

---

## “实用生物信息技术”课程小组讨论总结报告

组：G4 次：5 组长：高嘉禾 执笔：赵雪倩

**1.时间：2026年5月13日 18:00**

**2.方式：线上（腾讯会议，微信）**

**3.主题：深入了解 TBtools 软件以及其中的各个板块功能，完成课外练习等。**

**4.内容：**

A. 熟悉了解 TBtools 软件功能以及使用逻辑；

B. 对课外练习中 TBtools 的简单案例进行实操；例如 Fasta 序列操作，简单 BLAST 序列对比，表格快速筛选提取，热图绘制以及 Venn 图制作。

C. 小组成员讨论说明遇到报错时如何有效沟通解决；

D. 小组成员分别介绍在自己课题中如何使用 TBtools 软件；

E. 小组成员分别说明使用时的问题并探讨解决办法。

## TBtools 软件:

### 一、TBtools 软件简介:

1. 相比于其他生物信息软件, TBtools 重点关注湿实验科研工作者日常的数据分析需求。其主要特点是:

(1) 用户友好的操作界面和高兼容性(跨平台, 兼容不同操作系统);

(2) 统一的功能使用逻辑, 即 IOS (Input-Output-Start, 输入-输出-启动), 具体解释可看下图;

(3) 全面的多功能集成, 覆盖了绝大多数日常生信数据解读需要; 在软件使用体验上, 充分考量湿实验出身朋友的软件使用思维, 在功能实现效果上, 尽可能考量干实验数据分析的结果解读需求。

(4) 独立自主研发的绘图引擎 (JIGplot: Java Interactive Graph Plotting Engine)。

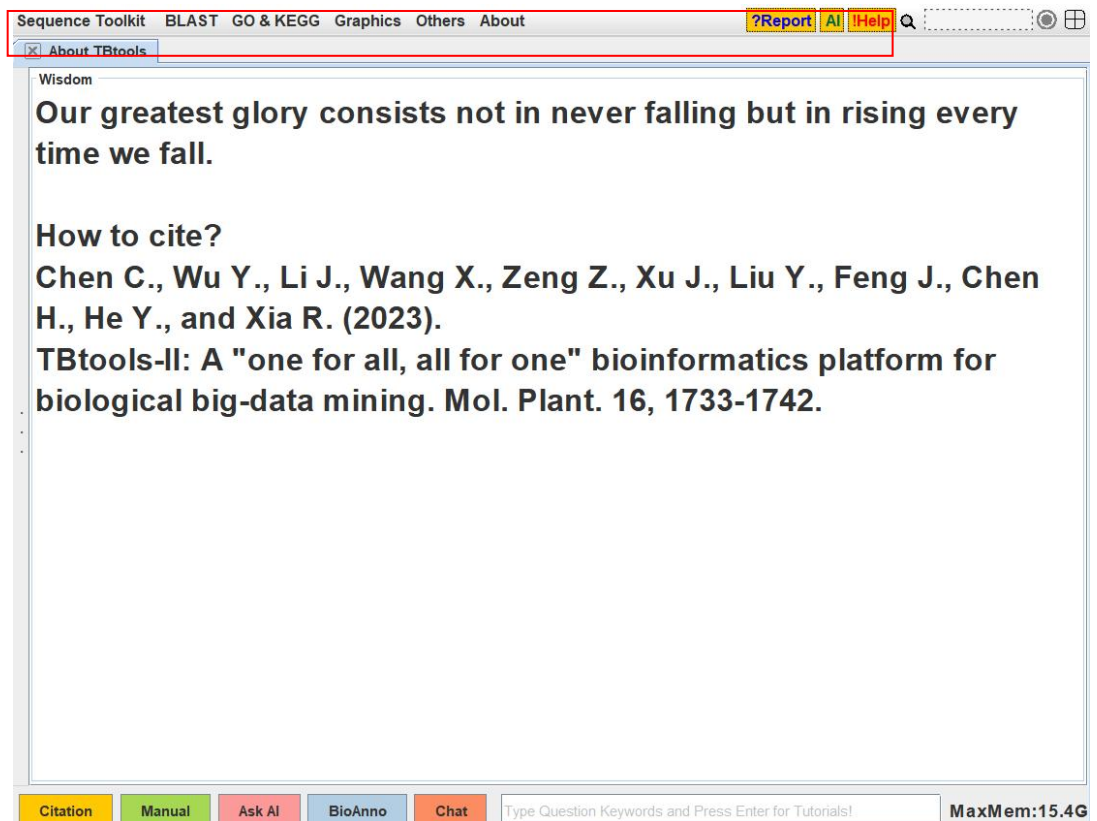


### 二、TBtools 软件的功能介绍:

目前, TBtools 具有超过 500 个功能, 其中界面化功能 300 余个。可以简单将功能划分如下:

- (1) Sequence Toolkits, 序列处理与操作
- (2) BLAST, 序列比对与可视化
- (3) GO & KEGG, 基因集合功能分析
- (4) Graphics, 生物信息学数据可视化
- (5) Others, 暂未划分功能

