krait

Release 1.5.1

Lianming Du

Sep 08, 2023

CONTENTS:

1 Preface **3**
- 1.1 Introduction . . . 3
- 1.2 License . . . 3
- 1.3 Requirements . . . 3
- 1.4 Installation . . . 3
- 1.5 User Interface . . . 5

2 Import Sequences **7**
- 2.1 Import One FASAT File . . . 7
- 2.2 Import Multiple FASTA Files . . . 7
- 2.3 Import Fasta from NCBI . . . 8

3 Search for Tandem Repeats **9**
- 3.1 Search for SSRs . . . 9
- 3.2 Search for cSSRs . . . 11
- 3.3 Search for iSSRs . . . 12
- 3.4 Search for VNTRs . . . 14
- 3.5 Settings for Search . . . 16

4 Mapping Repeats in Genes **19**
- 4.1 Start Mapping Repeats . . . 19
- 4.2 Locating Results . . . 19
- 4.3 Select by Color . . . 20
- 4.4 Remove Mapping Results . . . 21

5 Primer Design **23**
- 5.1 Start Primer Design . . . 23
- 5.2 Primer Design Results . . . 23
- 5.3 Show or Remove Primer Design Results . . . 24
- 5.4 Settings for Primer Design . . . 24

6 Filter Data in Table **27**
- 6.1 Start Filter Data . . . 27
- 6.2 Construct Filter Conditions . . . 28

7 Statistical Analysis **29**
- 7.1 Perform statistical Analysis . . . 29
- 7.2 Statistical Analysis Results . . . 29
- 7.3 Refresh Statistical Results . . . 30

8 Export Output Files **31**

8.1 Export Table 31
8.2 Export FASTA 31

9 Feedback **33**

10 Indices and tables **35**

Du L, Zhang C, Liu Q, Zhang X, Yue B (2018) Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics. 34(4):681-683. 10.1093/bioinformatics/btx665

CHAPTER

ONE

PREFACE

1.1 Introduction

Krait is a robust and ultrafast tool with a user-friendly graphic interface for genome-wide investigation of microsatellites, which attempts to overcome the limitations of the currently available tools. Krait is written in Python and can be run as a standalone desktop application on Windows, Linux or Mac systems without dependencies. The microsatellite search engine is written in C and compiled as Python modules for import into Krait. Krait has many features: 1) Identification of perfect SSRs, imperfect SSRs (iSSRs), compound SSRs (cSSRs) and VNTRs from extremely large genome. 2) Locating the SSRs in gene coding region. 3) Design primer for microsatellite 4) Statistical analysis and plotting. 5) Supporting gzip compressed fasta as input file. 6) Supporting export FASTA, GFF3 or CSV. 7) Downloading DNA sequence from NCBI database.

1.2 License

Krait is distributed under a license called GNU General Public License, version 2 (GPL2).

1.3 Requirements

Krait is an operating system independent application and works on Windows, Linux and Mac systems. We recommend at least 1GB of RAM and 5GB of available hard disk space. In order to process very large genomes, a faster processor (64 bit) and larger amount of physical memory will be needed.

1.4 Installation

1.4.1 Windows

On windows, the preferred way to install Krait is to download the installer directly from https://github.com/lmdu/krait/releases Then double click the downloaded installer to install the program following the on-screen instructions.

1.4.2 MacOS

On MacOS, you can download the dmg package from the following URL: https://github.com/lmdu/krait/releases and double click the dmg file to mount it, and then drag Krait to Applcations folder to install krait.

1.4.3 Linux

On Linux, you don't need to install krait, just download the package from the following URL: https://github.com/lmdu/krait/releases and then decompress the downloaded file. Just simply double click krait to run the program.

1.4.4 Build from source

If you can not install krait through above approaches. You can run krait from source code and build yourself binary executable. Prior to start krait, you should install python (3.6 and 3.7 recommended) required packages: PySide2, pyfastx, numpy, requests, jinja2, appdirs, primer3-py, Cython, pyinstaller.

