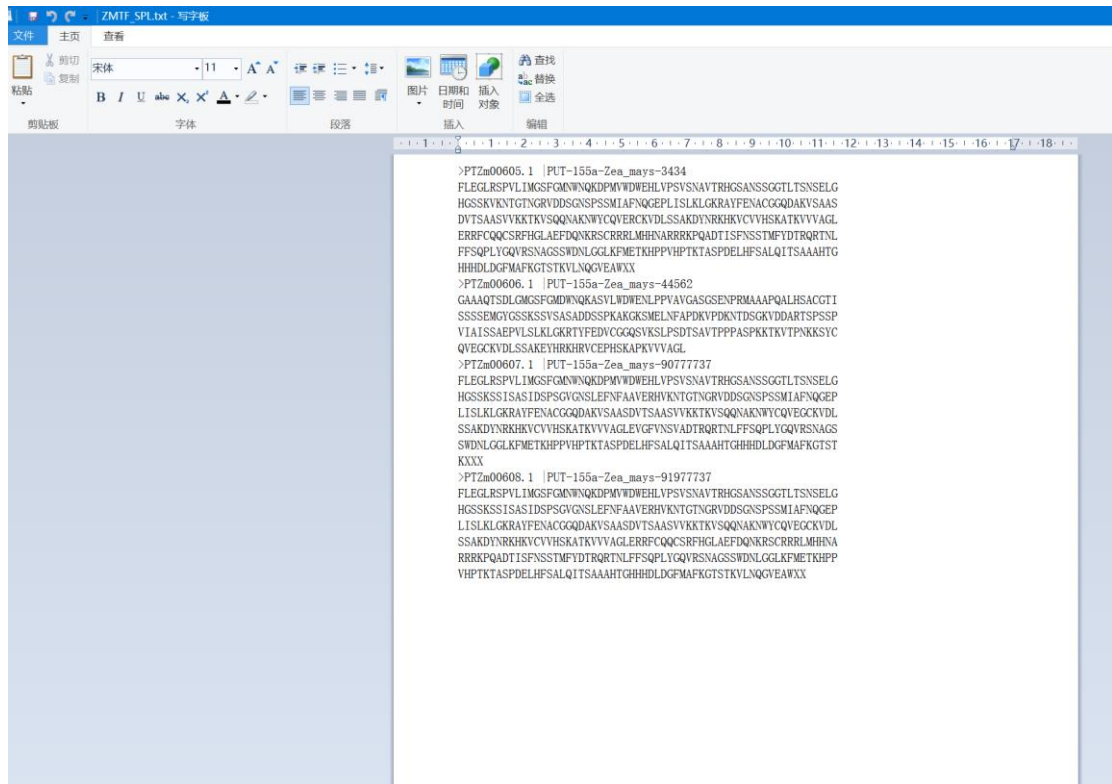
PTZm00606.1

PTZm00607.1

PTZm00608.1



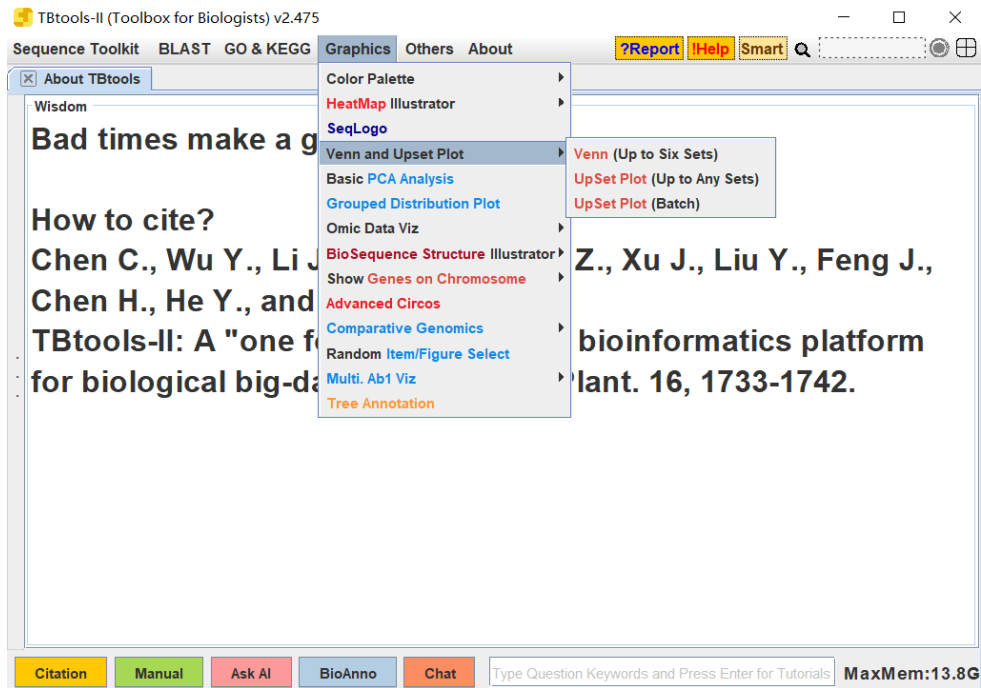
3) 点击开始 Start 按钮，用 WordPad 查看结果 ZMTF_SPL.txt;



