GMATA User's Guide

Genome-wide Microsatellite Analyzing Toward Application

(GMATA)

Version 2.0

Nov 2013

Scripts written by: Xuewen Wang, Ph.D

Manual prepared by: Meng Li (Msc),   Xuewen Wang , Ph.D

Contents

1 Overview

1.1 What is GMATA

Genome-wide Microsatellite Analyzing Toward Application (GMATA) is a software for Simple Sequence Repeats (SSR) analyses, and SSR marker designing and mapping in any DNA sequences. It has the following functions:

1. Accurate and fastest SSR mining in any large sequences;
2. Complete statistical analysis and plotting;
3. SSR loci and marker graphic displaying in Gbrowser with genome features;
4. Specific SSR marker designing, and simulated PCR;
5. Electronic mapping, and marker transferability investigation.

GMATA is accurate, sensitive and fast. It was designed to process large genomic sequence data sets, especially large whole genome sequences. In theory, genomes of any size can be analyzed by GMATA easily. Software GMATA works on sever, desktop or even laptop, and it can run in graphic interface with just clicks or run in command line or in automated pipeline. It is also cross-platform and supports Unix/Linux, Win and Mac. Results from software GMATA can be directly graphically displayed with genome or gene features in Gbrowser and easily integrated with any genomic database.

1.2 Workflow of GMATA

The following chart shows the workflow of GMATA.

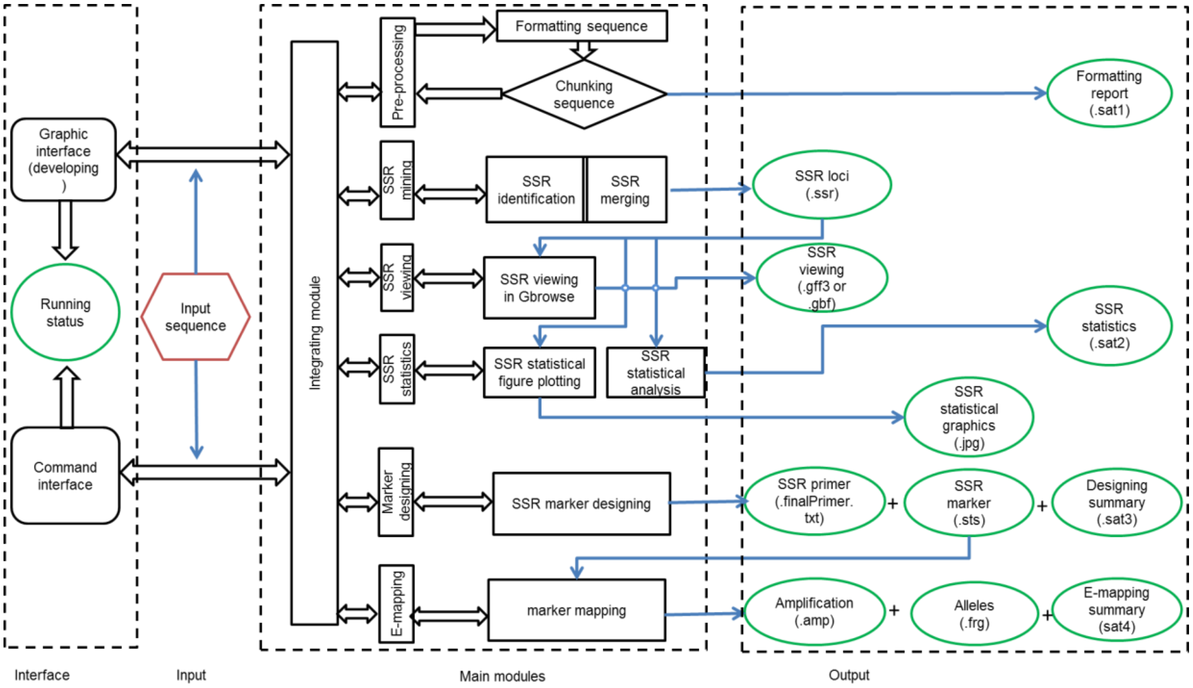


Figure 1. Workflow of GMATA.

2 Installation

2.1 Installation instructions

Download GMATA2.0.tar.gz from disk or web http://sourceforge.net/p/GMATA;

Uncompress this file:

For linux:

```
>tar -zxf GMATA2.0.tar.gz;
```

For win, simply extract the files with 7zip or WinRAR;

Keep all those files in the same directory. Then GMATA programs can execute directly without any further installing.

2.2 Hardware and software Environment

2.2.1 Hardware environment

The recommended requirements for the software GMATA is minimum RAM 1G, CPU 1G or higher. A minimum 5M disk space is needed. The software GMATA will use around 1.5x memory of the size of the input sequence files.

2.2.2 Software environment

The software GMATA was programmed in computing language Perl 5 , R 2.0 and Java 7.0 . The release GMATA is in the source code so the running needs Perl running environment, R environment and Java running environment.

Since "GMATA" relies on Perl, R and Java to work, you need to install them first. See these pages for download, <http://www.activestate.com/> (Perl for windows), <http://www.r-project.org/> (R), <http://www.java.com/> (Java).

