

# CFVisual\_V1.1

## 用户手册

作者：葛伟娜，陈辉龙，尚千

单位：华北理工大学

邮箱：[chenhuilong131@163.com](mailto:chenhuilong131@163.com)，[Qian Shang16302@163.com](mailto:Qian Shang16302@163.com)

著作权证书号：软著登字第 6089858 号



# 目录

CFVisual_V1.1.....	1
目录.....	2
1. 软件简介.....	3
2. 编写目的.....	4
3. 系统操作.....	5
3.1 首页.....	5
3.2 操作.....	5
3.2.1 基因结构图绘制.....	6
3.2.2 Motif 图绘制.....	12
3.2.3 结构域图绘制.....	15
3.2.4 启动子图绘制.....	17
3.2.5 基因结构与 Motif 的组合图.....	19
3.2.6 基因结构和结构域的组合图.....	21
3.2.7 Motif 和结构域的组合图.....	23
3.2.8 基因结构、Motif 和结构域的组合图.....	24
3.2.9 其他组合.....	25
4. 后序.....	26

Visual

## 1. 软件简介

CFVisual\_V1.1 软件是一款借助 Python 语言的 matplotlib 库进行生物序列结构绘图的数据可视化软件，主要用于生物信息学分析中的基因结构分析，相当于 GSDS 的本地版 (<http://gsds.gao-lab.org/index.php>)，也能绘制进化树图，Motif 图，结构域图，启动子图，BED 文件等多种图表的绘制和组合绘制。



## 2. 编写目的

该用户手册是用户使用 CFVisual\_V1.1 软件的操作说明，用户通过阅读用户手册，能够直接对 CFVisual\_V1.1 软件的功能、操作有一定的了解，按照用户手册上的说明，通过实际操作，用户能够迅速掌握 CFVisual\_V1.1 软件的使用方法。



## 3. 系统操作

### 3.1 首页

CFVisual\_V1.1 可以免费从 GitHub 网址 (<https://github.com/ChenHuilong1223/CFVisual>) 下载。目前，网址可获得的版本仅仅是支持 window 系统，在获取该安装程序后，双击运行安装在用户的 window 系统的电脑相应的位置即可。或点击 CFVisual\_V1.1 软件依赖文件包中的 CFVisual.exe 即可进入到 CFVisual\_V1.1 软件的首页界面中如下图所示：

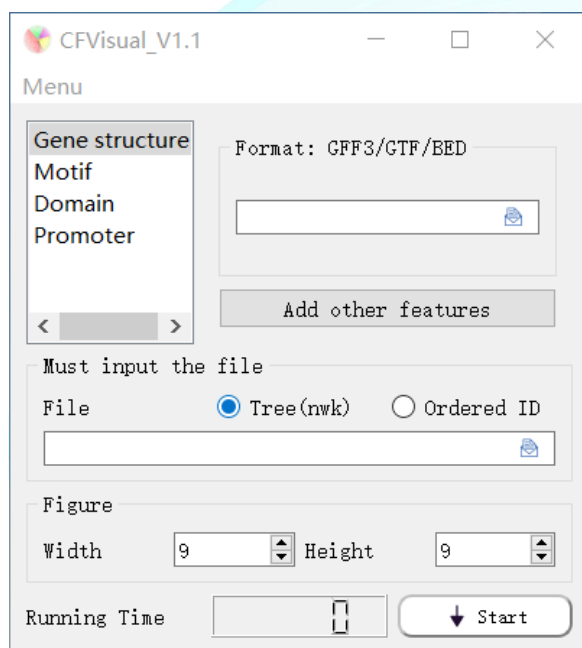


图 1 CFVisual\_V1.1 首页面

### 3.2 操作

使用须知：

1) 用户在首页按照软件设计规则可以进行文件的预选工作，我们默认您必须准备进化树文件（Si.PME.Tree.nwk）或者控制基因 ID 展示顺序文件（Si.PME.Ordered.ID.txt）。

2) CFVisual 中 Gene structure 和 Motif 功能中增加的信息 gff 格式文件不是标准的 GFF3 文件，而是 Gene, Start, End, Feature 的这样从左到右的四列——相当于 GFF3 文件中的这四列。（Si.PME.Structure.and.Motif.add.gff.txt）

### 3.2.1 基因结构图绘制

支持 GFF3 (Si.PME.Structure.GFF3.txt), GTF 和 BED (Si.PME.Structure.BED.txt) 文件格式, 输入树文件和基因结构文件, 点击 Start 按钮, 稍作等待即可出图 (如图 2, 图 3)。