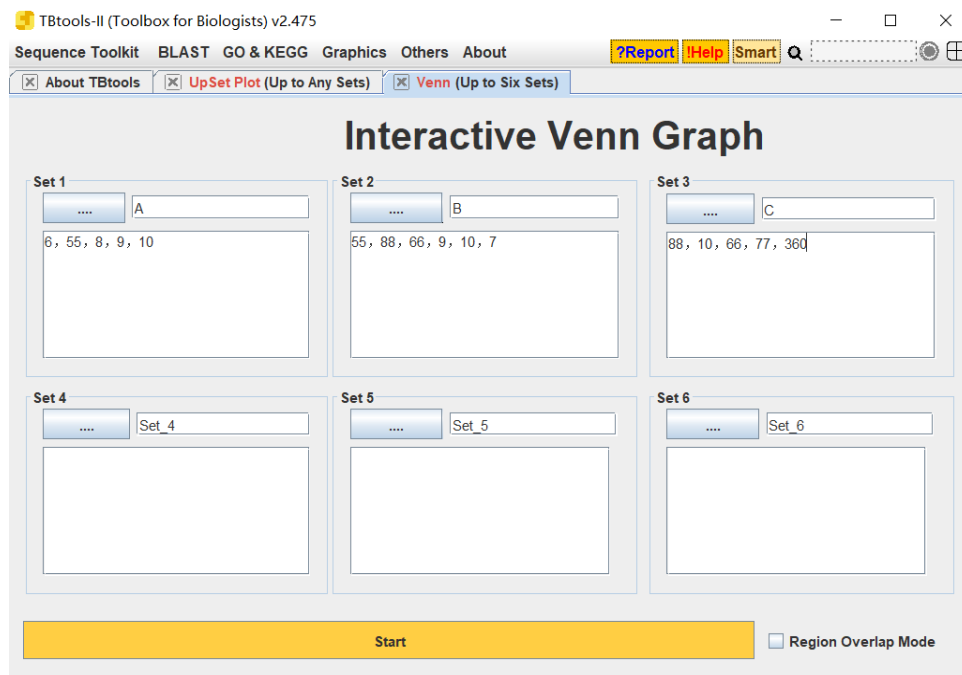
E TBtools 中关于韦恩图 Venn Diagram 分析，以简单的集合为例。

集合 A=(6, 55, 8, 9, 10), 集合 B=(55, 88, 66, 9, 10, 7), 集合 C=(88, 10, 66, 77, 360)

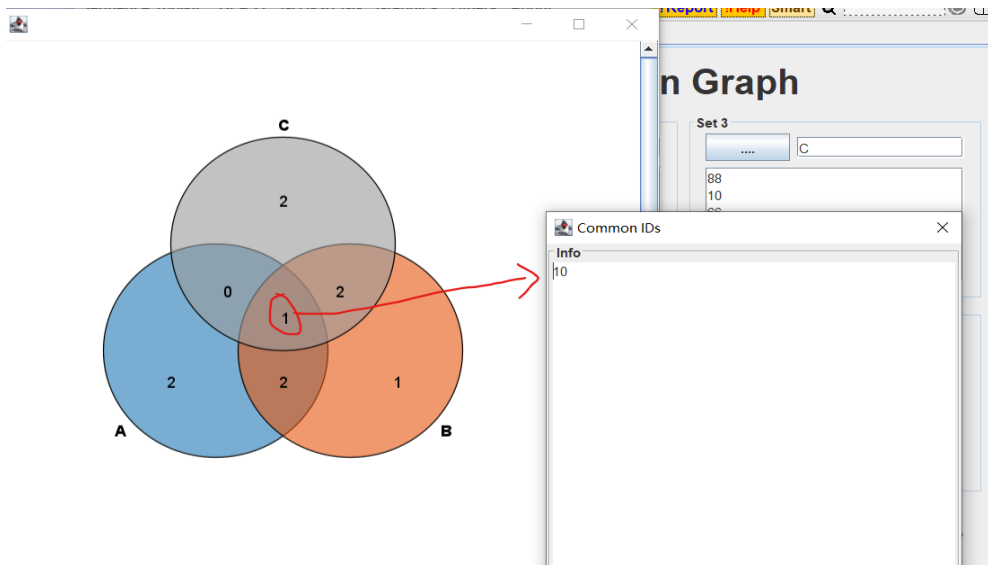
1) 打开 TBtools 选择 VENN and upset plot 中的 venn。