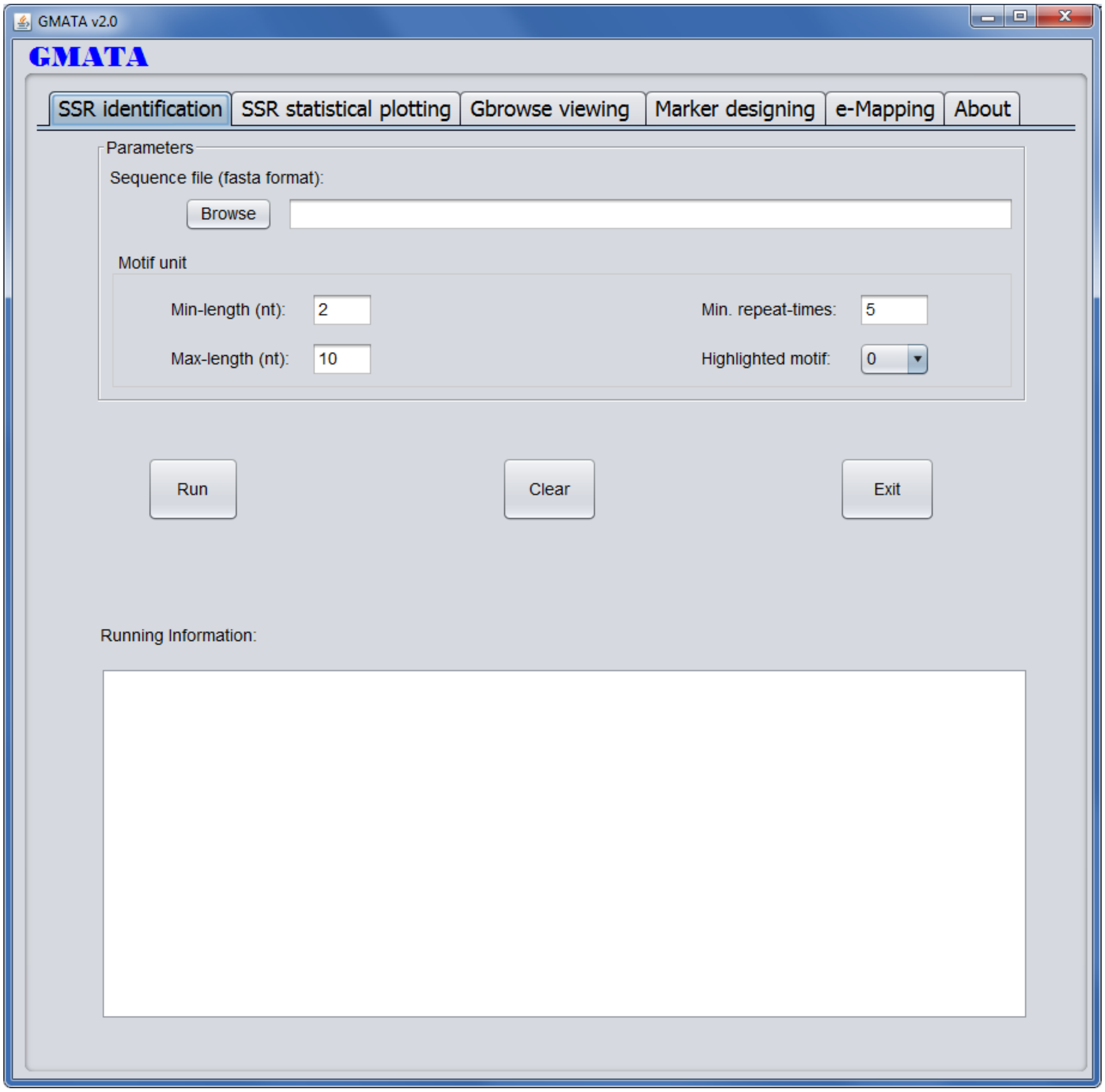
NOTE: Make sure Perl, R and Java have been added to system Environment Variables before starting "GMATA".

"GMATA" works on any computers that can run Perl (version 5, higher than 5.8), R (version 2.5 or higher) and Java (version 7 or higher) correctly. For a better experience, it is strongly recommended to update Perl, R and Java to the latest version.

This software has been verified in Win XP, 7, 8, MS DOS 6.1, Mac OS X 10.7 or higher, Unix/Linux 5.5 or higher. GMATA may work well in other platforms. Both 32-byte and 64-byte system are accepted by GMATA.

2.3 User-interfaces

GMATA has two distinct user-interfaces. It runs in either command line interface or graphic interface.



Min-length (nt): 2

Max-length (nt): 10

Min. repeat-times: 5

Highlighted motif: 0

Figure 2. Graphic interface of GMATA

The graphic interface (Figure 2) is user-friendly and is of easy to use. See part 3.2 for detailed guidelines.

The command-line interface is recommended for advanced users that are familiar with GMATA options or for running GMATA in batch mode.

2.3.1 Instruction for running GMATA in command line

GMATA can be invoked in a single command. For the running in command interface, just type the following command in the console of Win/DOS, MAC or Unix/Linux:

```
perl gmata.pl -c config_file -i SequenceFileName
```

GMATA can also be invoked in multiple steps in sequential. After these steps, the same results can be reached as in single step of gmata.pl. The command for each step is listed below.

Step 1 (SSR identification):

```
perl gmat.pl -i SequenceFileName [options]
```

Step 2 (SSR statistical plotting):

```
perl ssrfig.pl –i SequenceFileName.ssr.sat2 [options]
```

Step 3 (Gbrowse viewing):

```
perl ssr2gff.pl –i SequenceFileName.ssr
```

Or:

```
perl ssr2gff3.pl –i SequenceFileName.ssr
```

Step 4 (Marker designing):

```
perl doprimer_smt.pl –i SequenceFileName –sr SequenceFileName.ssr
[options]
```

Step 5 (e-Mapping):

```
perl elctPCR.pl –i SequenceFileName –mk SequenceFileName.
ssr.finalPrimer.txt.sts [options]
```

2.3.2 Tips for setting parameters in command line

Above *[options]* are the parameters which user can set the parameter-value pair. The parameter must be given in pair and the order of each parameter-value pair (e.g. -r 5, separated by space) can be in any order. For example:

```
perl gmat.pl -i SequenceFileName -r 5 -m 2 -x 6
```

```
perl gmat.pl -i SequenceFileName  -m 2 -x 6 -r 5
```

The above two commands are the same despite different orders of parameter pair.

The *[options]* can be omitted if users do not want to modify the default value. For example:

```
perl gmat.pl -i SequenceFileName -r 5 -m 2
```

There are only two parameter-value pairs in above command, which means the other parameters will be set to default.

```
perl gmat.pl -i SequenceFileName
```

Without any options given, the programs will use the default value in above command.

3 Tutorial

The *Oryza sativa* genome data for this tutorial could be downloaded from ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Osativa/assembly/.

3.1 Command line interface

3.1.1 SSR identification

To search *Oryza sativa* genome data for SSR, type the following command in the terminal of DOS/windows, MAC or Unix/Linux:

```
perl gmat.pl –i Osativa_204.fa
```

This operation generates following files:

Osativa_204.fa.fms	--	Formatted sequences for SSR identification
Osativa_204.fa.fms.sat1	--	Statistic summary of input sequence(s)
Osativa_204.fa.ssr	--	Microsatellite data
Osativa_204.fa.ssr.sat2	--	SSR statistic results

3.1.2 SSR statistical plotting

Use the .sat2 file (Osativa_204.fa.ssr.sat2) from previous step as the input file:

```
perl ssrfig.pl -i Osativa_204.fa.ssr.sat2
```

This operation generates 12 figures ending with .jpg:

A1. K-mer distribution of motifs.jpg

A2. Top k-mers.jpg

B1. Motif distribution.jpg

B2. Top motifs.jpg

C1. Grouped motif distribution.jpg

C2. Top grouped motifs.jpg

D1. SSR loci distribution.jpg

D2. Top SSR loci on chromosomes.jpg

D3. Top SSRs frequency on chromosomes.jpg

D4. Relationship of SSRs vs length.jpg

E1. SSR length distribution.jpg

E2. Top distribution of SSR length.jpg

3.1.3 Gbrowse viewing

Use the .ssr file (Osativa_204.fa.ssr) from 3.1.1 as the input file:

```
perl ssr2gff.pl -i Osativa_204.fa.ssr
```

This comes out .gbf file (Osativa_204.fa.ssr.gbf). For .gff3 file, type:

```
perl ssr2gff3.pl -i Osativa_204.fa.ssr
```

The output file (.gbf/.gff3) can be viewed in Generic Genome Browser (GBrowse) for manipulating and displaying annotations. e.g., http://webgbrowse.cgb.indiana.edu/cgi-bin/webgbrowse/uploadData, http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/.

