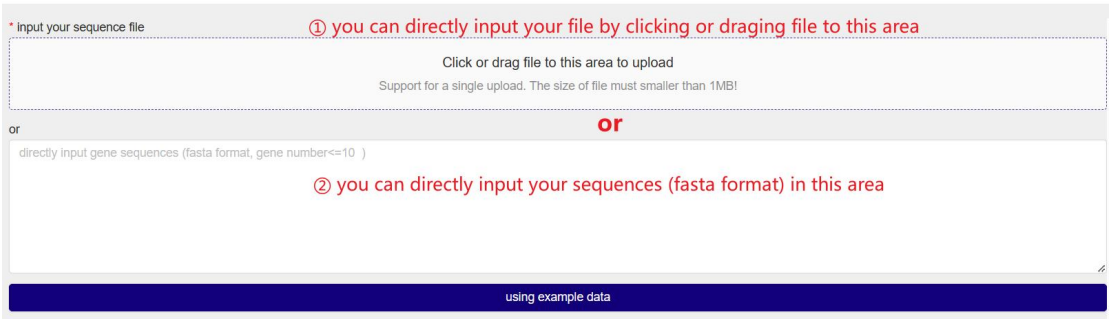
Quick start and notes of using GFAP

Quick start

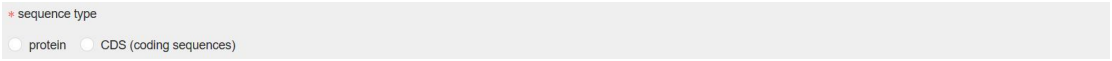
For functional annotation

These functions in 'Annotation-GO/KEGG/pfam', which can be used for annotating genes with GO, KEGG, protein domains and other protein functional information.

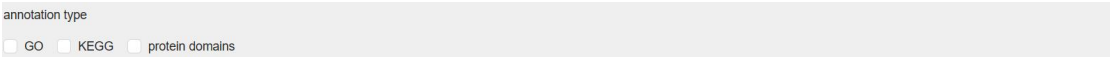
1. Upload your file (fasta format) into our website



2. Select sequence type ('coding sequences' means these sequences can be fully translated into protein sequences)



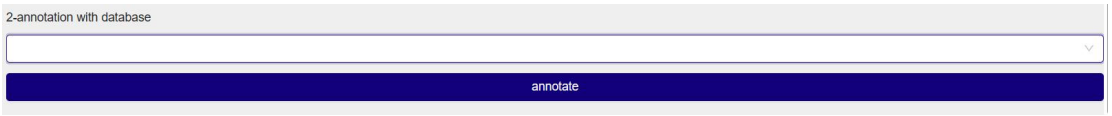
3. Select annotation type (if you want to annotate genes using GO, KEGG or protein domains, you need to select you required annotation types)



4-1. If you want to annotate genes using the model of closely related species

Select the closely related species and press annotate button to finish this process.

4-2 If it is hard to select a closely related species, you can utilize other databases to annotate genes.

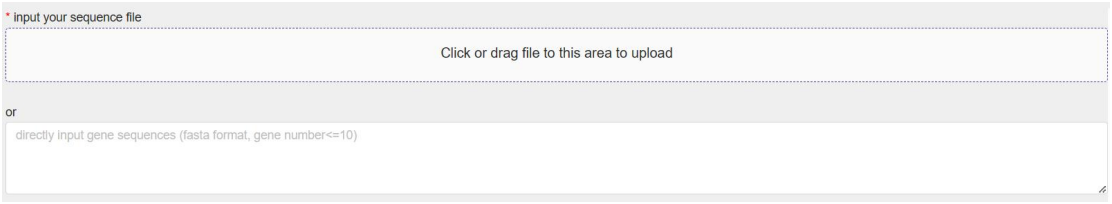


There are four databases provided for selection. The detail information can be found in our article. In a nutshell, plant-special and total databases can be used for annotating genes with GO, KEGG and protein domains. The swissprot and nr databases can be utilized for annotating genes with protein functions (not the information of GO, KEGG and protein domains)

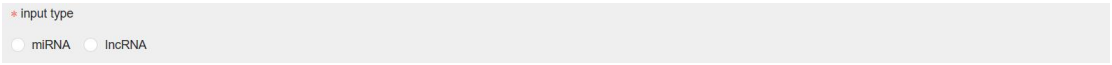
For ncRNA annotation

These functions in 'Annotation-miRNA-lncRNA', which were used for identifying which ncRNAs has been known.

