

CFVisual

User Manual

Author: Huilong Chen

E-mail: chenhuilong131@163.com



Content

CFVisual	1
Content.....	2
1. Software Description.....	3
2. Writing Purpose	4
3. System Operation	5
3.1 Front Page	5
3.2 Operation	6
3.2.1 Drawing of Gene Structure.....	7
3.2.2 Drawing of Motif	13
3.2.3 Drawing of Domain.....	16
3.2.4 Drawing of <i>cis</i> -acting Elements in Promoter Region	18
3.2.5 Combination Diagram of Gene Structure and Motif	20
3.2.6 Combination Diagram of Gene Structure and Domain	22
3.2.7 Combination Diagram of Motif and Domain.....	24
3.2.8 Combination Diagram of Gene Structure, Motif, and Domain.....	26
3.2.9 Other Combinations	27
3.3.0 Interaction, Classification, and Beautification of Tree Diagram	28

Visual

1. Software Description

CFVisual software is a data visualization software for biological sequence structure drawing with the help of Python language matplotlib library and PySide2 library. It is mainly used for gene structure analysis in bioinformatics analysis, which is equivalent to the local version of GSDS (<http://gsds.gao-lab.org/index.php>), can also draw phylogenetic tree diagrams, motif diagrams, structural domain diagrams, promoter diagrams, BED files and other diagrams and combinations of diagrams.



2. Writing Purpose

The user manual is the operation instructions for users to use CFVisual software. Users can have a certain understanding of the functions and operations of CFVisual software directly by reading the user manual. According to the instructions in the user manual and through practical operation, users can quickly master the use of CFVisual software.



3. System Operation

3.1 Front Page

CFVisual can be downloaded for free from the GitHub URL (<https://github.com/ChenHuilong1223/CFVisual>). After obtaining the installer, double-clicking to run it and install it in the corresponding location on the user's computer. Or clicking CFVisual.exe in the CFVisual software dependency file package to enter the home page of the CFVisual software. As shown below:

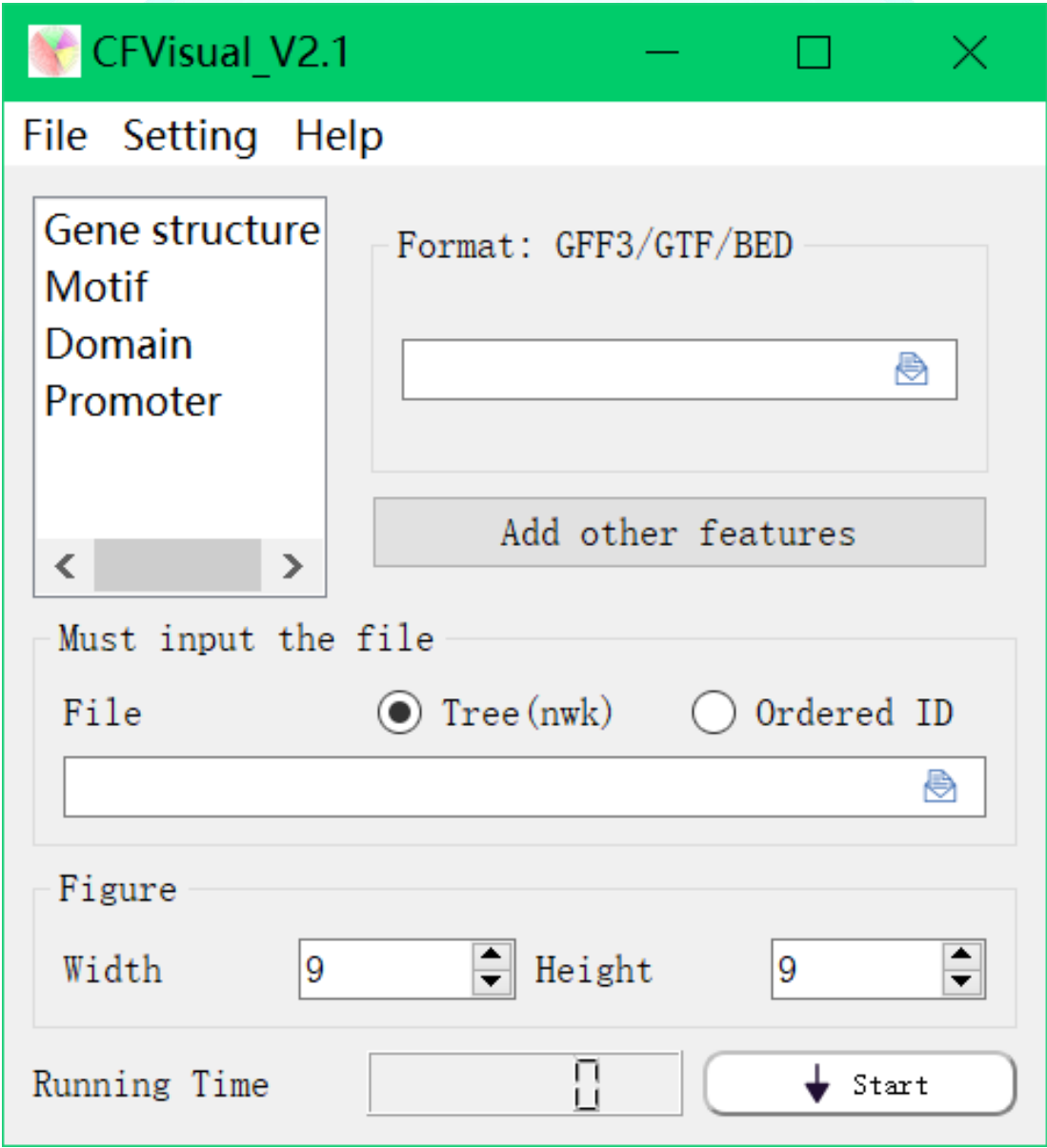
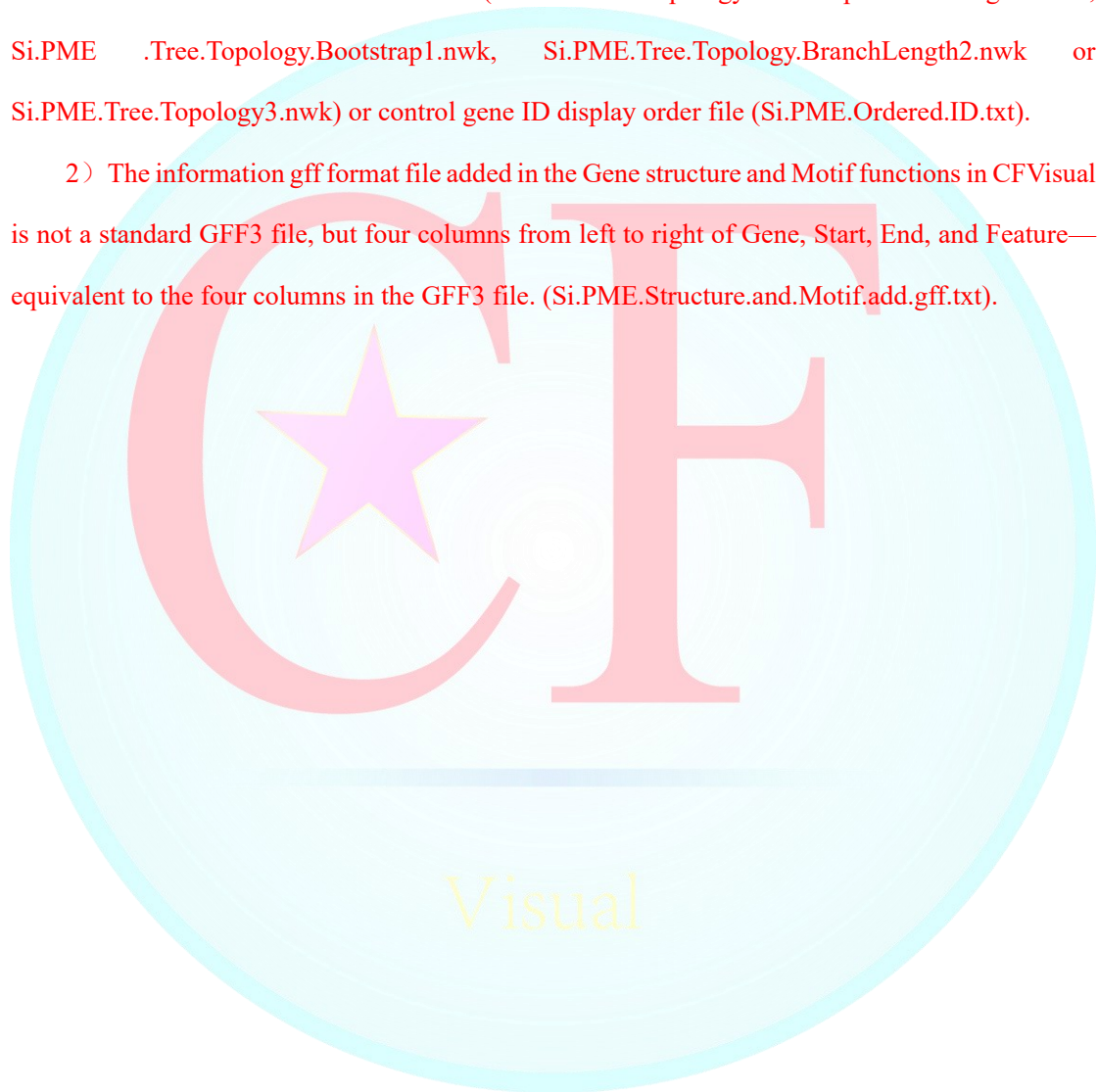


Figure 1 CFVisual home page

3.2 Operation

Terms and Conditions:

- 1) Users can pre-select files on the home page according to the software design rules. By default, you must prepare Newick format (hereinafter referred to as nwk format) evolutionary tree files (Si.PME.Tree.Topology.Bootstrap.BranchLength0.nwk, Si.PME .Tree.Topology.Bootstrap1.nwk, Si.PME.Tree.Topology.BranchLength2.nwk or Si.PME.Tree.Topology3.nwk) or control gene ID display order file (Si.PME.Ordered.ID.txt).
- 2) The information gff format file added in the Gene structure and Motif functions in CFVisual is not a standard GFF3 file, but four columns from left to right of Gene, Start, End, and Feature—equivalent to the four columns in the GFF3 file. (Si.PME.Structure.and.Motif.add.gff.txt).



3.2.1 Drawing of Gene Structure

Support GFF3 (Si.PME.Structure.GFF3.txt), GTF and BED (Si.PME.Structure.BED.txt) file formats, input tree file and gene structure file, click the “Start” button, wait for a while to draw the picture (See Figure 2, Figure 3).

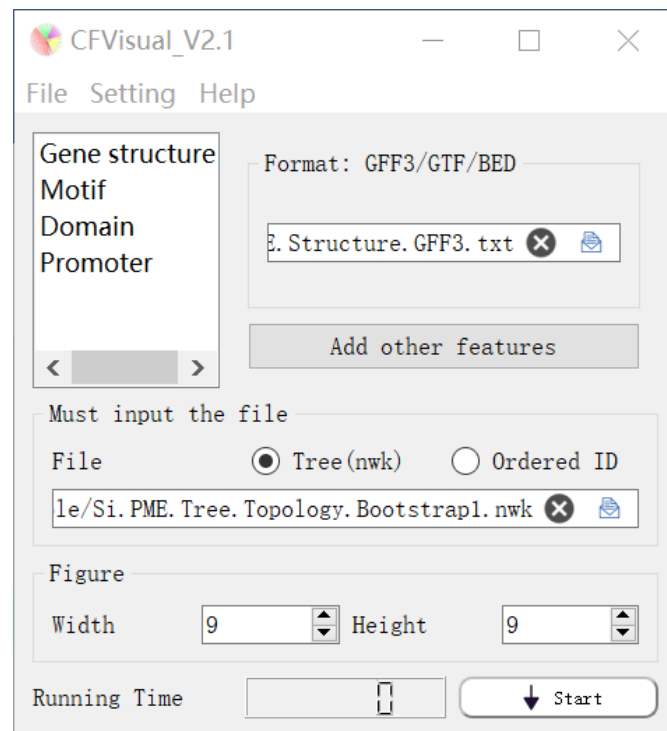


Figure 2 Input gene structure diagram preparation file

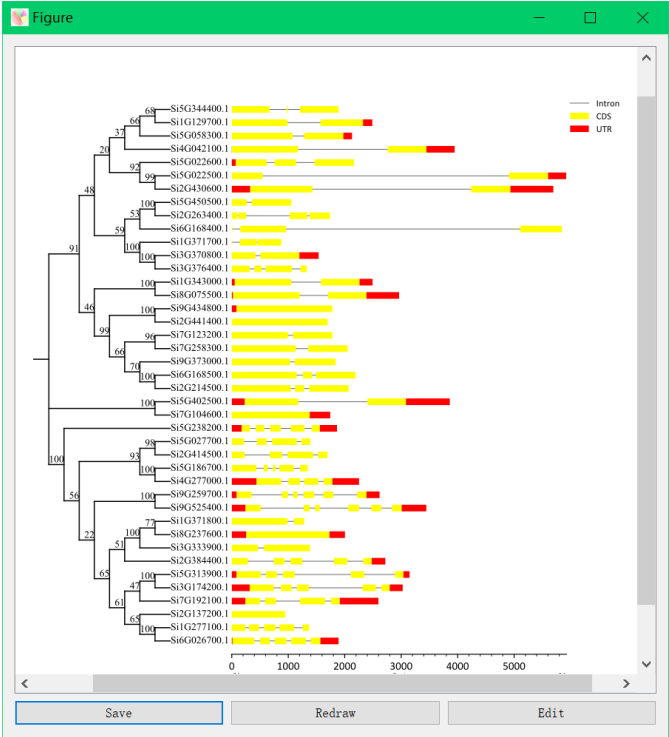


Figure 3 Pop-up window showing gene structure diagram

Click the “Edit” button to pop up the window for editing the picture (Figure 4, Figure 5)

