

“实用生物信息技术”课程小组讨论总结报告

组次: G5 组长: 韩佳凝 执笔: 韩佳凝

1. 时间

2026年4月29日

2. 方式

线上腾讯会议讨论

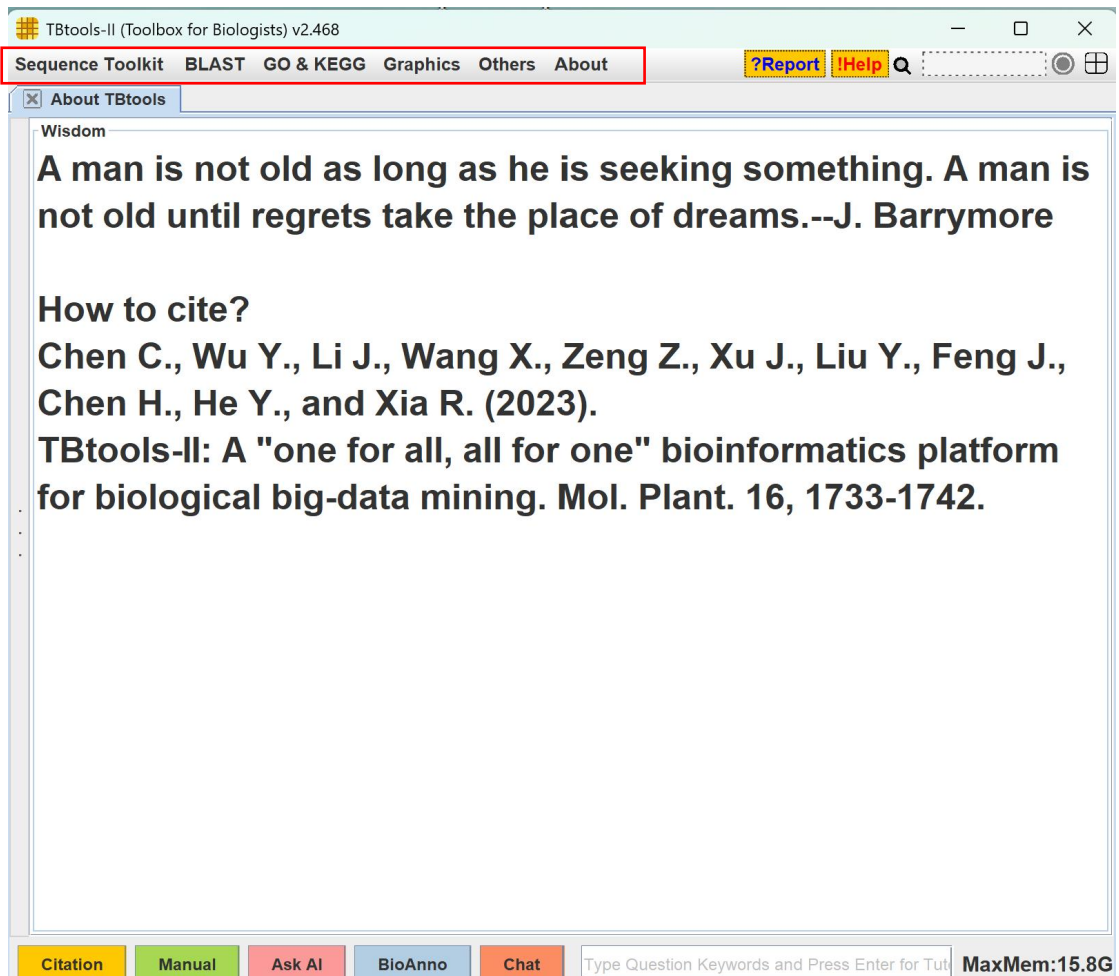
3. 主题

课后对 TBtools 的复习总结

4. 总结内容

(1) TBtools 简介

TBtools 以“一人为所有人，所有人为一人”（one for all, all for one）为开发理念，旨在为湿实验背景的生物学研究者提供无需编程基础的生物大数据分析解决方案。软件的逻辑可以总结为 IOS——Input（输入文件）、Output（输出路径）、Start（开始运行），用户只需拖放文件、选择输出位置、点击启动，即可完成数据分析。



红色框内是不同模块的操作页面，点击可进入相应分析的操作页面。

- **Sequence Toolkit (序列工具集)**: 核心是序列的编辑、处理与信息提取。包括按 ID 或区间提取序列、序列的统计与信息获取（如长度和 GC 含量）、批量重命名或拆分合并序列以及序列的格式化（如反向互补）等。
- **BLAST (序列比对)**: 将复杂的 BLAST 功能图形化。可以进行本地比对、远程比对（直接比对到 NCBI 数据库）以及比对结果的可视化与转换。
- **GO/KEGG (功能注释与富集分析)**: 用于理解基因或蛋白的生物学功能。可以进行 GO/KEGG 富集分析并一站式进行差异基因的功能富集与结果可视化。
- **Graphics (可视化绘图)**: 支持多种图表（包括热图和火山图等）的绘制。

(2) TBtools 的安装问题

小组成员多数在 GitHub 网站下载链接，常见的安装流程和使用说明可从语雀网站查看

TBtool

简介

1. 主要特点 — 功能齐全，界面友好，操作方便，文档丰富，自动更新
2. 开发团队 — 中国热带农业科学院热带作物品种资源研究所 陈程杰课题组 华南农业大学园艺学院 夏瑞课题组
3. 下载地址 — [GitHub](#) 本地下载
4. 帮助文档 — [语雀网站](#)
5. 网络视频 — [哔哩哔哩](#)