## (6) About, TBtools 相关信息与运行设置



### 三、TBtools 软件课外练习：

#### 实例 1 — 癌胚抗原家族成员双序列比对

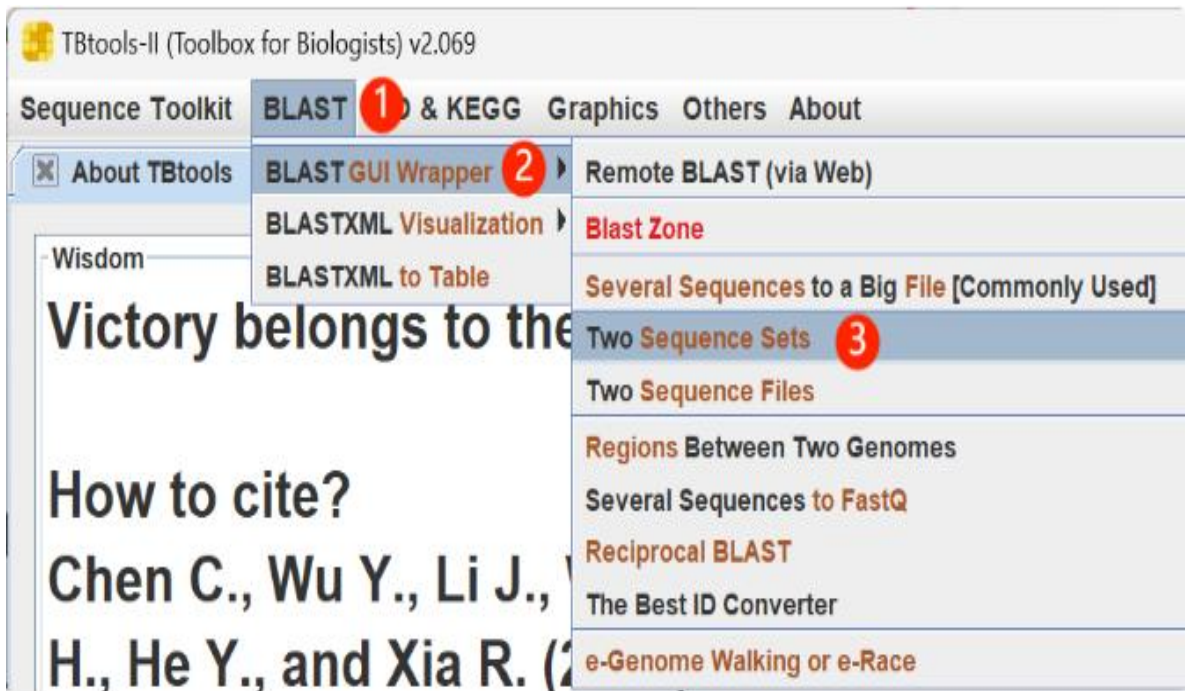
研究背景：

1. 人的癌胚抗原家族成员 CEA5 中三个与 CEA6 恒定结构域 A 相似的区域
2. TBtools 整合的 BLAST 双序列局部比对程序 Blast2Seq 用于寻找多个重复区域

使用 BLAST 比对两条或者两组序列，可以使用 Two Sequence Sets，此功能会自动选择合适的比对方式。从研究背景可知，BLAST 双序列局部对比 CEA5 中三个与 CEA6 恒定结构域 A 相似的区域，故使用 Two Sequence Sets。（注意，如果比对两个序列集合，甚至是两个物种的转录组序列集合，那么可以使用 Two Sequence Files 这一功能。）

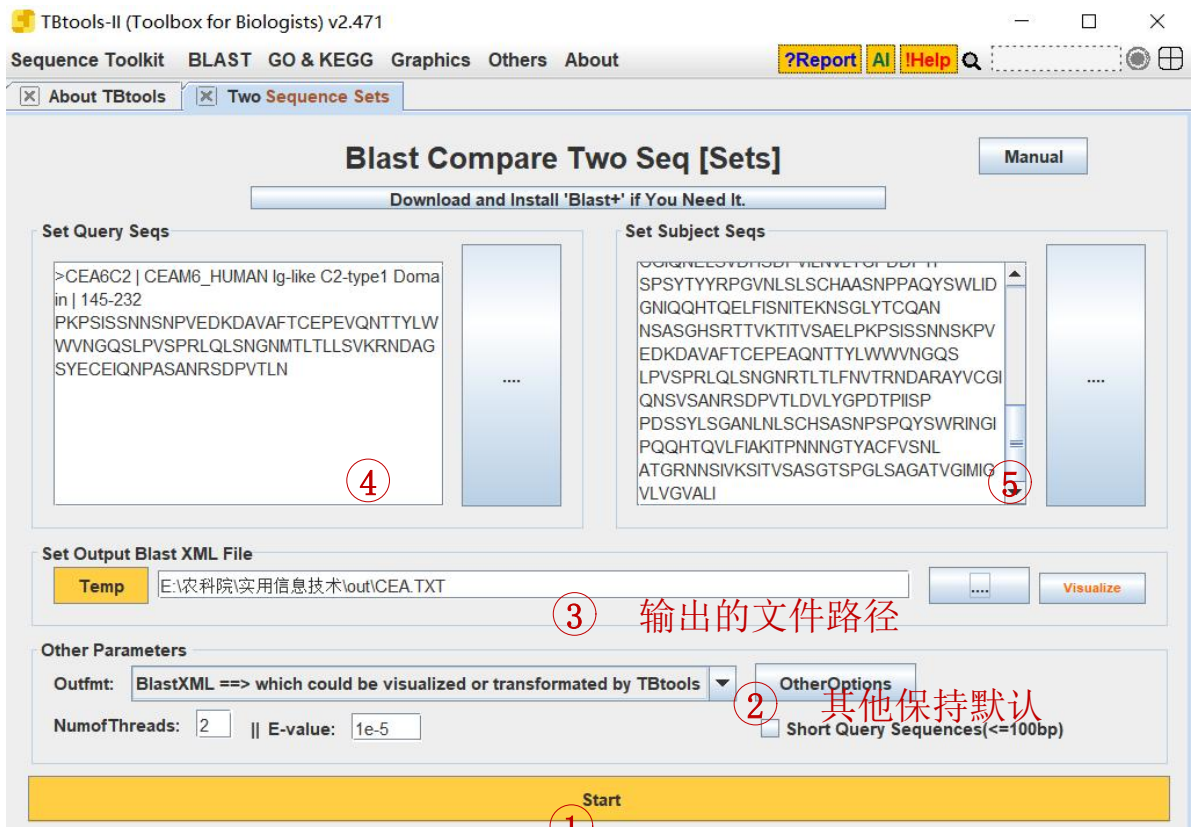
#### 实际操作

打开功能;



设置输入文件信息并设置输出文件，点击 start 即可。

注意该功能的使用逻辑，即 IOS（Input-Output-Start，输入-输出-启动），其中 Output 输出时注意设置输出文件的命名格式以及位置，方便后期找到，若忘记输出文件的命名，后续软件会出现报错。



---

注意：设置输出文件路径，也可以点击「Temp」按钮自动获取临时输出文件路径

**Outfmt**：设置输出文件格式；其中 **BlastXML**：适合可视化以及 TBtools 其他功能的文件格式；**Pairwise**：适合阅读的文件格式；**Table**：适合用于其他生信软件的文件格式。

## 实例 2 — 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

注意：该目的是比对拟南芥 SPL7\_ARATH 与粳稻中的直系同源基因的序列集合，上文中 TBtools 软件的 BLAST 序列对比中，如果比对两个序列集合，甚至是两个物种的转录组序列集合，那么可以使用 Two Sequence Files 这一功能。其中：

