A Five Minutes Crash Course for using PMGmap

Last updated on October 9th, 2023

PMGmap can draw genes at exon levels, draw cis- and trans-splicing gene maps, draw non-coding features, draw repetitive sequences, scale the genic regions using a scaling the genic regions on the genome (SGM) algorithm, and draw multiple chromosomes simultaneously. The web server can be accessed at http://www.1kmpg.cn/pmgmap. This help document will help you use the PMGmap website quickly and correctly.

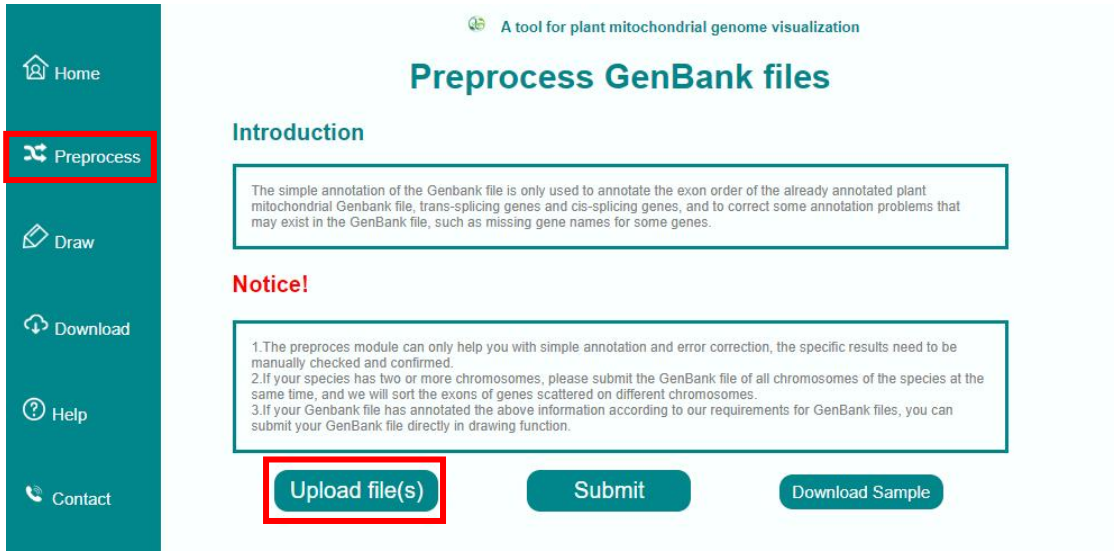
Authors: Xinyi Zhang1, Haimei Chen1, Yang Ni1, Bin Wu1, Jingling Li1, Artur Burzyński 2 Chang Liu1*

Sample inputs and output website link (Arabidopsis thaliana, Clematis acerifolia)
: http://47.96.249.172:16086/drawing/download_sample

Notice: If your species has two or more chromosomes, please submit the GenBank file of all chromosomes of the species at the same time.