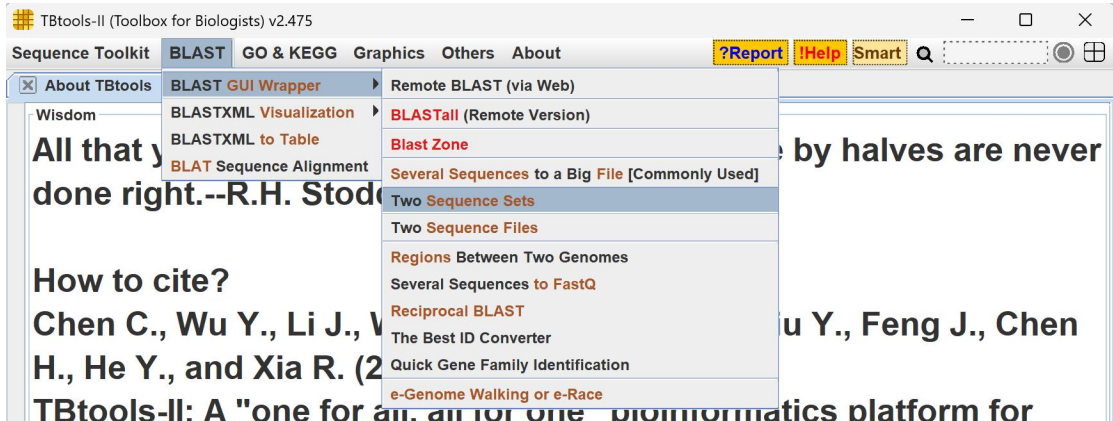
(3) 课外练习

A. 癌胚抗原家族成员双序列比对

操作步骤

从 UniProt 数据库下载 CEAM6_HUMAN 第 145-232 位第一个恒定结构域 Ig-like C2-type1 序列 CEA6C2 和 CEAM5_HUMAN 全长序列 CEA5（也可以在 abc 网站直接下载，右键点击链接后选择“链接另存为”保存至文件夹中）

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:\TBtool_Example\CEA\out\CEA.txt (该文件夹需提前建好)

在输出格式 Outfmt 选择框选择成对比较 Pairwise

点击开始 Start 按钮，比对结果保存到输出文件夹 CEA\out\中： CEA.txt



结果如下图所示

(1) 结果中中有一段主要的比对结果 (Score = 153 bits, 146 bits, 143 bits)，这意味着你的短序列 (Query) 在 CEAM5 全长序列 (Sbjct) 上找到了三个非常像的区域。

	CEAM5 序列区间	相似度
1	501-561	85%
2	145-205	81%
3	323-383	80%

Database: TBtoolsBlastSubject15097649318027750981fa
1 sequences; 702 total letters

Query= CEA6C2 | CEAM6_HUMAN Ig-like C2-type1 Domain | 145-232

Length=88

E	Score
Sequences producing significant alignments:	(Bits)
Value	
CEAM5_HUMAN P06731 Cell adhesion molecule CEACAM5	153
2e-49	

>CEAM5_HUMAN | P06731 | Cell adhesion molecule CEACAM5
Length=702

Score = 153 bits (386), Expect = 2e-49, Method: Compositional matrix adjust.

Identities = 75/88 (85%), Positives = 78/88 (89%), Gaps = 0/88 (0%)

Query 1 PKPSISSNNSNPVEDKDAVAFTCEPEVQNTTYLWVWNGQSLPVPRLQLSNGNMTLLS
60
PKPSISSNNS PVEDKDAVAFTCEPE QNTTYLWVWNGQSLPVPRLQLSNGN TLTL +
Sbjct 501 PKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWVWNGQSLPVPRLQLSNGNRTLTLFN
560

Query 61 VKRNDAGSYECEIQNPASANRSDPVTLN 88
V RND +Y C IQN SANRSDPVTL+
Sbjct 561 VTRNDARAYVCGIQNSVSNRSDPVTLD 588

Score = 146 bits (368), Expect = 6e-47, Method: Compositional matrix adjust.

Identities = 71/88 (81%), Positives = 74/88 (84%), Gaps = 0/88 (0%)

Query 1 PKPSISSNNSNPVEDKDAVAFTCEPEVQNTTYLWVWNGQSLPVPRLQLSNGNMTLLS
60
PKPSISSNNS PVEDKDAVAFTCEPE Q+ TYLWVWN QSLPVPRLQLSNGN TLTL +
Sbjct 145 PKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWVWNNQSLPVPRLQLSNGNRTLTLFN
204

Query 61 VKRNDAGSYECEIQNPASANRSDPVTLN 88
V RND SY+CE QNP SA RSD V LN
Sbjct 205 VTRNDTASYKCETQNPVSARRSDSVILN 232

Score = 143 bits (361), Expect = 6e-46, Method: Compositional matrix adjust.