2) 将集合 A、B、C 放入 set1、2、3 中，注意每个数字为一行，否则将会将整个集合作为一个整体进行比较，之都点击 start。



3) 结果显示，他们共有的有一个，点击是数字“1”，便可查看他们共有的元素。同理，将他们换成基因集合，进行相同操作，便可找到他们之间的相同基因。



F 热图 Heatmap 分析

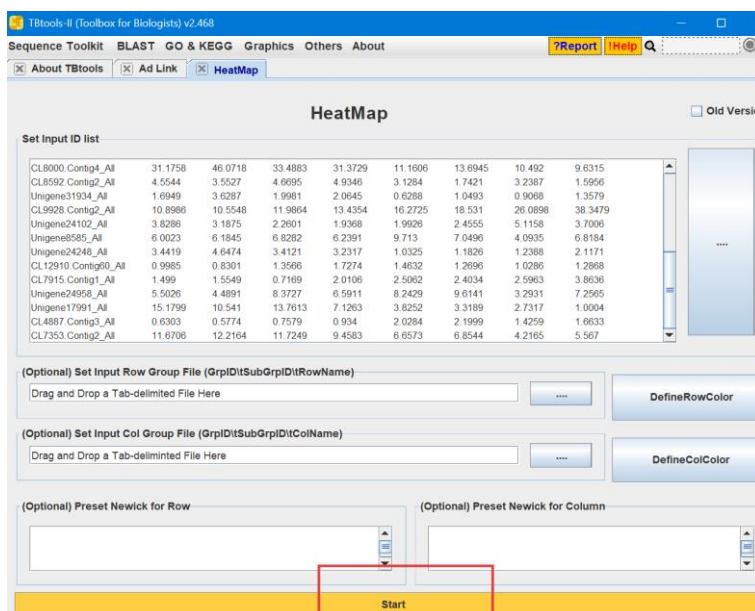
实例 6.1 — TBtools 自带热图演示实例

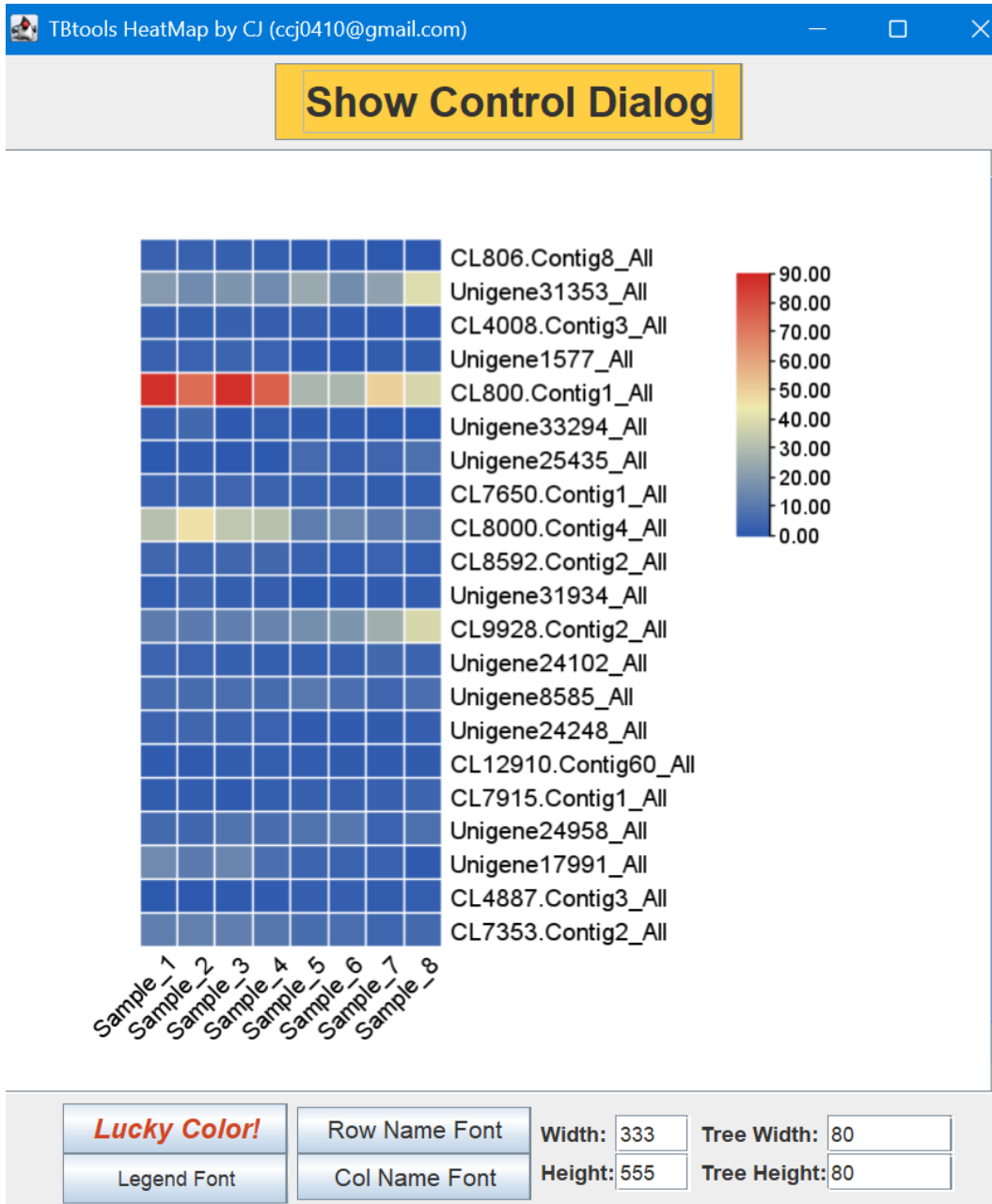
1. 研究背景

热图分析的实际用途:热图就是用颜色深浅或颜色冷暖来表示数值大小的二维矩阵图，颜色越红越深代表数值越高，越蓝越浅代表数值越低，一眼就能看出差异和规律。

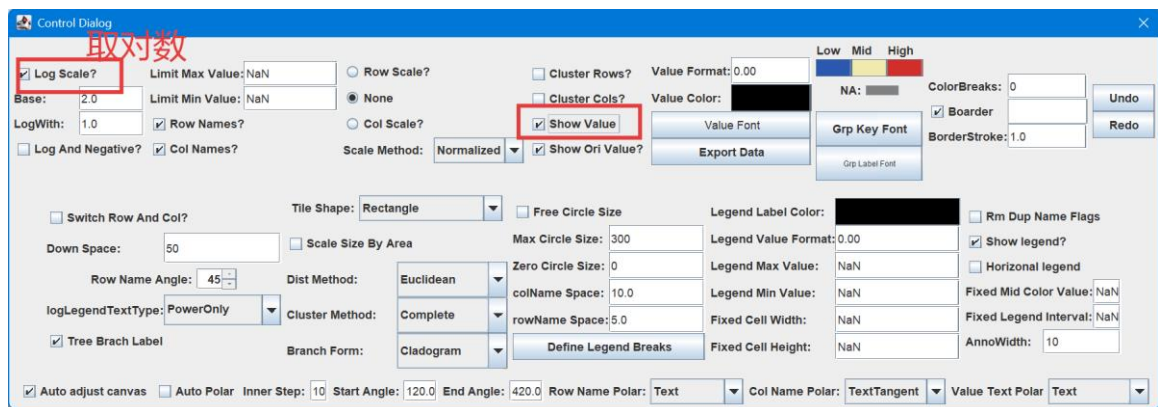
2. 操作步骤

1) 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;





2) 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;