3.1.4 Marker designing

The input files for this program include the original data (Osativa_204.fa) and the output file .ssr (Osativa_204.fa.ssr) from 3.1.1.

```
perl doprimer_smt.pl -i Osativa_204.fa –sr Osativa_204.fa.ssr
```

This operation generates following files:

Osativa_204.fa.seq - Prepared template sequence for the primer designing

Osativa_204.fa.ssr.finalPrimer.txt - Detailed final primer information

Osativa_204.fa.ssr.finalPrimer.txt.sts - Marker's primer sequence and size information

Osativa_204.fa.ssr.finalPrimer.txt.sat3 - Statistic summary of this operation

3.1.5 e-Mapping

The input files for this program include the original data (Osativa_204.fa) and the output file .sts (Osativa_204.fa.ssr.finalPrimer.txt.sts) from 3.1.4.

```
perl elctPCR.pl -i Osativa_204.fa –mk
Osativa_204.fa.ssr.finalPrimer.txt.sts
```

This operation generates following files:

Osativa_204.fa.eMap - Mapped markers position on each chromosome

Osativa_204.fa.eMap.amp - Amplification information of each marker

Osativa_204.fa.eMap.frg - Number of alleles of each marker

Osativa_204.fa.eMap.frg.sat4 - Summary of ePCR/eMapping

3.1.6 Run this tutorial in all-in-one step

Just type the following command in terminal of Windows, MAC or Unix/Linux:

```
perl gmata.pl -c default_cfg.txt -i Osativa_204.fa
```

This will get all the results which are the same as above steps (3.1.1 - 3.1.5).

The file “default_cfg.txt” can be found in the “bin” directory.

3.2 Graphic interface

This section gives a simple introduction to the GMATA Graphic mode. For detailed information about this tutorial (e.g., input/output files, parameters), refer to section 3.1.

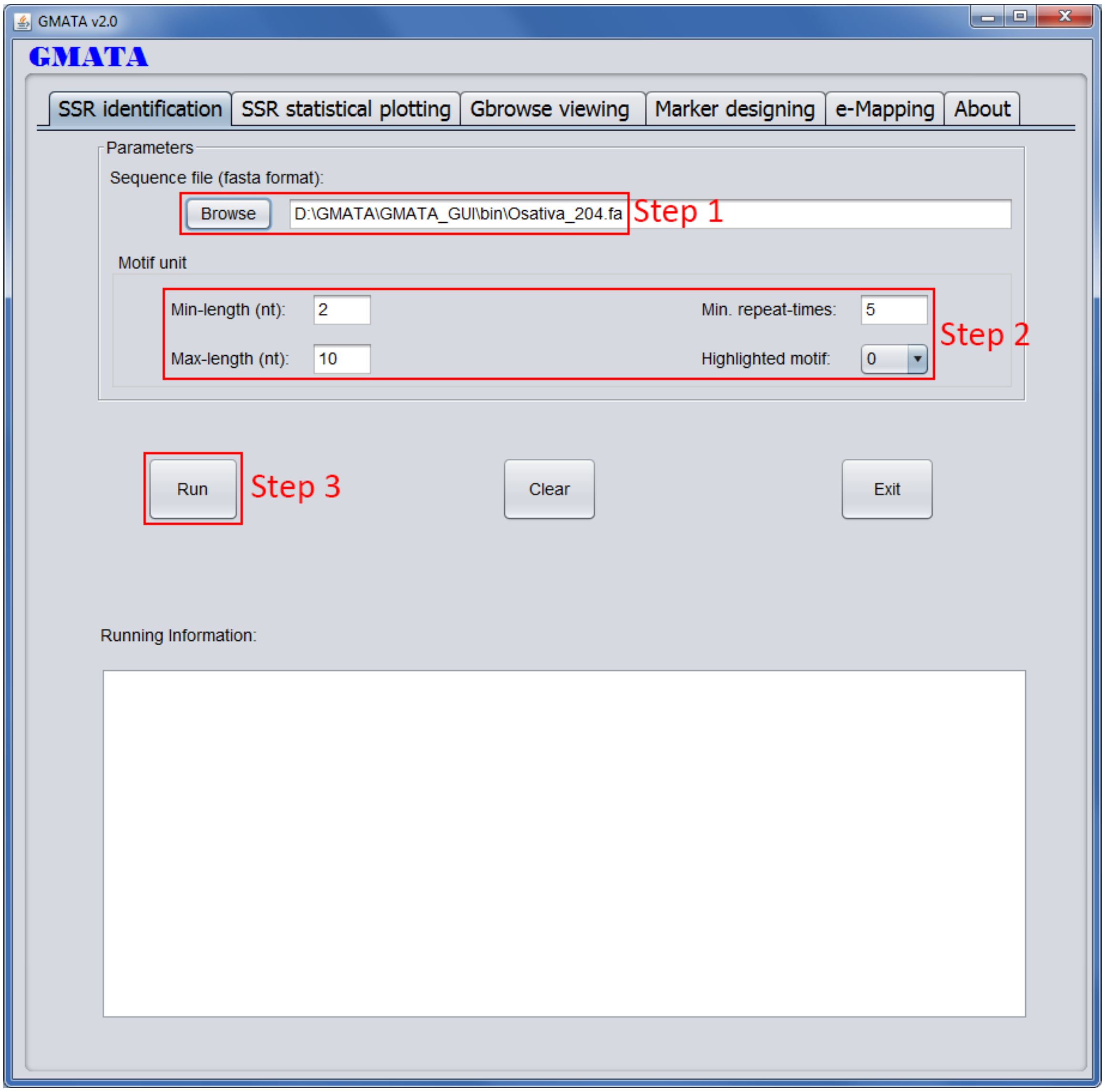
3.2.1 SSR identification

Launch GMATA and choose “SSR identification” tab (Figure 3);

Step 1, Click “Browse” to choose the input file;

Step 2, Set the parameters;

Step 3, Click “Run” to perform the analysis.



Parameters

Sequence file (fasta format): D:\GMATA\GMATA_GUI\bin\Osativa_204.fa

Motif unit

Min-length (nt): 2

Max-length (nt): 10

Min. repeat-times: 5

Highlighted motif: 0

Running Information:

Figure 3. SSR identification in graphic interface.