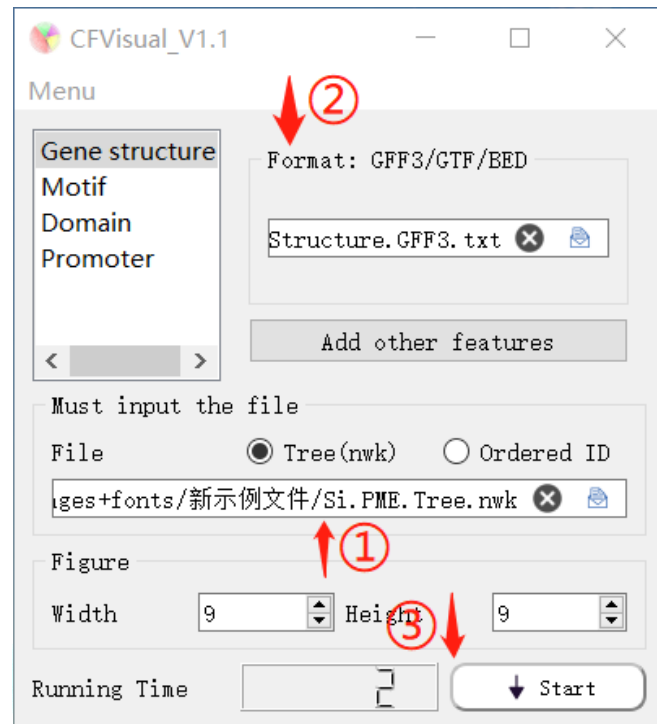


图2 输入基因结构图准备文件

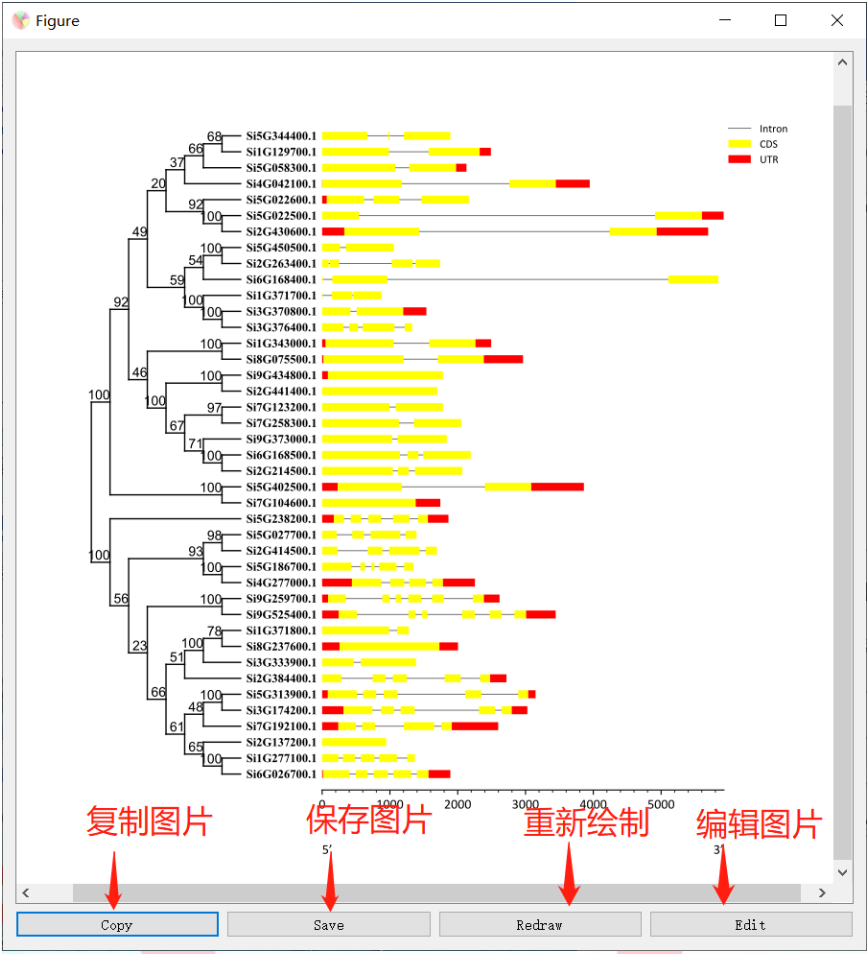


图 3 弹出显示基因结构图的窗口

点击 Edit 按钮，弹出编辑图片的窗口（图 4，图 5）

Visual



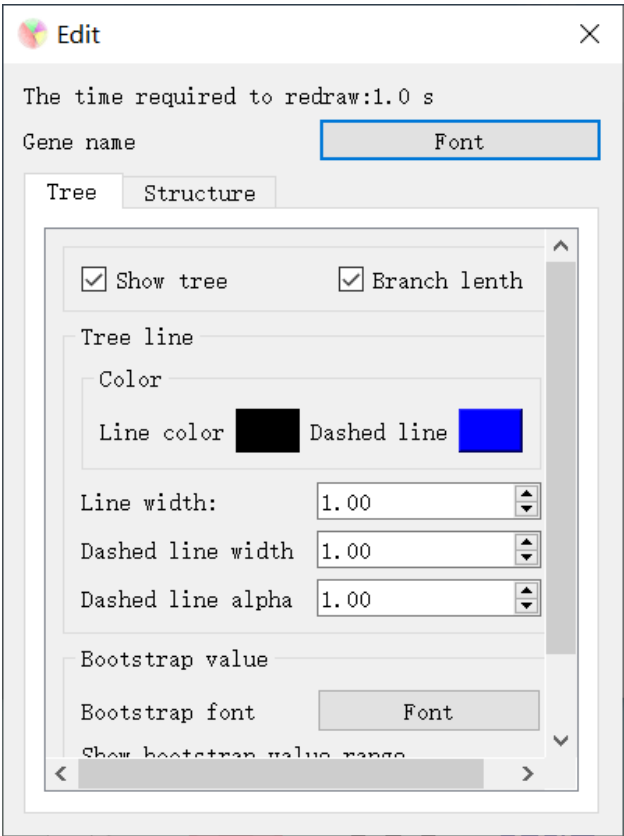


图 4 编辑树图

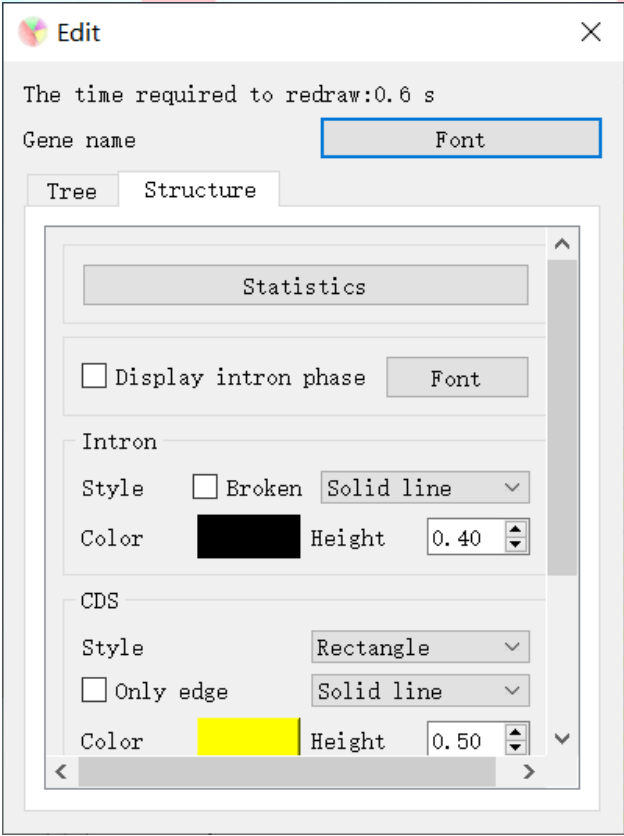
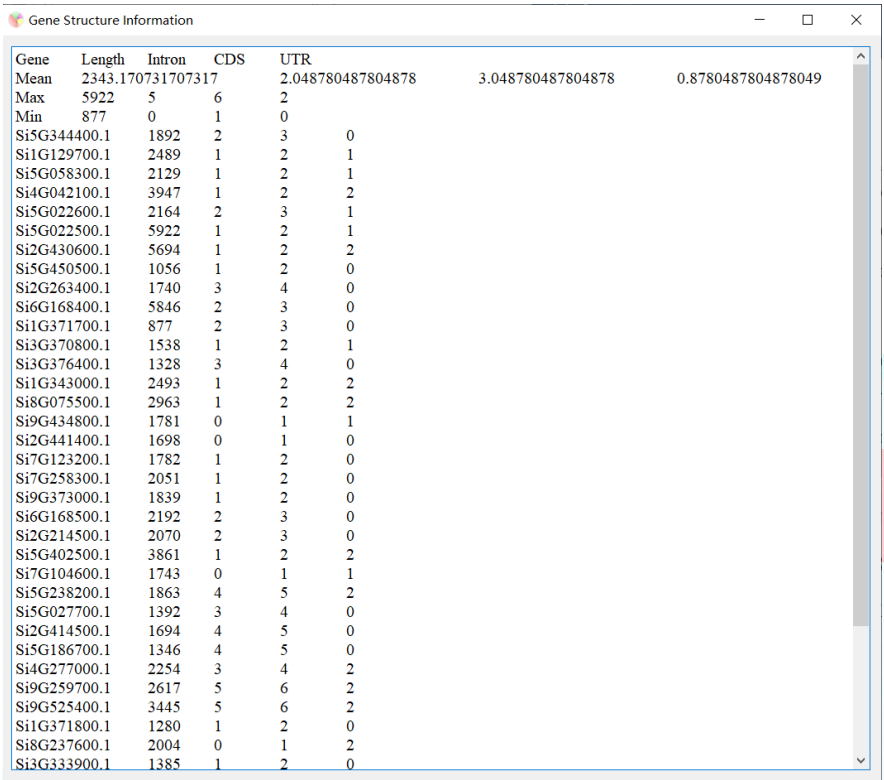


图 5 编辑基因结构图



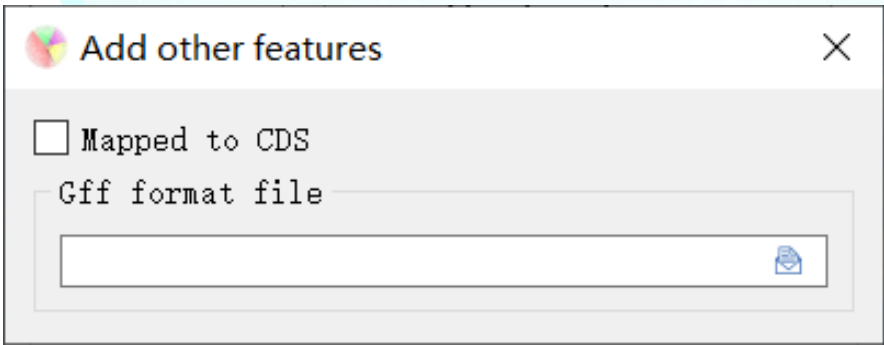
编辑基因结构图特色：点击 **Statistics** 按钮，弹出显示基因结构长度，内含子个数，UTR，CDS，exon 等数量信息（图 6）。（方便写论文时，对基因结构图的数字描述。）



Gene	Length	Intron	CDS	UTR
Mean	2343.170731707317		2.048780487804878	3.048780487804878
Max	5922	5	6	2
Min	877	0	1	0
Si5G344400.1	1892	2	3	0
Si1G129700.1	2489	1	2	1
Si5G058300.1	2129	1	2	1
Si4G042100.1	3947	1	2	2
Si5G022600.1	2164	2	3	1
Si5G022500.1	5922	1	2	1
Si2G430600.1	5694	1	2	2
Si5G450500.1	1056	1	2	0
Si2G263400.1	1740	3	4	0
Si6G168400.1	5846	2	3	0
Si1G371700.1	877	2	3	0
Si3G370800.1	1538	1	2	1
Si3G376400.1	1328	3	4	0
Si1G343000.1	2493	1	2	2
Si8G075500.1	2963	1	2	2
Si9G434800.1	1781	0	1	1
Si2G441400.1	1698	0	1	0
Si7G123200.1	1782	1	2	0
Si7G258300.1	2051	1	2	0
Si9G373000.1	1839	1	2	0
Si6G168500.1	2192	2	3	0
Si2G214500.1	2070	2	3	0
Si5G402500.1	3861	1	2	2
Si7G104600.1	1743	0	1	1
Si5G238200.1	1863	4	5	2
Si5G027700.1	1392	3	4	0
Si2G414500.1	1694	4	5	0
Si5G186700.1	1346	4	5	0
Si4G277000.1	2254	3	4	2
Si9G259700.1	2617	5	6	2
Si9G525400.1	3445	5	6	2
Si1G371800.1	1280	1	2	0
Si8G237600.1	2004	0	1	2
Si3G333900.1	1385	1	2	0

图 6 基因结构数值信息

增加其它信息：用户可以增加其它信息与基因结构图进行结合（图 7），增加信息文件格式为 gff 格式文件。如果用户勾选中 **Mapped to CDS**（图 8），那么位置信息将映射到 CDS 结构位置上（图 9）（方便了解哪段 CDS 编码该信息）。