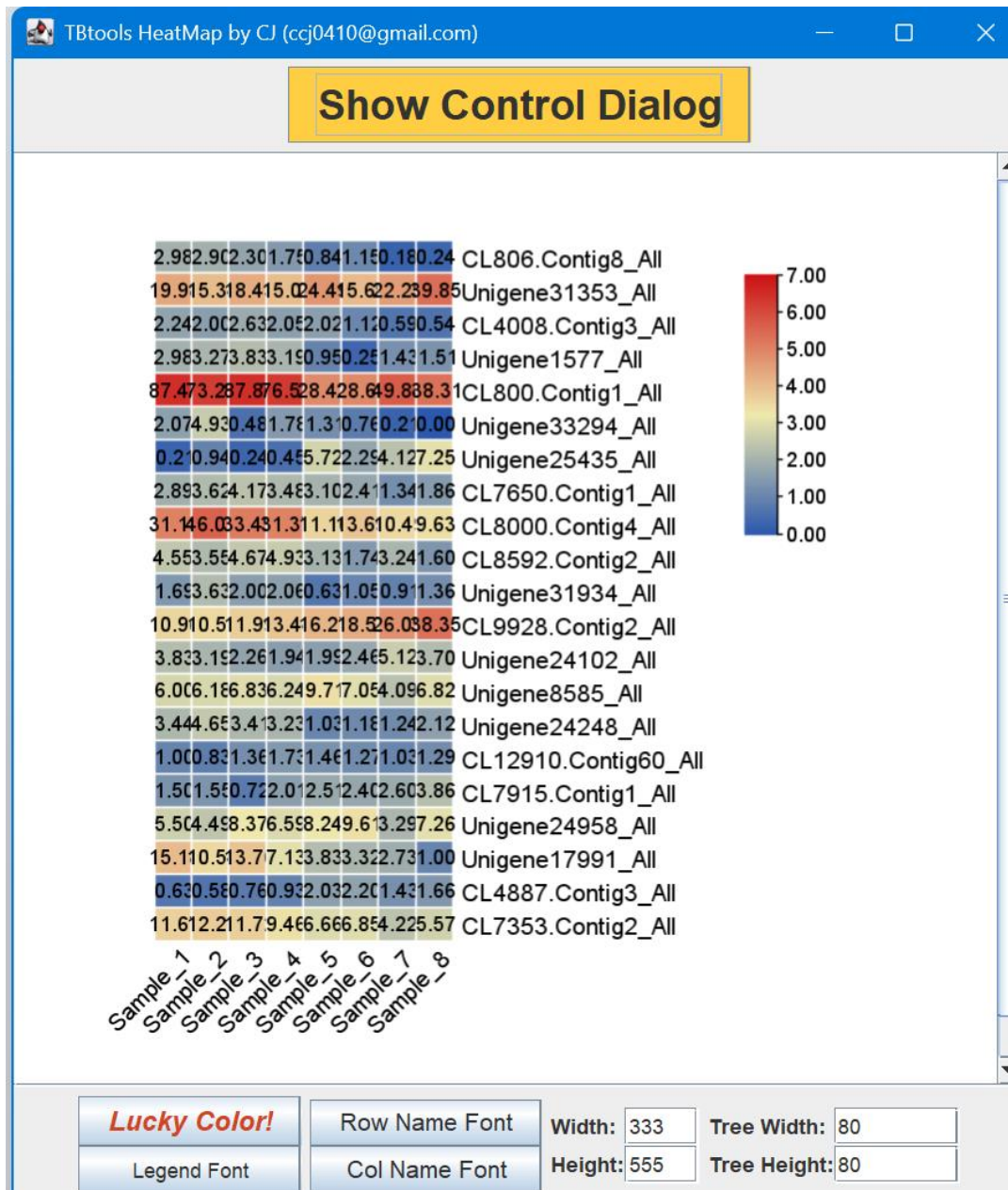


Log Scale? 已勾选 (启用对数刻度)

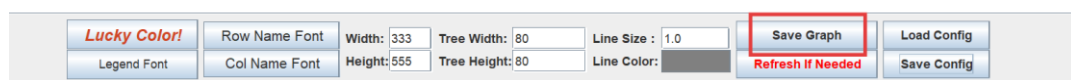
Base: 2.0 (以 2 为底的对数, \log_2)

LogWith: 1.0 (对数前先加 1, 避免 $\log_2(0)$ 无定义)

取对数的意义: 低数值的差异也能被热图清晰展示, 避免高数值的颜色掩盖掉低数值的变化。

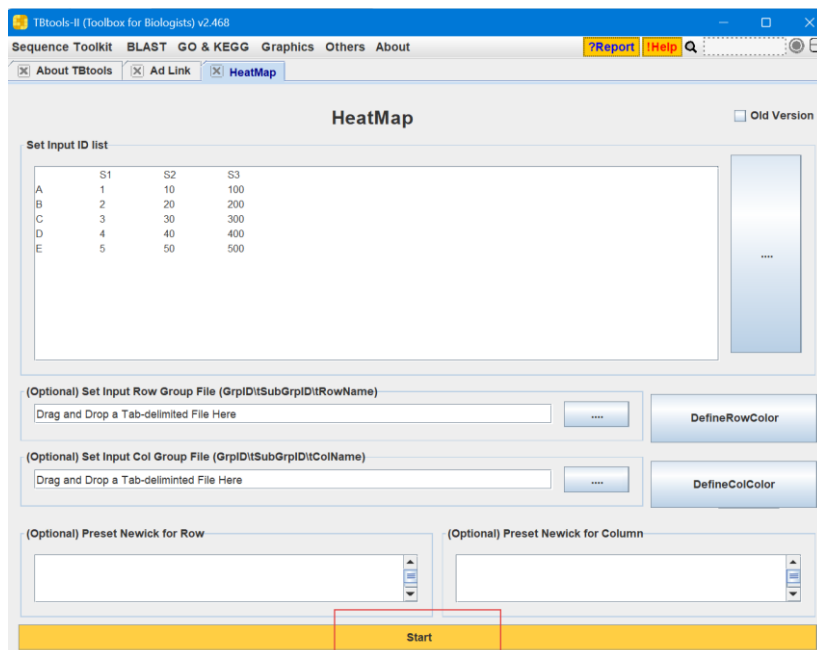


3) 在图形窗口中点击 Save Graph 按钮, 选择 SVG 格式, 保存新的热图:
Heatmap1

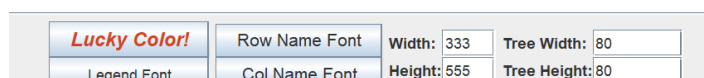
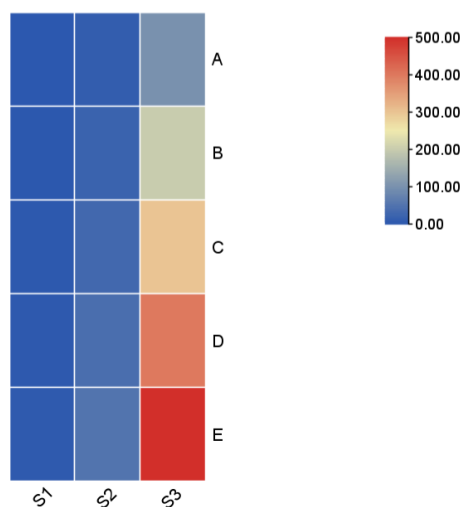
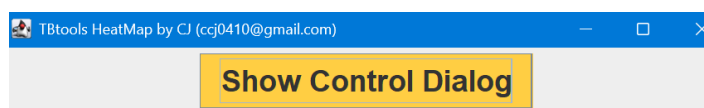