Identities = 70/88 (80%), Positives = 74/88 (84%), Gaps = 0/88 (0%)

Query 1 PKPSISSNNSNPVEDKDAVAFTCEPEVQNTTYLWVWNGQSLPVPRLQLSNGNMTLLS
60
PKP I+SNNSNPVED+DAVA TCEPE+QNTTYLWVWN QSLPVPRLQLSN N TLTL
Sbjct 323 PKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWVWNNQSLPVPRLQLSNDNRTLTL
382

Query 61 VKRNDAGSYECEIQNPASANRSDPVTLN 88
V RND G YEC IQN S + SDPV LN
Sbjct 383 VTRNDVGPYECGIQNELSVDHSDPVILN 410

(2) CEA6C2 与 CEAM5 的 C 端第三个恒定结构域（对应序列 501-560 这一段）相似性最高。

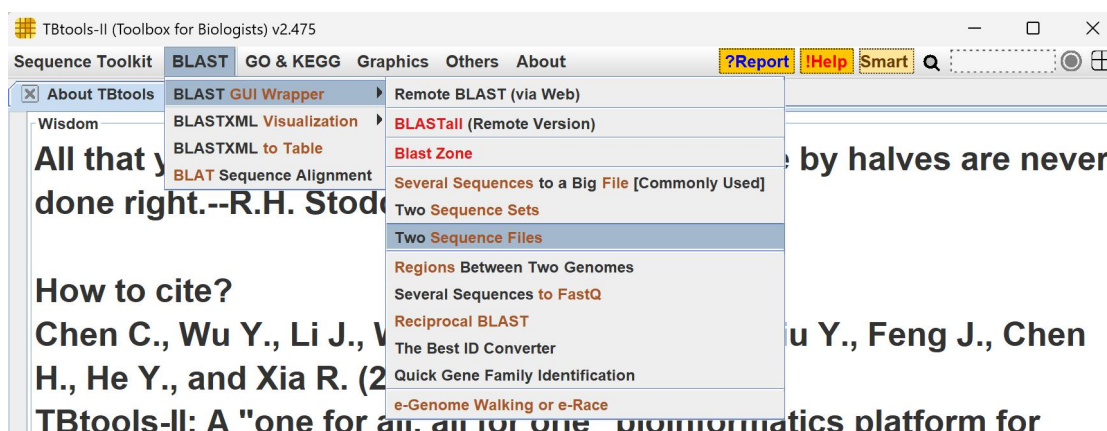
(3) TBtools 的页面简洁，操作界面比较简单，适合批量或快速验证。

B. 拟南芥和水稻 SBP 转录因子家族成员比对

操作步骤

从 ABC SBP Project 网站下载 17AT_SPL.FAS 和 19OJ_SPL.FAS 保存到 TBtools_Example\SBP 文件夹中（右键点击并选择链接另存为）

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 中选择双序列比对 Two Sequence Files



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

在输出文件选择框中设定比对结果存放文件夹和文件名 C:\TBtool_Example\SBP\Out\SPL.txt

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，找出水稻中拟南芥 SPL7_ARATH 的直系同源基因 SPL.txt

Blast Compare Two Seqs [Sets] <Big File>

Set Query Seqs

_win64_portable.2.466\Example\SBP\17AT_SPL.FAS

粘贴 17AT_SPL.FAS

Set Subject Seqs

win64_portable.2.466\Example\SBP\190J_SPL.FAS

粘贴 190J_SPL.FAS

Set Output Blast Result File

_win64_portable.2.466\Example\SBP\SBP17+19

选择输出路径

Start

Other Parameters

Outfmt: Table

NumofThreads:

E-value:

NumofHits:

NumofAligns:

Short Query Sequences(<=100bp)

OtherOptions

SPL7_ARATH 的比对结果如下图所示

注：结果第三列是序列一致性，第四列为比对长度，第十列为期望值（越小越好），最后一列为比特得分，综合了相似性和比对长度的打分（越大越好）

序列名称	序列一致性	比对长度	期望值	Bit score
SPL9_ORYSJ	37.17%	436	1.64e-169	502

则拟南芥 SPL7_ARATH 在水稻（粳稻）中的直系同源基因是 SPL9_ORYSJ。

```

# BLASTP 2.12.0+
# Query: SPL7_ARATH
# Database: 190J_SPL.FAS.TBtoolsDB
# 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 hits found
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

```

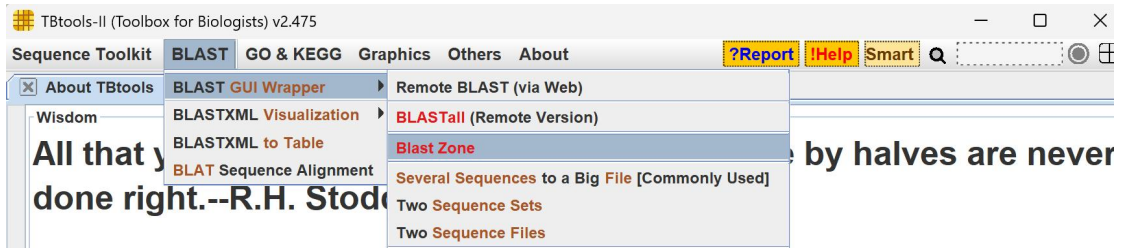
SPL7_ARATH	SPL6_ORYSJ	29.675246	156	4	307	548	432	664
1.20e-26	106							
SPL7_ARATH	SPL6_ORYSJ	47.664107	53	2	125	228	138	244
9.88e-25	100							
SPL7_ARATH	SPL14_ORYSJ	57.33375	32	0	137	211	103	177
2.34e-25	100							
SPL7_ARATH	SPL2_ORYSJ	42.453106	53	3	127	224	77	182
1.73e-23	94.4							
SPL7_ARATH	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							