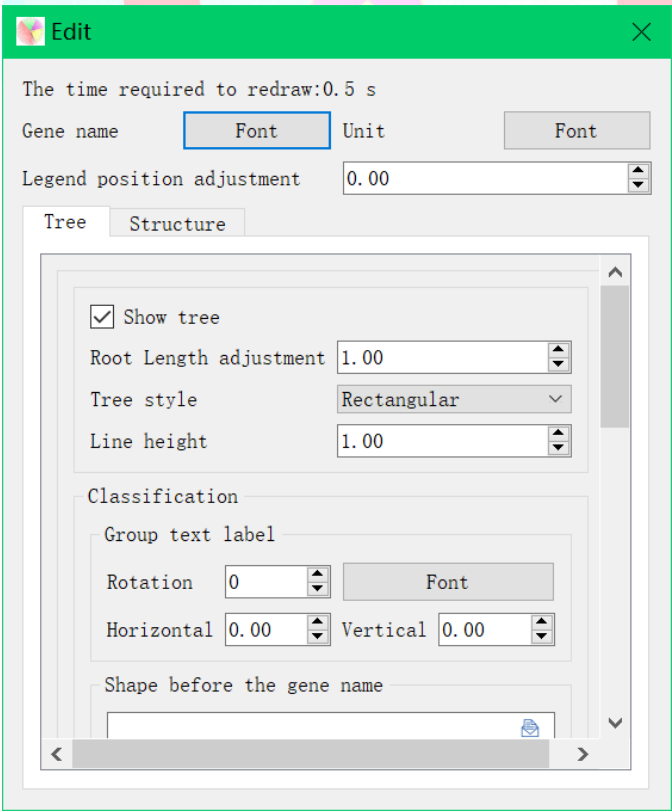


Figure 4 Editing tree diagram

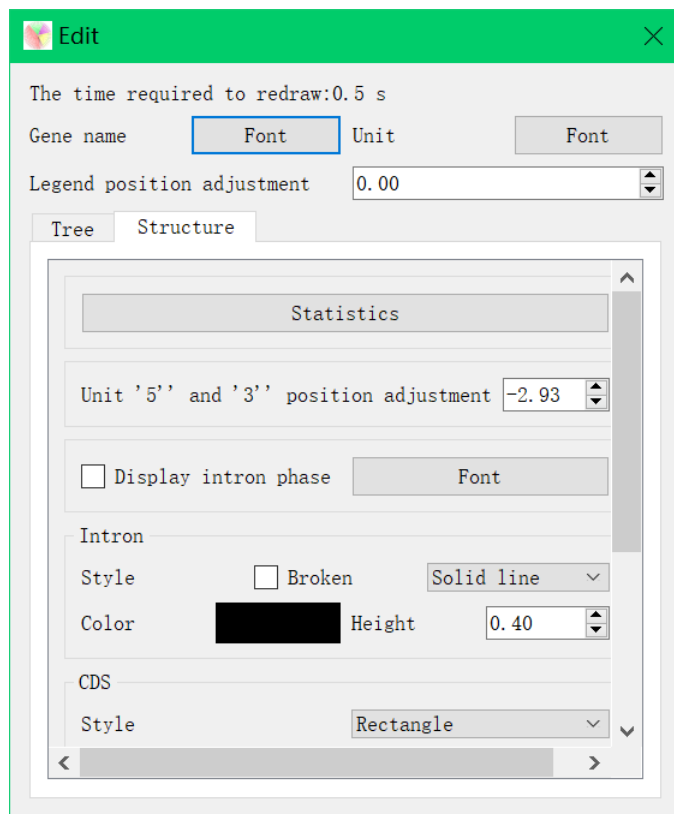
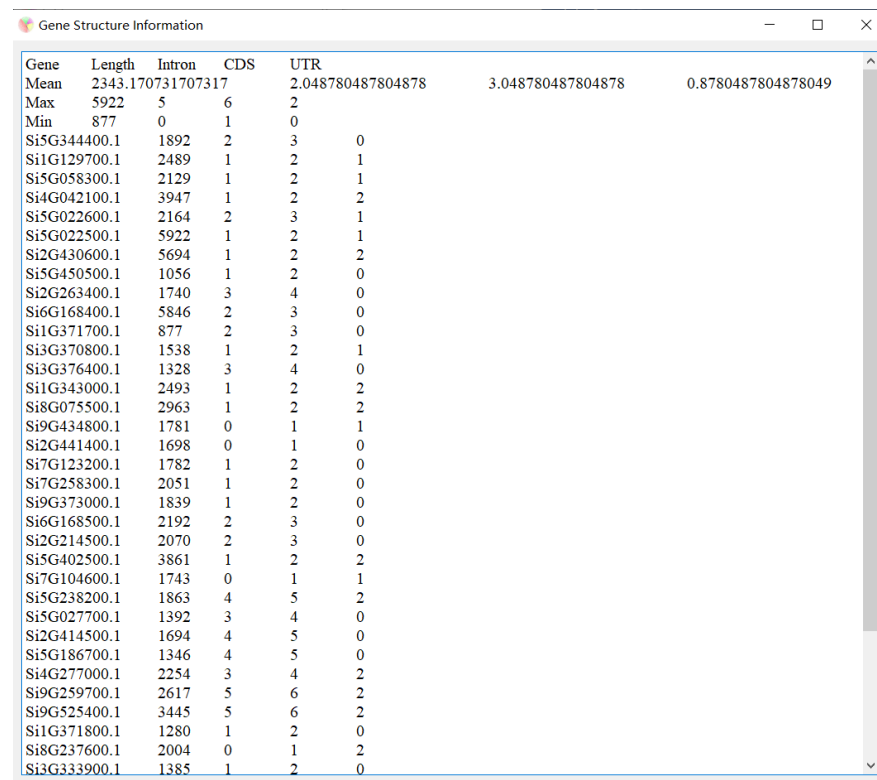


Figure 5 Editing gene structure diagram

Features of editing gene structure diagram: Clicking the “Statistics” button to display the gene structure length, intron number, UTR, CDS, and other quantitative information (Figure 6). (The numerical description of the gene structure map is convenient when writing a paper.)

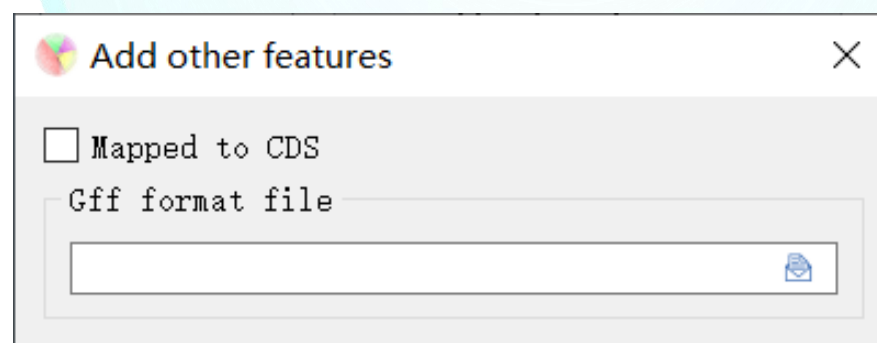
Visual



Gene	Length	Intron	CDS	UTR		
Mean	2343.170731707317			2.048780487804878	3.048780487804878	0.8780487804878049
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		

Figure 6 Numerical information on gene structure

Adding other information: Users can add other information to combine with the gene structure diagram (Figure 7), and the added information file format is gff format file. If the user selects “Mapped to CDS” (Figure 8), then the location information will be mapped to the CDS structure position (Figure 9) (convenient to know which segment of CDS encodes this area).



Add other features

☐ Mapped to CDS

Gff format file

Figure 7

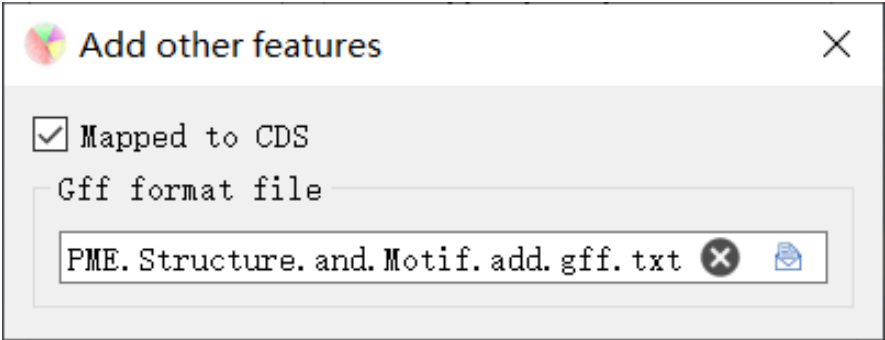


Figure 8

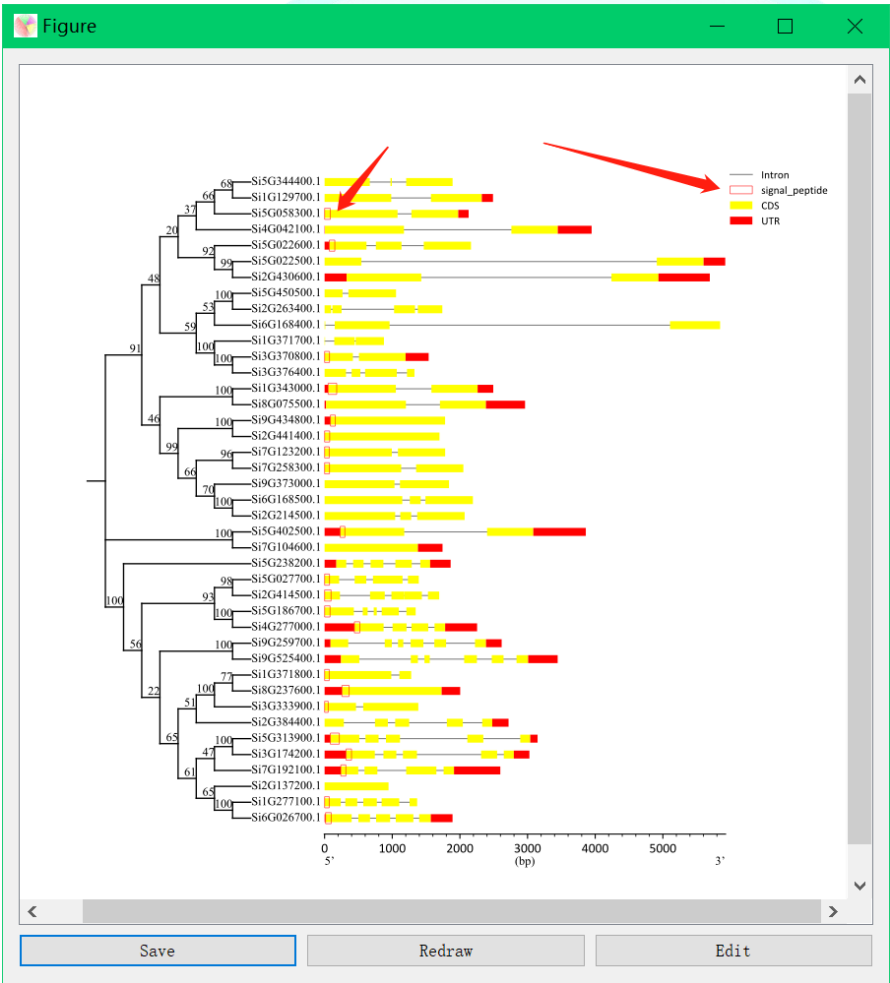


Figure 9

Saving images: Currently, six image formats are supported: png (500dpi), pdf, svg, svgz, raw, and rgba (Figure 10). Note: If the user does not select the path, the CFVisual software will automatically save it to the folder where the tree file or ID sequence file input by the user is located (Figure 11) (the worrying operation of selecting a folder is omitted).