Add other features

×

☐ Mapped to CDS

Gff format file

图 7

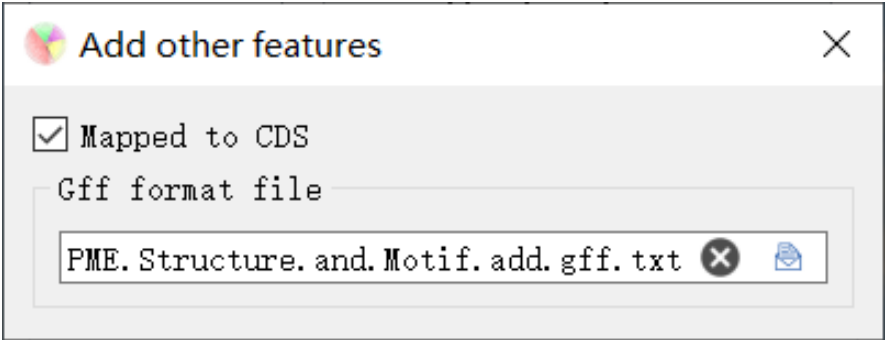


图 8

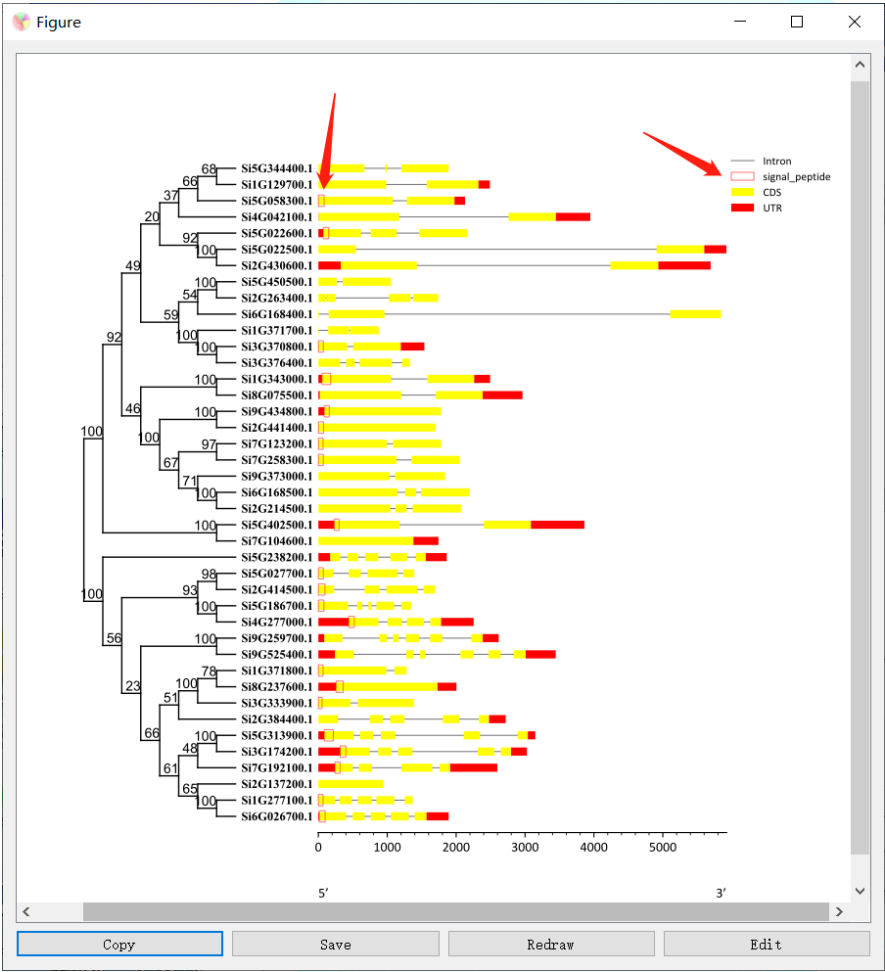


图 9

保存图片：目前支持 png（500dpi），pdf，svg，svgz，raw，rgba 这六种图片个格式（图 10）。注意：如果用户不选择路径，CFVisual 软件自动保存到用户输入的树文件或 ID 顺序文件所在文件夹中（图 11）（省去选文件夹的揪心操作）。

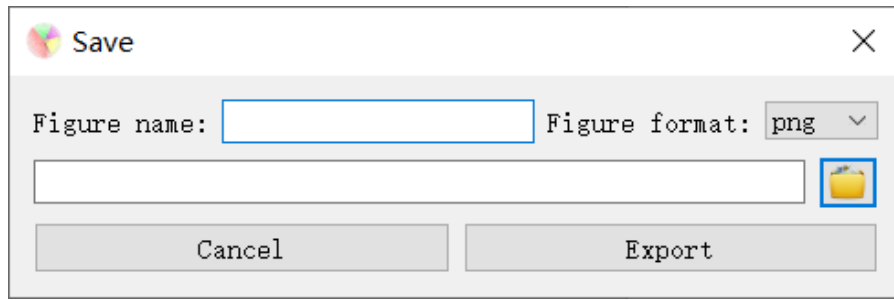


图 10 CFVisual 的图片输出窗口

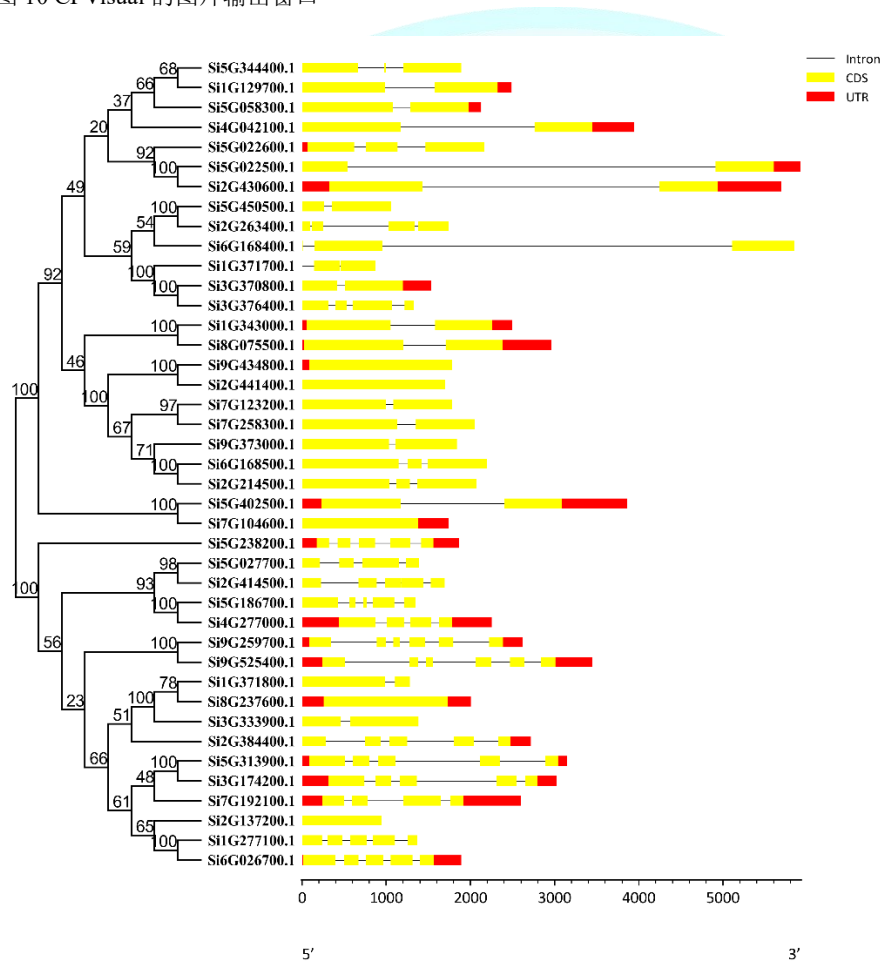


图 11 保存之后的 png 格式图片

### 3.2.2 Motif 图绘制

准备文件为 MEME 网站 (<https://meme-suite.org/meme/>) 或软件分析之后获取的 meme.xml (Si.PME.Motif.meme.xml) 或者 mast.xml (Si.PME.Motif.mast.xml) 文件。

注意: CFVisual 完全还原了 MEME 官网的 Motif 图结果。当用户输入的文件为 meme.xml 将有两种结果的展示, 分别为: Only Motif Sites (图 12) 和 Motif Sites+Scanned Sites (图 13)。可通过 Motif 编辑选项卡上的控件进行切换 (图 14)。另外, CFVisual 也支持用户选择性地隐藏用户不想要展示的 motif。