On windows, firstly you should install mingw64 and add it to your PATH enviroment. Download the source code from github and then open the cmd.

```
pip3 install apsw-wheels
pip3 install -r requirements.txt
cd krait\src\libs\src
python3 setup.py build_ext --inplace -c mingw32
move *.pyd ..
cd ..\..\..
python3 main.py
```

On MacOS and Linux, open the terminal.

```
pip3 install apsw
pip3 install -r requirements.txt
cd krait/src/libs/src
python3 setup.py build_ext --inplace
mv *.so ..
cd ../../..
python3 main.py
```

Build binary executable

```
# On windows
pyinstaller win.spec

# On Linux
pyinstaller linux.spec

# On MacOS
pyinstaller mac.spec
```

1.5 User Interface

The general components of the Krait main user interface are shown in Figure 1.

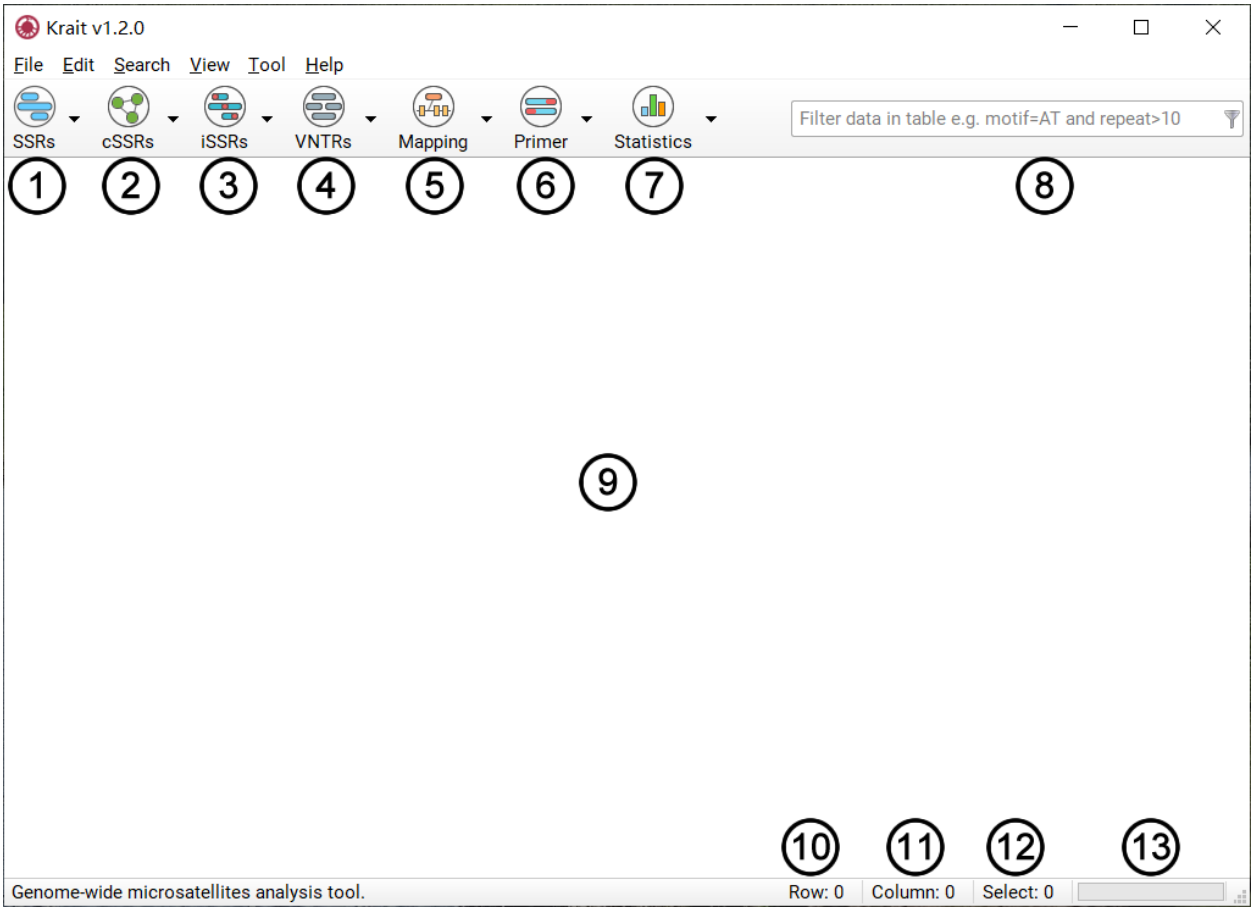


Figure 1: Krait main user interface

1. Search for perfect microsatellites or show results if you have searched.
2. Search for compound microsatellites or show results if you have searched.
3. Search for imperfect microsatellites or show results if you have searched.
4. Search for VNTRs or show results if you have searched.
5. Mapping perfect, compound, imperfect microsatellites or VNTRs to coding region of genes.
6. Design primers for microsatellites screened from results.
7. Perform statistical analysis and save to file.
8. Filter results using user inputted conditions.
9. Display the result of analysis in table.
10. The number of table rows.
11. The number of table columns.
12. The number of selected rows.
13. The task progressing.