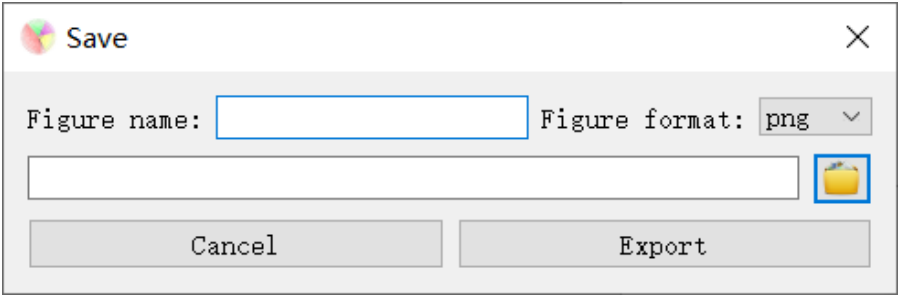


Figure 10 Image output window of CFVisual

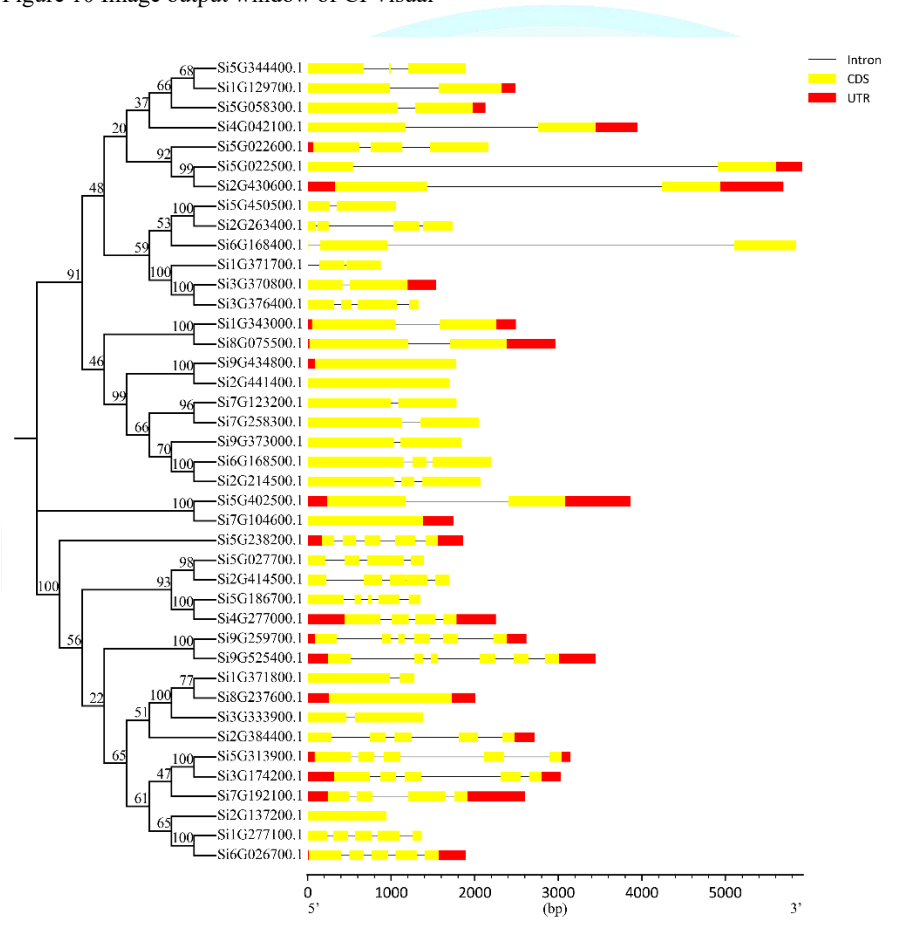


Figure 11 Saving the image in png format

3.2.2 Drawing of Motif

The preparation file is the meme.xml (Si.PME.Motif.meme.xml) or mast.xml (Si.PME.Motif.mast.xml) file obtained after the MEME website (<https://meme-suite.org/meme/>) or software analysis.

Note: CFVisual completely restores the motif results of the MEME official website. When the file entered by the user is meme.xml, there will be two results displayed, namely: Only Motif Sites (Figure 12) and otif Sites+Scanned Sites (Figure 13). This can be toggled through the controls on the Motif Edit tab (Figure 14). In addition, CFVisual also supports the user to selectively hide the motifs that the user does not want to display.

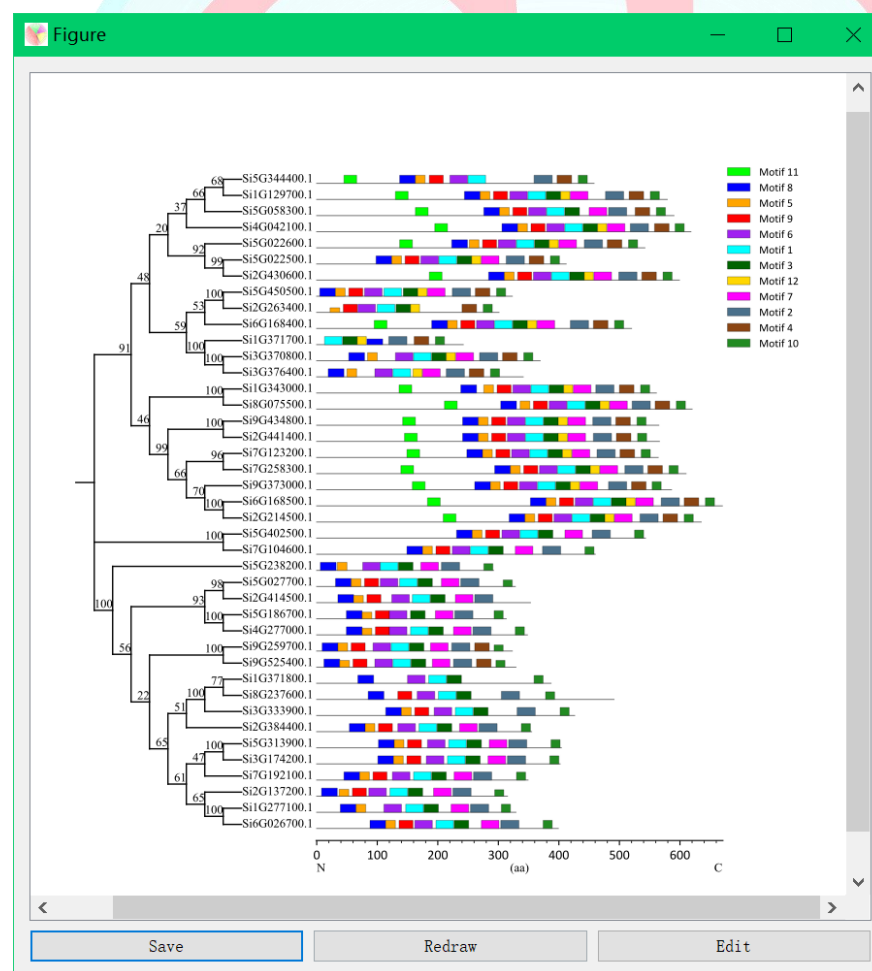


Figure 12 The same initialization result as MEME official website (Only Motif Sites)

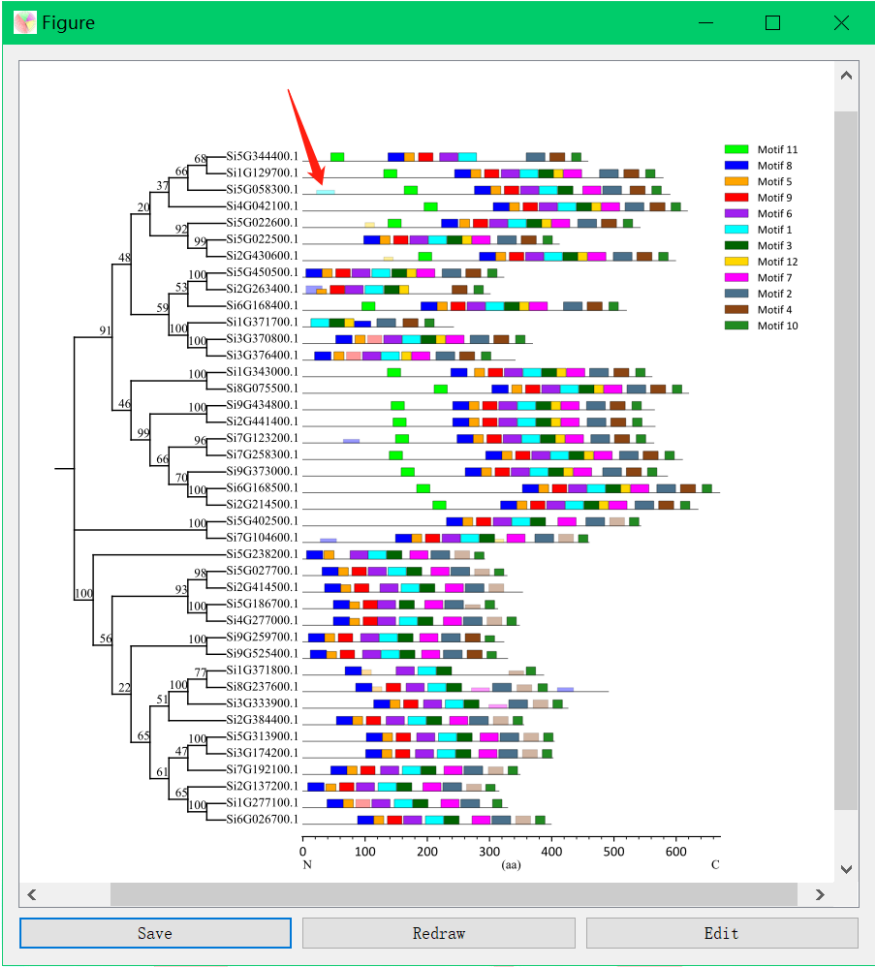


Figure 13 Results of Motif Sites+Scanned Sites

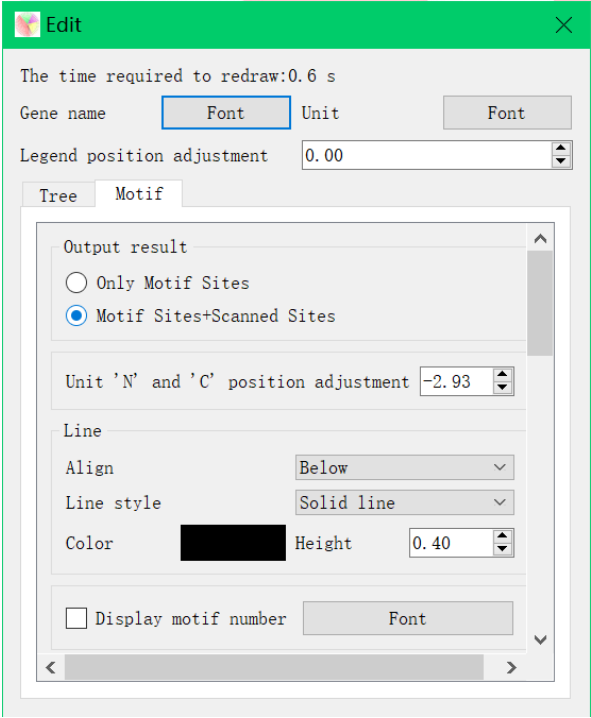


Figure 14

Add other information: Users can add other information to combine with motif (Figure 15, Figure 16), and the file type of the added information is gff format file.

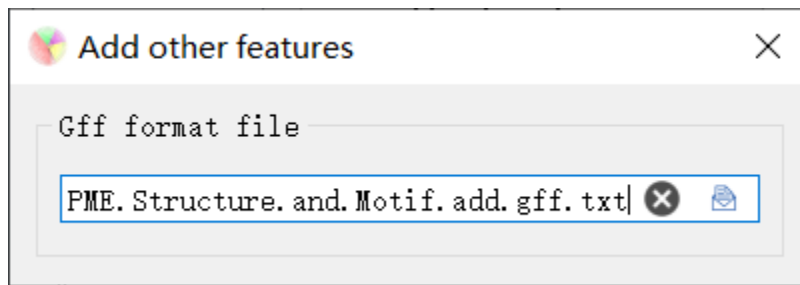


Figure 15

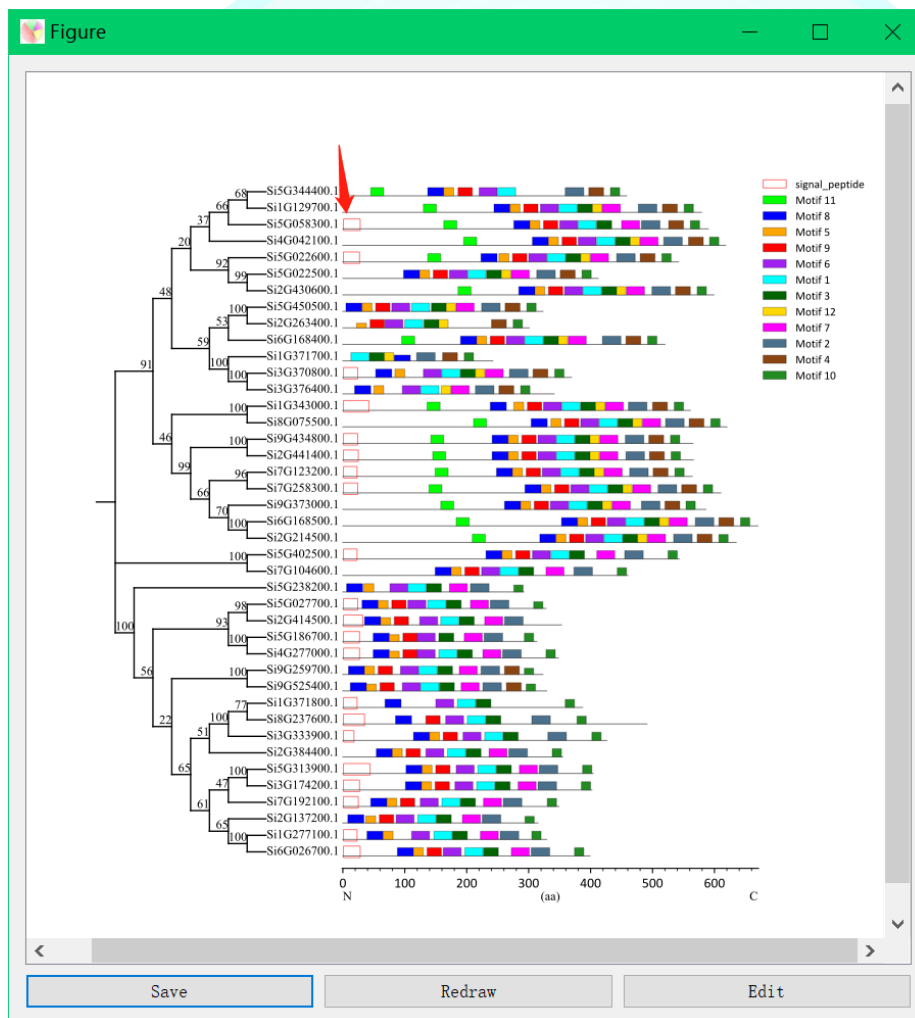


Figure 16

3.2.3 Drawing of Domain

The prepared file is the Pfam result file (Si.PME.Domain.Pfam.txt) obtained from the mailbox after analysis on the Pfam website (<http://pfam.xfam.org/search#tabview=tab1>) or the CDD website (<https://www.ncbi.nlm.nih.gov/cdd>) downloaded CDD results file (Si.PME.Domain.CDD.txt) after analysis (Figure 17). **Note: When using the CDD result file, it is recommended that the user first manually alter the name of the same type of structure domain in the “Short name” column to a unified name.**

The results (Figure 18) and user interaction functionality (Figure 19) are the same as above.