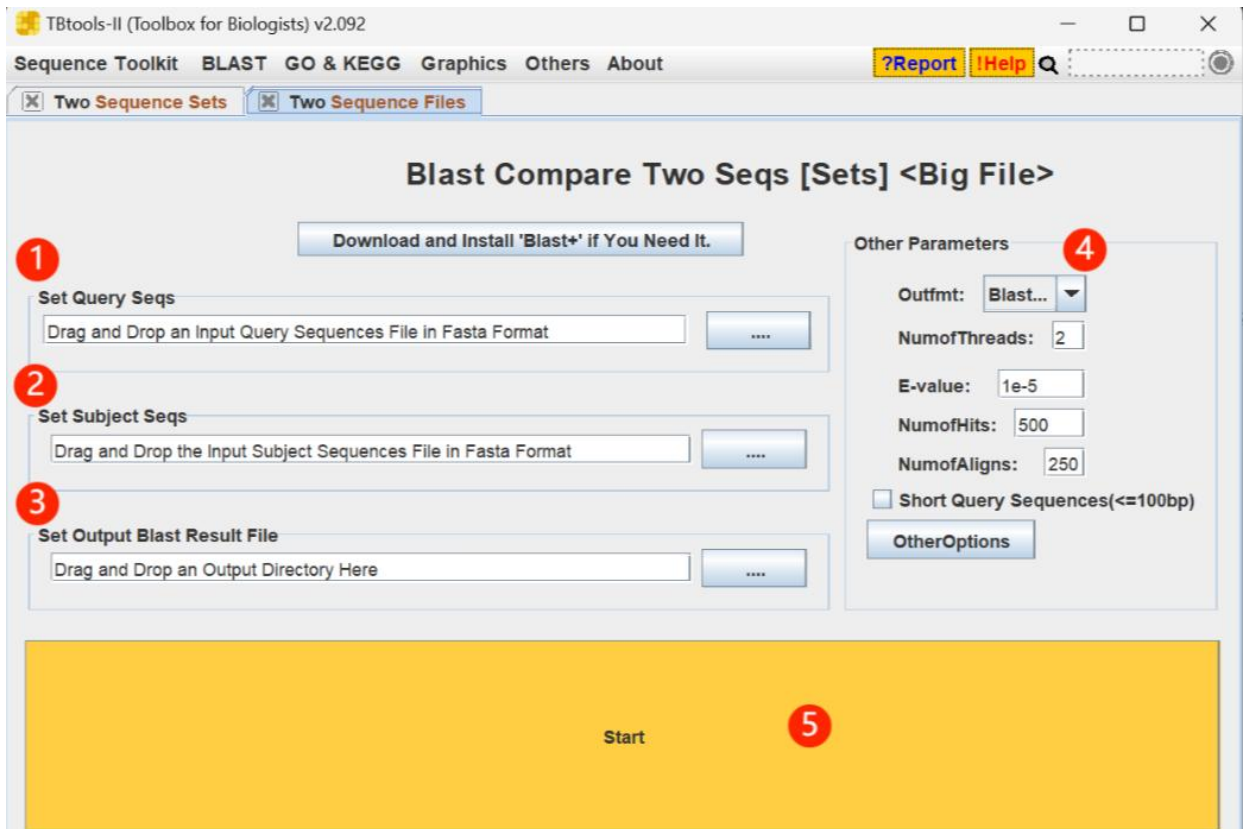
- **Outfmt**：设置输出文件格式

BlastXML：适合可视化以及TBtools其他功能的文件格式

Pairwise：适合阅读的文件格式

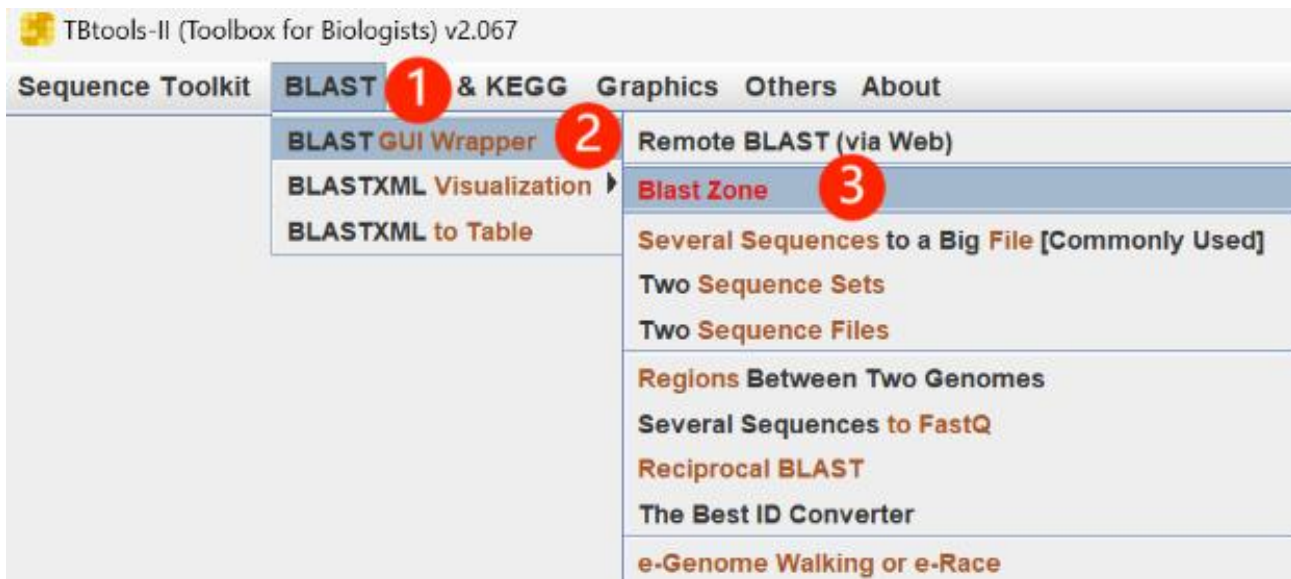
Table：适合用于其他生信软件的文件格式

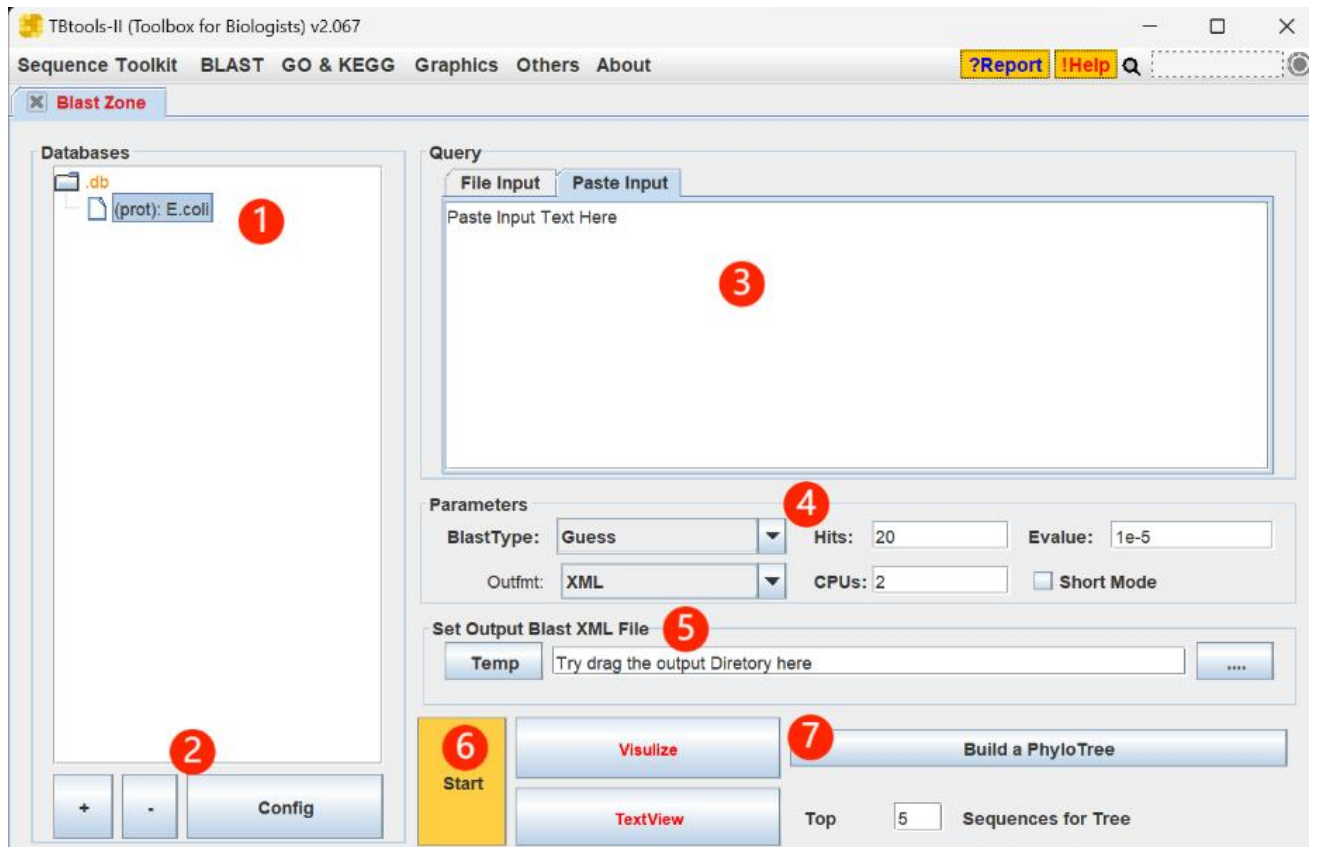
- **Number of Threads**：设置比对最高使用线程数，默认为2
- **E-value**：设置最高比对结果期望值，默认为1e-5
- **NumofHits**：默认为500
- **NumofAligns**：默认为250
- **Short Query Sequences(<= 100bp)**：使用针对短序列的敏感比对模式
- **Other Options**：其他可选参数



### 实例 3 — 玉米 SBP 转录因子数据库构建和搜索

研究背景：构建 764 个预测所得玉米转录因子数据库，搜索拟南芥 SPL3\_ARATH 相似序列





其中参数如下：