CHAPTER

TWO

IMPORT SEQUENCES

To start a new analysis you just have to import your sequence data from files into Krait. The application accepts text or gzip compressed files containing one or more DNA sequences in FASTA format (Figure 2). These files may have the extension .fasta, .fna, .fa, .txt, .fasta.gz, .fna.gz, .fa.gz. A sequence in FASTA format begins with a single-line description or header starting with a "**>**" character. The rest of the header line is arbitrary but should be informative. Avoid strange characters in the sequence header, such as '&' or '\' and use 'N' to denote in-determinations in the sequences.

Figure 2: An example for the FASTA format

2.1 Import One FASAT File

1. Go to **File Menu -> Import Fasta**.
2. Select a FASTA formatted file containing your sequences.
3. Click **Open** to import FASTA file.

2.2 Import Multiple FASTA Files

1. Put your FASTA files in a folder
2. Go to **File Menu -> Import Fastas in Folder**.
3. Select the folder to load FASTA files in that folder.

2.3 Import Fasta from NCBI

Krait provides a tool for user to download the FASTA formatted DNA sequence from NCBI nucleotide database (Figure 3).

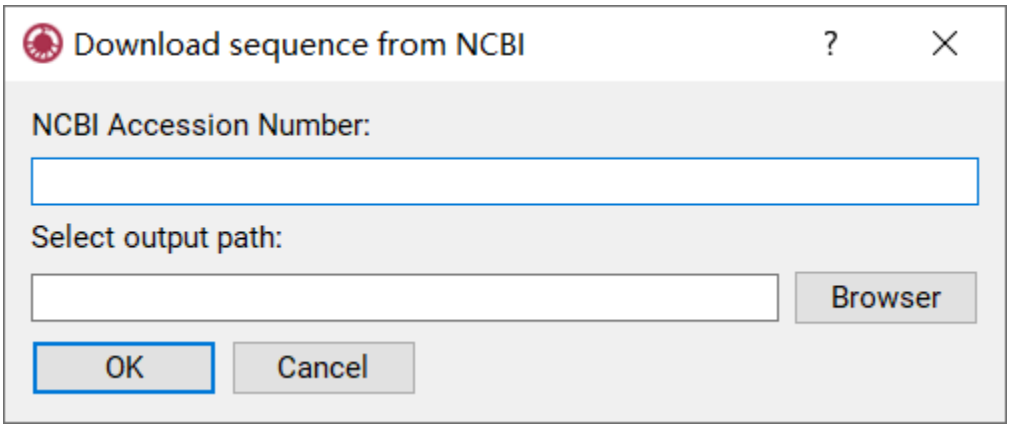


Figure 3: Downloading sequence from NCBI database

1. Go to **Tool Menu -> Download Sequence from NCBI** to open download dialog.
2. Input accession, version or GI number of a sequence in NCBI nucleotide database.
3. Click **Browser** to select an output file path.
4. Click **OK** to start download file and wait for finish.
5. Once download finished, go to **File Menu -> Import Fasta** to select your downloaded file to import.

CHAPTER

THREE

SEARCH FOR TANDEM REPEATS

Krait allows user to identify perfect microsatellites (SSRs), compound microsatellites (cSSRs) and imperfect SSRs (iSSRs) as well as VNTRs.