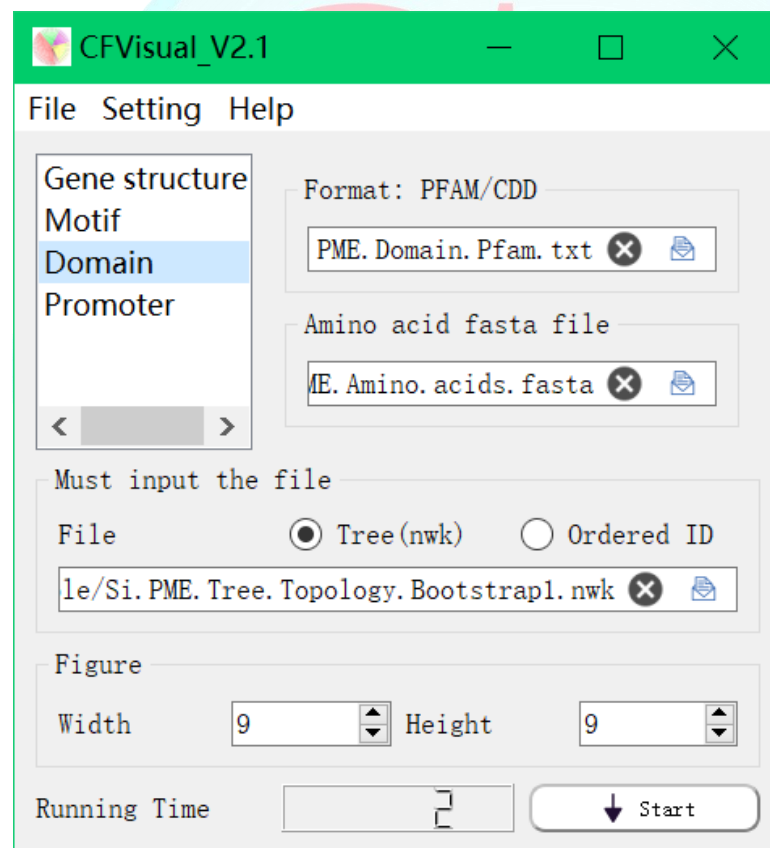


Figure 17

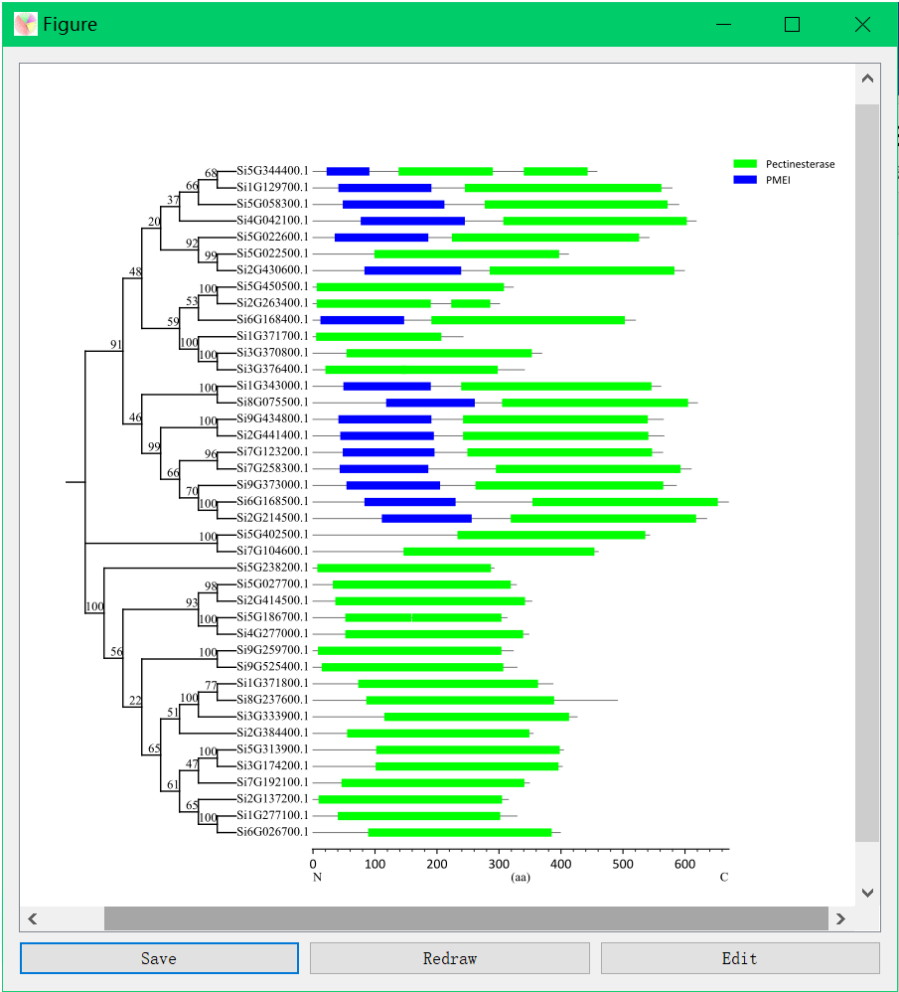


Figure 18

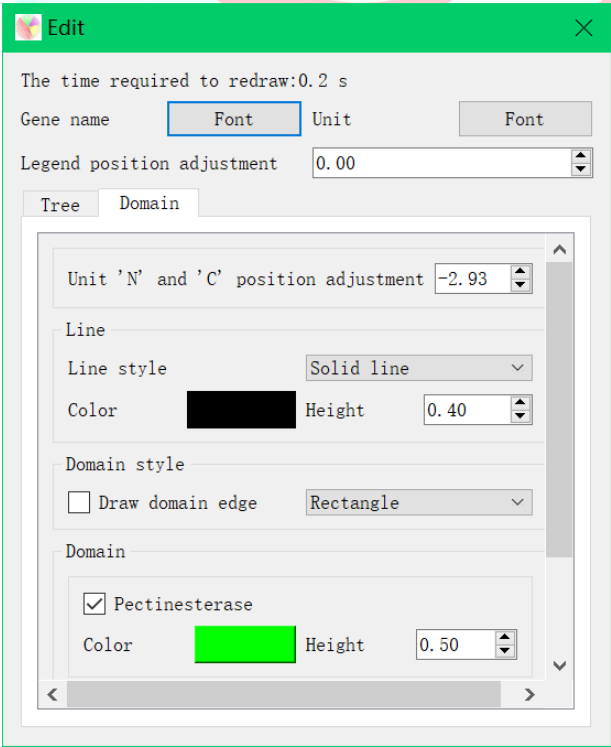


Figure 19

3.2.4 Drawing of *cis*-acting Elements in Promoter Region

Preparing the file as four columns (Si.PME.PlantCare.result.txt) in the results file obtained after analysis on the PlantCare website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Figure 20) Or the results downloaded from other promoter prediction websites, as long as the result format can be processed into the Si.PME.PlantCare.result.txt file type, it can be drawn. **Note:** When using the PlantCare website to analyze the results, you need to do some processing according to your own project needs, and only keep the *cis*-acting element prediction results on the + chain. Flexible use of CFVisual's promoter drawing function to selectively hide *cis*-acting element controls that the user does not want to display can achieve the effect of filtering *cis*-acting elements. In other words, after the user obtains the results of PlantCare, You only need to delete the redundant columns and the results of the row where the - chain is located, and then you can draw it directly with CFVisual, and you no longer need to spend tedious efforts to filter the *cis*-acting component categories.

The results (Fig. 21) and user interaction functions (Fig. 22) are the same as above.