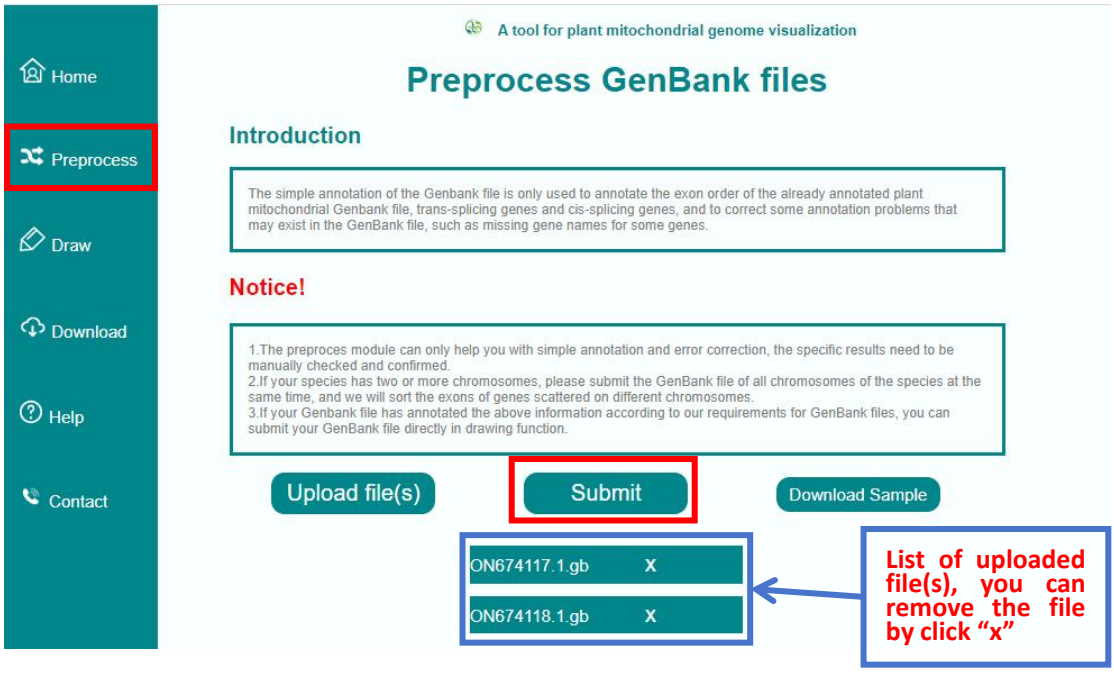
Step1:

In the module "Preprocess", select the GenBank file(s) that needed to upload



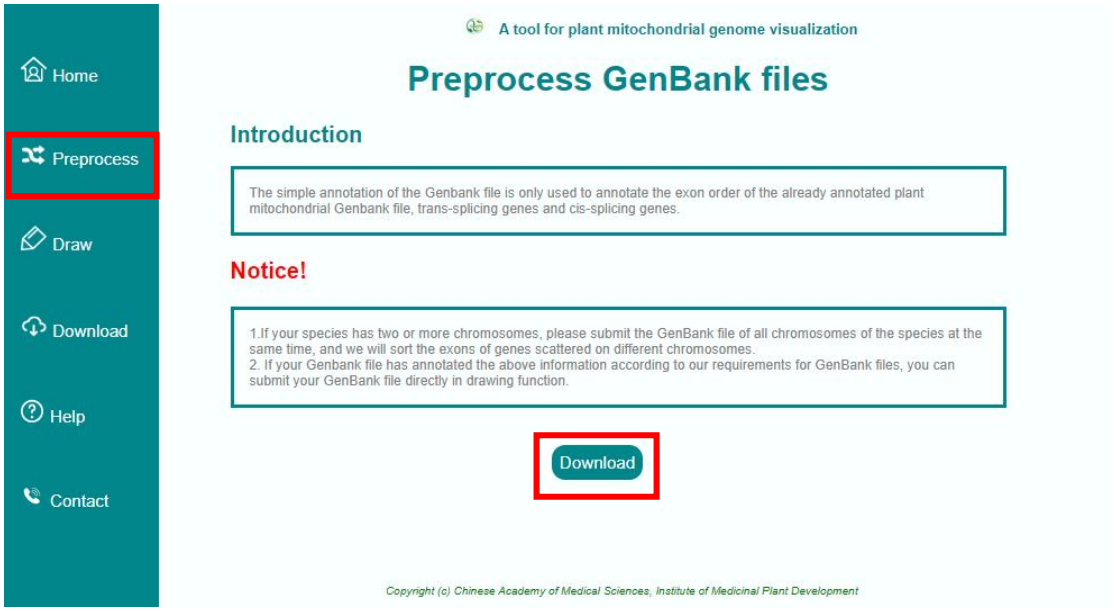
Step2:

In the module “Preprocess”, submit the uploaded GenBank files



Step3:

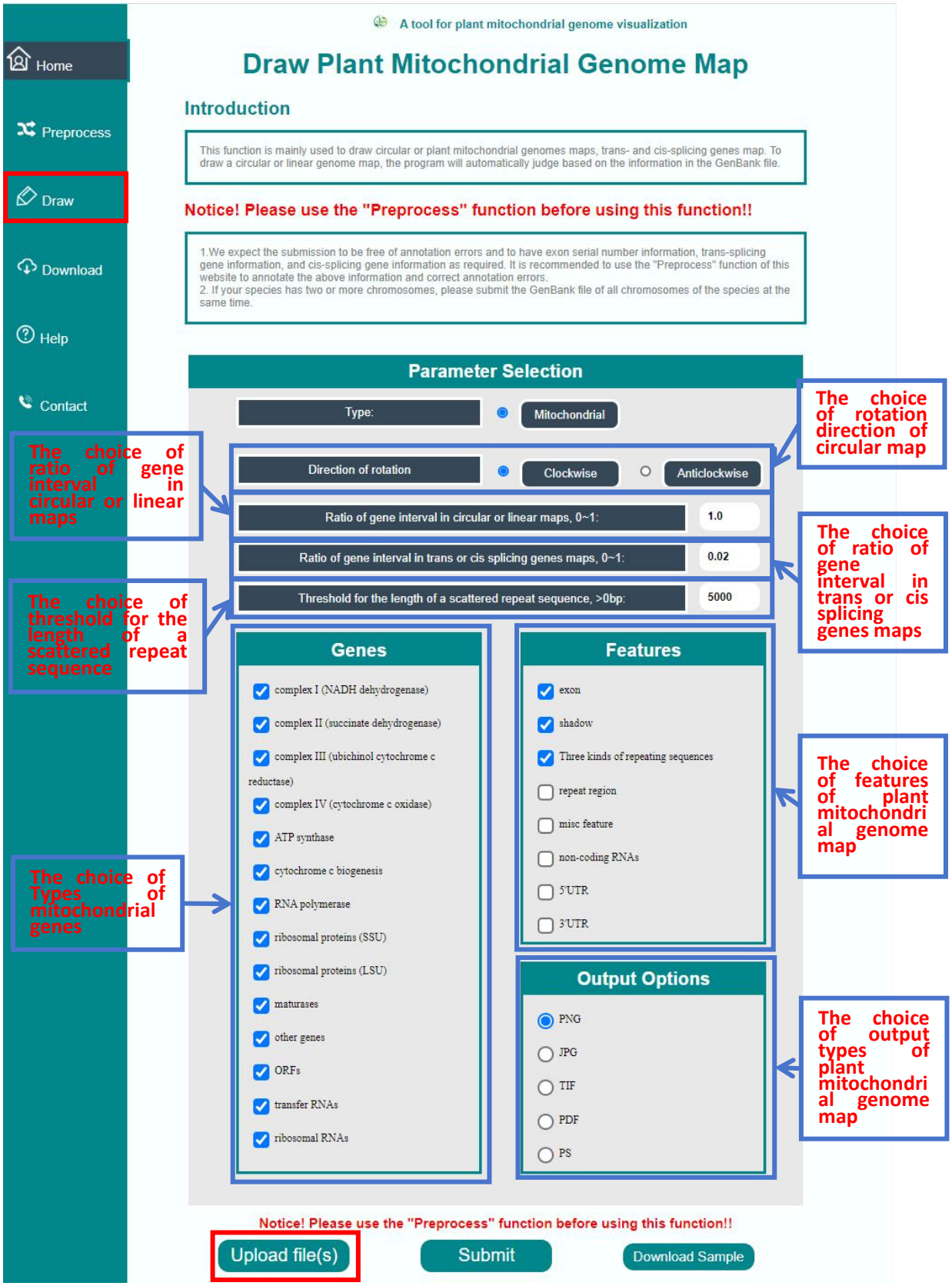
Download files processed by “Preprocess”. In this step, PMGmap returns the same number of GenBank files as were submitted.