C.玉米 SBP 转录因子数据库构建和搜索

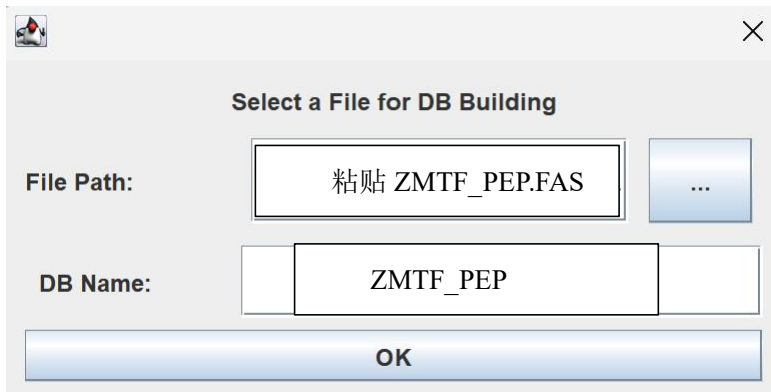
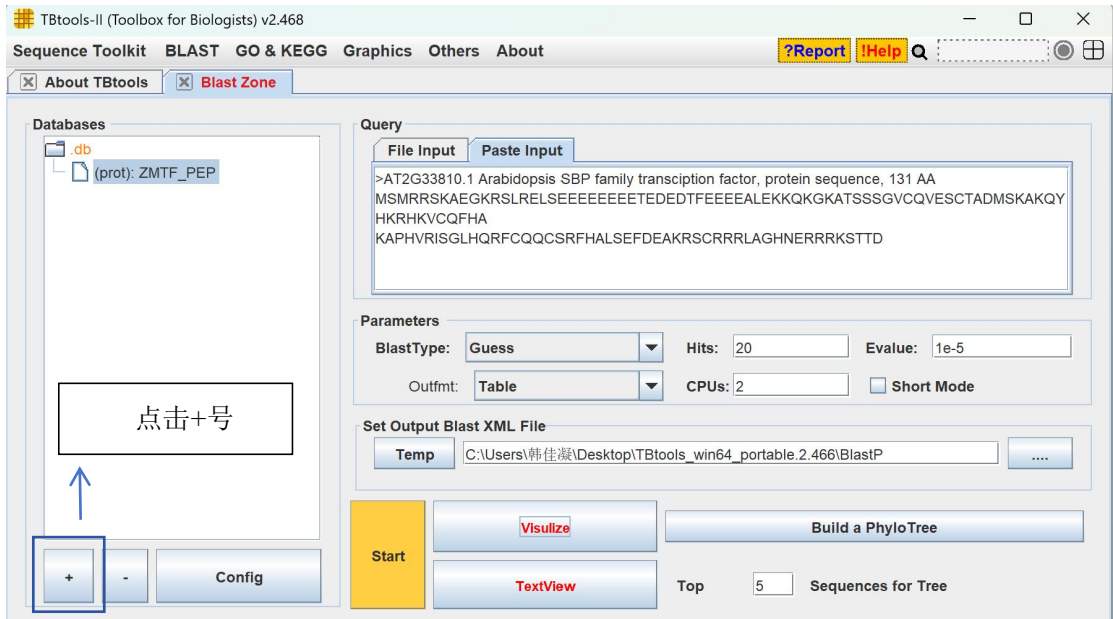
操作步骤

从 ABC SBP Project 网站下载 764 个玉米转录因子蛋白质序列数据集 ZMTF_PEP.FAS 和拟南芥 SPL3 转录因子蛋白质序列 SPL3_PEP.FAS 保存到 TBtools_Example\SBP 文件夹中

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



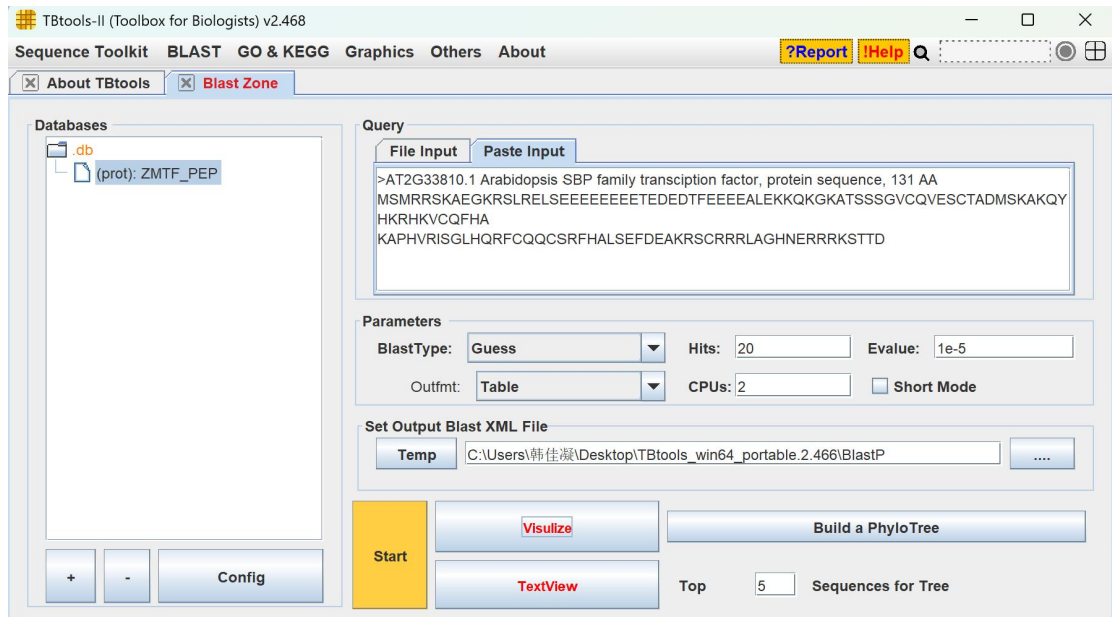
点击 BLAST Zone 用户界面左侧数据库 Database 窗口下方添加数据库按钮"+", 在弹出窗口中选择数据集 ZMTF_PEP.FAS, 在数据库名称输入框中输入 ZMTF_PEP, 点击 OK, 数据库构建完毕