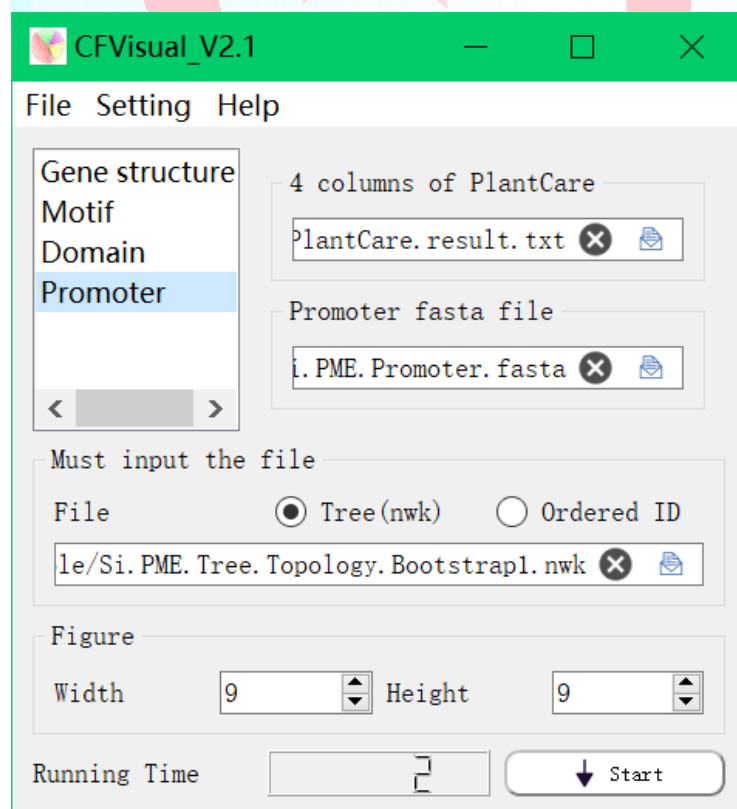


Figure 20

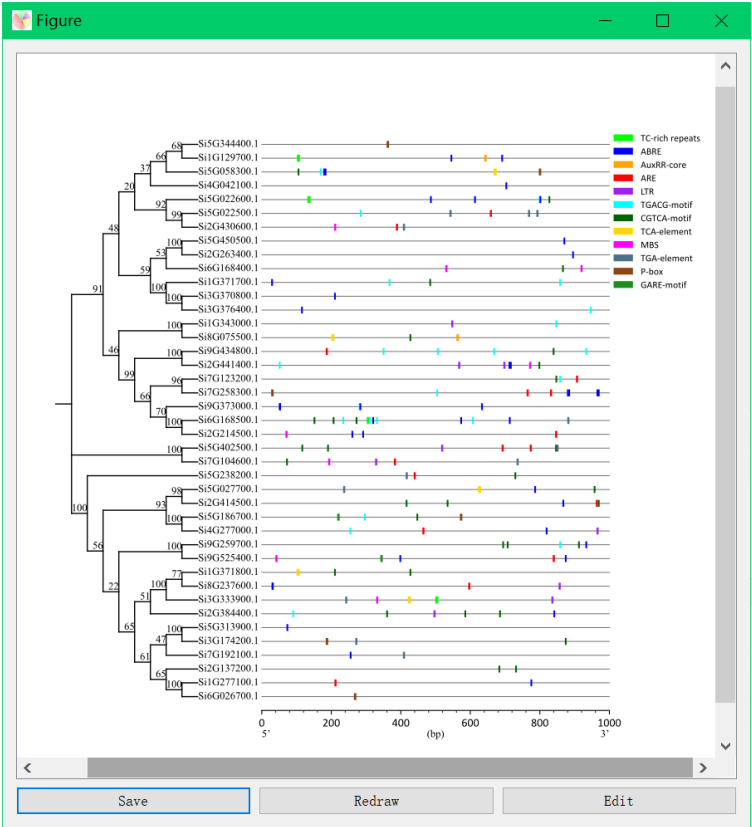


Figure 21

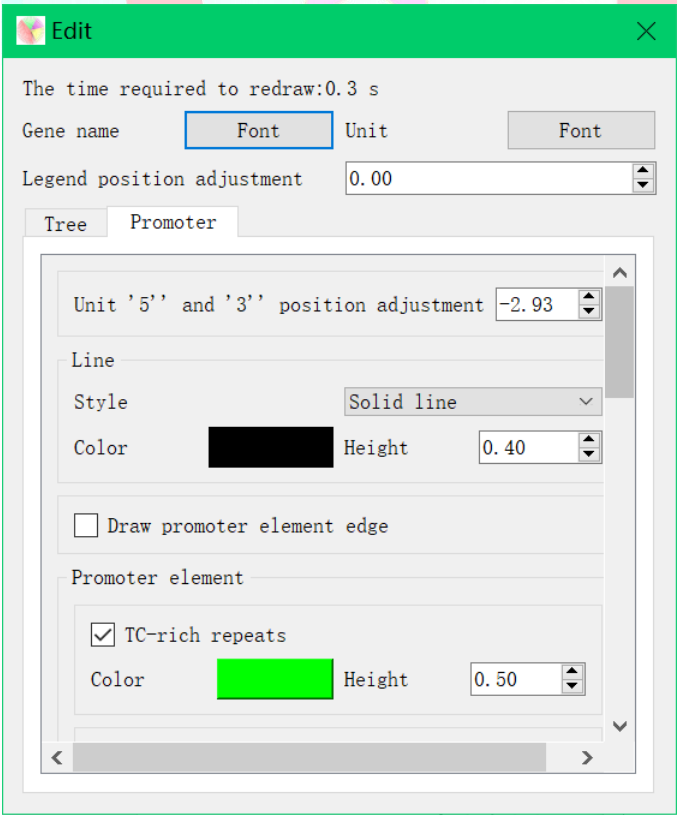


Figure 22

3.2.5 Combination Diagram of Gene Structure and Motif

The author has been in graduate school for three years, and the most bioinformatics analysis he has done is gene family analysis. The author knows that these two analyses are indispensable for gene family analysis, and most people like the combination of tree diagram, gene structure, and motif diagram. To display a picture, we usually use Ai or Ps software to stitch manually. The author knows the tediousness and pain of editing pictures, so programming to achieve the display of this combination picture. One is convenient for myself, but also convenient for teachers and siblings.

Prepare files: same as preparation file for gene structure map and motif map drawing (Figure 23, Figure 24). The editing method is the same as the gene structure map and the motif map (Figure 25, Figure 26). The author considers this situation, and some people like to put the gene structure map in the middle, so this function (Exchange coordinates in the Motif editing tab) is specially added (Figure 25).

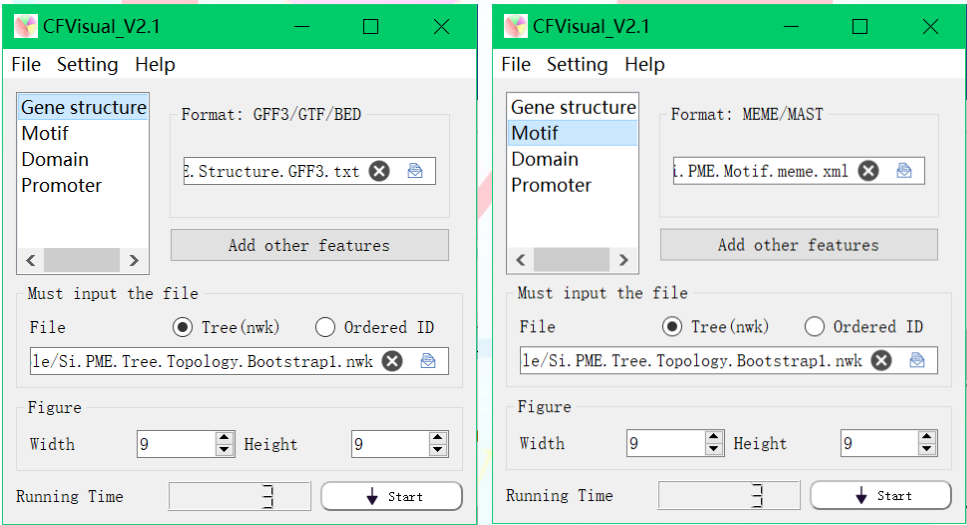


Figure 23 and Figure 24

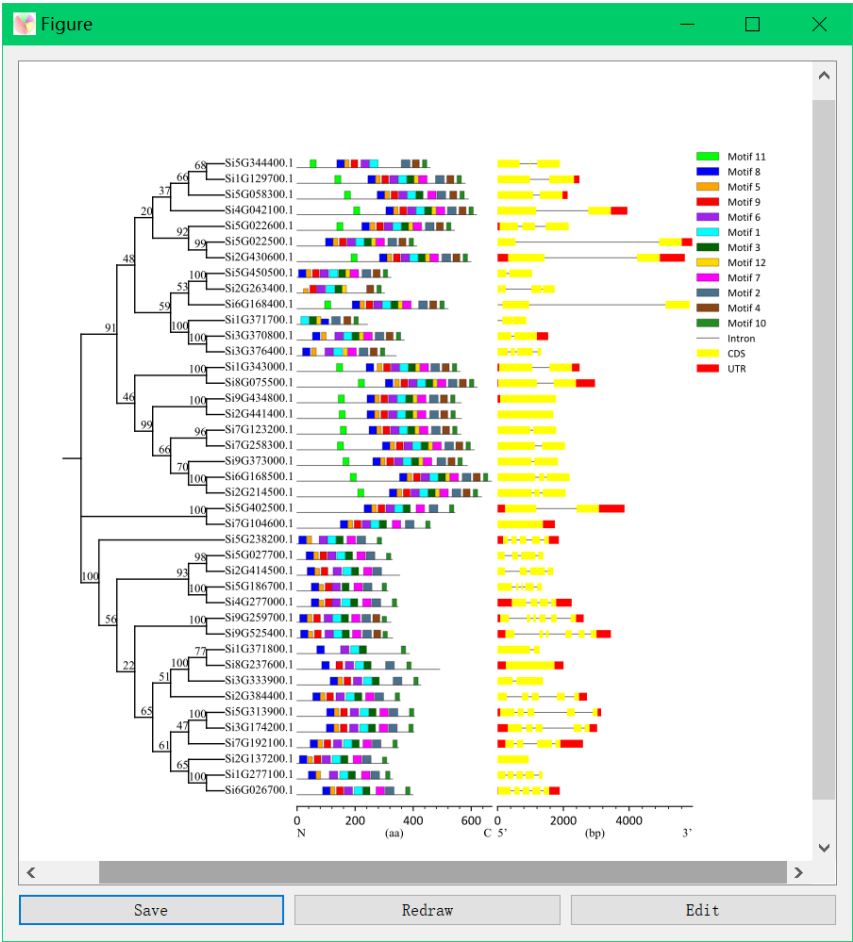


Figure 25

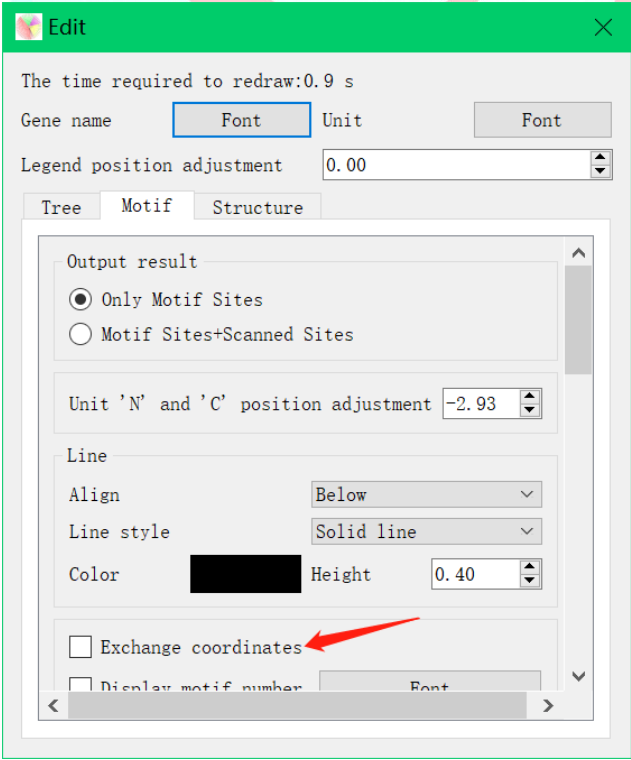


Figure 26

3.2.6 Combination Diagram of Gene Structure and Domain

The role of this combination map is to facilitate researchers to understand which cds encode domains. The preparation file is the same as gene structure diagram and domain diagram drawing (Figure 27), but it should be noted that when CFVisual draws the combination diagram, it is no longer necessary to add the amino acid fasta file (Figure 27) when preparing the domain file. The function of the gene structure binding domain is already available in GSDS, but it is not convenient to prepare the amino acid position file by yourself. The author directly reads and manipulates the Pfam or CDD result file to obtain the position information. The most important point is that CFVisual's domain representation is not the same as GSDS, but the essence is the same. The difference is that the domain of CFVisual is represented by a box, and the intron of the interval is also boxed (Figure 28), while GSDS is a solid block and does not color introns. The reason why the author does this is to try to open up a new idea of gene structure analysis. Since the tutor pays more attention to introns, the author believes that the intron insertion or loss in CDS of the domain can be easily viewed by showing it in this way without losing its original purpose.

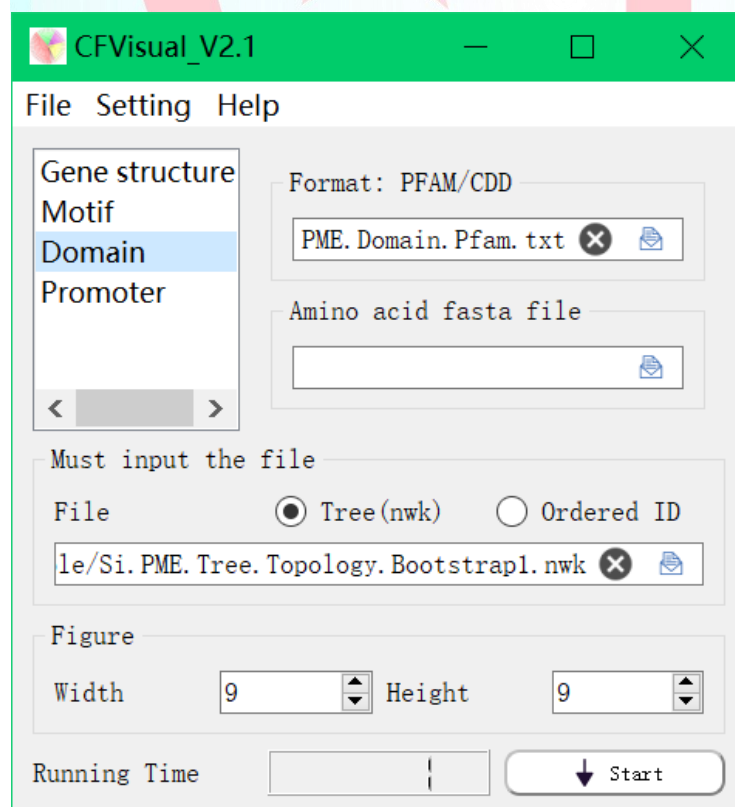


Figure 27

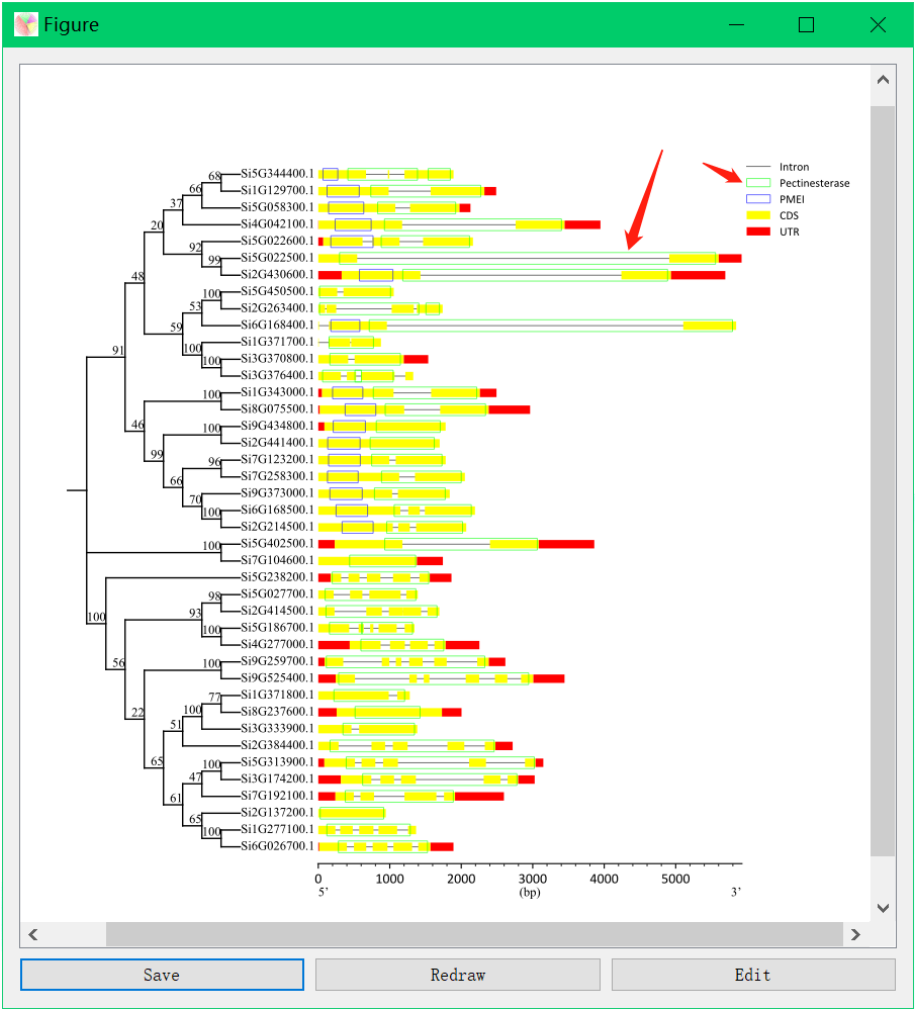


Figure 28

Visual

3.2.7 Combination Diagram of Motif and Domain

This combination originated from a sister who worked in the NBS-LRR family of disease resistance genes. The reviewer asked her to analyze the combination of motifs and domains **to see which motifs the domains contain or the positional distribution relationship between domains and motifs**. Since no software with this function has been found, she can only mark it with the naked eye according to the position of the structural domain, which is troublesome and the position anchoring is not necessarily accurate. Fortunately, the reviewer only asked her to mark a few positions as a schematic diagram, otherwise a family with such a large disease resistance gene would have to be marked in what year and month. Based on this, the author implements this combined requirement (Figure 29).