Step4:

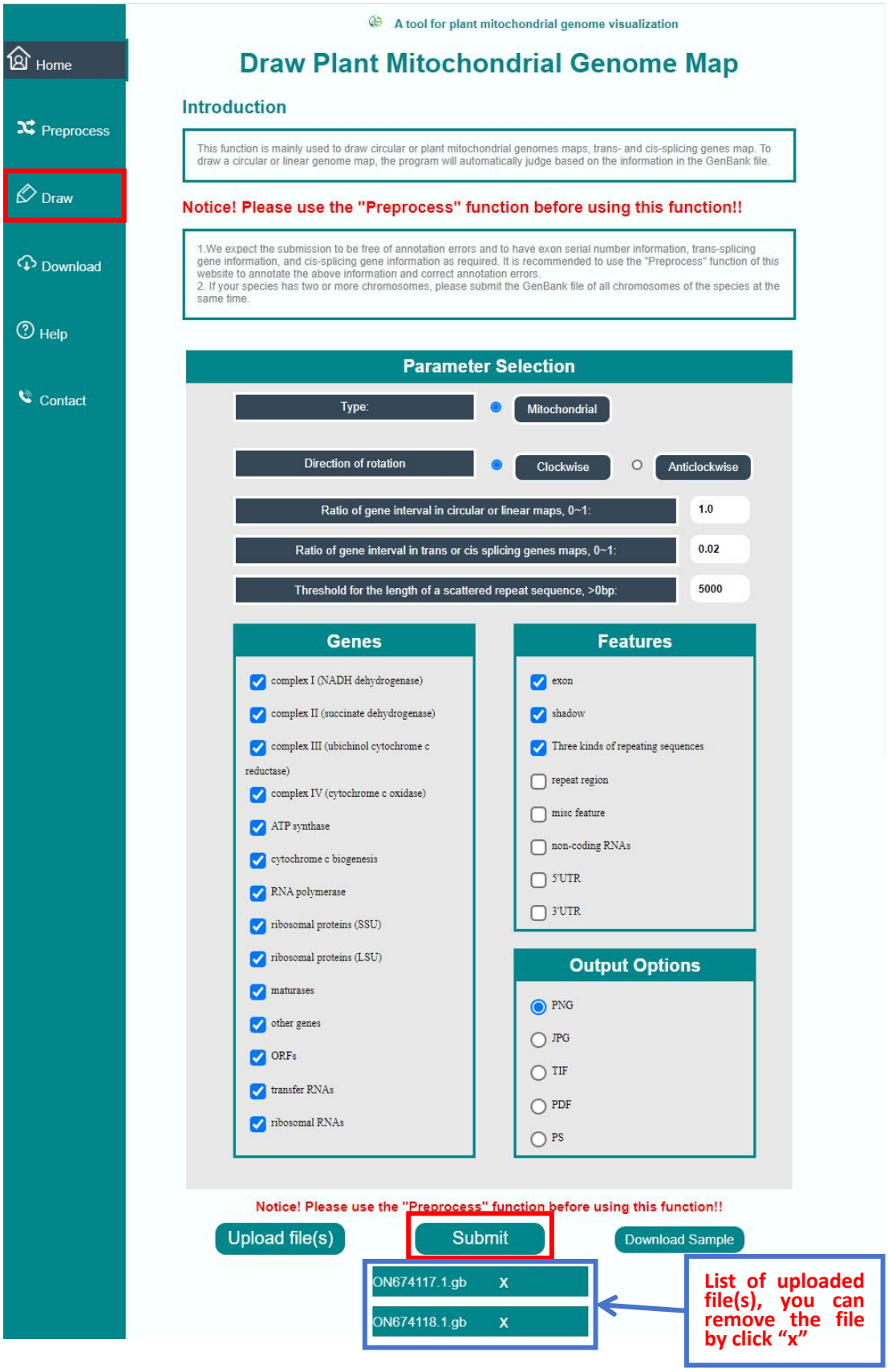
In the module "Draw", submit the GenBank files processed by "Preprocess" or the GenBank files with exon and splicing gene annotation information added.