1. Input fasta-format sequences



2. Select ncRNA types



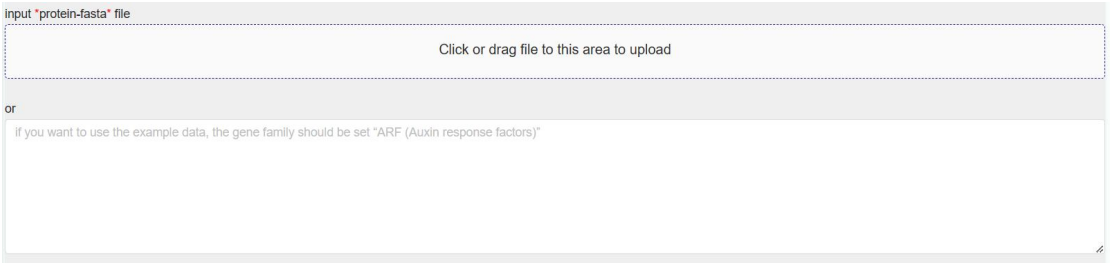
3. Press 'annotate' button to finish this process

For gene family annotation

Functions in 'Annotation-gene families', which can be utilized for identifying members of a selected gene family, and can be used for annotating genes with the domains and names of gene families.

If you want to identify members of a selected family

1. Input fasta-format protein sequences



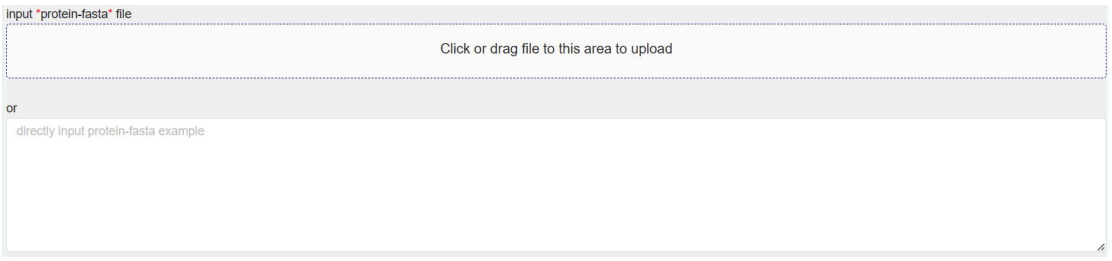
2. Select an interested gene family

3. Press the button of "show members of a single family" to finish this process

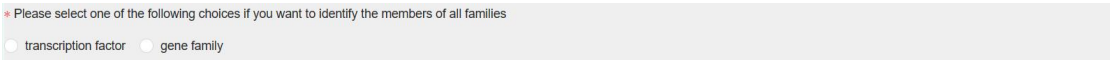
In addition to output member IDs of the selected gene family, the related sequences will also be provided simultaneously.

If you want to annotate genes with the information of gene families

1. Input fasta-format protein sequences



2. Select an annotation type



3. Press the button of 'show genes containing domains of families'

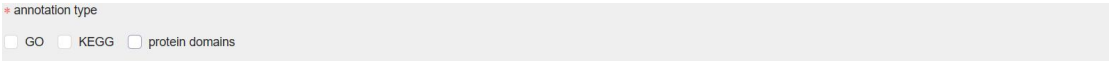
For drawing annotation results

If you want to draw annotation results using bar chart and heatmap

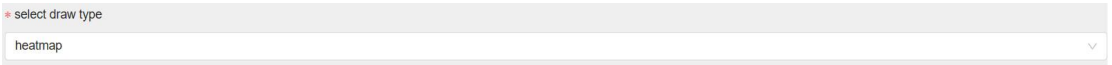
Functions in 'Draw-statistics' for showing annotation results using bar chart and heatmap

1. Input annotation results

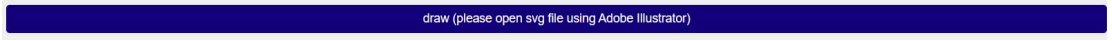
2. Select annotation type



3. Select draw type ('bar_chart' or 'heatmap')



4. Press 'draw (please open svg file using Adobe Illustrator)' to draw results

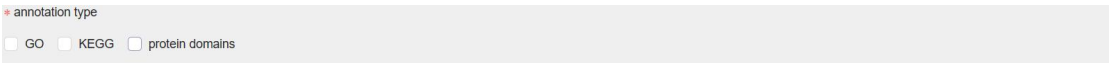


If you want to draw annotation results using bubble chart and pathway

Functions in 'Draw-pathway' for showing annotation results using bubble chart and pathway