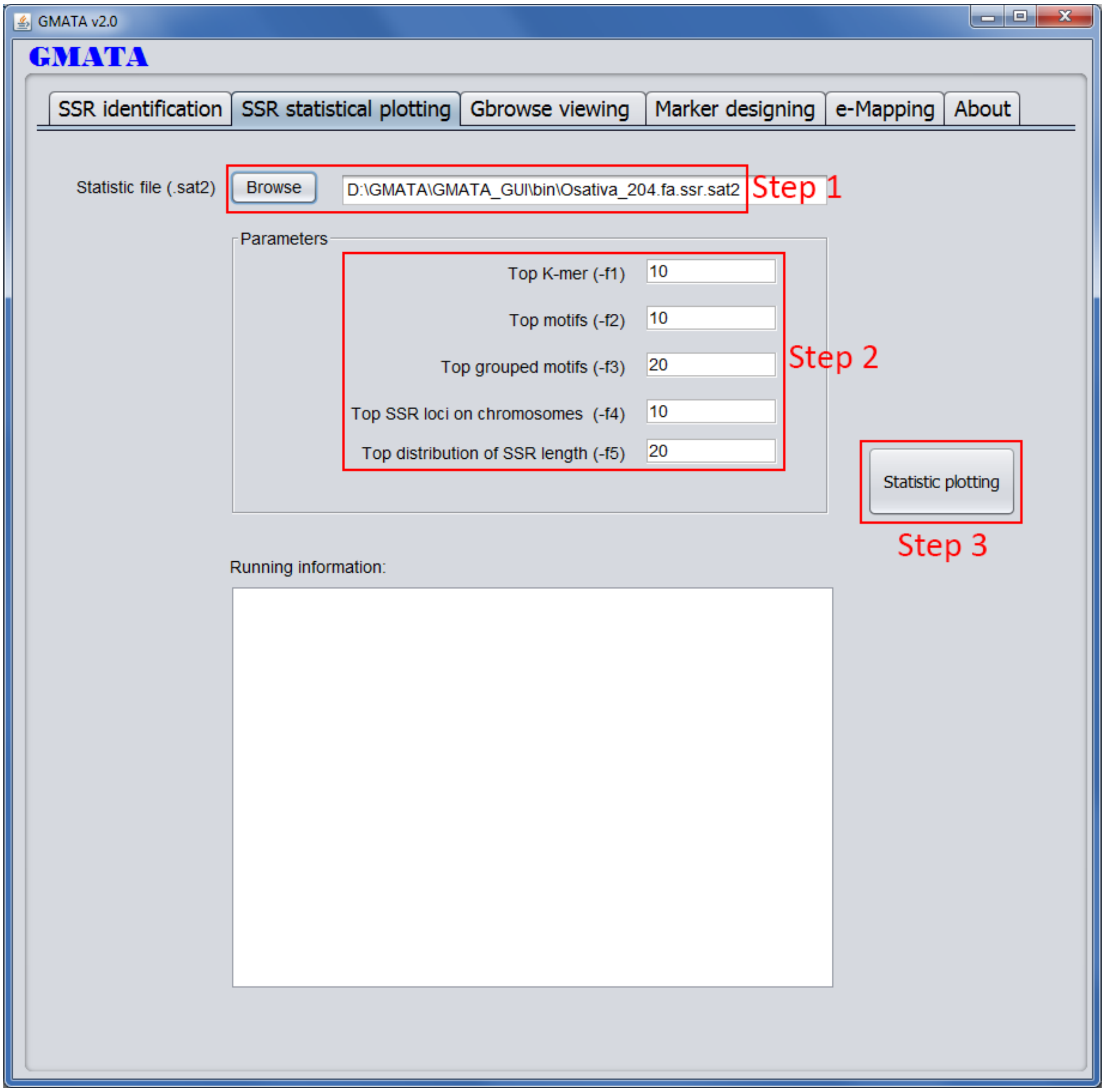
3.2.2 SSR statistical plotting

Choose “SSR statistical plotting” tab (Figure 4);

Step 1, Click “Browse” to choose the input file;

Step 2, Set the parameters;

Step 3, Click “Statistic plotting” to run.



Statistic file (.sat2): D:\GMATA\GMATA_GUI\bin\Osativa_204.fa.ssr.sat2

Top K-mer (-f1): 10

Top motifs (-f2): 10

Top grouped motifs (-f3): 20

Top SSR loci on chromosomes (-f4): 10

Top distribution of SSR length (-f5): 20

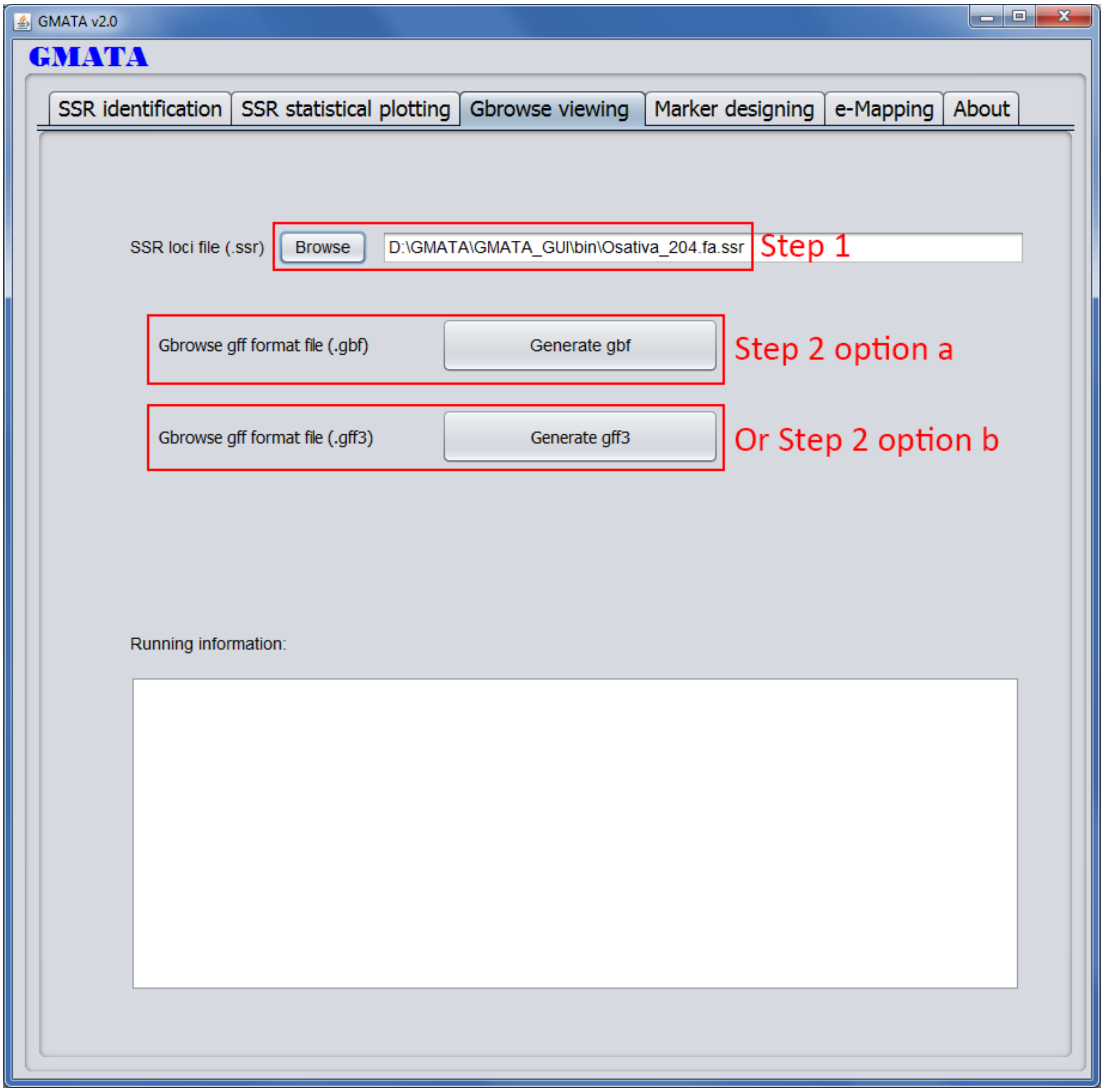
Figure 4. SSR statistical plotting in graphic interface.