User can specofy four types of parameters: 1.general parameters; 2.Genes; 3.Features; 4.Output image format.


A tool for plant mitochondrial genome visualization
Home
Preprocess
Draw
Download
Help
Contact
Draw Plant Mitochondrial Genome Map
Introduction
This function is mainly used to draw circular or plant mitochondrial genomes maps, trans- and cis-splicing genes map. To draw a circular or linear genome map, the program will automatically judge based on the information in the GenBank file.
Notice! Please use the "Preprocess" function before using this function!!
1.We expect the submission to be free of annotation errors and to have exon serial number information, trans-splicing gene information, and cis-splicing gene information as required. It is recommended to use the "Preprocess" function of this website to annotate the above information and correct annotation errors.
2. If your species has two or more chromosomes, please submit the GenBank file of all chromosomes of the species at the same time.
Parameter Selection
Type:
Mitochondrial
Direction of rotation
Clockwise
Anticlockwise
Ratio of gene interval in circular or linear maps, 0~1:
1.0
Ratio of gene interval in trans or cis splicing genes maps, 0~1:
0.02
Threshold for the length of a scattered repeat sequence, >0bp:
5000
Genes
complex I (NADH dehydrogenase)
complex II (succinate dehydrogenase)
complex III (ubichinol cytochrome c reductase)
complex IV (cytochrome c oxidase)
ATP synthase
cytochrome c biogenesis
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
maturases
other genes
ORFs
transfer RNAs
ribosomal RNAs
Features
exon
shadow
Three kinds of repeating sequences
repeat region
misc feature
non-coding RNAs
5'UTR
3'UTR
Output Options
PNG
JPG
TIF
PDF
PS
Notice! Please use the "Preprocess" function before using this function!!
Upload file(s)
Submit
Download Sample
The choice of rotation direction of circular map
The choice of ratio of gene interval in circular or linear maps
The choice of ratio of gene interval in trans or cis splicing genes maps
The choice of threshold for the length of a scattered repeat sequence
The choice of features of plant mitochondrial genome map
The choice of Types of mitochondrial genes
The choice of output types of plant mitochondrial genome map

Step5:

In the module “Draw”, select the parameters that are needed, submit the uploaded GenBank files



Step6:

Download files processed by “Draw”. In this step, PMGmap returns the same number of plant mitogenome feature maps as the commit, a cis-splicing gene map, a trans-splicing gene map, and some process files.

A tool for plant mitochondrial genome visualization

Results

The genome map

ON674117_linear.png
ON674118_circular.png

For a circular genome map, it is represented from inside out: (1) the relationship between dispersed repeat sequences; (2) the distribution of dispersed repeat sequences on the chromosome, where yellow represents direct dispersed repeat sequences and green represents inverted dispersed repeat sequences; (3) the distribution of tandem repeat sequences on the chromosome; (4) the distribution of tandem repeat sequences on the chromosome; (5) the distribution of GC content on the chromosome; (6) the scale coordinate axis; (7) genes located on the negative strand; (8) genes located on the positive strand; (9) Orange shadows represent forward scattered repeats that are greater than the selection threshold; (10) blue shadows represent reverse scattered repeats that are greater than the selection threshold
For linear genome map, from top to bottom: (1) genes located on the sense strand; (2) the distribution of microsatellite repeat sequences on the chromosome; (3) the distribution of tandem repeat sequences on the chromosome; (4) the distribution of dispersed repeat sequences on the chromosome, where yellow represents direct repeat sequences and the green represents inverted repeat sequences; (5) the association between dispersed repeat sequences; (6) the distribution of GC content on the chromosome; (7) the scale coordinate axis; (8) genes located on the negative strand; (9) Orange shadows represent forward scattered repeats that are greater than the selection threshold; (10) blue shadows represent reverse scattered repeats that are greater than the selection threshold