1. Input annotation results

2. Select annotation type



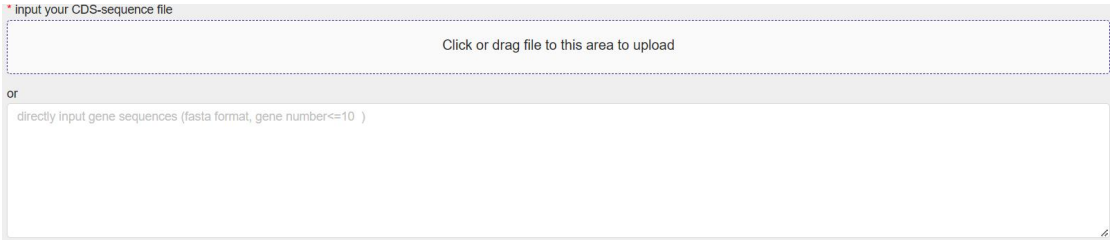
3. Select the model of closely related species

4. Press 'draw (please open svg file using Adobe Illustrator)' to draw results

For translate coding sequences into protein sequences

Function in 'Others-translation'

1. Input the coding sequences

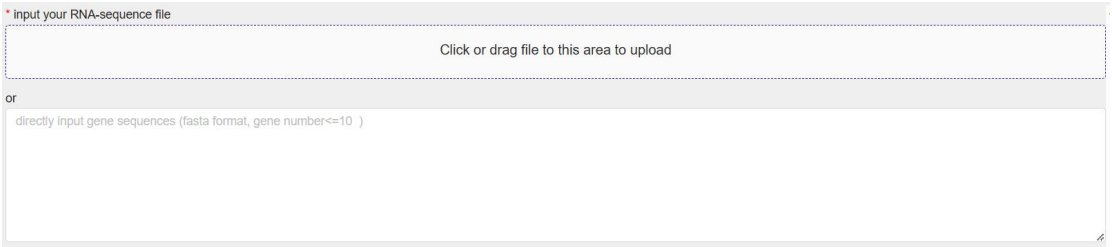


2. Press the button of 'translate' for finishing this function

For converting RNA to DNA sequences

Function in 'Others-RNA2DNA'

1. Input fasta-format RNA sequences



2. Press the 'translate' to convert RNA to DNA

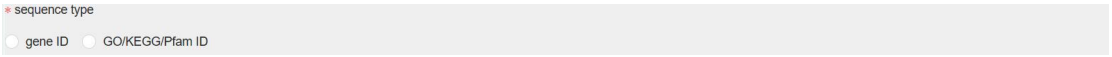
For extracting information from input files

Functions in 'Others-extraction'

If you want to extract interested contents of annotation results based on input IDs

1. Input annotation results and IDs

2. Select ID type (GFAP will extract contents based on the selected ID type, one ID per line)



3. Press the button of 'extract' to run the function

If you want to extract coding sequences of transcripts

1. Input fasta-format transcripts

2. Press the button of 'extract coding sequence' to run this function.

For converting format of transcriptome results

Function in 'Others-conversion'

Traditional format of transcriptome results is not conducive to downstream analysis, and this function will extract your interested contents from transcriptome results to organize them into new format that can be analyzed by GFAP and other downstream analysis tools.

1. Input transcriptome results (the format will be showed by clicking 'using example data')

2. Input indexes in the transcriptome results of the following contents

3. Press the button of 'conversion' to finish the function

NOTE:

1. If you want to annotate transcripts (not coding sequences), you can utilize the function in 'Others-extraction-Extract Coding Sequences from Transcripts' to extract possible coding sequences from transcripts, and then run annotation functions to finish annotation processes. Why the extracted sequences is the 'possible' coding sequences? In addition to non-coding regions, some introns may also be contained in transcripts because of alternative splicing. Therefore, different coding sequences may be predicted from a same transcript.

Furthermore, as limited server size of our website, ten transcripts can be analyzed at one time using GFAP website version. If you want to analyze large amounts of transcripts, you can utilize Linux version of GFAP.

2. If you want to annotate genes with the information of gene families, the input file or sequences should be organized by fasta format, and the contained sequences should be protein sequences. If you have not proteins sequences, you can utilize the translation function to convert DNA to proteins (in 'other-translation').
3. All input sequences should be organized as fasta format.
4. We encouraged users to organize their data according to the format of examples, and run functions using examples before performing GFAP functions.