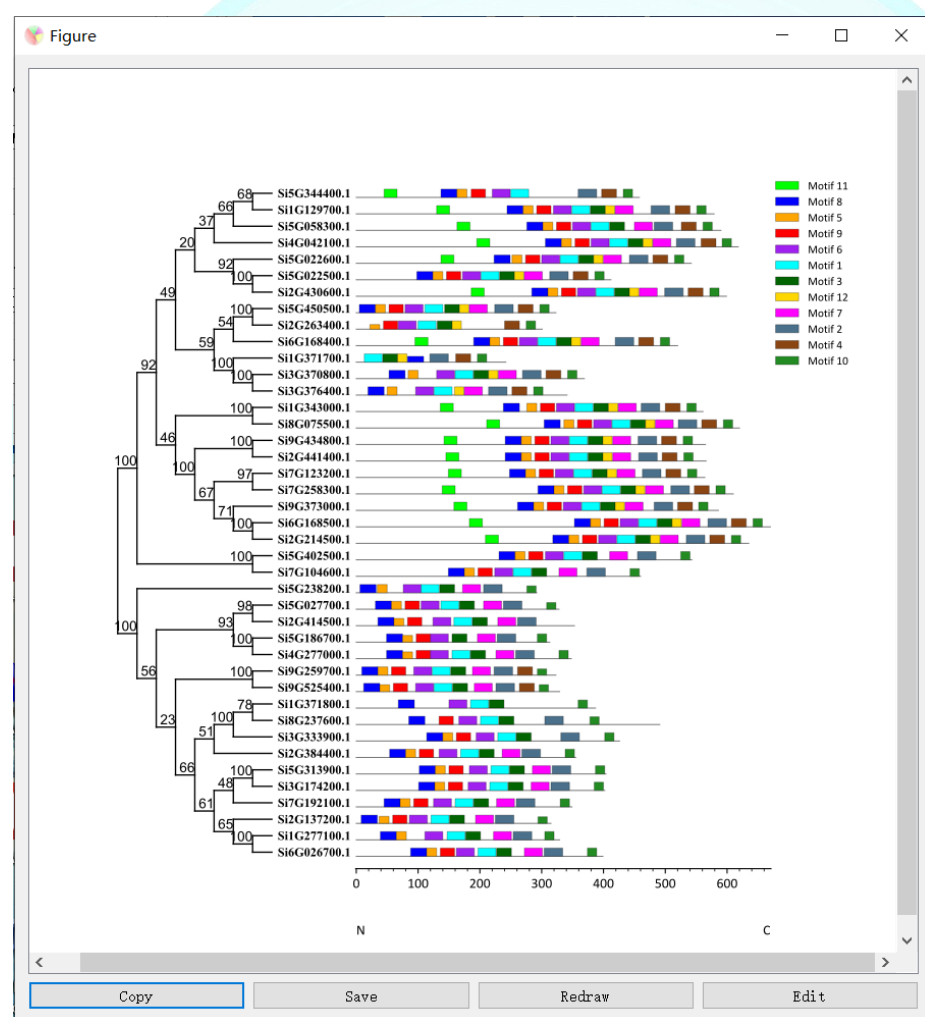


图 12 跟 MEME 官网一模一样的初始化结果 (Only Motif Sites)。

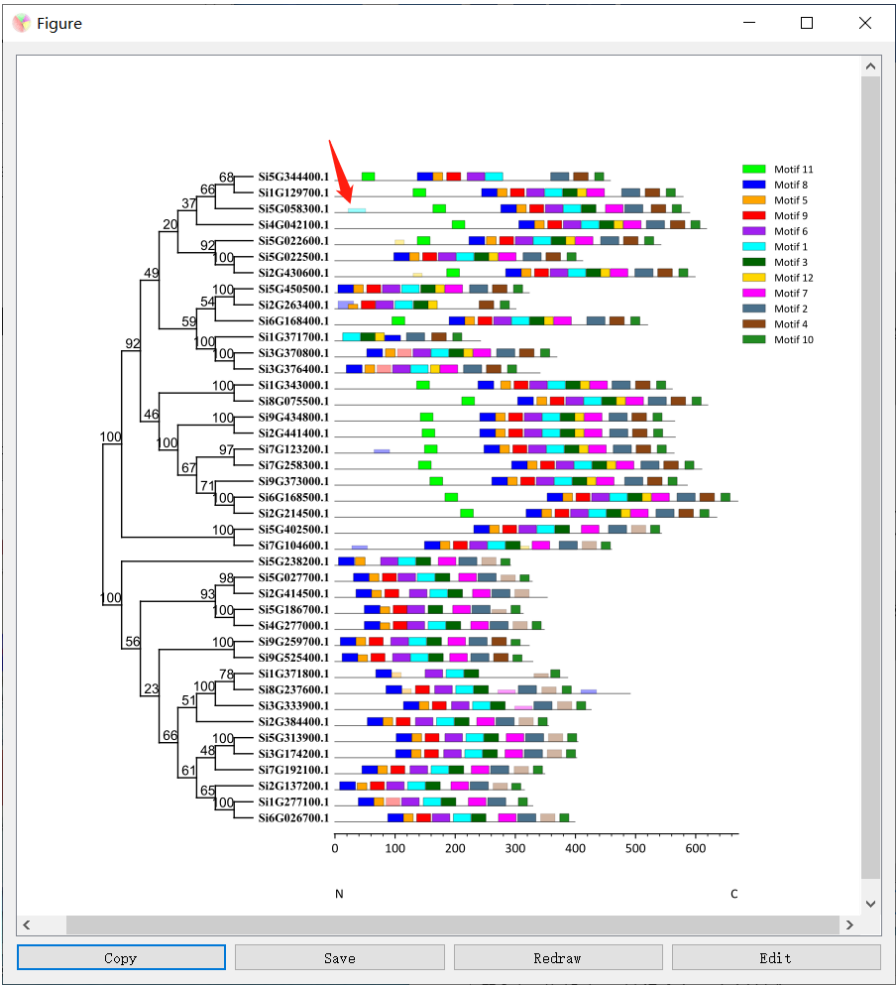


图 13 Motif Sites+Scanned Sites 的结果

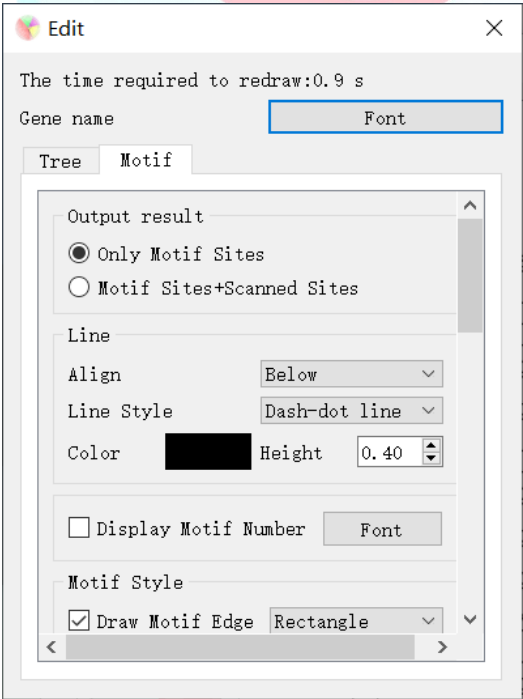


图 14

增加其它信息：用户可以增加其它信息与 Motif 图进行结合（图 15，图 16），增加信息文件类型为 gff 格式文件。

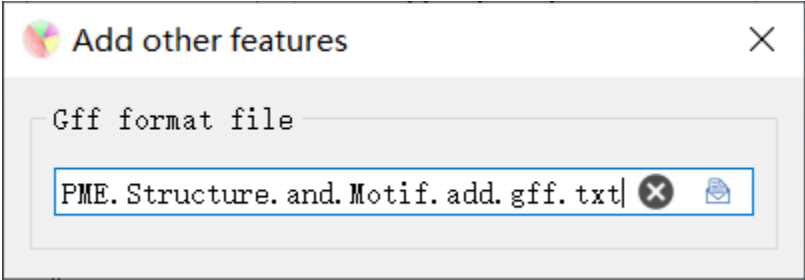


图 15

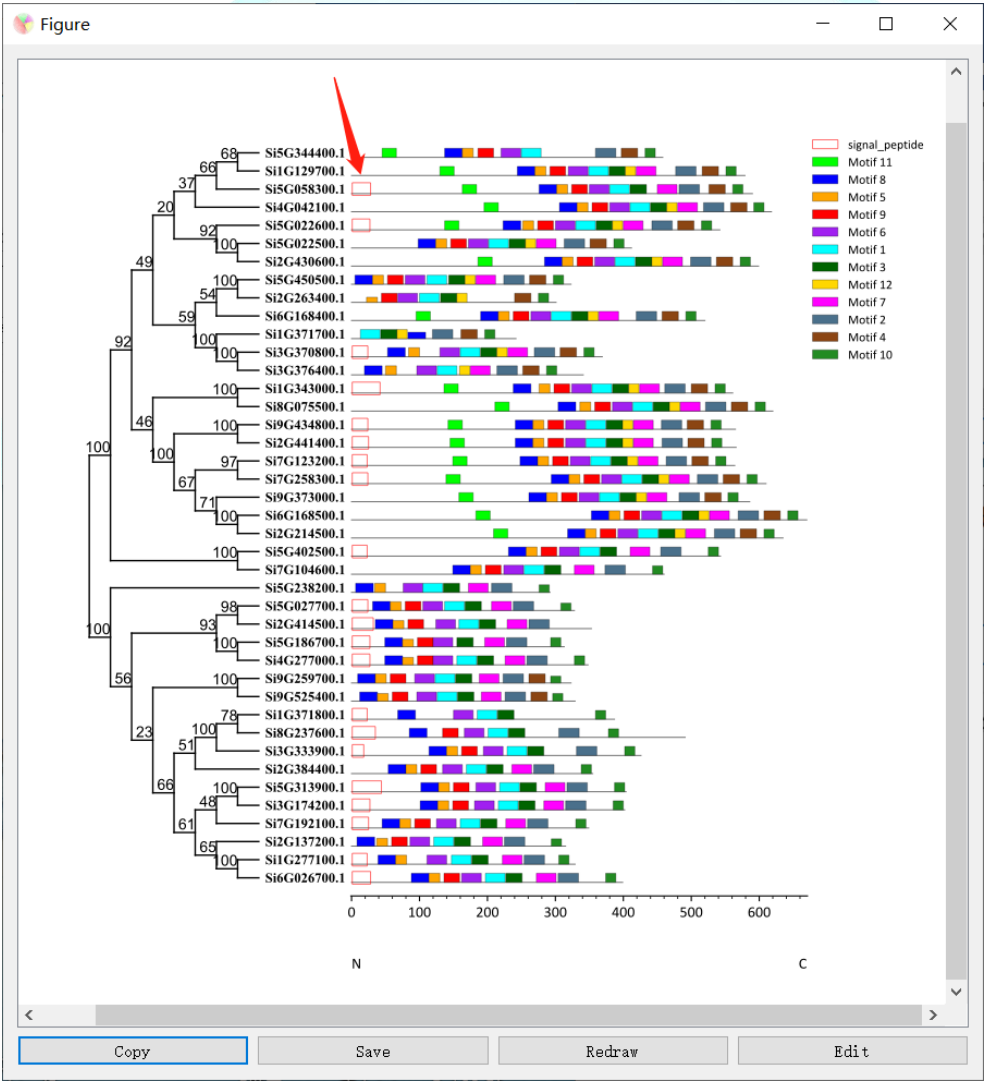


图 16

### 3.2.3 结构域图绘制

准备文件为 Pfam 网站 (<http://pfam.xfam.org/search#tabview=tab1>) 分析之后从邮箱获取的 Pfam 结果文件(Si.PME.Domain.Pfam.txt)或者 CDD 网站(<https://www.ncbi.nlm.nih.gov/cdd>) 分析之后下载的 CDD 结果文件 (Si.PME.Domain.CDD.txt) (图 17)。注意：在使用 CDD 结果文件的时候,建议用户先把 Short name 这一列的同一类结构域名字手动改成统一的名字。