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, 点击已建好的数据集

在输出文件选择框中设定比对结果存放文件夹和文件名 C:\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



结果如下图所示

```
# 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
```

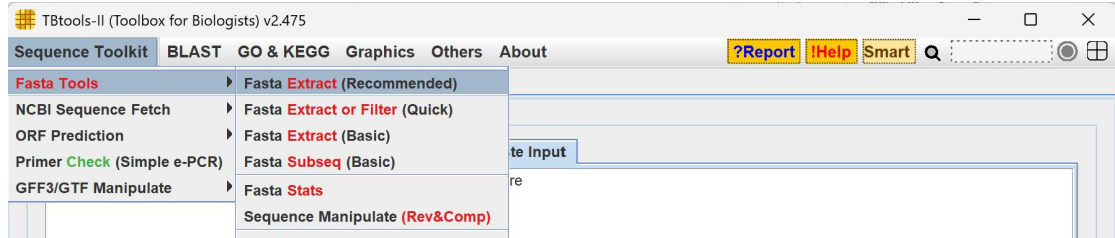
四个玉米蛋白与拟南芥 SPL3 蛋白具有显著的序列相似性，它们很可能是玉米中 SPL3 的同源基因。

序列	序列一致性	比对长度	期望值	Bit score
PTZm00608.1	45.08%	67	9.03e-34	115
PTZm00605.1	45.08%	67	1.32e-33	114
PTZm00606.1	56.76%	16	1.36e-12	57.0
PTZm00607.1	30%	63	9.62e-12	55.1

D.玉米 SBP 转录因子数据集中序列提取

操作步骤

打开 TBtools，在工具选择菜单中选择 Sequence Toolkit，在下拉菜单中选择 FASTA 格式序列提取 Fasta Extract (Recommended)