实例 6.2 — 热图简例

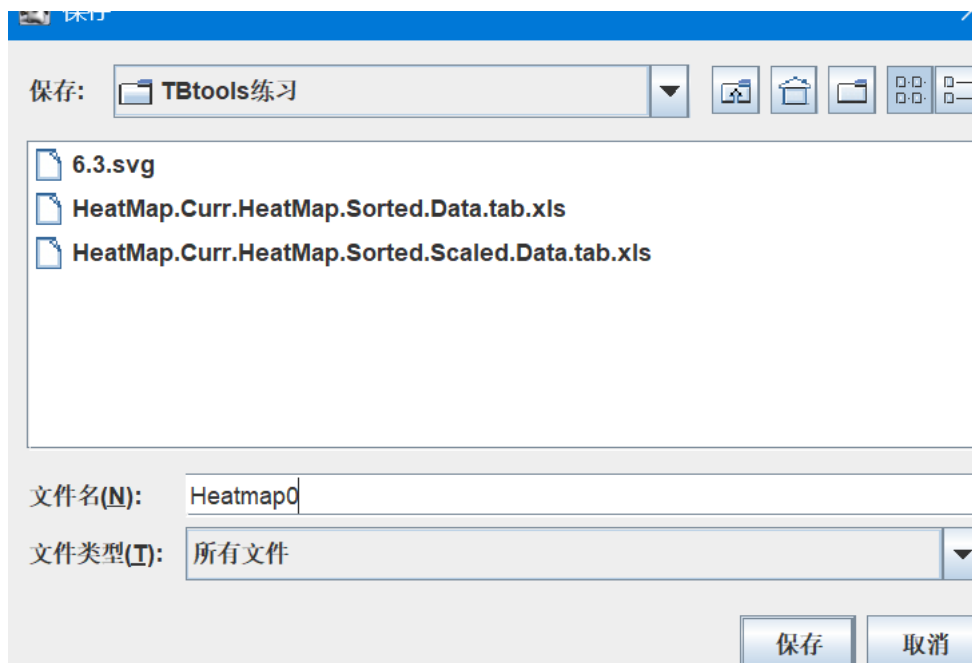
1) 打开 TBtools, 选择图形 Graphics 热图分析 Heatmap; 在输入窗口中输入: heatmap-data;



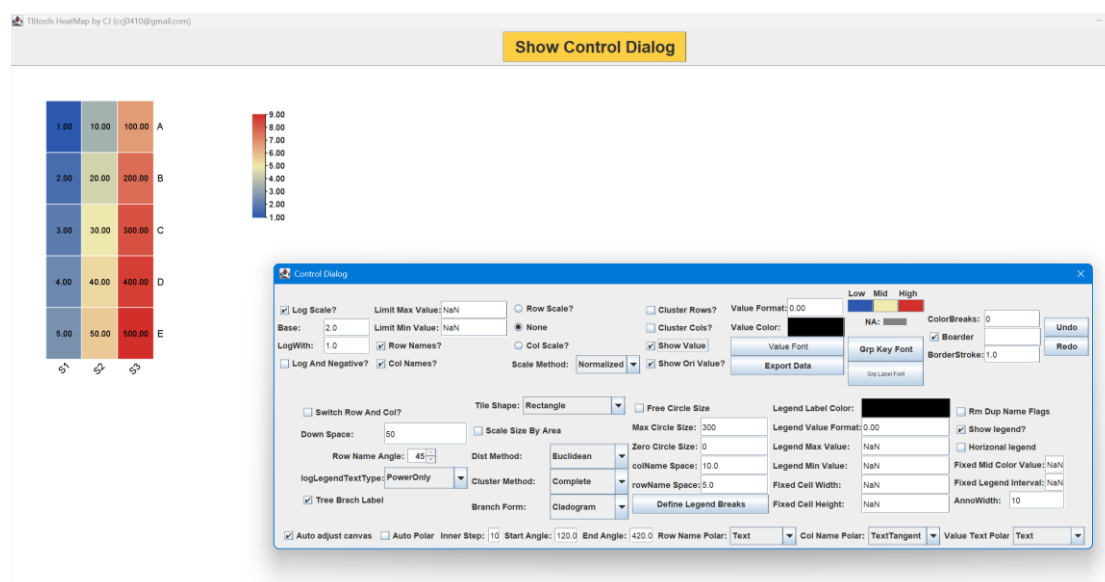
2) 点击开始按钮 Start, 生成初始热图;



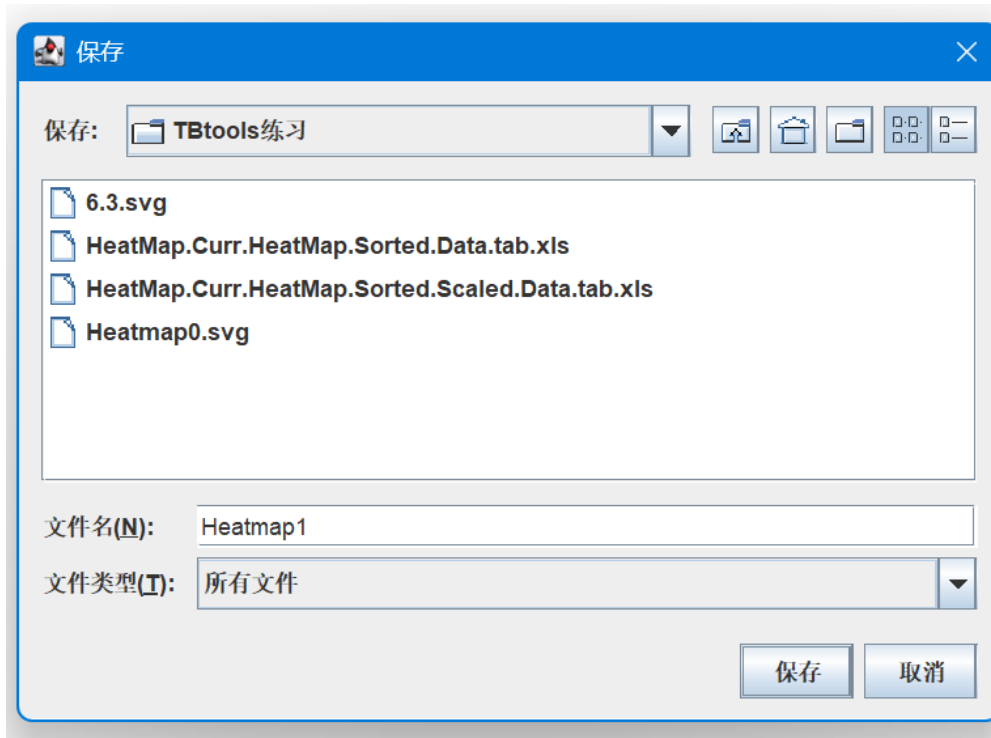
3) 在图形窗口中点击 Save Graph 按钮, 选择 SVG 格式, 保存初始热图 Heatmap0;