1. 输出文件：点击 **temp** 按钮后，不会输出新文件；若想得到输出文件，可以设置输出文件
2. 点击 **start**，输出结果
3. 点击 **start** 后对输出结果已有的本地数据库列表
4. 从左至右分别为创建数据库(+), 删除数据库(-), 配置数据库路径(**Config**: 打开可查看你的本地数据库路径，可调整数据库存放路径)
5. 输入文本框：可选择 **File Input** 拖入文件，也可选择 **Paste Input** 粘贴序列（若>10 个序列，建议使用文件模式）
6. 可选参数

**BlastType**: 比对类型设置，默认为 **Guess**，也可选择 **blastn**、**blastp**、**blastx**、**tblastn**、**tblastx**

**Hits**: 选择保留的 **hit** 数目

**Evaluate**: 比对结果最高期望值，一般保持默认值 **1e-5** 即可

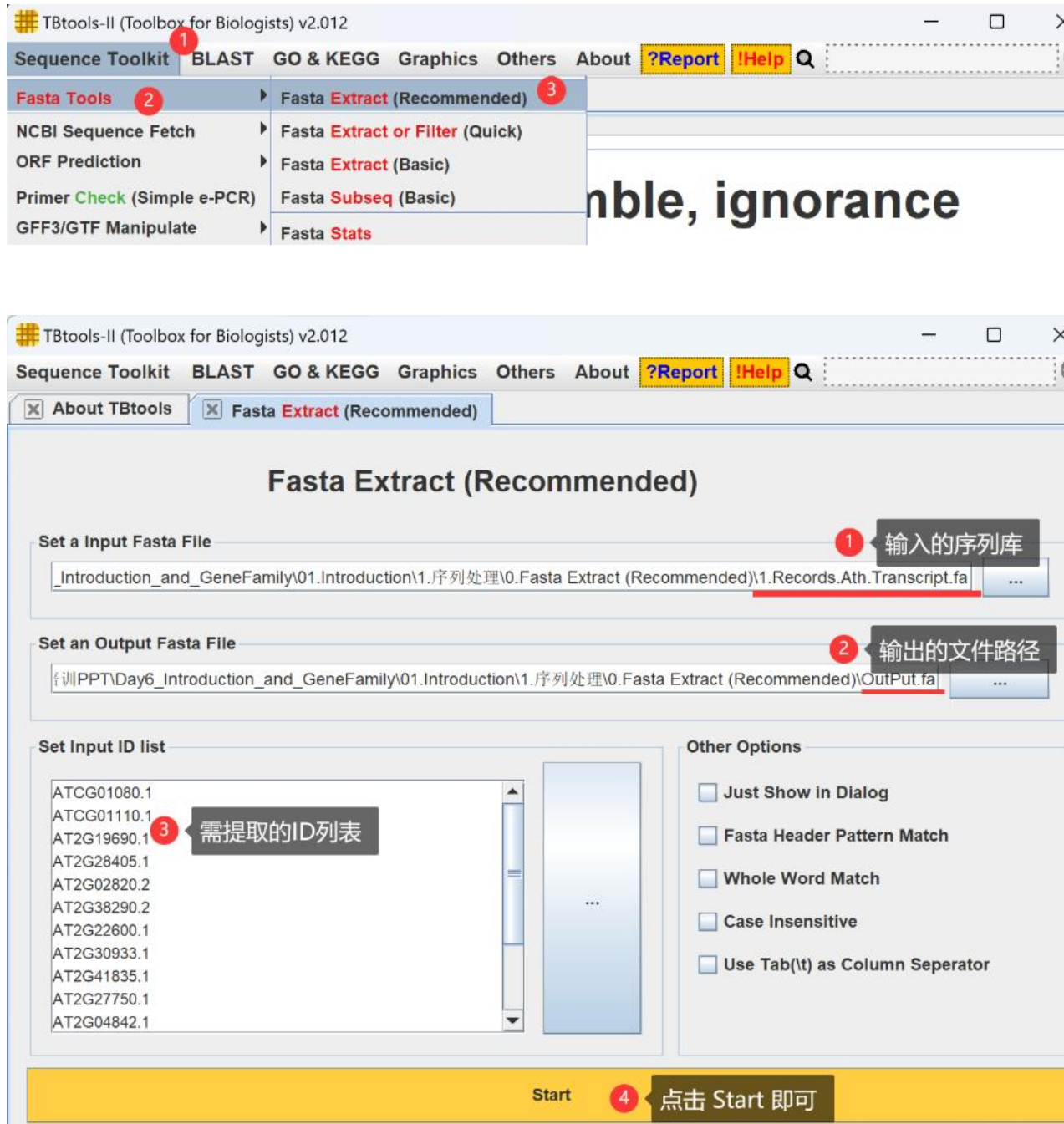
**Outfmt**: 输出的文件格式，可选择 **XML**、**Pairwise**、**Table**