3.2.3 Gbrowse viewing

Choose “Gbrowse viewing” tab (Figure 5);

Step 1, Click “Browse” to choose the input file;

Step 2, Click “Generate gbf” or “Generate gff3” to run.



SSR loci file (.ssr): D:\GMATA\GMATA_GUI\bin\Osativa_204.fa.ssr

Figure 5. Gbrowse viewing in graphic interface.

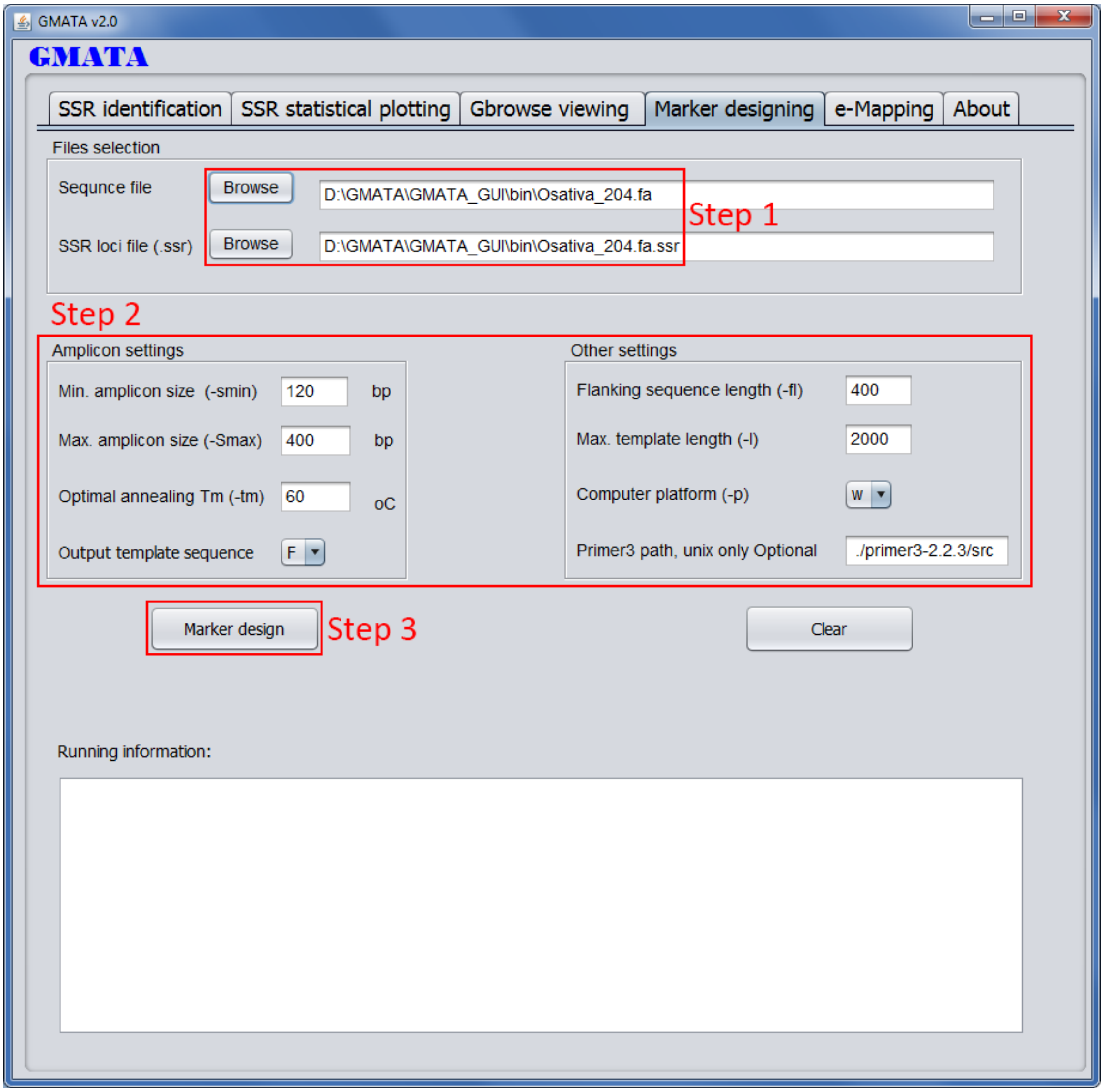
3.2.4 Marker designing

Choose “Marker designing” tab (Figure 6);

Step 1, Click “Browse” to choose the input files;

Step 2, Set the parameters;

Step 3, Click “Marker design” to run.



Files selection

Sequnce file D:\GMATA\GMATA_GUI\bin\Osativa_204.fa

SSR loci file (.ssr) D:\GMATA\GMATA_GUI\bin\Osativa_204.fa.ssr

Step 1

Step 2

Amplicon settings

Min. amplicon size (-smin) 120 bp

Max. amplicon size (-Smax) 400 bp

Optimal annealing Tm (-tm) 60 oC

Output template sequence F

Other settings

Flanking sequence length (-fl) 400

Max. template length (-l) 2000

Computer platform (-p) w

Primer3 path, unix only Optional ./primer3-2.2.3/src

Step 3

Running information:

Figure 6. Marker designing in graphic interface.