3.1 Search for SSRs

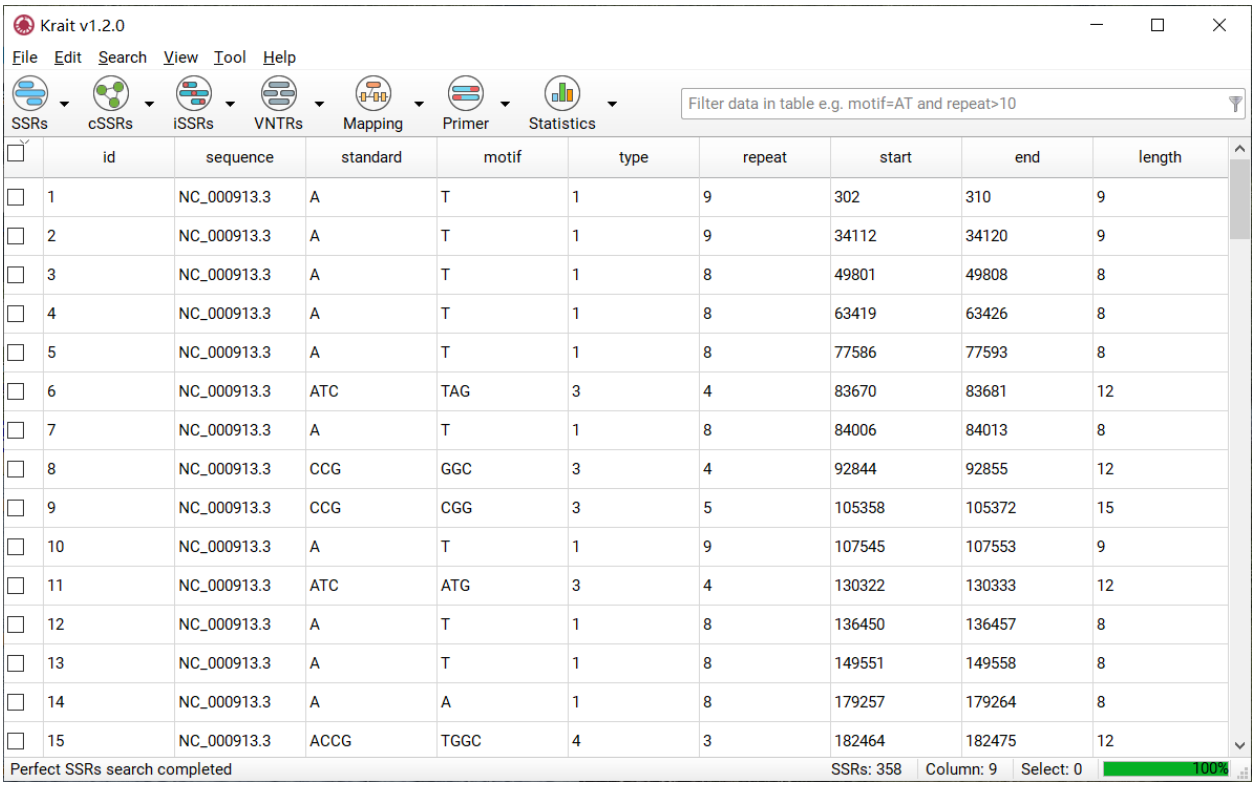
3.1.1 Start SSR Search

1. Import fasta sequence file (See 2 Input Files).
2. Go to **SSRs (toobar) -> Specify Minimum Repeats** to specify minimum repeats.
3. Click SSR search bubtton to start search SSRs.

Note: **Search Menu -> Search for SSRs** and **SSRs (toolbar) -> Search for SSRs** will remove the previous searched SSR results and then search for SSRs again.

3.1.2 SSR Search Results

After SSR search finished, a table containing results will be displayed. An example was shown in Figure 4. The result table contains 9 columns.



id	sequence	standard	motif	type	repeat	start	end	length
1	NC_000913.3	A	T	1	9	302	310	9
2	NC_000913.3	A	T	1	9	34112	34120	9
3	NC_000913.3	A	T	1	8	49801	49808	8
4	NC_000913.3	A	T	1	8	63419	63426	8
5	NC_000913.3	A	T	1	8	77586	77593	8
6	NC_000913.3	ATC	TAG	3	4	83670	83681	12
7	NC_000913.3	A	T	1	8	84006	84013	8
8	NC_000913.3	CCG	GGC	3	4	92844	92855	12
9	NC_000913.3	CCG	CGG	3	5	105358	105372	15
10	NC_000913.3	A	T	1	9	107545	107553	9
11	NC_000913.3	ATC	ATG	3	4	130322	130333	12
12	NC_000913.3	A	T	1	8	136450	136457	8
13	NC_000913.3	A	T	1	8	149551	149558	8
14	NC_000913.3	A	A	1	8	179257	179264	8
15	NC_000913.3	ACCG	TGGC	4	3	182464	182475	12