注意: Build a Phylo Tree 为一键构建进化树,按钮下方可以指定使用 BLAST 结果的前多少个序列拿来建树

#### 实例 4 — 玉米 SBP 转录因子数据集中序列提取

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。

Fasta Extract (Recommended)的使用场景: 从一个比较大的序列文件中提取一些序列记录 (Fasta 格式)。比如, 我们有一个基因 ID 列表 (差异表达基因, 或者是其他方式), 要提取这些转录本/基因序列。



待提取序列信息设置框, 参考界面说明, 接受三种类型的提取模式:

(1) 基于 ID 的完整序列记录提取, 如输入 `Unigene_1 ...` 等完整序列 ID, 每行一个, 即可提取完整序列记录;

(2) 基于序列坐标信息, 进行序列区间截取, 如提取染色体 `Chr1` 上第 10000 个碱基到 20000 个碱基的一段序列, 那么输入如下。如果需要提取反向互补序列, 使起始坐标大于终止坐标即可。

```
# 注意, 制表符[\t]分隔, 而非空白[Space]分隔
Chr1      10000  20000
# 提取反向互补序列, 则翻转碱基坐标
Chr1      20000  10000
```

(3) 提取坐标信息的功能, 重命名区间, 如我们需要提取 `Peak` 或者 `Promoter` 序列信息, 并指定输出时序列名字 (如下, 提取出的序列被命名为 `peak_1`)

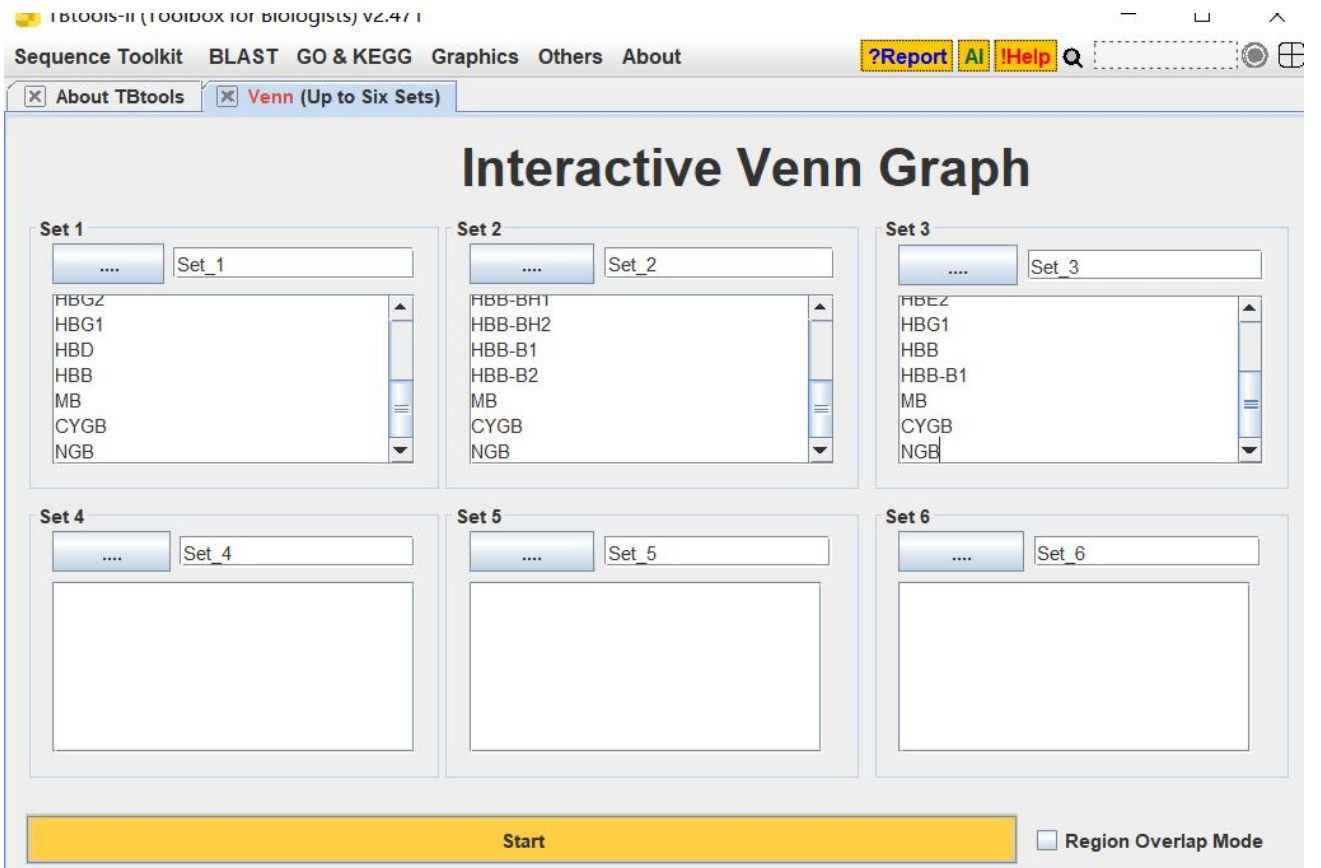
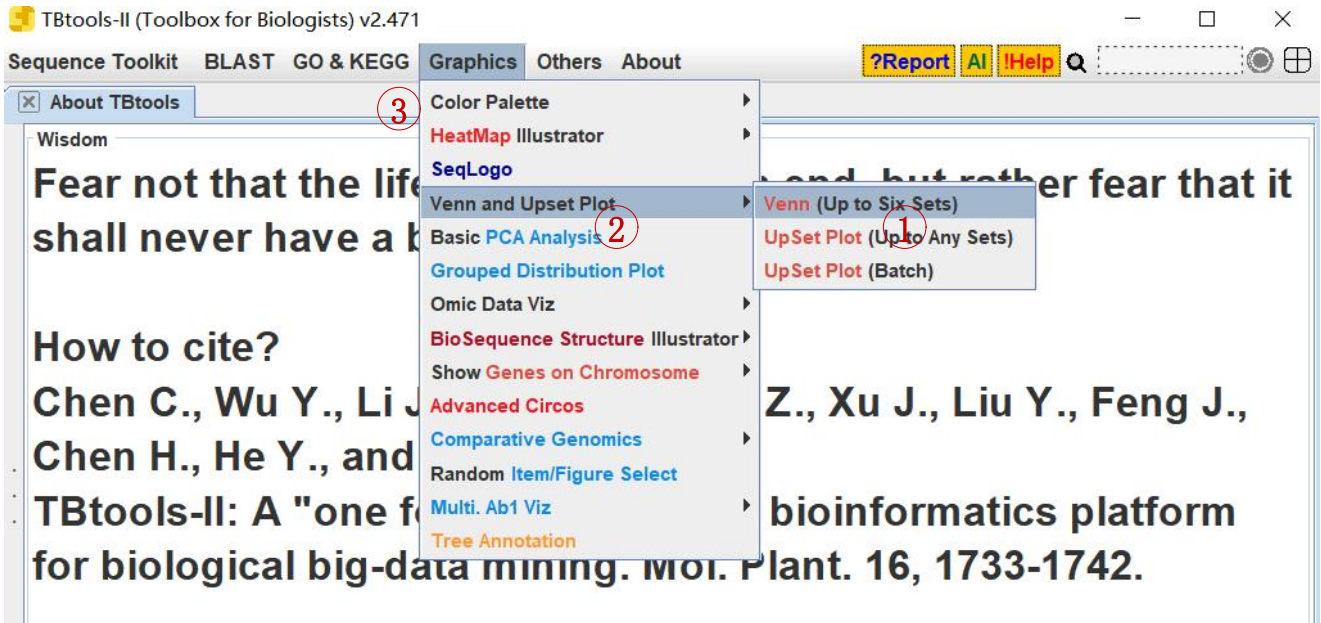
Shell 运行代码复制代码