4) 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;

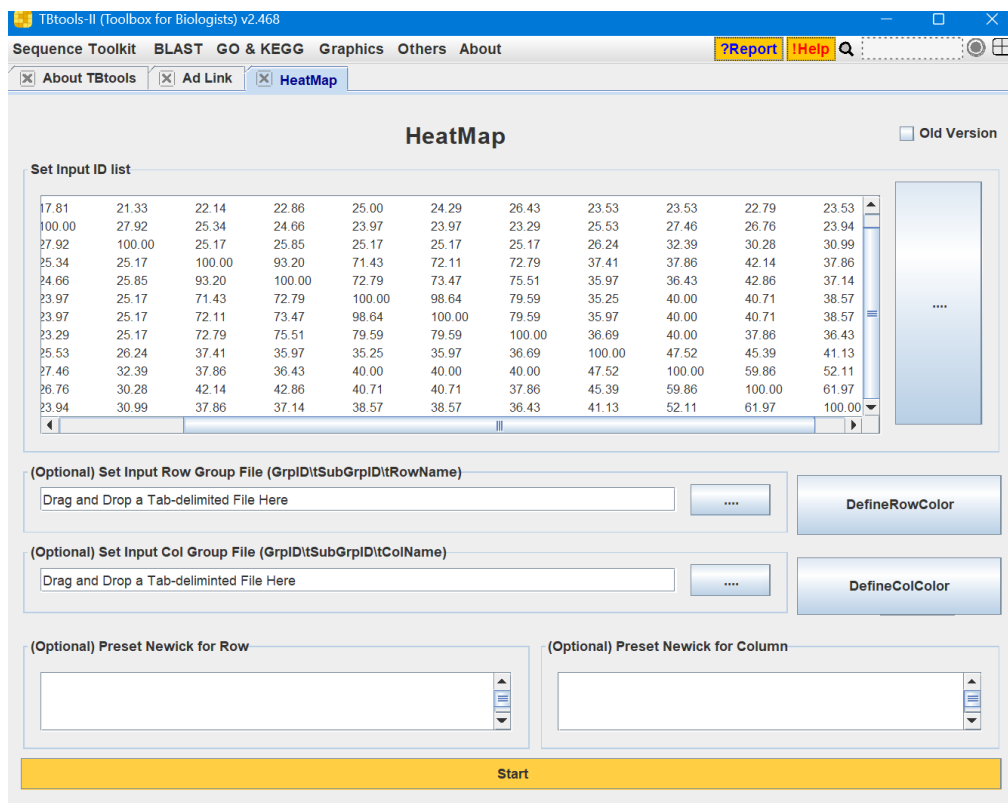


5) 在图形窗口中点击 Save Graph 按钮, 选择 SVG 格式, 保存新的热图 Heatmap1;

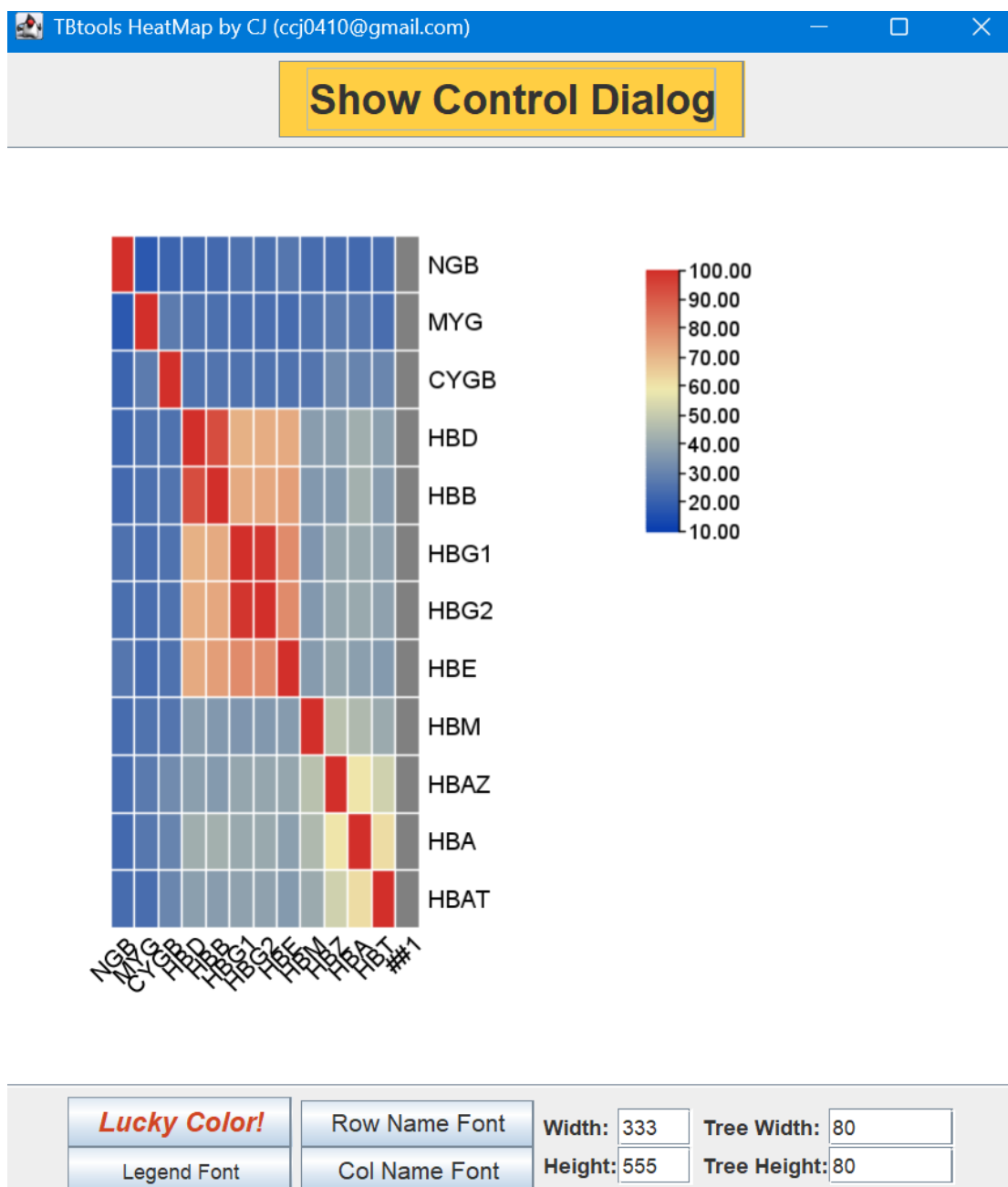


实例 6.2 — 珠蛋白家族相似性距离矩阵

1) 打开 TBtools, 选择图形 Graphics 热图分析 Heatmap; 在输入窗口中输入: Globin Matrix;