结果 (图 18) 和用户交互功能 (图 19) 同上。

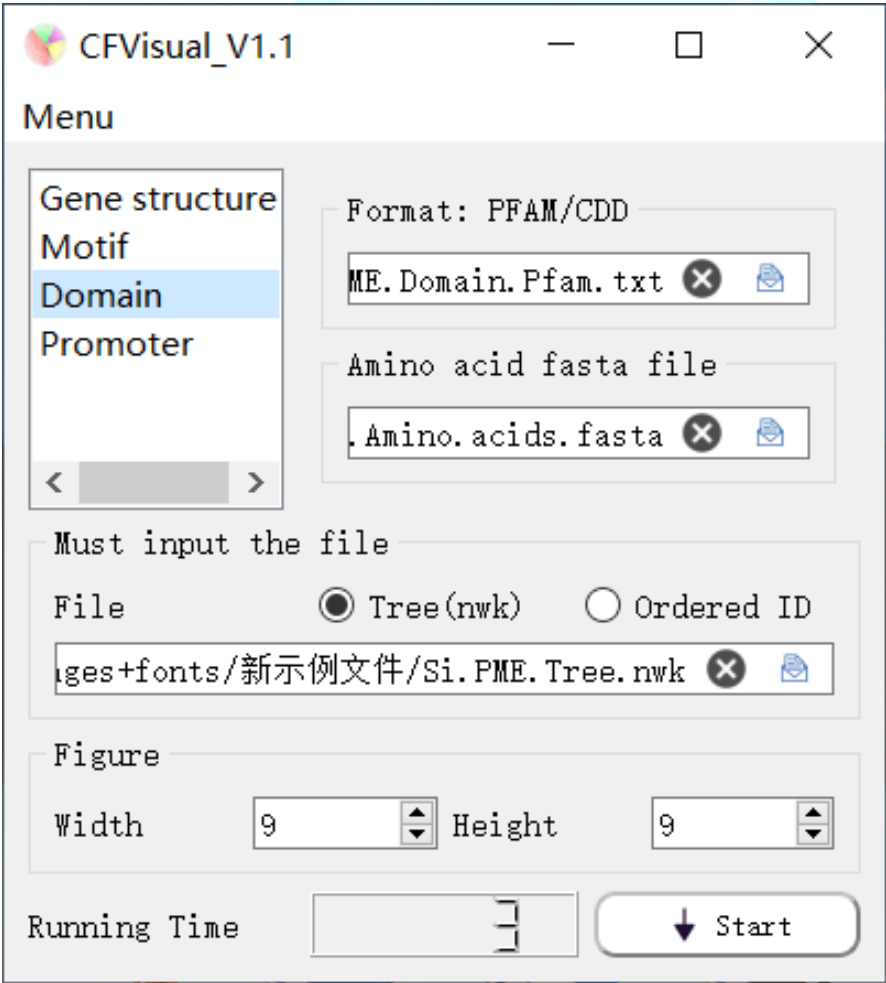


图 17



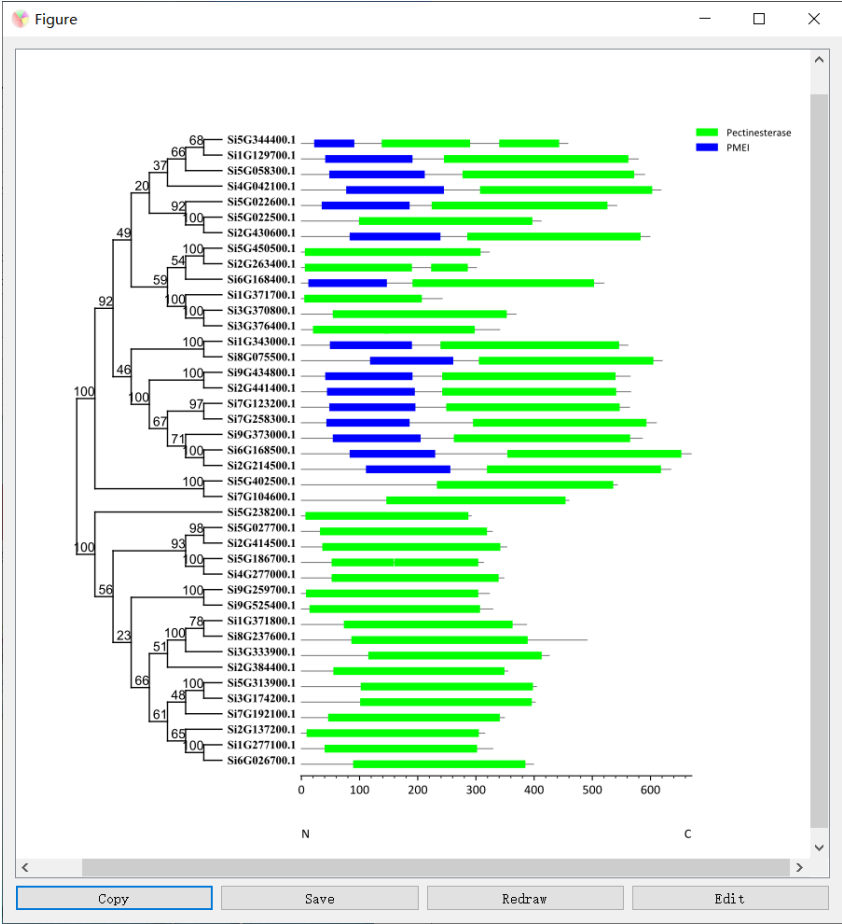


图 18

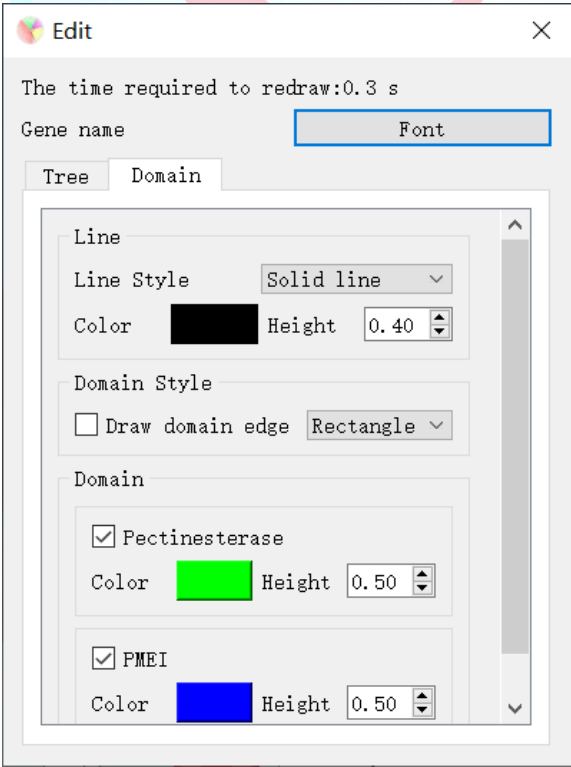


图 19

### 3.2.4 启动子图绘制

准备文件为 PlantCare 网站 (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 分析之后获取的结果文件中的四列 (Si.PME.PlantCare.result.txt) (图 20) 或者其他启动子预测网站下载的结果, 只要结果格式能够处理成 Si.PME.PlantCare.result.txt 文件类型, 都能绘制 (作者没有用过其他启动子预测网站, 欢迎补充。)。注意: 使用 PlantCare 网站分析结果中, 要根据自己的课题需求稍作处理, 且只保留+链上的顺式作用预测结果。灵活地使用 CFVisual 的启动子绘图功能中的用户选择性地隐藏用户不想要展示的顺式作用元件控件, 可以达到用户筛选顺式作用元件的效果——换句话说, 用户获得 PlantCare 的结果之后, 只需要删掉多余的列和-链所在行的结果, 即可直接用 CFVisual 绘制, 不再需要花费繁琐的功夫去筛选顺式作用元件类别了。