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

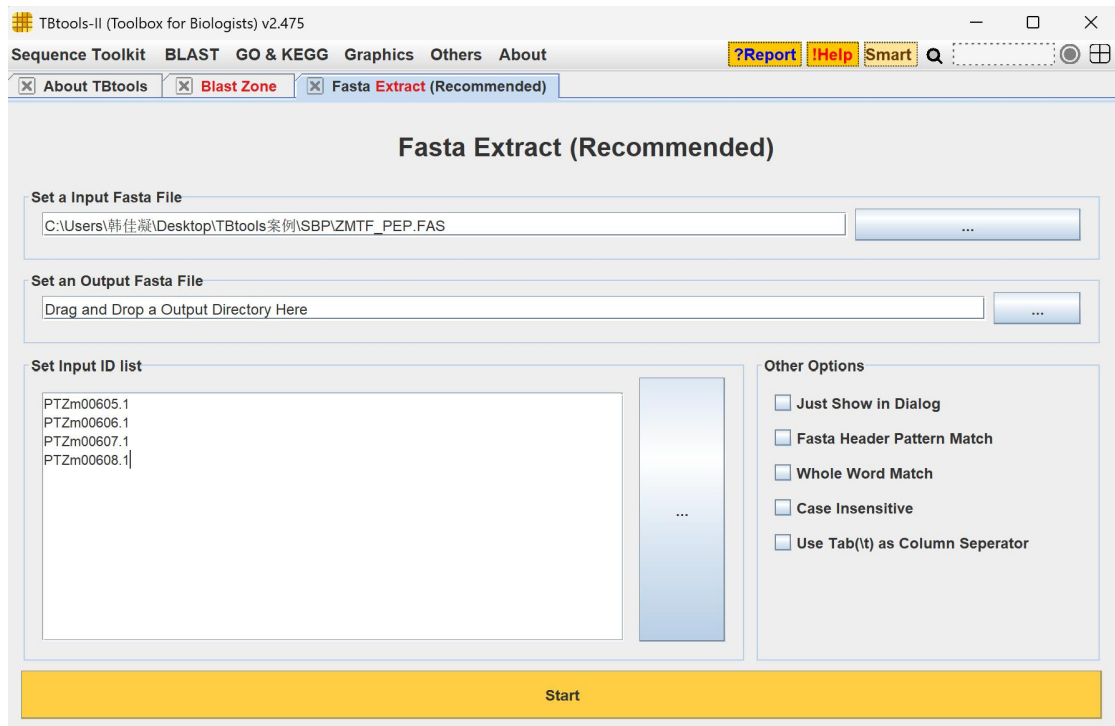
PTZm00605.1

PTZm00606.1

PTZm00607.1

PTZm00608.1

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



结果如下：

```

PTZm00605.1 |PUT-155a-Zea_mays-3434
FLEGLRSPVLIMGSFGMNWNQKDPMVWDWEHLVPSVSNVTRHGSANSSGGTLTNSNELG
HGSSKVKNTGTNGRVDDSGNSPSSMIAFNQGEPLISLKLKRAYFENACGGQDAKVSAAS
DV TSAASVVKTKVSQQNAKNWYCQVERCKVDLSSAKDYNRKHKVCVVHSKATKVVVAGL
ERRFCQQCSRFGHGLAEFDQNKRSRRRLMHHNARRRKPQADTISFNSSTMFYDTRQRTNL
FFSQPLYGQVRSNAGSSWDNLGGLKFMETKHPPVHPTKTASPDELHFSALQITSAAAHTG
HHHDLDFMAFKGTSTKVLNQVEAWXX
>PTZm00606.1 |PUT-155a-Zea_mays-44562
GAAQTSDLGMGSFGMDWNQKASVLWDWENLPPVAVGASGENPRMAAAPQALHSACGTI
SSSSEMGYSSKSSVSASADDSSPKAKGKSMELNFAPDKVPDKNTDSGKVDDARTSPSSP
VIAISSAEPVLSLKLKGRTYFEDVCGGQSVKSLPSDTSAVTPPPASPCKTKVTPNKKSYC
QVEGCKVDLSSAKEYHRKHRVCEPHSKAPKVVVAGL
>PTZm00607.1 |PUT-155a-Zea_mays-90777737
FLEGLRSPVLIMGSFGMNWNQKDPMVWDWEHLVPSVSNVTRHGSANSSGGTLTNSNELG
HGSSKSSISASIDSPSGVNSLEFNFAAVERHVKNKTGTNGRVDDSGNSPSSMIAFNQGEPL
LISLKLKRAYFENACGGQDAKVAASDV TSAASVVKTKVSQQNAKNWYCQVEGCKVDL
SSAKDYNRKHKVCVVHSKATKVVVAGLEVGFNVSADTRQRTNLFFSQPLYGQVRSNAGS
SWDNLGGLKFMETKHPPVHPTKTASPDELHFSALQITSAAAHTGHHHDLDFMAFKGTST
KXXX
>PTZm00608.1 |PUT-155a-Zea_mays-91977737
FLEGLRSPVLIMGSFGMNWNQKDPMVWDWEHLVPSVSNVTRHGSANSSGGTLTNSNELG
HGSSKSSISASIDSPSGVNSLEFNFAAVERHVKNKTGTNGRVDDSGNSPSSMIAFNQGEPL
LISLKLKRAYFENACGGQDAKVAASDV TSAASVVKTKVSQQNAKNWYCQVEGCKVDL
SSAKDYNRKHKVCVVHSKATKVVVAGLEERRFCQQCSRFGHGLAEFDQNKRSRRRLMHHNA
RRRKPQADTISFNSSTMFYDTRQRTNLFFSQPLYGQVRSNAGSSWDNLGGLKFMETKHPP
VHPTKTASPDELHFSALQITSAAAHTGHHHDLDFMAFKGTSTKVLNQVEAWXX

```

E. 韦恩图 Venn Diagram 分析

操作步骤一

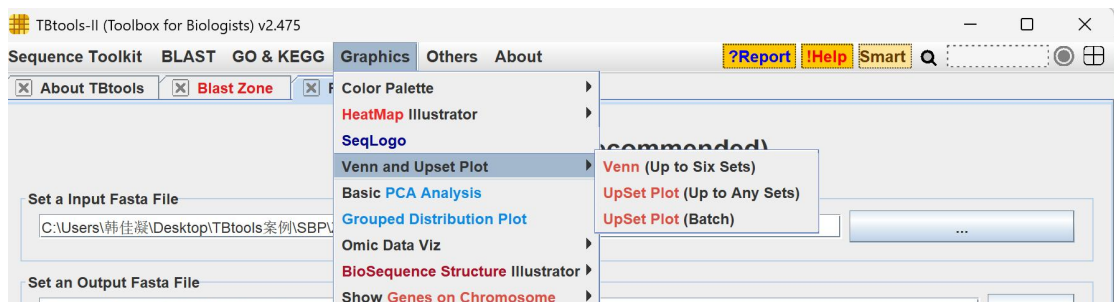
打开 TBtools，选择图形 Graphics 中韦恩图分析 Venn

在 Set1 输入窗口中逐行输入大写字母 A, B, C

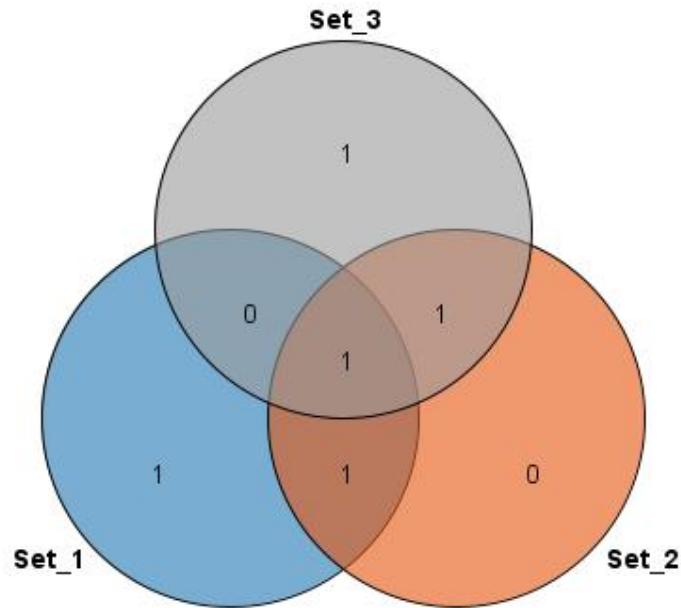
在 Set2 输入窗口中逐行输入大写字母 B, C, D

在 Set1 输入窗口中逐行输入大写字母 C, D, E

点击开始按钮 Start，生成韦恩图 Venn



结果如下：



操作步骤二

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

Sequence Toolkit BLAST GO & KEGG Graphics Others About ?Report !Help Smart Q

About TBtools Blast Zone Fasta Extract (Recommended) Venn (Up to Six Sets)