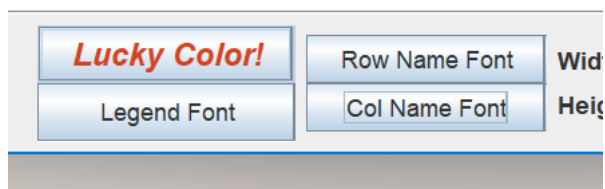


2) 点击开始按钮 Start, 生成初始热图;

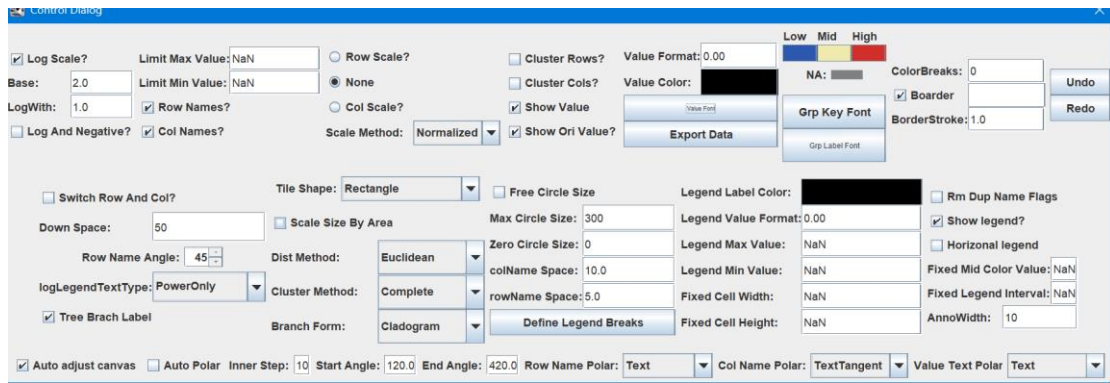


3) 在图形窗口中打开会话窗口 Show Control Dialog, 设置并调节参数, 生成热图;

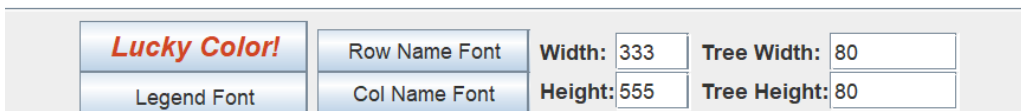
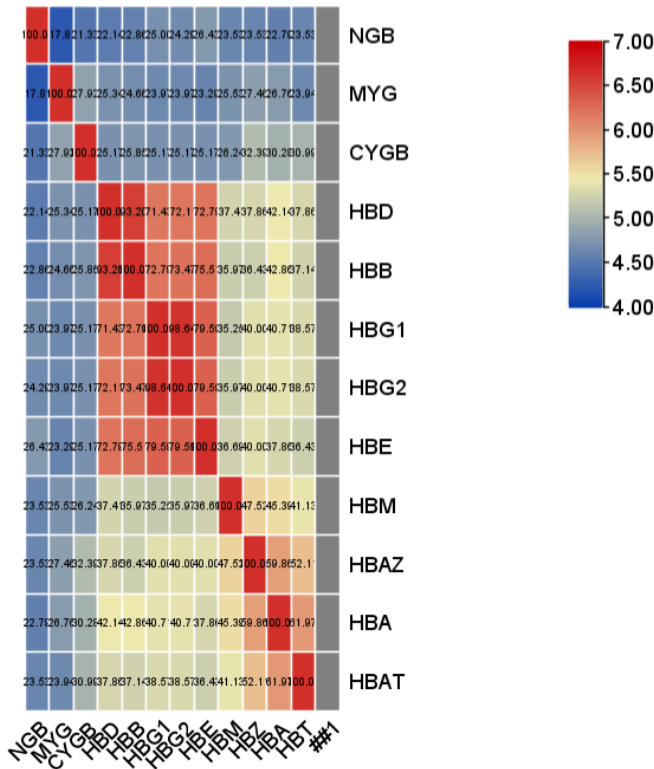
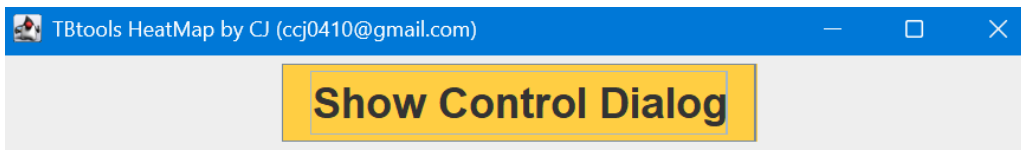
调整字体:



调整其它参数:



4) 在图形窗口中点击 Save Graph 按钮, 选择 SVG 格式, 保存新的热图
Globin Matrix Heatmap



5) 结果分析: 本研究构建了珠蛋白家族 12 个成员的序列相似性距离矩阵热图, 结果显示: α -和 β -珠蛋白亚家族内部成员间序列相似性极高, 交叉区域呈明显的红色或橙黄色, 而神经珠蛋白 (NGB)、肌红蛋白 (MYG)、细胞珠蛋白 (CYGB) 与其他成员的序列相似性极低, 交叉区域呈深蓝色, 提示它们在进化上属于更早分化的独立分支, 与经典血红蛋白存在显著的序列与功能分化。

个人总结

G2A 边汉青个人总结

通过这段时间的学习，我主要获得了一下收获：

1.基本掌握了 TBtools 的常用功能。在生信分析实践中，我逐步掌握了 TBtools 的几项常用功能。

序列比对功能。传统在线比对工具在高峰时段常因网络拥堵而影响效率，而 TBtools 内置的 BLAST 功能使用户可以直接在本地环境下完成序列比对，将查询序列与数据库中的序列进行比较，快速找到高度相似的序列。此外，TBtools 还新增了 BLAT 功能，其轻量化设计仅需一个小程序即可实现类似 BLAST 的比对，特别适用于近源序列的比对分析。