结果 (图 21) 和用户交互功能 (图 22) 同上。

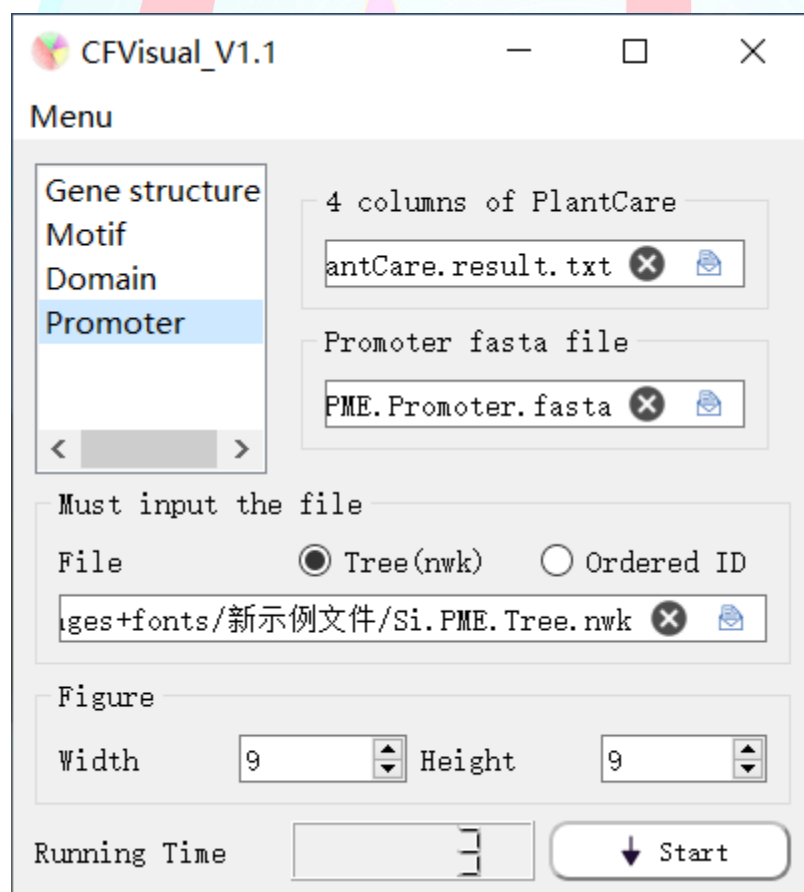


图 20

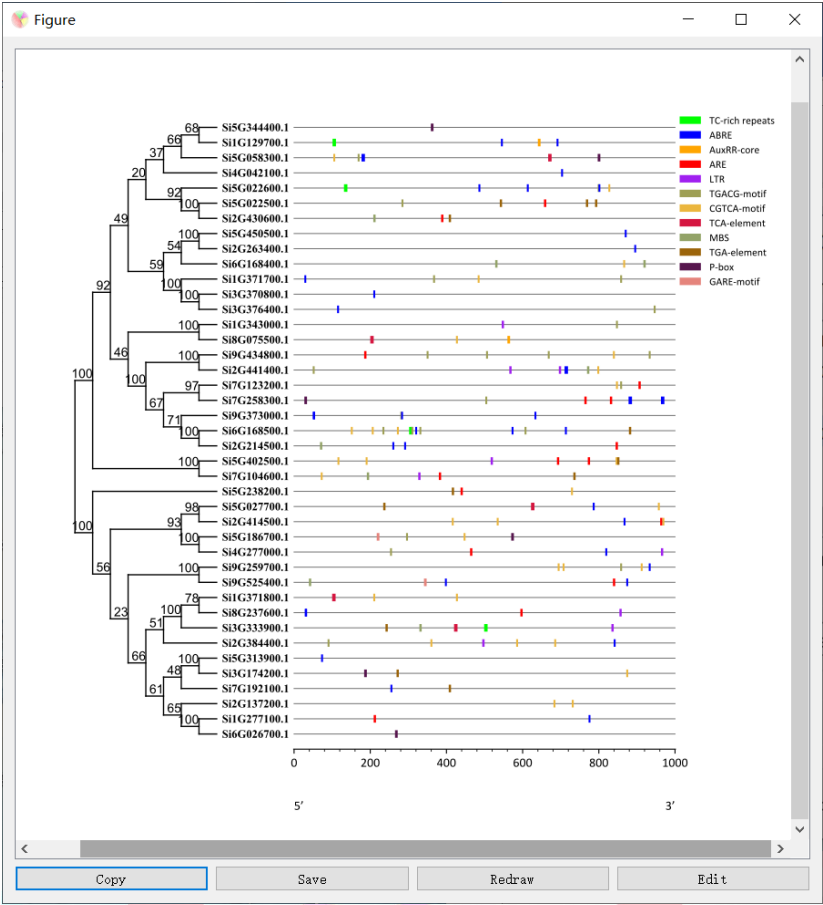


图 21

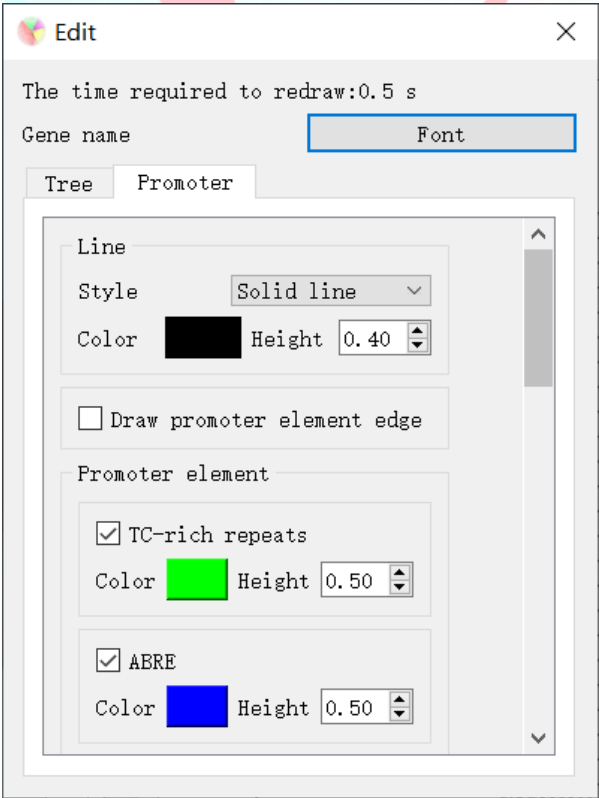


图 22

3.2.5 基因结构与 Motif 的组合图

作者读研三年，做的最多的生信分析就是基因家族分析了，深知做基因家族分析少不了这两个分析，而大多数人最喜欢的还是将树图、基因结构和 Motif 图进行组合成一张图展示，我们一般都是用 Ai 或 Ps 手动进行拼接，作者深知修图的繁琐和痛苦，因此就自己编程实现了这组合图的展示，一是方便自己，也方便师兄妹们。

准备文件：同基因结构图和 Motif 图绘制准备文件（图 23，图 24）。编辑方法同基因结构图和 Motif 图绘制（图 25，图 26）。作者考虑到这种情况，有的人喜欢把基因结构图放在中间，因此，特意增加该功能（Motif 编辑选项卡中的 Exchange coordinates）（图 25）。