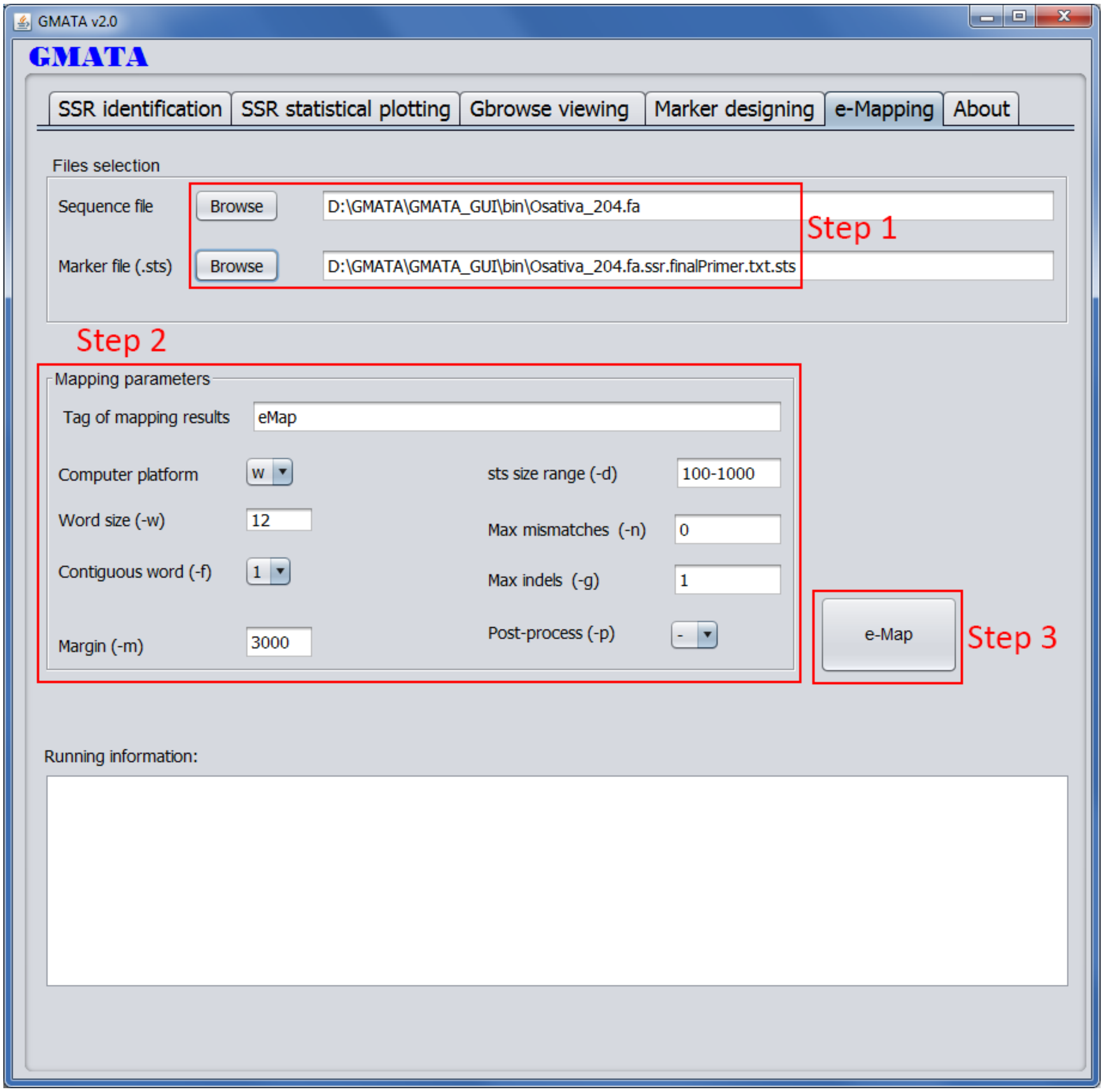
3.2.5 e-Mapping

Choose “e-Mapping” tab (Figure 7);

Step 1, Click “Browse” to choose the input files;

Step 2, Set the parameters;

Step 3, Click “e-Map” to perform the analysis.



Step 1

Sequence file: D:\GMATA\GMATA_GUI\bin\Osativa_204.fa

Marker file (.sts): D:\GMATA\GMATA_GUI\bin\Osativa_204.fa.ssr.finalPrimer.txt.sts

Step 2

Tag of mapping results: eMap

Computer platform: w

Word size (-w): 12

Contiguous word (-f): 1

Margin (-m): 3000

sts size range (-d): 100-1000

Max mismatches (-n): 0

Max indels (-g): 1

Post-process (-p): -

Step 3

Figure 7. e-Mapping in graphic interface.

4 Manual pages

4.1 SSR identification

Description

The module *gmat.pl* is a better and faster tool for Simple Sequence Repeats (SSR), also called microsatellite, characterization in any huge genome and for facilitating SSR marker designing in genome scale.

Usage

```
perl gmat.pl -i SequenceFileName [options]
```

e.g.:

```
perl gmat.pl -i Osativa_204.fa
```

Options "i" defines the name of input sequence file (plain text file, similar to fasta format). The following is the detailed description of the parameters.

Options

-r Minimum repeated times (Repeat-times) of motif, integer value, >0, the default is 5.

-m Minimum length (Mini-length) of motif, integer value, >0, the default is 2.

-x Maximum length (Maxi-length) of motif, the default is 10, value of x greater or equal to value of m, if value x= value m, mining motif at only this given length, meaning one length of motif.

-s Highlighting repeat sequence or not, the value should be 0 or 1.

1:display original DNA sequence in lowcase and repeated sequence in upcase;

0:no flanking sequence output in SSR loci file.

Inputs

The input file should be in plain text format. Sequences downloaded from the NCBI or a genome sequence database can be directly used as an input file. If the sequences come from experiment, you should create the file in the described below. The format of input file is similar to fasta format, beginning with a ">" sign for the sequence name in one line identified by line return sign and then followed by one or multiple lines of DNA sequences. What is different for GMATA from standard fasta format is that the DNA sequence may be ATGC or any nucleotide letter, number, space, table and empty lines, and line return, which is very useful for some DNA format from web page such as Arabidopsis web TAIR.

e.g.:

```
>MethylFiltered.Decon.masked_contig_365|3276:3524|249/249-249
CACCCAAACGGTCCTCTACAGACACAGAATTCCAAGAATGCCCTTGCCTCAGCCTAGTCGAACTCGTCCTC
GAACTCTGATTATTCTTTATAGGGTTTGGTAAATTAGGATCTTTCAAAGGGTATTTCTGTCTTTTCCTTTC
CTGGTTTGGTTTCTCTAACCTTCGACTCCGCAGCTGAAGGTAGGACCCACCACCACCACCACCCTCACCGC
TTCCGTTTCCGGTTGAGTTCGATTTCTGCAAGCGTT
>MethylFiltered.Decon.masked_contig_366|1628:1882|255/249-249
CACCCAAACGCTCCTCTACAGACCCAGAATTCCAAGAATGCCCTTTCCTCAGCCTAGTAGAACTCGTCCTC
GAACTCTGATTATTCTTTACAGGGTTTGGTAAATTAGGATCTTTCAAAGGGTATTTCTGTCTTTTCCCTTC
ATGGTTTGGTTTCTCTAACCTTCGGCTCCGCAGCTGAAGGTAGGACNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNGCTTCCGTTTCCGTTTGAGTTCGATTTCTGCAGGCGTT
>MethylFiltered.Decon.masked_contig_479|1595:1901|307/335-335
GAGAAACAACTCCCACAAACACAAACTAAGATCTTGTAACCTTAAAATGGTGATTTACAGAAAATGCTGTT
GTTGTAATTGGTGATTATTTTTTTGGGGGGGGGTGTGGGGAGGGGGGAGGGGTCTTTGTTTTTGTGAATTG
```

```
GAAGTGAAAAGGTCTGAACTGTGAAGAGGGGCTACACACACTCTCTCTCTACTTCTCTGTTTTTTCAAA
TCAAAAATTGCAGTGGGGGAGGTTTTTGCAGCTTTGGGAGAGGTTTTGGGGGTGTTTTTACAATTTCTTAA
AGTTGATGTGCCAGCGATATTCA
```