The preparation file is the same as the motif diagram and the domain diagram (Figure 27). It should also be noted that when CFVisual draws the combination diagram, it is no longer necessary to add the amino acid fasta file when preparing the domain file (Figure 27).

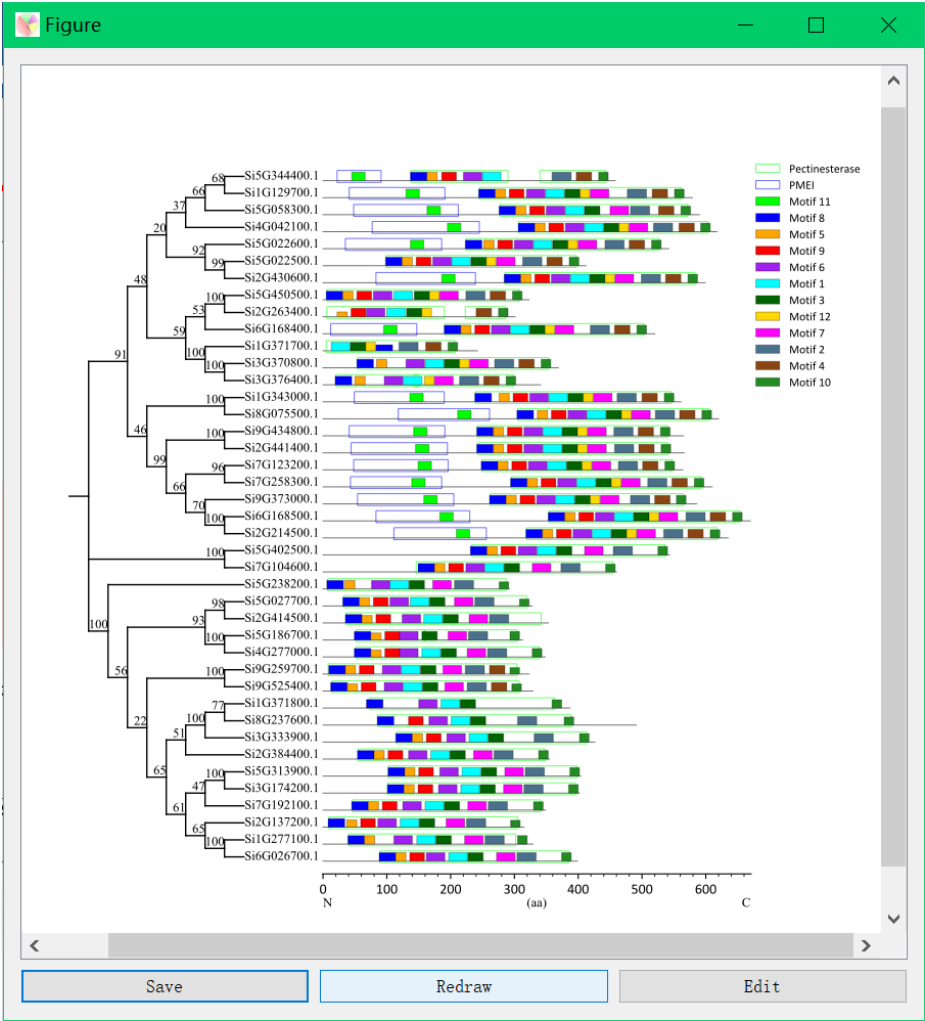


Figure 29

Visual

3.2.8 Combination Diagram of Gene Structure, Motif, and Domain

Preparing the file is the same as gene structure drawing, motif drawing and domain drawing. Similarly, when drawing the combination graph, the domain drawing does not need to prepare the amino acid fasta file. The drawing results are shown in Figure 30, and the interpretation of the combination chart is the same as above.

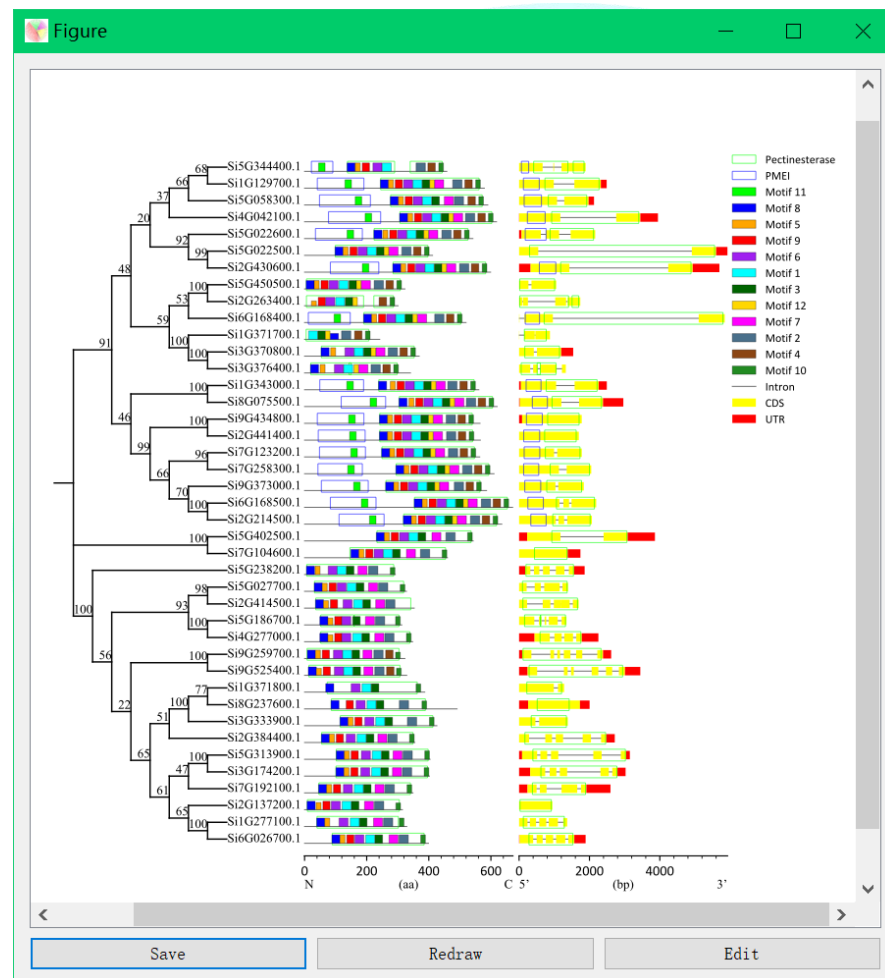


Figure 30

3.2.9 Other Combinations

The author thought about other combination maps, such as the combination of promoter map and gene structure map, etc., but considering that this combination is not common, this combination is not supported for the time being.



3.3.0 Interaction, Classification, and Beautification of Tree Diagram

CFVisual only supports tree files in nwk format. If the user's tree file format is not nwk file, please use other software to convert the tree file into nwk format tree file. Here we recommend FigTree software (<http://tree.bio.ed.ac.uk/software/figtree/>), as shown in Figure 31 and Figure 32:

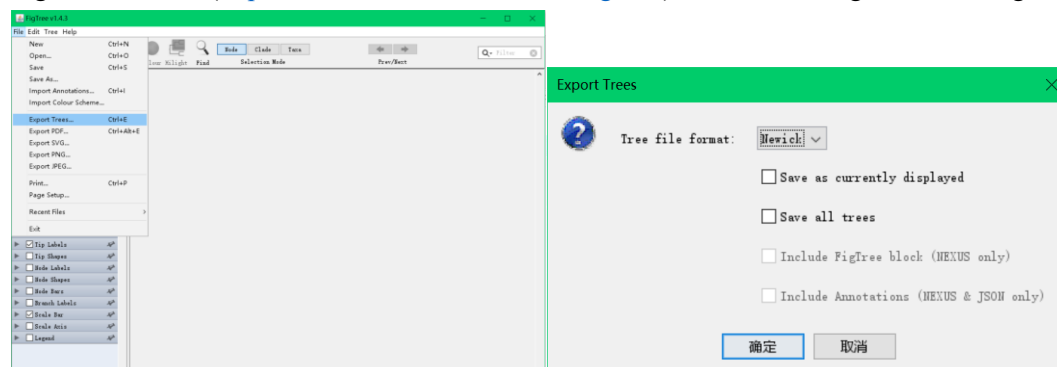


Figure 31 and Figure 32 Steps to convert tree files to nwk format using FigTree

The tree files in nwk format can be divided into **four categories** in terms of information, namely nwk files containing bootstrap value and branch length information (Si.PME.Tree.Topology.Bootstrap.BranchLength0.nwk) (Figure 33), nwk file containing only bootstrap value (Si.PME.Tree.Topology.Bootstrap1.nwk) (Figure 34), nwk file containing only branch length information (Si.PME.Tree.Topology.BranchLength2.nwk) (Figure 35) and the simplest topology nwk file (Si.PME.Tree.Topology3.nwk) (Figure 36).

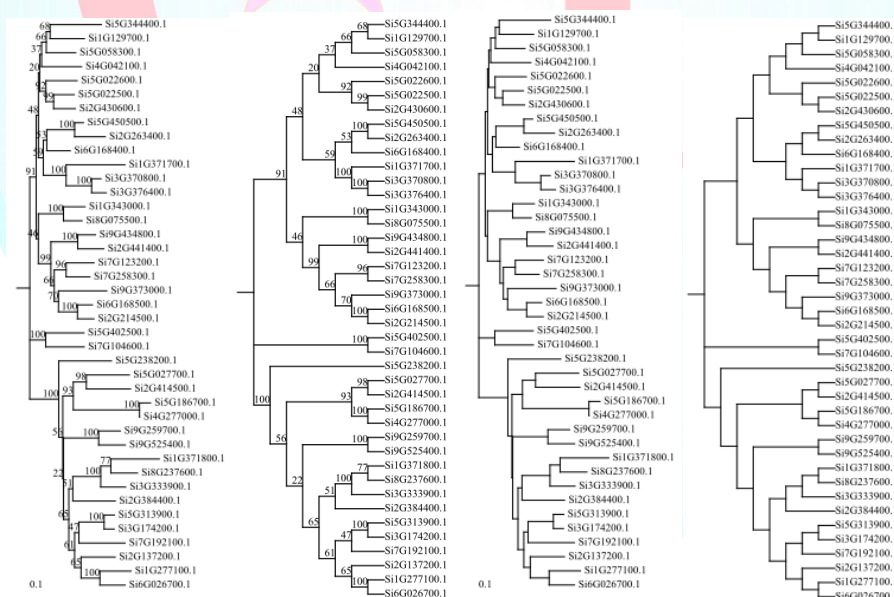


Figure 33, Figure 34, Figure 35, and Figure 36 Four types of files in nwk format

The boundary window that interacts with the treemap is shown in the following figures (Figure 37 and Figure 38), and users can experience it by themselves:

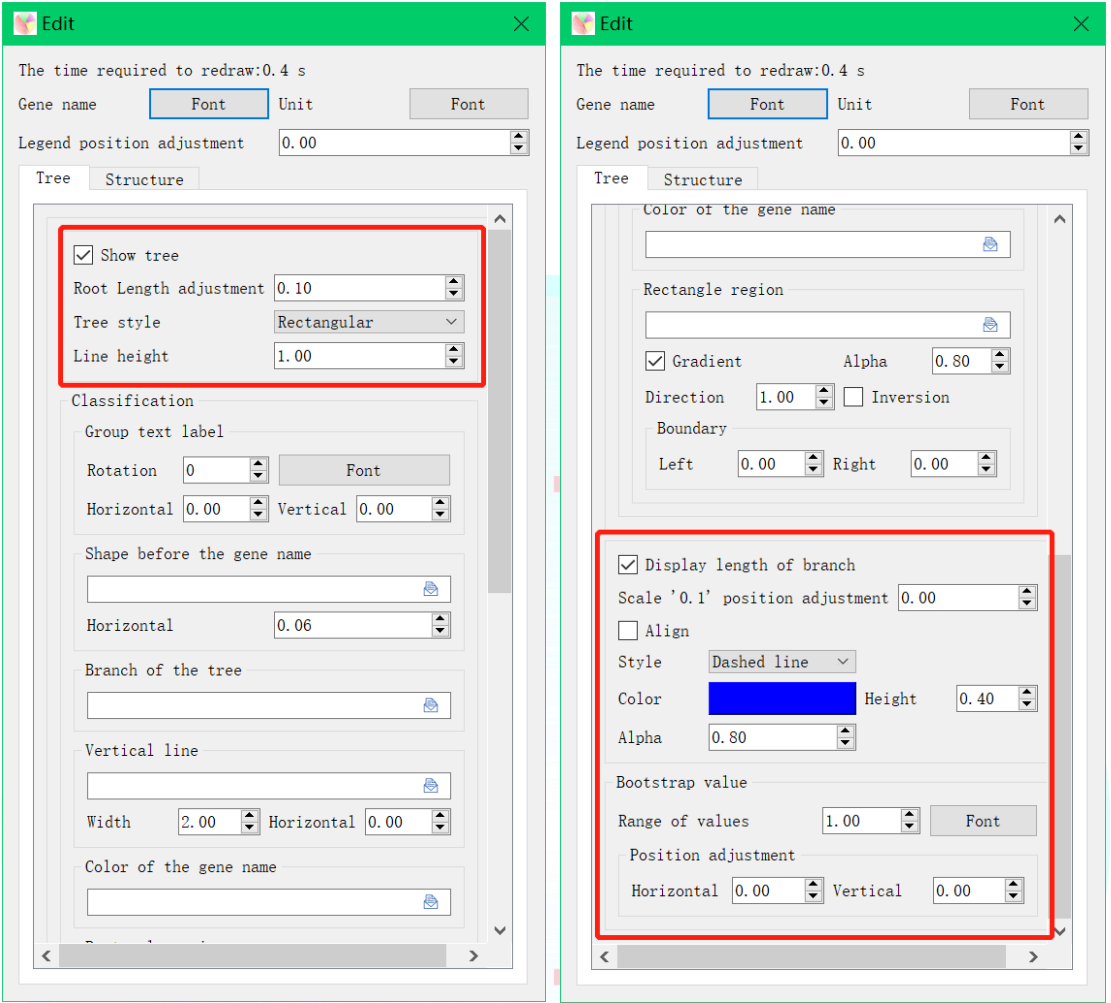


Figure 37 and Figure 38 Treemap interactive interface

Since the author knows that in order to truly realize the multi-picture integration in gene family analysis, the classification and beautification functions of tree diagrams are necessary. Here, it is worth explaining carefully about **the classification and beautification of tree diagrams**.

First of all, CFVisual supports two recognizable classification and beautification files, **namely the shape control file before the gene name (Si.PME.Tree.Shape.before.gene.name.txt) and the treemap classification control file (Si.PME. Tree.Classification.txt).**

The format of the shape control file before the gene name is as shown in the figure below (Figure 39). Users can use EXCEL to make an interactive format file that can be recognized by CFVisual according to their own needs: **the columns from left to right are expressed as: gene name, shape symbol, shape size, shape fill color and shape outline color.** Among them, the shape symbol is understood as shown in Figure 40, and the color parameter supports 8 commonly used words or abbreviations representing colors (Figure 41) and hexadecimal color-coded strings (Figure 42) —here the author recommends one of the rookie tools Section (<https://c.runoob.com/front-end/55>), because it not only supports color search, but also supports mutual conversion between RGB format and hexadecimal color-coded strings, it is worth collecting.



Si1G371700.1	o	5	blue	#00FF00
Si2G414500.1	v	10	cyan	#00FF01
Si9G259700.1	^	15	green	#00FF02
Si1G277100.1	<	20	black	#00FF03
Si2G263400.1	>	25	magenta	#00FF04
Si3G333900.1	8	30	red	red
Si2G384400.1	s	35	white	white
Si5G450500.1	p	40	yellow	yellow
Si9G525400.1	*	50	b	b
Si6G026700.1	h	55	c	#00FF01
Si8G237600.1	D	60	g	#00FF02
Si5G186700.1	X	55	k	#00FF03
Si3G376400.1	o	50	m	#00FF04
Si5G022500.1	v	45	r	#00FF05
Si2G137200.1	^	40	w	#00FF06
Si5G313900.1	<	35	y	y
Si5G027700.1	>	30	#FF0000	#FF0000
Si7G192100.1	8	25	#00FF00	red
Si3G370800.1	s	20	#0000FF	red
Si1G371800.1	p	15	#FFA500	#FFA500
Si5G238200.1	*	10	#A020F0	#FFA500
Si7G104600.1	h	5	#00FFFF	red
Si3G174200.1	D	10	#006400	red
Si4G277000.1	X	15	#FFD700	#FFD700
Si5G402500.1	o	20	#FFD700	#FFD700
Si5G344400.1	v	25	#FFD700	#FFD700
Si5G058300.1	^	30	#FFD700	#FFD700
Si5G022600.1	<	35	#FFD700	#FFD700
Si9G434800.1	>	40	#FFD700	b
Si9G373000.1	8	45	#FFD700	b

Figure 39

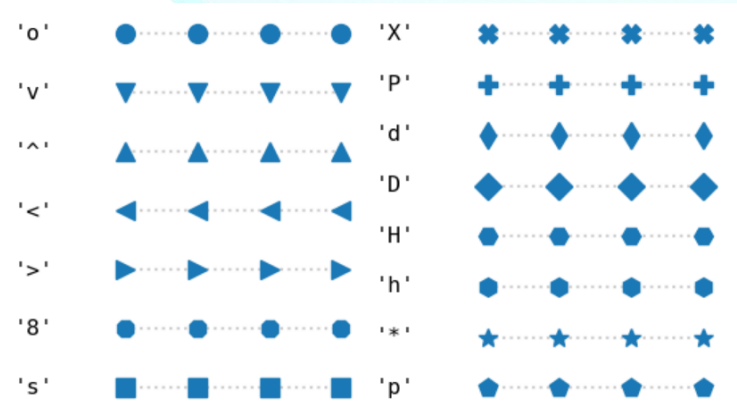


Figure 40



Figure 41

RGB转16进制工具用于将RGB颜色值与十六进制字符串相互转换，工具使用简单，你只需要在以下三个输入框：红(R)、绿(G)、蓝(B)中输入RGB的颜色值及会自动回转换十六进制的字符串，反之你只想要在右侧十六进制输入框输入16进制字符串即可转换为 RGB 数值。

RGB 与 HEX 颜色对照表

实色效果	英文名称	R.G.B	16色	实色效果	英文名称	R.G.B	16色
	Snow	255 250 250	#FFFAFA		PaleTurquoise1	187 255 255	#BBFFFF
	GhostWhite	248 248 255	#F8F8FF		PaleTurquoise2	174 238 238	#AEEEEE
	WhiteSmoke	245 245 245	#F5F5F5		PaleTurquoise3	150 205 205	#96CDCD
	Gainsboro	220 220 220	#DCDCDC		PaleTurquoise4	102 139 139	#668B8B
	FloralWhite	255 250 240	#FFFAF0		CadetBlue1	152 245 255	#98F5FF
	OldLace	253 245 230	#FDF5E6		CadetBlue2	142 229 238	#8EE5EE
	Linen	250 240 230	#FAF0E6		CadetBlue3	122 197 205	#7AC5CD
	AntiqueWhite	250 235 215	#FAEBD7		CadetBlue4	83 134 139	#53868B
	PapayaWhip	255 239 213	#FFEFD5		Turquoise1	0 245 255	#00F5FF
	BlanchedAlmond	255 235 205	#FFEBCD		Turquoise2	0 229 238	#00E5EE

Figure 42

Visual

After preparing the shape control file before the gene name, you can enter the file in the tree diagram editing window (Figure 43), and then click the “Redraw” button in the Figure window to redraw it (Figure 44).

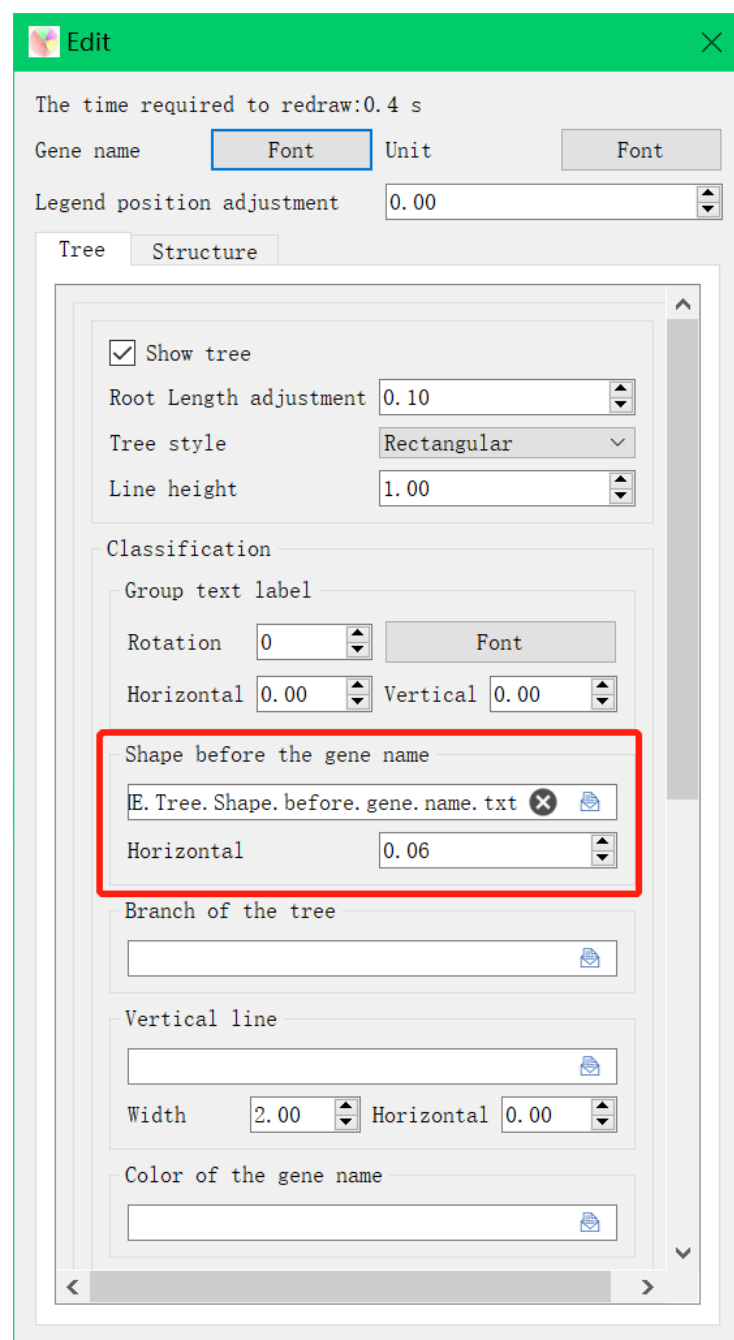


Figure 43

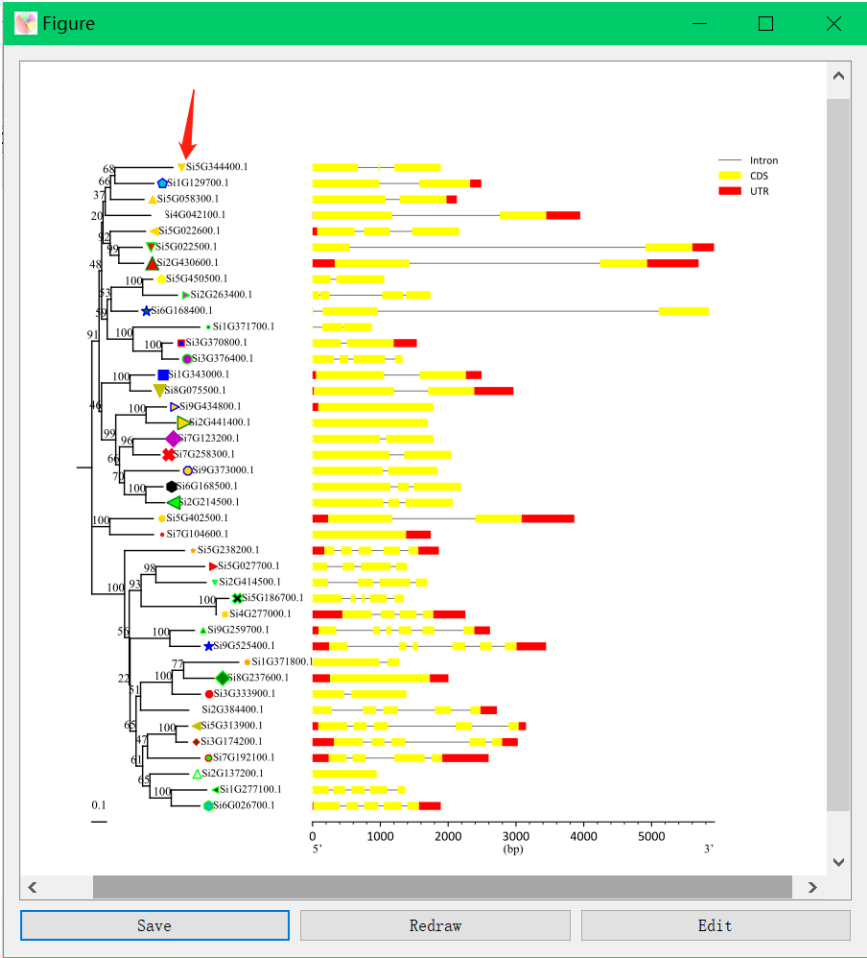


Figure 44

The format of the treemap classification control file is as follows (Figure 45): The columns from left to right represent: the gene name at the upper or lower end of the classification region, the gene name corresponding to the lower or upper end of the classification region, the color, and the label of the classification. Note: Each row corresponds to a classification area; the fourth column may not be provided.