Interactive Venn Graph

Set 1: Human
HBZ
HBM
HBA2
HBA1
HBQ1
HBE1
HBG2

Set 2: Mouse
HBB-BH1
HBB-BH2
HBB-B1
HBB-B2
MB
CYGB
NGB

Set 3: Rat
HBE2
HBG1
HBB
HBB-B1
MB
CYGB
NGB

Set 4: Set_4

Set 5: Set_5

Set 6: Set_6

Start Region Overlap Mode

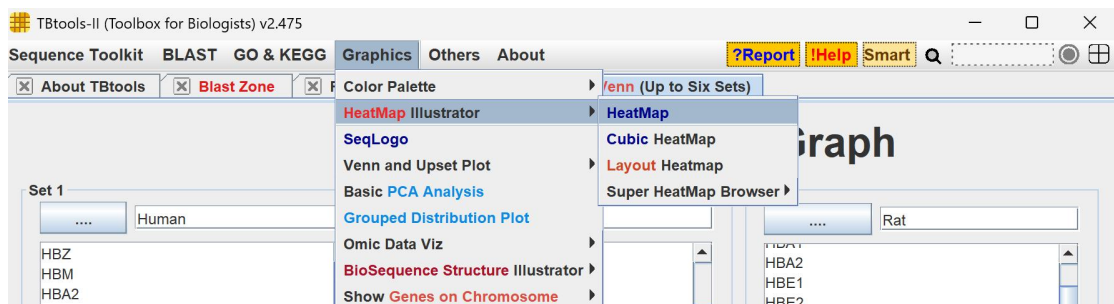
点击开始按钮 Start，生成韦恩图 Globin.svg



双击中间的交集区，可以看到共同拥有的基因具体名称。

即人、大鼠、小鼠三个物种中共有基因共有三个：MB、NGB、CYGB

F.热图 Heatmap 分析



操作步骤一

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

TBtools-II (Toolbox for Biologists) v2.475

Sequence Toolkit BLAST GO & KEGG Graphics Others About ?Report !Help Smart Q

About TBtools Blast Zone Fasta Extract (Recommended) Venn (Up to Six Sets) HeatMap

HeatMap Old Version

Set Input ID list

Drag Input Expression Matrix
 # With Column Names and Row Names
 # For Example:
 # =====

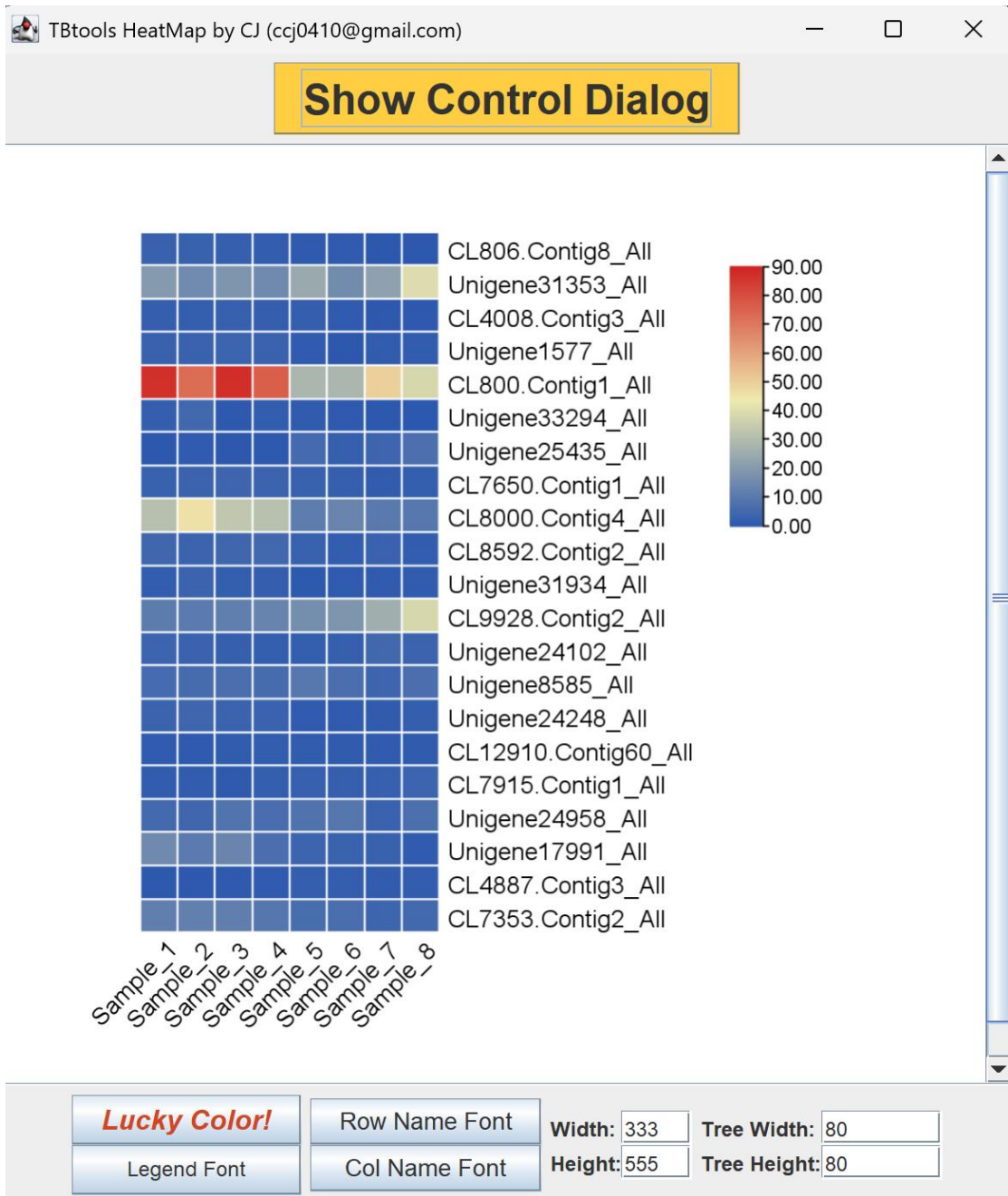
GeneID	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6	Sample_7	Sample_8	
CL806.Contig8_All	2.9811	2.8991	2.2991	1.7494	0.8362	1.1523	0.1844	0.2401	
Unigene31353_All	19.9784	15.3412	18.4279	15.0576	24.487	15.6895	22.2459	39.8472	
CL4008.Contig3_All	2.2377	1.9977	2.6257	2.0544	2.0217	1.1162	0.5906	0.5447	
Unigene1577_All	2.9771	3.2668	3.8315	3.1946	0.9526	0.248	1.4287	1.5114	
CL800.Contig1_All	87.4159	73.2224	87.7985	76.5201	28.4157	28.6147	49.8389	38.3058	
Unigene33294_All	2.0745	4.93	0.4763	1.7809	1.3144	0.7603	0.2053	0	