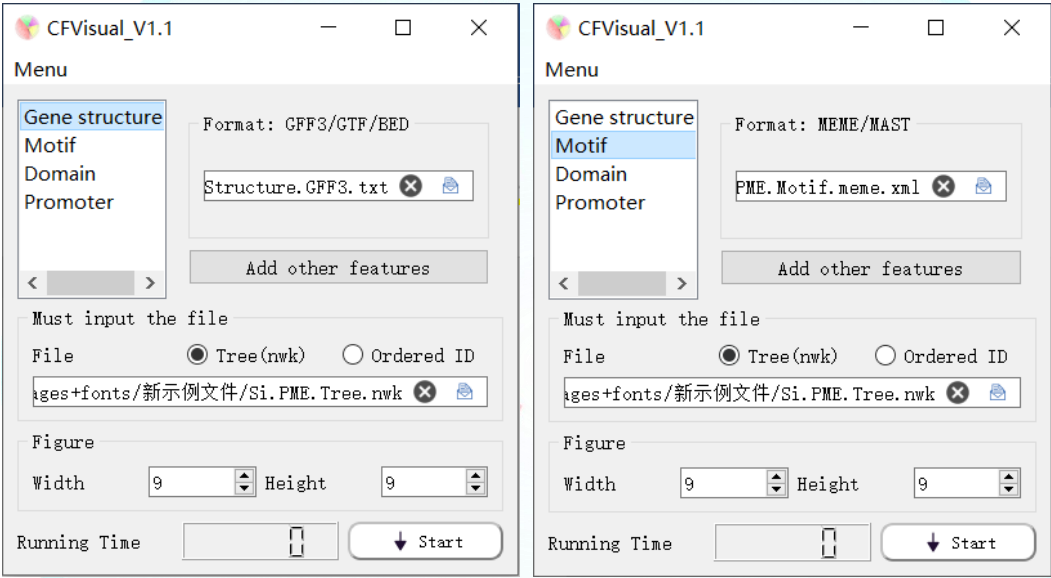


图 23 和图 24

Visual

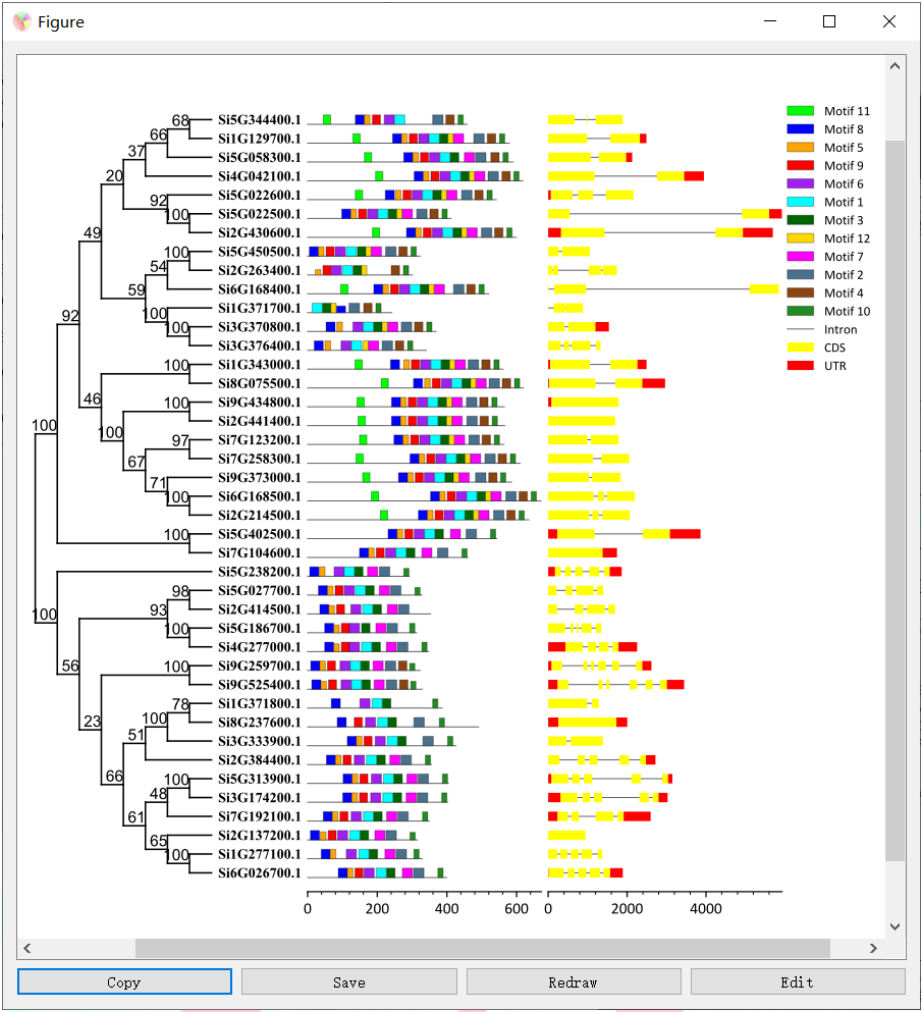


图 25

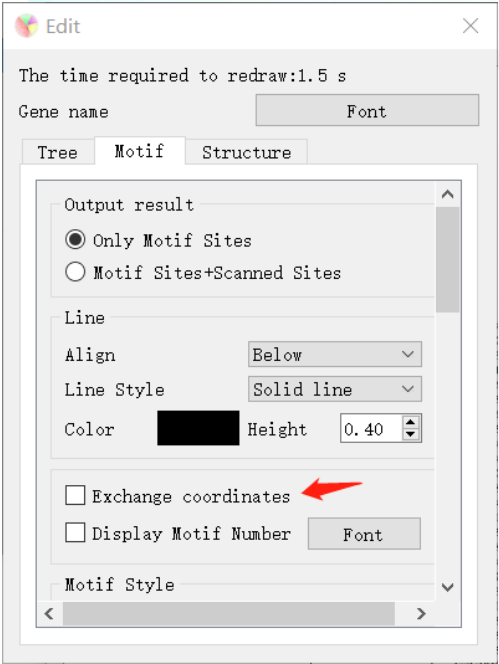


图 26

### 3.2.6 基因结构和结构域的组合图

这种组合图的作用在于方便科研人员了解是哪些 cds 编码结构域。准备文件同基因结构图和结构域图绘制（图 27），但是要注意一下，CFVisual 绘制组合图的时候，准备结构域文件时，不再需要增加氨基酸 fasta 文件（图 27）。基因结构结合结构域的功能 GSDS 其实已经有了，但是得自己准备氨基酸位置文件，不够方便，作者这里直接读取并操作 Pfam 或 CDD 结果文件获取位置信息，最需要注意的一点是，CFVisual 的结构域表现形式与 GSDS 不太一样，但本质一样，区别点在于，CFVisual 的结构域用的方框表示，把区间的内含子也框中了（图 28），而 GSDS 是实心区块并不给内含子上色。作者这样做的理由是试图打开一个新的基因结构分析思路，由于导师比较关注内含子，因此，作者认为，这样展示，在不失去原本用途的前提下，能比较方便地查看编码这段结构域的 CDS 中内含子的插入情况或者说丢失情况。

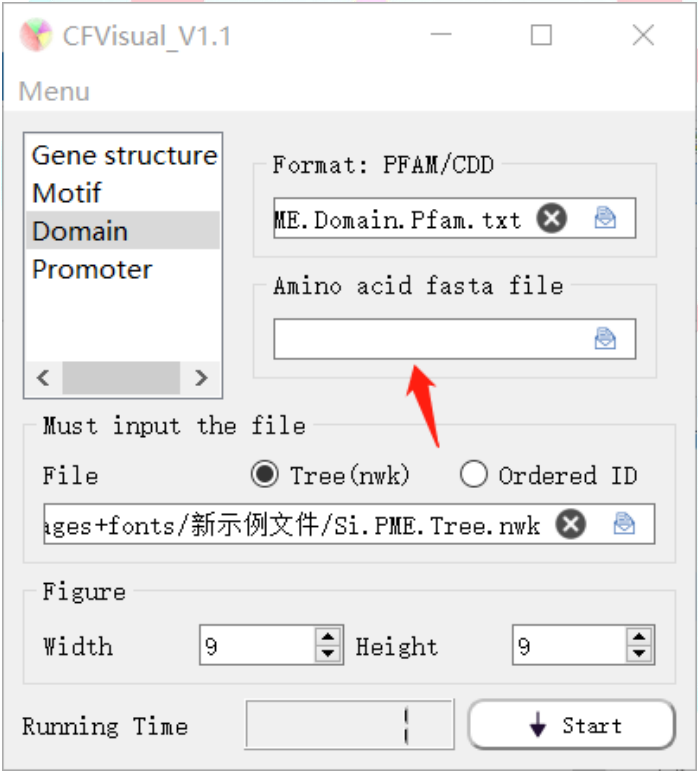


图 27

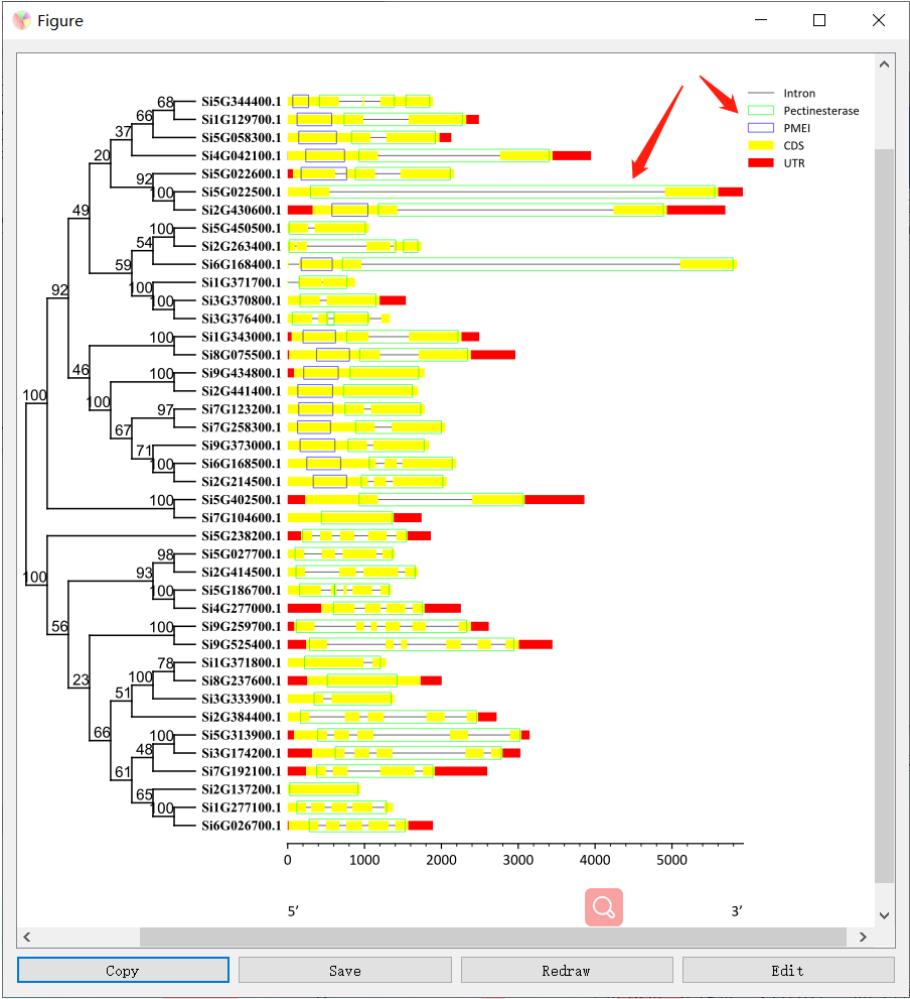


图 28

Visual



### 3.2.7 Motif 和结构域的组合图

这种组合源于一个师妹做抗病基因 NBS-LRR 家族，审稿人要求她把 motif 与结构域组合分析，看看结构域由哪些 motif 包含或者说结构域与 motif 的位置分布关系。由于目前没有找到拥有此功能的软件，她只能肉眼根据结构域位置标记，这样做，既麻烦，位置锚定还不一定准确。好在审稿人只要求她标记几个位置作为示意图即可，不然抗病基因那么大的家族，得标记到何年何月。基于此，作者实现了这种组合需求（图 29）。