The cis-splicing gene map

cis_splicing_genes.png

This diagram illustrates the distribution of all cis-splicing genes across one or more chromosomes, along with the composition of their mRNAs. The map includes this information: (1) the gene name; (2) the GenBank accession number of the corresponding chromosome; (3) the start and end positions of each exon; (4) the positive strand; (5) the negative strand; (6) individual exon on the chromosome; (7) the exon number; (8) the exon name and length; (9) the exon arrangement on the mRNA.

The trans-splicing gene map

trans_splicing_genes.png

This diagram illustrates the distribution of all trans-splicing genes across one or more chromosomes, along with the composition of their mRNAs. The map includes this information: (1) the gene name; (2) the GenBank accession number of the corresponding chromosome; (3) the start and end positions of each exon; (4) the positive strand; (5) the negative strand; (6) individual exon on the chromosome; (7) the exon number; (8) the exon name and length; (9) the exon arrangement on the mRNA.

Short Tandem Repeats

ON674118/Repetitive_sequence/123.fas.misa
ON674117/Repetitive_sequence/123.fas.misa

The STR/SRR sequences were identified using the following parameters: "1-10 2-6 3-5 4-5 5-5 6-5". The number before "-" is the size of the repeat unit. The second number after "-" is the minimum numbers of the repeat units. The identified STR sequences are shown in a table containing seven columns.
Column 1: the name of the reference genome submitted;
Column 2: the number of SSRs;
Column 3: the type of SSRs;
Column 4: the unit of the SSRs and the number of repeats;
Column 5: the total length of the SSRs;
Columns 6,7: the starting and ending points of the SSRs on the genome, respectively.

Long Tandem Repeats

ON674118/Repetitive_sequence/123.fas.2.7.7.80.10.50.500.dat
ON674117/Repetitive_sequence/123.fas.2.7.7.80.10.50.500.dat

The Long Tandem Repeats were identified using the following parameters: "2 7 7 80 10 50 500 -f -d -m". The results are shown in a table containing 15 columns.
Column 1: the repeat indices of the repeat relative to the start of the sequence;
Column 2: the period size of the repeat;
Column 3: the size of the repeat unit;
Column 4: the number of copies aligned to the consensus pattern;
Column 5: the size of the consensus pattern (may differ slightly from the period size);
Column 6: the overall percentage match between adjacent copies;
Column 7: the percent of indels between adjacent copies overall;
Column 8: the alignment score;
Columns 9,10,11,12: the percent composition for each of the four nucleotides 'A', 'C', 'G' and 'T';
Column 13: the entropy measure based on the composition percentage;
Column 14: the sequence of the repeat unit;
Column 15: the sequence of the entire repeat.

Dispersed Repeats

ON674118/Repetitive_sequence/123_vmatch.txt
ON674117/Repetitive_sequence/123_vmatch.txt

The dispersed repeats were identified using the following parameters: "-f -p -h 3 -l 30". The results of the dispersed repeats were shown in a table containing seven columns.
Column 1: the length of the left instance of the match;
Column 2: the relative position of the left instance of the match in sequence;
Column 3: the repeat type (D: direct matches; P: palindromic matches);
Column 4: the length of the right instance of the match;
Column 5: the length of the right instance of the match; The distance of the match;
Column 6: an exact match with a distance of 0. A k-mismatch match with k > 0;
Column 7: the E-value of the match.

The output files and the intermediate files in zip format

20231007095619_45.zip

Thank you for reading the help document. If you have any further questions or comments, please contact cliu@implad.ac.cn.