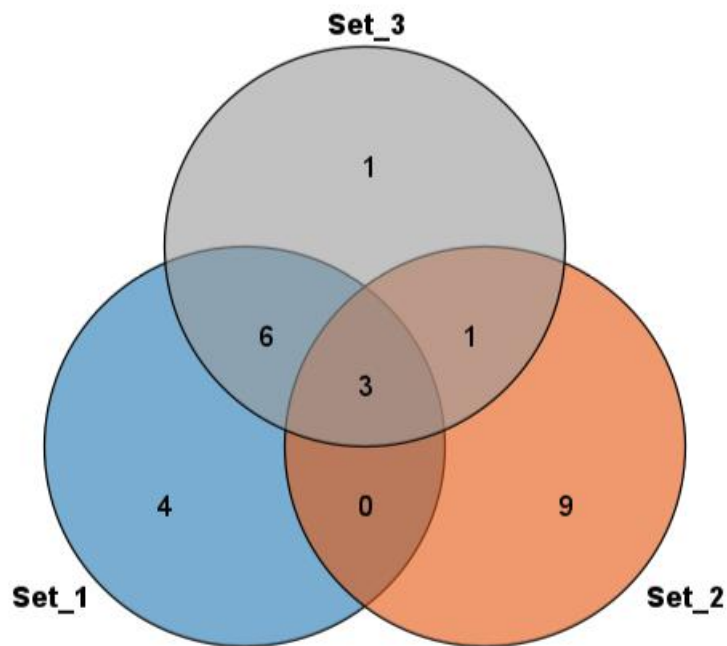
```
1
2
peak_1 Chr1 10000 10200
promoter_ATG8 Chr2 20300 22300
```

## 实例 5 — 韦恩图 Venn Diagram 分析

Venn 图是用图形表示数据集之间的逻辑关系, 包括并集、交集和补集, 由英国数学家 John Venn 于 1881 年提出; 常用于 2-6 组数据集之间共有或特有元素, 如基因名、蛋白质名

优点: 直观明了; 缺点: 超过 3 组数据集时图形很难解读。





示例：人、小鼠、大鼠珠蛋白基因家族包括血红蛋白 Hemoglobin (HBA HBB 等)、肌红蛋白 Myoglobin (MB)、胞红蛋白 Cytoglobin (CYGB)和脑红蛋白 Neuroglobin (NGB)，