Figure 4: An example for SSR search result

column	description
id	unique identifier generated by Krait
sequence	the name of sequence where SSR was found
standard	the standardized motif
motif	repeat unit of SSR
type	SSR type, mononucleotide, dinucleotide etc corresponding to motif length
repeat	number of repeats
start	start position of SSR in original sequence, 1-based
end	end position of SSR in original sequence, 1-based
length	the length of SSR (bp)

3.1.3 Show SSR Results

If you have searched SSRs, you can click SSR search button or go to **SSRs (toolbar) -> Show Perfect SSRs** or go to **View Menu -> Show Perfect SSRs** to display SSR results in table.

3.1.4 Remove SSR Results

You can go to **SSRs (toolbar) -> Remove Perfect SSRs** or **View Menu -> Remove Perfect SSRs** to remove searched SSR results.

3.2 Search for cSSRs

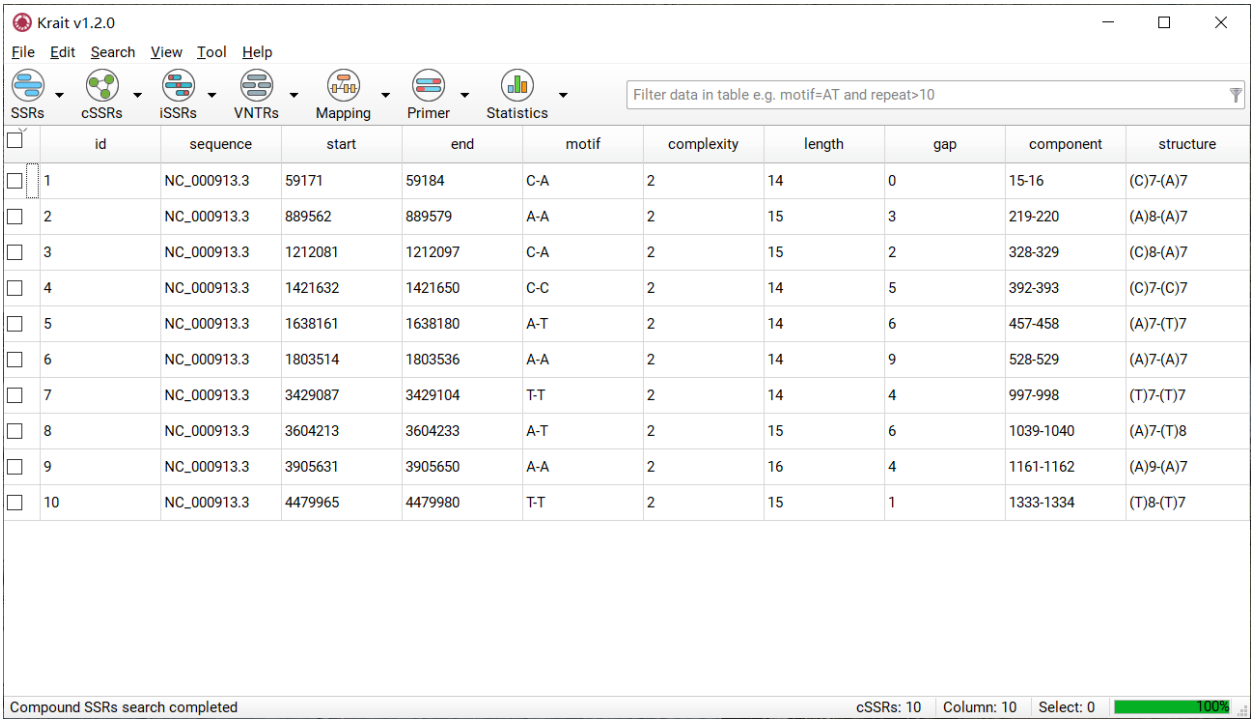
3.2.1 Start cSSR Search

1. Search perfect SSRs (See 3.1.1).
2. Go to **cSSRs -> Specify Maximum Distance** to specify a maximum distance (dMAX) allowed between two perfect SSRs.
3. Click cSSR search button to start search cSSRs.