序列提取功能。TBtools 的 Fasta Extract 功能支持从 Fasta 序列文件中提取某一个或多个特定基因，操作简便；同时，软件还提供了“获取最长代表性序列”功能，帮助用户从含有多条序列的文件中提取最长转录本，用于下游分析。这些功能将原本需要手动筛选的基因数据快速提取出来，大大减少了时间成本。

序列信息统计功能。利用 Fasta Stats 工具可以一次性获取序列总长（Total Len）、染色体数（Total Seq Num）和未测通碱基数（Total N Counts）等关键信息，帮助快速掌握数据全貌，显著提升信息处理效率。

本地数据库建立功能。通过 TBtools 可以构建自己专属的本地 BLAST 数据库，并且数据库可以随时“重用”——关闭 TBtools 重新开启后，数据库列表与上一次关闭时保持一致。这一功能既实现了比对等多种应用，又从源头上杜绝了敏感数据上传至公共服务器可能引发的外泄风险。

2.学习途径丰富多样，多平台资源互为补充。

一方面可以通过软件创建者提供的 PPT 和讲解回放系统性地学习常用功能；另一方面，哔哩哔哩等平台上大量用户分享的特定功能教程，便于按需检索和学习。

3.深刻认识到基础知识的重要性。

仅仅停留在软件操作层面是远远不够的，必须认识到所做分析的生物学背景，理解每个步骤的科学意义。例如，在构建系统发生树时，需要深刻理解序列相似性与同源性的区别、直系同源与旁系同源的差异、核苷酸替换模型和氨基酸替换

模型的适用条件等基本概念；还需要掌握不同建树方法（如邻接法、最大简约法和最大似然法）的基本原理和特点，理解自举法（**Bootstrap**）检验进化树稳定性的原理。只有具备了这些基础知识，才能在选择分析方法和参数时做出合理判断，而非盲目套用默认设置；在得到结果后，才能够正确解读结果的含义，判断其可靠性；进而将生信分析结果与具体的生物学问题相联系，得出具有科学价值的结论。

G2B 刘奇个人总结

TBtools 不仅是零散的工具集成，而是一条可以串起来用的分析链条。

首先序列比对，这是最基础也最常用的一步，并且操作简便：导入 **FASTA** 序列，一键运行，即可获得对齐后的结果。通过比对，能看到哪些位点是高度保守的，哪些物种间存在特异性变异。这在研究同源基因功能、寻找 **SNP** 位点或为进化树准备输入文件时非常实用。例如，当我从不同物种中获取某个基因家族成员的序列后，通过比对可以快速判断该基因的关键结构域是否完整，从而决定哪些序列可以进入下一步分析。

家族成员比对，跨物种去找直系同源关系。比如想找拟南芥和水稻 **SPL** 家族之间的直系同源关系，就把两个物种的所有 **SPL** 蛋白序列放在一起，用 **TBtools** 做多序列比对。比对结果输出为 **table** 格式，打开后是一张清晰的表格，通过这张表，你可以直观地看哪位拟南芥成员与哪位水稻成员同源性最高。

数据库构建和搜索，也就是本地 **BLAST**。这个功能在两种场景下特别好用：一是你手头有一个参考基因组或者蛋白集，想快速找出里面属于某个基因家族的所有成员；二是测了一批新序列，想知道它们跟已知的哪个基因最像。**TBtools** 的好处是把建库和搜索都做成了图形界面，不用敲命令行，点几下就能跑完。拿到检索结果后，就要用到序列提取了。**TBtools** 这个功能非常直接：输入一个 **ID** 列表，或者一段基因组区间，指定好源文件，它就能把对应的序列一次性取出来，很便捷也不容易出错。

当你有了几个不同的基因集，想看看它们之间有多少重叠、多少独有的，这时候可以使用韦恩图。**TBtools** 做韦恩图很简单，把几个列表文件导进去，就会

自动出图，能帮助快速把几个数据集之间的关系理清楚。

最后是热图，主要是用来展示表达模式或者其它定量数据。比如有多个样本、多个基因的表达量，做成矩阵导进去，TBtools 会直接生成一张颜色渐变的图。如果日常工作中将这些工具结合串联起来使用，就不需要总是更换软件或网站，比较省力。

G2C 高倩个人总结

过去两周的学习主要围绕 TBtools 软件的生信分析与可视化展开，从基础功能到课题实例实操，练习了从序列比对、序列提取到结果可视化。

课堂上，陈程杰老师与仝宗军老师亲自授课，讲解 TBtools 的开发历程、功能模块与课题应用，陈老师从用户需求到开发理念的分享，让我深受鼓舞，也对国产生物信息学工具的发展有了更深的理解。TBtools 是一款功能齐全、界面友好的国产生物信息学软件，涵盖序列处理、BLAST 数据库搜索、基因注释与富集、维恩图、热图、进化树构建等核心工具，为后续课题研究很有帮助。在基础功能学习中，我逐步熟悉了软件的主要模块：从序列比对、BLAST 数据库构建与搜索，到维恩图（Venn）、热图（Heatmap）、等可视化工具，理解了不同模块的适用场景与操作逻辑，当然还有更多模块值得我去挖掘。

在课题实例实操环节，以课程练习为载体，将理论知识转化为实操能力：完成了癌胚抗原家族成员双序列比对、拟南芥与水稻 SBP 转录因子家族成员比对，掌握了序列比对的参数设置与结果解读；构建并搜索玉米 SBP 转录因子数据库，练习了序列提取与批量处理；通过维恩图分析与热图分析练习，学会了用颜色梯度和交集展示基因和序列的相似性与差异特征，能直观呈现不同数据集的分布规律。

通过课程实例练习，我体会到 TBtools “一站式、交互式” 分析的优势，也意识到软件功能的强大与实操的灵活性。但是目前我的学习仍存在不足：一是对部分工具的参数设置逻辑理解不够透彻，复杂数据的批量处理与高级可视化操作不够熟练，后续会通过强化实操练习，将 TBtools 与本学期所学的序列比对、BLAST、系统发育树构建等内容结合，为课题提供更高效的分析支持。