准备文件同 Motif 图和结构域图绘制（图 27），同样要注意一下，CFVisual 绘制组合图的时候，准备结构域文件时，不再需要增加氨基酸 fasta 文件（图 27）。

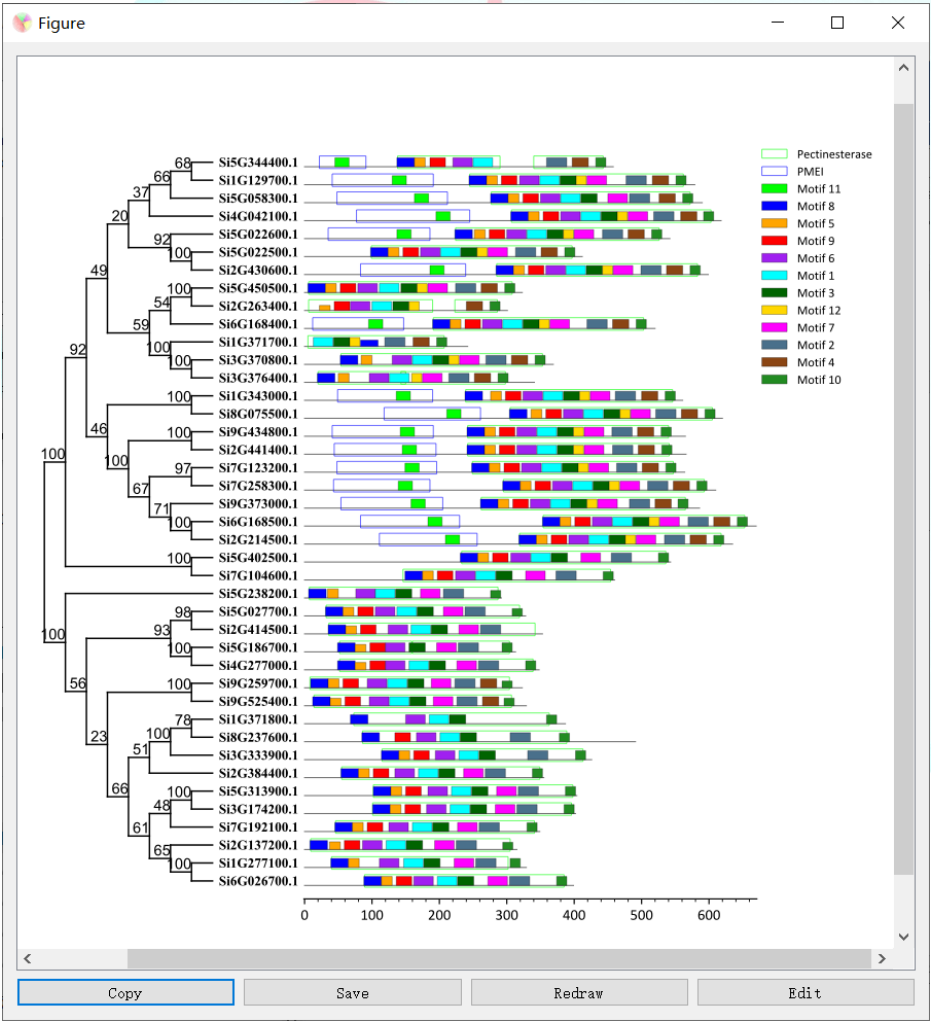


图 29

3.2.8 基因结构、Motif 和结构域的组合图

准备文件同基因结构图绘制，Motif 图绘制和结构域图绘制，同样的，绘制组合图时，结构域图不需要再准备氨基酸 fasta 文件。绘制结果如图（图 30），组合图的解读同上。

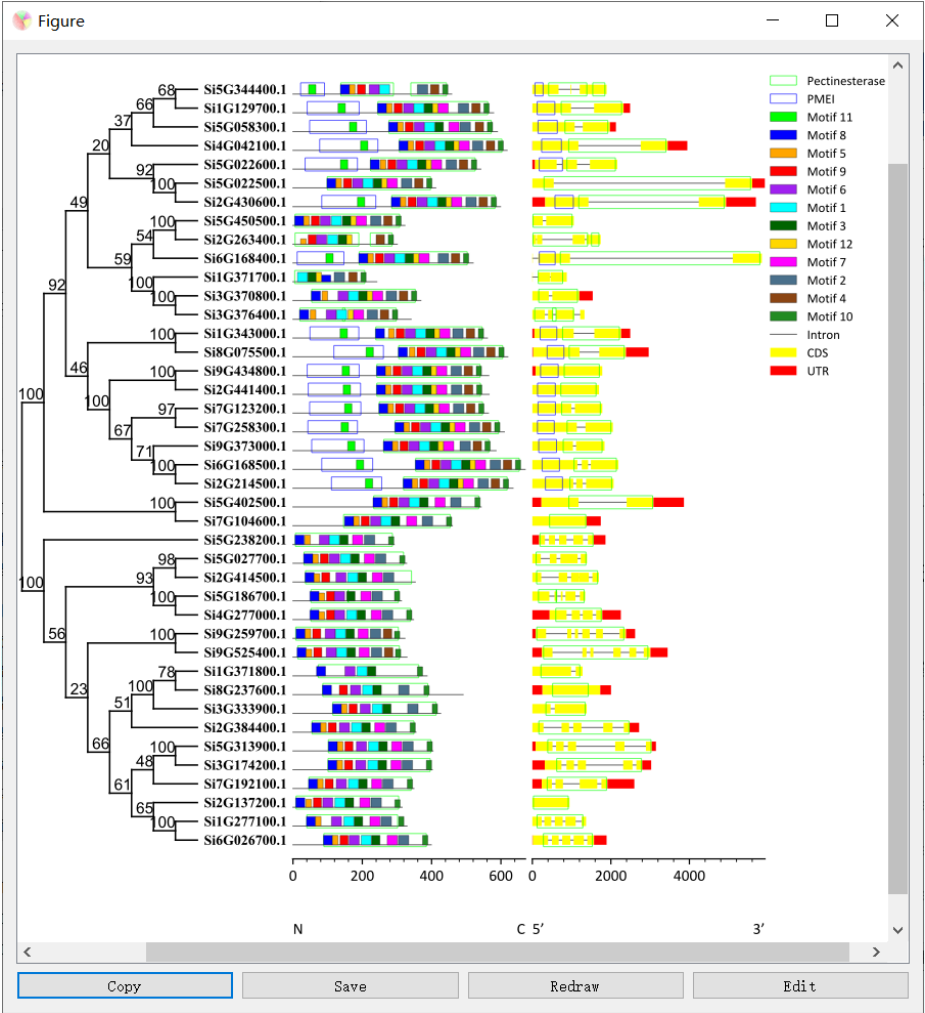


图 30

### 3.2.9 其他组合

作者想过其他组合图，比如启动子图与基因结构图的组合等，但是考虑到这种组合不常见，因此，暂不支持这种组合。



## 4. 后序

开发 CFVisual 心路历程，作者以前从来没想过写软件，因为基因结构图和 Motif 图分别可以通过 GSDS 和 MEME 绘制并通过 Ai 或 Ps 拼接修图，虽然过程繁琐，但是效果满意。是研二之初，要做启动子分析，发现市面上并没有一款软件可以绘制启动子图，研一时期的作者，编程能力几乎为 0，因此，无奈只能寻找已有的工具，作者小恋一下，通过 EXCEL 成功实现了启动子图的绘制，作者很高兴，但是绘制过程太麻烦，太繁琐，耗费自己太多时间，又不容易传授给师弟师妹。因此下定决心，自己去写一个绘制启动子图的程序。当写出来了第一个启动子图，慢慢想着自己能不能去实现其他做基因家族的图，经过努力和做家族的经验，绘图原理都已破解，需要的就是时间，于是乎，不到一年，作者成功将基因家族常见的图都自己用 Python 编程实现，比如 Circos，染色体定位图，聚类热图，共线性图等等，所以作者现在做基因家族已经不再需要使用其他的绘图的软件了，本来想着自己把这些都封装到一个软件里，但现实不允许，任务繁多，无人支持，也没有那个精力和时间去做这个。因此，这些程序一直自己使用，后来听老师的意见，利用空余时间，把自己最需求的启动子图和组合图的功能封装到软件里，方便自己所在课题组的师弟师妹使用。对于开发软件，得感谢老师和朋友的支持，帮我弄到第一个软件著作权，特此感谢。这里需要强调一下，CFVisual 的树图绘制是由作者的小师弟尚千创造编写，正因有了小师弟的加入（实际上，CFVisual 由作者跟尚千两人共同编写，葛伟娜老师支持并指导），作者才能在繁忙的科研生活中抽出时间来开发 CFVisual。作者从小白一步步利用空余时间成长，最终整出一个稍微能看的小软件 CFVisual，一直是自己和身边做家族的朋友使用，今，勇敢推出，不惧批评，软件虽小，可窥心血，还望珍惜！后，感谢使用，提出意见。

陈辉龙

2021 年 4 月 1 日