Outputs

By default, the SSR loci file has the same name as the input file plus suffix .ssr in the end of file name. There are 6 columns of data with clear title names including source sequence name, source sequence length, repeated sequence starting position, repeated sequence ending position, repetitions (repeated times) and motif of repeated sequence.

SSR distribution statistical file:the SSR distribution statistical file has the same name as the input file plus suffix .sat2 in the end of file name. The format of this output file is in tabular plain text. The statistical distribution file provides four different types of classification and statistical results. A summary is generated in the end of each classification. The last two classifications are specialized for whole genome sequence.

There are 4 files provided by this step:

*.fms	-	Formatted sequences for SSR identification
*.fms.sat1	-	Statistic summary of input sequence(s)
*.ssr	-	Microsatellite data
*.ssr.sat2	-	Statistic results

4.2 SSR statistical plotting

Description

The statistic data (.suffix sat2) is produced -in “SSR identification” section by default and this statistic plotting will generate SSR statistical graphics for the identified SSRs.

Usage

```
perl ssrfig.pl –i SequenceFileName.ssr.sat2 [options]
```

e.g.:

```
perl ssrfig.pl -i Osativa_204.fa.ssr.sat2 -f1 10 -f2 20 -f3 20 -f4 10 -f5 20
```

Options "i" defines the name of input file, default is the ssr.sat2 file generated previously, or you can use files with similar format with .sat2.

Options

-f1 The number of top k-mers to be displayed in bar chart (integer value). Default value is 10.

-f2 The number of top motifs to be displayed in bar chart (integer value). Default value is 20.

-f3 The number of top grouped motifs to be displayed in bar chart (integer value). Default value is 20.

-f4 The number of top SSR loci on chromosomes to be displayed in bar chart (integer value). Default value is 10.

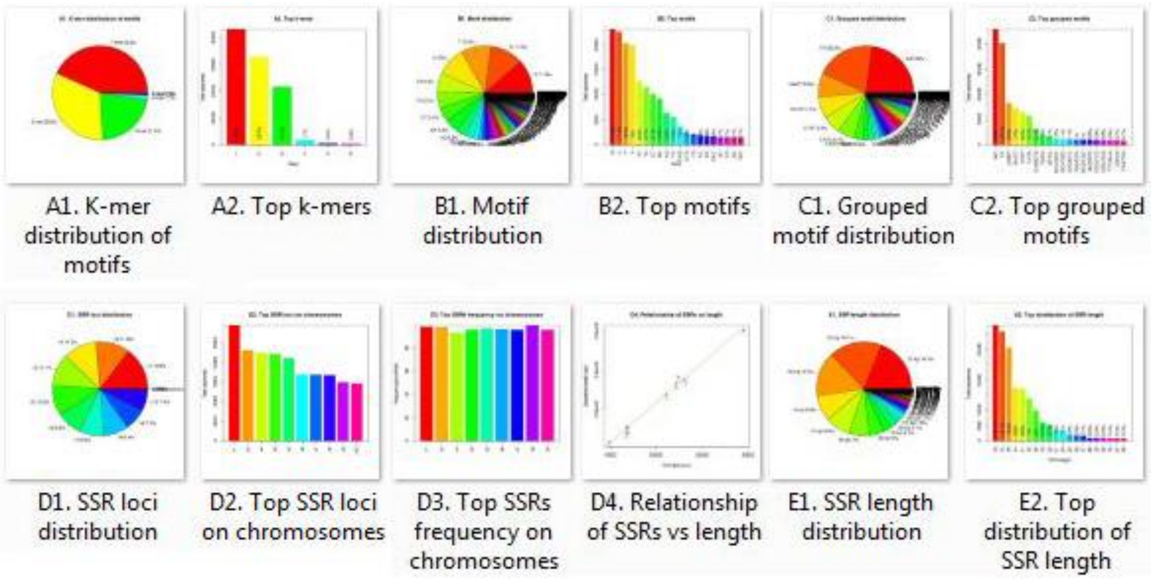
-f5 The number of top distribution of SSR length to be displayed in bar chart (integer value). Default value is 20.

Inputs

Use the .sat2 file from “SSR identification” (part 4.1) as the input file. You can also use other files whose formats are similar to .sat2.

Outputs

This operation generates the statistical plotting (Figure 8), including A1. K-mer distribution of motifs, A2. Top k-mers, B1. Motif distribution, B2. Top motifs, C1. Grouped motif distribution, C2. Top grouped motifs, D1. SSR loci distribution, D2. Top SSR loci on chromosomes, D3. Top SSRs frequency on chromosomes, D4. Relationship of SSRs vs length, E1. SSR length distribution, E2. Top distribution of SSR length.



A1. K-mer distribution of motifs

A2. Top k-mers

B1. Motif distribution

B2. Top motifs

C1. Grouped motif distribution

C2. Top grouped motifs

D1. SSR loci distribution

D2. Top SSR loci on chromosomes

D3. Top SSRs frequency on chromosomes

D4. Relationship of SSRs vs length

E1. SSR length distribution

E2. Top distribution of SSR length

Figure 8. Charts produced by SSR statistical plotting

4.3 Gbrowse viewing

Description

This module, *ssr2gff.pl/ssr2gff3.pl* , creates standard gbf/gff3 file from the identified SSRs for Generic Genome Browser (GBrowse, http://gmod.org/wiki/GBrowse).

Usage

```
perl ssr2gff.pl –i SequenceFileName.ssr
```

Or:

```
perl ssr2gff3.pl –i SequenceFileName.ssr
```

e.g.:

```
perl ssr2gff.pl -i Osativa_204.fa.ssr
```

Options "i" defines the name of input file.

Inputs

The input file is the .ssr file from SSR identification or files with formats similar to .ssr.

Outputs

This step produces the .gbf/gff3 file for browsing in Gbrowse.

You can view the gff3/gbf file on a local computer with GBrowse installed, or **just** upload them to Gbrowse online sever (e.g. http://webgbrowse.cgb.indiana.edu/cgi-bin/webgbrowse/uploadData, http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/).