利用 TBtools 韦恩图分析方法，找出三个物种中共有基因。

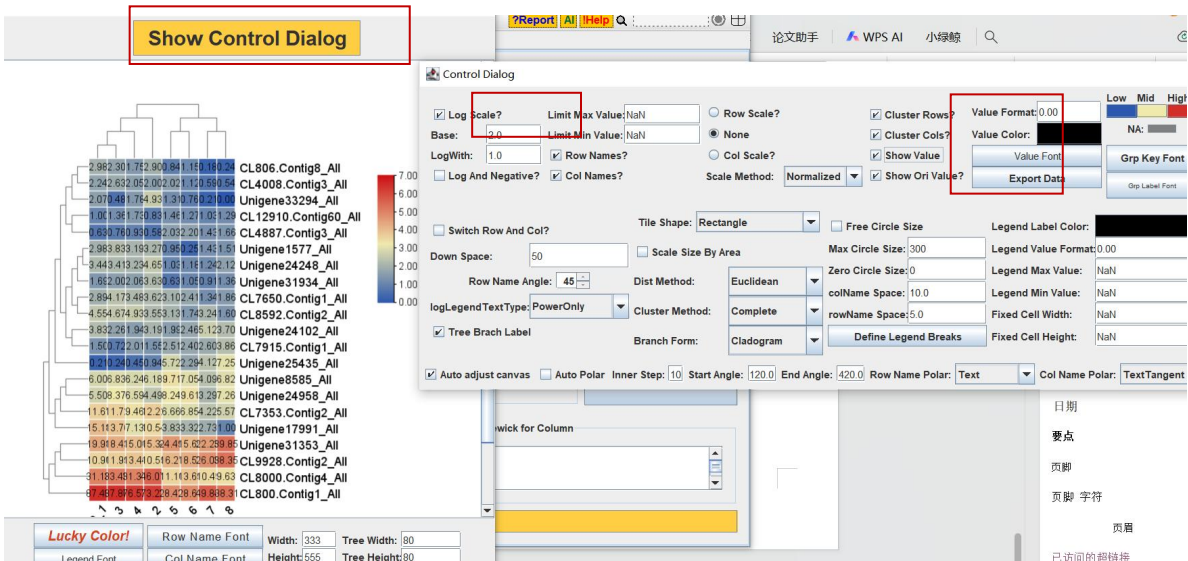
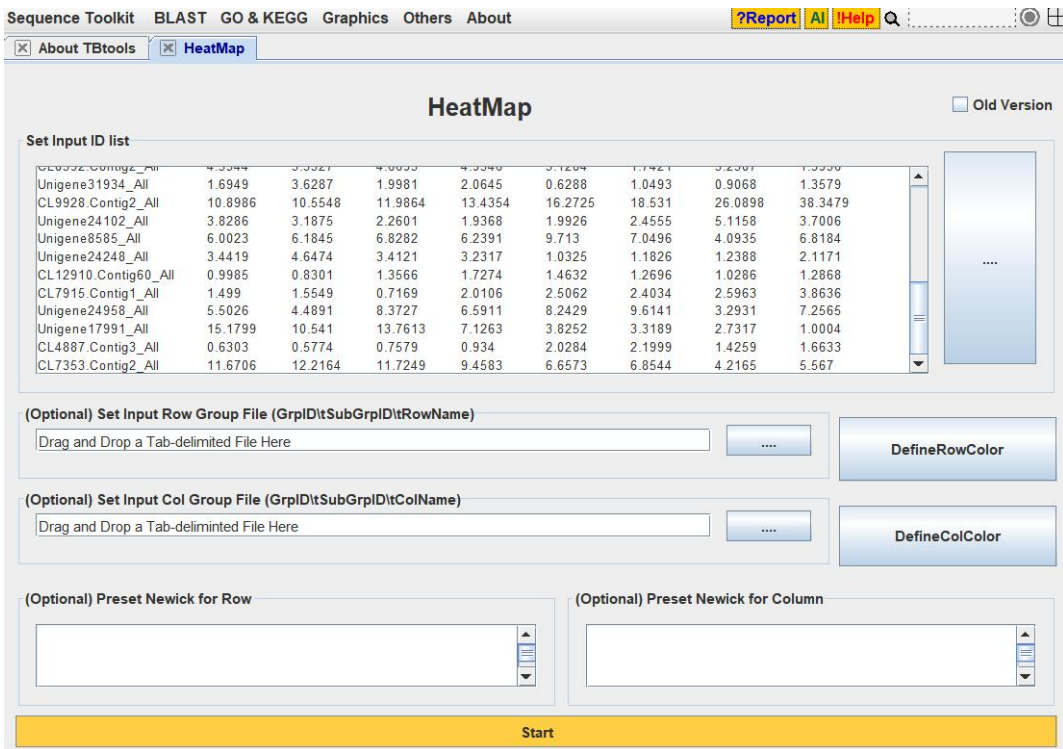
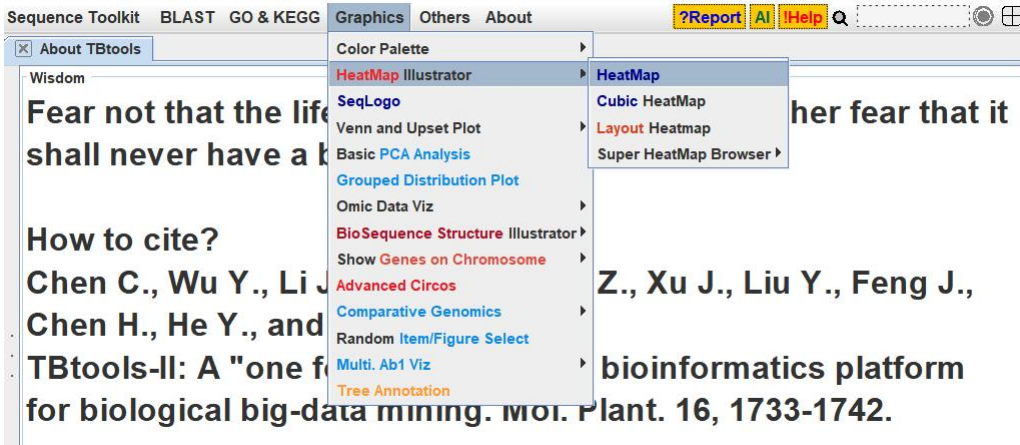
结果分析：

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

### 实例 6 — 热图 Heatmap 分析

图 Heatmap 中通常用不同颜色表示数据表中行和列数值大小，颜色越深数值越大，颜色越浅数值越小；常用于基因表达量高低

优点：直观明了；缺点：不够精确，数值差异过大时，不易分别



#### 四、TBtools 软件的疑问解答：

TBtools 使用问题可通过三大渠道高效解决：一是查阅语雀上的《TBtools Cookbook》及 BLAST 专题文档；二是观看哔哩哔哩上的“新手 TBtools ”等视频教程；三是遵循“两图一段话”原则向用户社群求助，并可借助软件的“? Report”按钮生成环境报告供他人快速诊断。

以 BLAST 序列比对为例：如果我们想将手头的序列与本地数据库比对查找同源基因，发现结果返回异常或为空。

解决办法：首先，进入软件的 BLAST Zone 功能，检查设置的本地数据库是否导入正确。若数据库缺失，可下载 Swissprot（约 86Mb）等蛋白序列文件，通过 BLAST Zone 快速构建本地库。其次，在待比对序列输入区添加目标 FASTA 文件，选择合适的 Blast Type（如查询蛋白质用 blastp，查询核酸用 blastn），设置 e-value 阈值（如默认 1e-5），点击 Start 开始运行。另外，需特别注意输入文件及文件夹名称须为英文或数字，因为早期 BLAST+版本对中文字符支持有限，可能导致任务卡死或报错。按照以上步骤，BLAST 比对问题通常能够得到解决。