(Optional) Set Input Row Group File (GrpID\tSubGrpID\tRowName)
 Drag and Drop a Tab-delimited File Here

(Optional) Set Input Col Group File (GrpID\tSubGrpID\tColName)
 Drag and Drop a Tab-delimited File Here

(Optional) Preset Newick for Row

(Optional) Preset Newick for Column

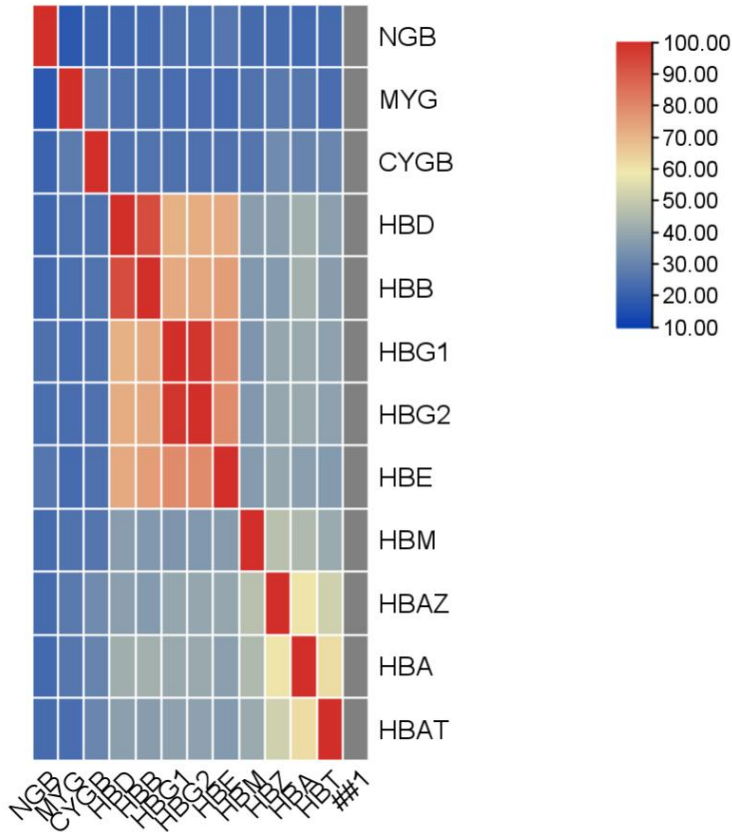


操作步骤二

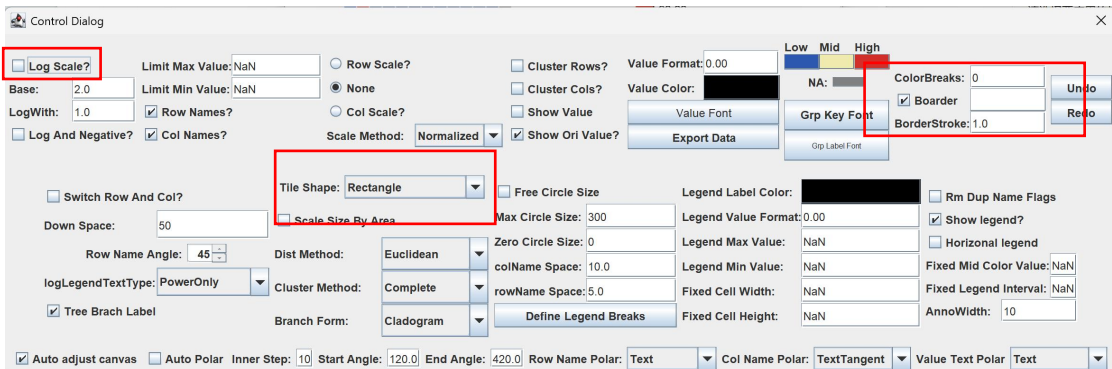
打开 TBtools，选择图形 Graphics 热图分析 Heatmap

在输入窗口中输入： Globin Matrix

点击开始按钮 Start，生成初始热图



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
 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



- 左上角 Log Scale?: 对原始数据进行对数转换, 有助于使偏态数据分布更对称, 突出相对变化而非绝对差异
- Tile Shape: 可以改变热图中每个单元格的形状, Rectangle 是矩形, 可变为 Circle (圆形) 等
- ColorBreaks: 调整热图颜色

5. 问题

(1) TBtools 中 Fasta Extract (Recommended) 和 Fasta Extract (Basic) 这两个版块有什么区别?

基本是一样的, 只不过 Fasta Extract (Recommended) 功能更为强大, 支持提取完整 ID 序列和指定区间序列的同时还能对提取的区间进行重命名, Fasta Extract (Basic) 是基础版, 但课上所用的功能二者处理结果完全相同。