Note: **Search Menu -> Search for cSSRs** and **cSSRs (toolbar) -> Search for cSSRs** will remove the previous searched cSSR results and then search for cSSRs again.

3.2.2 cSSR Seach Results

After cSSR search finished, a table containing results will be displayed. An example was shown in Figure 5. The result table contains 10 columns.



id	sequence	start	end	motif	complexity	length	gap	component	structure
1	NC_000913.3	59171	59184	C-A	2	14	0	15-16	(C)7-(A)7
2	NC_000913.3	889562	889579	A-A	2	15	3	219-220	(A)8-(A)7
3	NC_000913.3	1212081	1212097	C-A	2	15	2	328-329	(C)8-(A)7
4	NC_000913.3	1421632	1421650	C-C	2	14	5	392-393	(C)7-(C)7
5	NC_000913.3	1638161	1638180	A-T	2	14	6	457-458	(A)7-(T)7
6	NC_000913.3	1803514	1803536	A-A	2	14	9	528-529	(A)7-(A)7
7	NC_000913.3	3429087	3429104	T-T	2	14	4	997-998	(T)7-(T)7
8	NC_000913.3	3604213	3604233	A-T	2	15	6	1039-1040	(A)7-(T)8
9	NC_000913.3	3905631	3905650	A-A	2	16	4	1161-1162	(A)9-(A)7
10	NC_000913.3	4479965	4479980	T-T	2	15	1	1333-1334	(T)8-(T)7

Figure 5: An example for cSSR search result

column	description
id	unique identifier generated by Krait
sequence	the name of sequence where cSSR was found
start	start position of cSSR in original sequence
end	end position of cSSR in original sequence
motif	the motifs of individual SSR in a cSSR
complexity	the number of individual SSR in a cSSR
length	the length of cSSR
gap	total bases of distance between two SSR
component	the SSR id of individual SSRs comprised cSSR
structure	the detail of SSRs comprised cSSR

3.2.3 Show cSSR Results

If you have searched cSSRs, you can click cSSR search button or go to **cSSRs (toolbar) -> Show Compound SSRs** or **View Menu -> Show Compound SSRs** to dispay cSSR result in table.

3.2.4 Remove cSSR Results

You can go to **cSSRs (toolbar) -> Remove Compound SSRs** or **View Menu -> Remove Compound SSRs** to remove searched cSSR results.