4. 帮助文档 — 语雀网站  
5. 网络视频 — 哔哩哔哩





## 组员个人总结:

### A. G4A 高嘉禾个人总结:

课堂上系统学习了 TBtools 软件的核心功能与操作逻辑，学习了双序列和多序列的 BLAST 比对、数据提取、韦恩图与热图绘制等常用模块。在课堂上听了陈程杰老师和全宗军老师的学术汇报，让我对 TBtools 在今后学习工作中的应用和价值有了更深入的认识。不过目前我仍处于专业基础知识学习阶段，尚未在自身课题中实际使用该软件，但通过本次学习了解，我对 TBtools 在遗传育种研究中的潜在应用价值有了更加清晰的认识。

在陈程杰老师的汇报中，我了解到 TBtools 的核心特点是用户友好的操作界面、IOS（输入-输出-启动）的使用逻辑，以及覆盖日常生信分析需求的多样化功能。这些设计理念在后续课程的实操练习中得到了充分验证。

在序列处理方面，我掌握了 Fasta 序列提取与整理、BLAST 比对、获取候选基因序列、本地建库等操作流程。如果未来需要筛选与猪肉质、生长或抗病性状相关的候选基因，可以利用 TBtools 的“Fasta Extract”功能从参考基因组或转录组文件中快速提取目标序列，或借助 BLAST Zone 构建本地数据库进行同源比对分析。

在数据可视化方面，学习了韦恩图和热图的基本绘制方法以及通过控制面板调整参数美化图形。韦恩图有助于识别不同比较组之间共有或特有的基因集合，热图可以直观展示不同组织、不同品种或不同处理条件下基因表达的变化趋势。这些功能在育种数据分析中，可以对差异表达基因进行筛选、对比候选基因集的交叉等。

---

此外，课程中还讨论了 TBtools 使用中的常见问题及解决策略，如输入输出路径设置、文件命名规范、BLAST 结果为空时的排查方法等。结合陈程杰老师之前在汇报中提到的软件使用反馈机制（如“两图一段话”原则和“？”Report 按钮），我也学会了如何高效解决在使用中遇到的报错问题。

通过本次 TBtools 软件的系统学习，虽然我还没有将该软件投入实际课题使用，但已经掌握了软件的核心操作逻辑和常见应用场景，同时掌握了自主开展序列处理、同源比对和基本数据可视化的能力，为今后课题的数据分析做好技术储备。

#### **B. G4B 赵雪倩个人总结：**

在本次“实用生物信息技术”课程的小组讨论中，我们系统学习了 TBtools 软件的功能逻辑与实际操作，重点围绕序列处理、BLAST 比对、数据提取、Venn 图与热图绘制等常用模块展开练习。作为一名从事抗菌肽研究的研究生，我在本次学习过程中深刻体会到 TBtools 在湿实验科研工作中的实用价值，并对如何将其应用于抗菌肽的序列挖掘、功能预测与表达分析有了更清晰的思路。

抗菌肽是一类具有广谱抗菌活性的小分子多肽，广泛存在于动植物及微生物中。其研究通常涉及大量序列信息，如抗菌肽前体序列、成熟肽序列、同源比对、结构域识别以及表达量分析等。TBtools 所集成的多种功能模块，恰好能够高效支持这些分析需求。

首先，在序列操作方面，TBtools 提供的 Fasta 序列处理与提取功能非常实用。例如，在我们从转录组或基因组数据中筛选出候选抗菌肽基因后，常需要根据基因 ID 列表从大规模序列文件中快速提取对应序列。借助“Fasta Extract”功能，我可以轻松完成这一操作，极大提高了数据处理的效率。此外，在抗菌肽的多序列比对中，我也可以通过“BLAST”模块（如 Two Sequence Sets 或 Two Sequence Files）快速识别不同物种间抗菌肽的同源序列，进而推测其保守结构域与功能位点。

其次，在数据库构建与本地比对方面，TBtools 的 BLAST Zone 功能为我提供了便捷的本地数据库管理方式。例如，我可以将已知的抗菌肽序列构建为本地 BLAST 数据库，随后对实验室新测序获得的候选序列进行快速比对，判断其是否为已知抗菌肽或潜在新肽。这种方式不仅避免了频繁使用在线 BLAST 带来

---

的网络限制，也便于批量处理和大规模筛选。

在数据可视化方面，韦恩图与热图功能对我分析抗菌肽基因表达模式也具有重要价值。例如，在我所在的研究中，常需要比较不同处理条件下（如不同病原刺激、不同组织来源）抗菌肽基因的表达差异。通过热图，我可以直观地展示不同样品中多个抗菌肽基因的表达水平高低；而韦恩图则有助于识别不同处理组之间共有或特有的抗菌肽基因，帮助我聚焦关键候选分子。

此外，课程中我们还讨论了 TBtools 使用中的常见问题及解决策略，如输入输出路径设置、文件命名规范、BLAST 结果为空时的排查方法等。这些实践经验对我今后的科研工作尤其实用。尤其是在处理大批量序列时，遵循 IOS（输入-输出-启动）逻辑，养成规范的文件夹与文件命名习惯，能有效避免不必要的报错，提高工作效率。

总而言之，通过本次 TBtools 软件的学习与实践，我不仅掌握了一套高效、易用的生物信息分析工具，更深刻理解了“湿实验室+干分析”相结合的研究范式。未来，我将继续将 TBtools 融入到抗菌肽的基因挖掘、功能注释与表达可视化等环节中，提升研究的系统性与效率。

### **C. G4C 张俊鑫个人总结：**

经过本次 TBtools 的完整学习，我对这款国产一站式生物信息学分析平台有了更加全面且深入的认识。TBtools 功能覆盖面十分全面，从基础的序列处理、BLAST 数据库检索、同源基因快速查找、凝胶电泳模拟，到基因 GO 功能注释、KEGG 通路富集分析，再到韦恩图、热图、最大似然法进化树构建、全基因组数据可视化等高级分析与绘图，几乎可以满足大部分基础生信研究从原始数据处理到结果可视化的全套需求。在学习过程中，我借助课程配套的演示数据、课上的讲解案例，结合一系列针对性的实操练习，亲手熟悉了各项功能的使用逻辑，切实体会到 TBtools 无需复杂代码、可视化界面友好、操作便捷高效的突出优势，大大降低了生信分析的入门难度。同时，通过查阅相关参考文献，我也了解到这款软件的开发背景、迭代历程与学术影响力，明白它在大数据时代生物信息挖掘中的实用价值。整体学习下来，我不仅掌握了软件的具体操作方法，也梳理清楚了基础生物信息分析的整体思路，拓宽了自己在分子生物学研究工具上的视野。