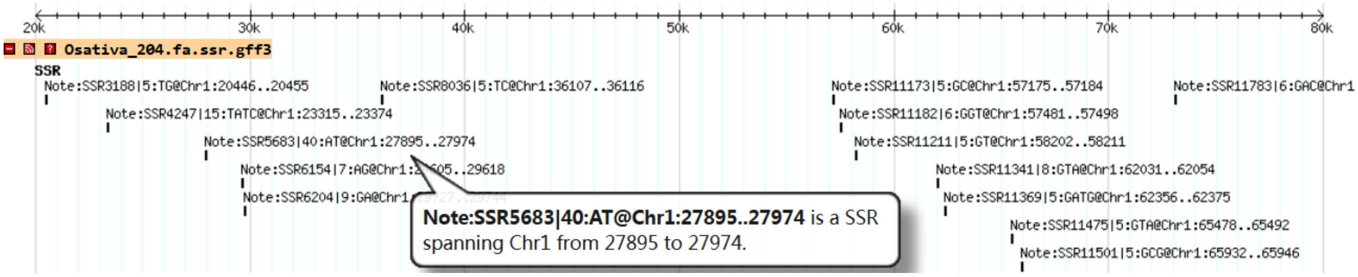

20k
30k
40k
50k
60k
70k
80k
Osativa_204.fa.ssr.gff3
SSR
Note:SSR3188|5:TG@Chr1:20446..20455
Note:SSR8036|5:TC@Chr1:36107..36116
Note:SSR11173|5:GC@Chr1:57175..57184
Note:SSR11783|6:GAC@Chr1
Note:SSR4247|15:TATC@Chr1:23315..23374
Note:SSR11182|6:GGT@Chr1:57481..57498
Note:SSR5683|40:AT@Chr1:27895..27974
Note:SSR11211|5:GT@Chr1:58202..58211
Note:SSR6154|7:AG@Chr1:29605..29618
Note:SSR11341|8:GTA@Chr1:62031..62054
Note:SSR6204|9:GA@Chr1:29727..29744
Note:SSR11369|5:GATG@Chr1:62356..62375
Note:SSR11475|5:GTA@Chr1:65478..65492
Note:SSR11501|5:GCG@Chr1:65932..65946
Note:SSR5683|40:AT@Chr1:27895..27974 is a SSR spanning Chr1 from 27895 to 27974.

Figure 9. Browsing Osativa_204.fa.ssr.gff3 with Gbrowse (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/).

4.4 Marker designing

Description

The module *doprimer_smt.pl* prints out primers for amplifying identified SSRs. It also generates - unique markers across whole input sequences based on ssr loci information.

Usage

```
perl doprimer_smt.pl -i SequenceFileName -sr SequenceFileName.ssr
[options]
```

```
e.g.:
```

```
perl doprimer_smt.pl -i Osativa_204.fa -sr Osativa_204.fa.ssr
```

Options "i" defines the name of input sequence file; "sr" defines SSR loci file name, default value is <sequence filename>.ssr, or you can use files with formats similar to .ssr.

Options

-p Platform, the value can only be u or w (u: for Unix like, w: for windows).

-fl length of extracted SSR flanking sequence at either side of SSR loci(integer value). default value is 200.

-l Length of allowed maximum template for primer designing (integer value), default value is 2000.

-smin Minimum length of amplicon size (integer value), default value is 120.

-Smax Maximum length of amplicon size (integer value), default value is 400.

-tm Optimal annealing temperature for PCR primer (integer value), default value is 60.

-dir Directory or path of primer3_core program in unix like systems. The actual directory should be given. note: no "/" in the end.e.g. -dir=/home/SCE/wangxuewen/bin/primer3-2.2.3/src. This option works in unix system only.

-ts Logic T or F, output template sequence with primer (T) or not (F).

Inputs

The input files for this step include (1). the original sequence file and (2). the .ssr file used/generated by SSR identification, respectively.

Outputs

This step produces 4 files:

*.seq	-	Prepared template sequence for primer designing
*.ssr.finalPrimer.txt	-	Detailed final primer information
*.ssr.finalPrimer.txt.sts	-	marker's primer sequence and size information
*.ssr.finalPrimer.txt.sat3	-	Statistic summary of this operation

4.5 e-Mapping

Description

This module *elctPCR.pl* maps the designed markers to original sequence/chromosome (You can use another sequence/chromosome too). It produces the mapped maker position on all sequences, and calculates alleles of these markers on all sequences/chromosome. It also returns the amplification information of markers, and a summary of this mapping.

Usage

```
perl elctPCR.pl –i SequenceFileName –mk SequenceFileName.
ssr.finalPrimer.txt.sts [options]
```

Options "i" defines the name of input sequence file; "mk" defines the name of the marker file, default is the .sts file generated by Maker designing.

e.g.:

```
perl elctPCR.pl –i Osativa_204.fa –mk
Osativa_204.fa.ssr.finalPrimer.txt.sts
```

Options

-o Emapping tag of output files. Default tag is "eMap". Mapping result file will be the sequence file name plus ".emap". There are four files produced: $seq.eMap, $seq.eMap.amp, $seq.eMap.amp.frg, $seq.eMap.amp.frg.sat4.

-pf System platform, the value is u or w (u: Unix, w: windows).

-w Word size for ePCR, default is 12. For more details, refer to e-PCR software.

-f value takes 1 or 3. 1 for contiguous words and 3 for discontiguous. Default value is 1. For more details, refer to e-PCR software.

-m Integer value. Margin. Default is 3000. For more details, refer to e-PCR software.

-d Set allowed sts size range: value1-value2, e.g. 100-1000. default is 100-1000. For more details, refer to e-PCR software.

-n Max mismatches allowed (integer value, default 0, suggested value is 1). For more details, refer to e-PCR software.

-g Max indels allowed (integer value, default 0, suggested value is 1). For more details, refer to e-PCR software.

-p value is -|+. Turn hits post-process on/ for e-PCR. Default is -. For more details, refer to e-PCR software.

Inputs

The input files for this step include (1). the original sequence file and (2). the .sts file used/generated by Marker designing, respectively.

Outputs

There are totally 4 output files for this step:

*.eMap	-	Mapped markers position on each chromosome
*.eMap.amp	-	Amplification information of each marker
*.eMap.frg	-	Number of alleles of each marker
*.eMap.frg.sat4	-	Summary of ePCR/eMapping

4.6 All-in-one step to run GMATA

Above multiple steps (4.1 - 4.5) can be replaced by a single command:

```
perl gmata.pl -c config_file -i SequenceFileName
```

This is the easiest way to run GMATA modules.

The config_file (e.g. default_cfg.txt, it is attached with GMATA) is necessary for this operation. It contains all the parameters/options used in steps 4.1 - 4.5. They can be edited if you want or left unrevised (all values will be the default value).

----end of user's guide