Si5G344400.1	Si4G042100.1	r	Group 1
Si5G022600.1	Si2G430600.1	blue	Group 2
Si5G450500.1	Si6G168400.1	#FFA500	
Si1G371700.1	Si3G376400.1	green	Group 4
Si1G343000.1	Si7G258300.1	red	Group 5
Si9G373000.1	Si2G214500.1	#00FF05	Group 6
Si5G402500.1	Si7G104600.1	b	Group 7
Si5G238200.1	Si7G192100.1	yellow	Group 8
Si2G137200.1	Si6G026700.1	#FFD700	Group 9

Figure 45

Four classification forms are supported (as shown in Figure 46), namely branch color, classification vertical line color, gene name color and classification area color. The user can drag the prepared treemap classification control file into the text box corresponding to any one or more classification forms that the user likes (Figure 47). After setting the classification parameters, clicking the “Redraw” button, and the result is shown in Figure 48. It is worth mentioning that by making full use of the settings of Direction and Inversion in the color classification interface, you can adjust the direction of the regional gradient. For example, the default Direction is set to 1, Inversion is not selected, and the regional gradient direction is from right to left; Inversion is selected, the area gradient direction is opposite from left to right. Direction is set to 0, Inversion is not selected, the area gradient direction is from bottom to top; Inversion is selected, the area gradient direction is opposite from top to bottom. In addition, uncheck the Gradient control, the area is completely filled, and the gradient effect is not displayed.

Visual

Branch of the tree

Vertical line

Width 2.00 Horizontal 0.00

Color of the gene name

Rectangle region

☒ Gradient Alpha 0.80

Direction 1.00 ☐ Inversion

Boundary

Left 0.00 Right 0.00

Figure 46

Branch of the tree

le/Si.PME.Tree.Classification.txt

Vertical line

le/Si.PME.Tree.Classification.txt

Width 2.00 Horizontal 0.00

Color of the gene name

le/Si.PME.Tree.Classification.txt

Rectangle region

le/Si.PME.Tree.Classification.txt

☒ Gradient Alpha 0.40

Direction 1.00 ☐ Inversion

Boundary

Left 0.00 Right 0.00

Figure 47

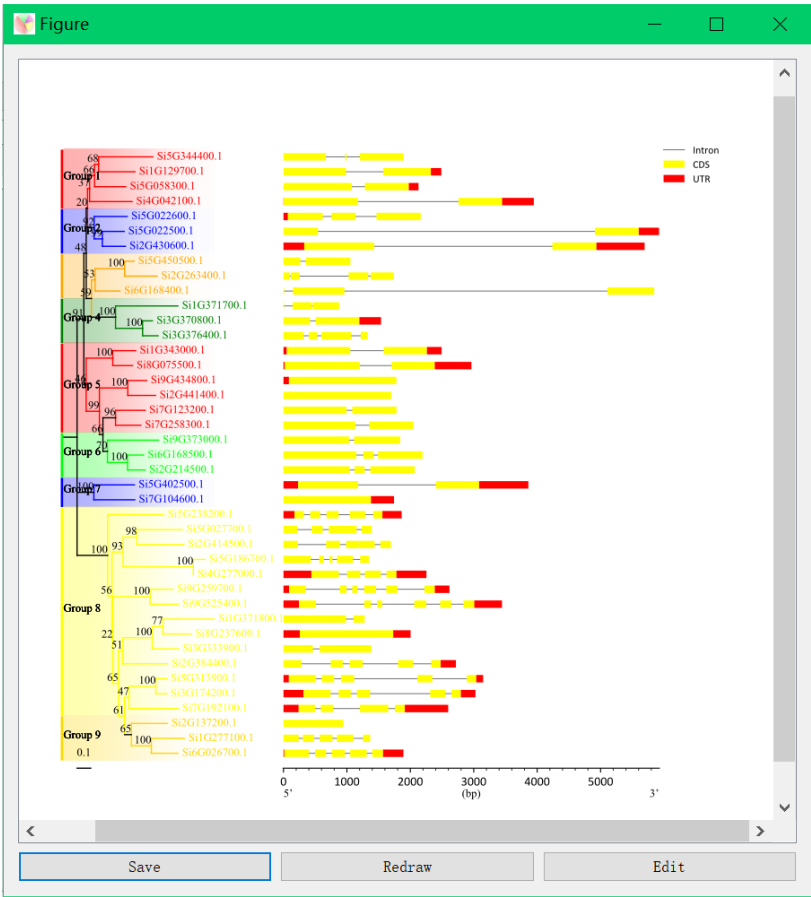


Figure 48

All in all, as long as the user makes full use of the treemap classification function of CFVisual, there is really no need to use Ps or Ai and other retouching tools to modify. The authors then present two sample images for reference (Figure 48 and Figure 49).

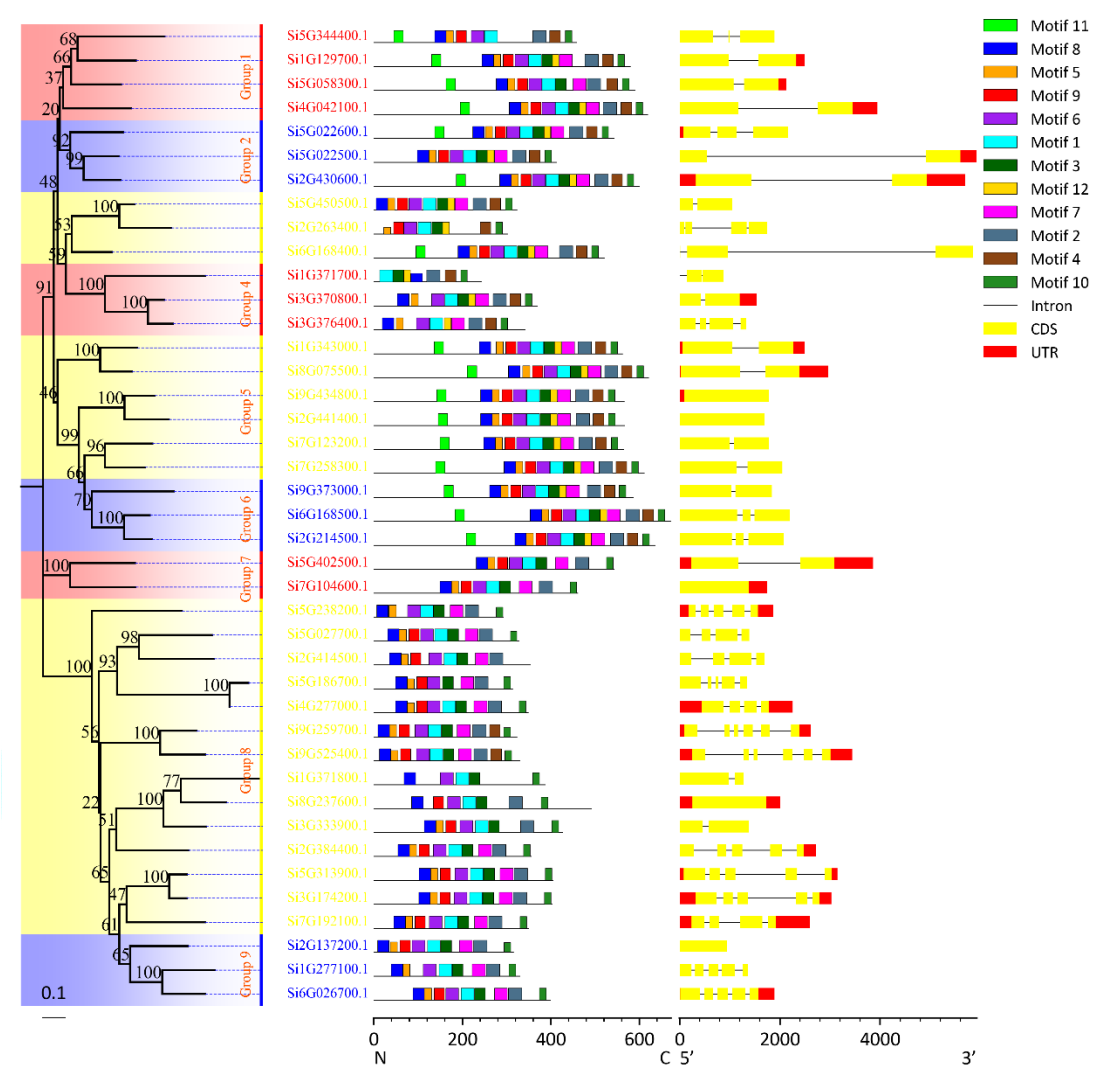


Figure 49

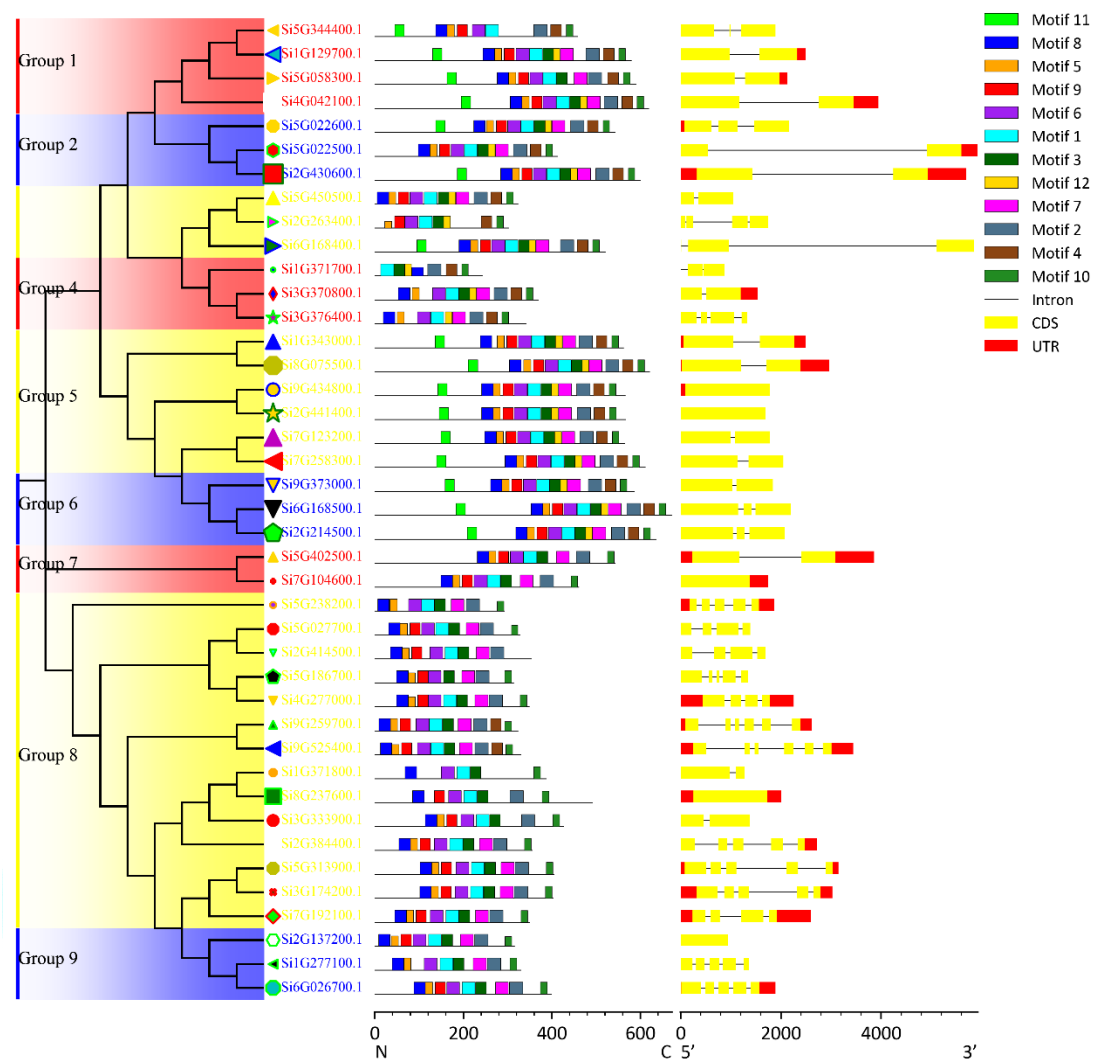


Figure 50

Visual