3.3 Search for iSSRs

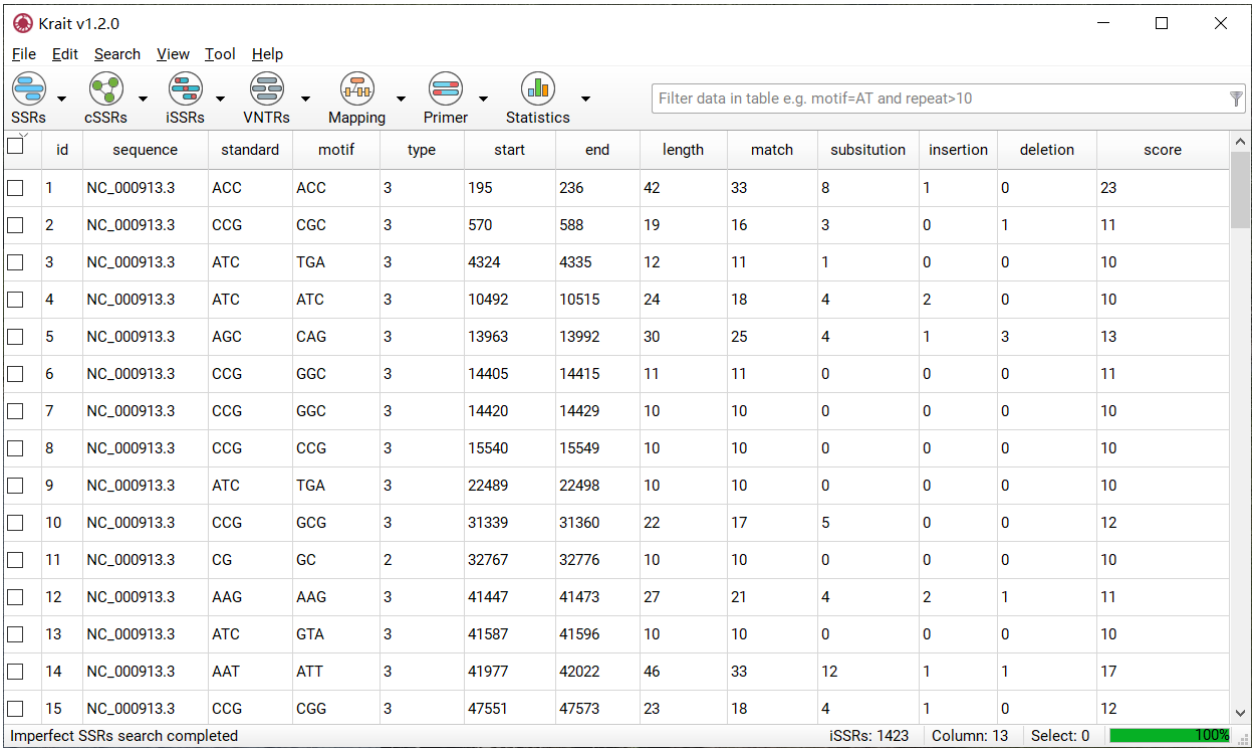
3.3.1 Start iSSR Search

1. Import fasta sequence file (See 2 Input Files).
2. Go to **iSSRs (toobar) -> Specify Search Parameters** to specify minimum length and repeats of seed, maximum consecutive edits, gap penalty and minimum score required.
3. Click cSSR search bubtton to start search cSSRs.

Note: **Search Menu -> Search for iSSRs** and **iSSRs (toolbar) -> Search for iSSRs** will remove the previous searched iSSR results and then search for iSSRs again.

3.3.2 iSSR Search Results

After iSSR search finished, a table containing results will be displayed. An example was shown in Figure 6. The result table contains 13 columns.



id	sequence	standard	motif	type	start	end	length	match	subsitution	insertion	deletion	score
1	NC_000913.3	ACC	ACC	3	195	236	42	33	8	1	0	23
2	NC_000913.3	CCG	CGC	3	570	588	19	16	3	0	1	11
3	NC_000913.3	ATC	TGA	3	4324	4335	12	11	1	0	0	10
4	NC_000913.3	ATC	ATC	3	10492	10515	24	18	4	2	0	10
5	NC_000913.3	AGC	CAG	3	13963	13992	30	25	4	1	3	13
6	NC_000913.3	CCG	GGC	3	14405	14415	11	11	0	0	0	11
7	NC_000913.3	CCG	GGC	3	14420	14429	10	10	0	0	0	10
8	NC_000913.3	CCG	CCG	3	15540	15549	10	10	0	0	0	10
9	NC_000913.3	ATC	TGA	3	22489	22498	10	10	0	0	0	10
10	NC_000913.3	CCG	GCG	3	31339	31360	22	17	5	0	0	12
11	NC_000913.3	CG	GC	2	32767	32776	10	10	0	0	0	10
12	NC_000913.3	AAG	AAG	3	41447	41473	27	21	4	2	1	11
13	NC_000913.3	ATC	GTA	3	41587	41596	10	10	0	0	0	10
14	NC_000913.3	AAT	ATT	3	41977	42022	46	33	12	1	1	17
15	NC_000913.3	CCG	CGG	3	47551	47573	23	18	4	1	0	12

Figure 6: An example for iSSR search result

column	description
id	unique identifier generated by Krait
sequence	the name of sequence where SSR was found
standard	the standardized motif
motif	repeat unit of SSR
type	SSR type, mononucleotide, dinucleotide etc. corresponding to motif length
start	start position of SSR in original sequence
end	end position of SSR in original sequence
length	the length of SSR (bp)
match	the number of matches
substitution	the number of substitutions
insertion	the number of insertions
deletion	the number of deletions
score	the score of iSSR

3.3.3 Show iSSR Results

If you have searched iSSRs, you can click iSSR search button or go to **iSSRs (toolbar) -> Show Imperfect SSRs** or **View Menu -> Show Imperfect SSRs** to dispay iSSR result in table.

3.3.4 Remove iSSR Results

You can go to **iSSRs (toolbar) -> Remove Imperfect SSRs** or **View Menu -> Remove Imperfect SSRs** to remove searched cSSR results.

3.4 Search for VNTRs

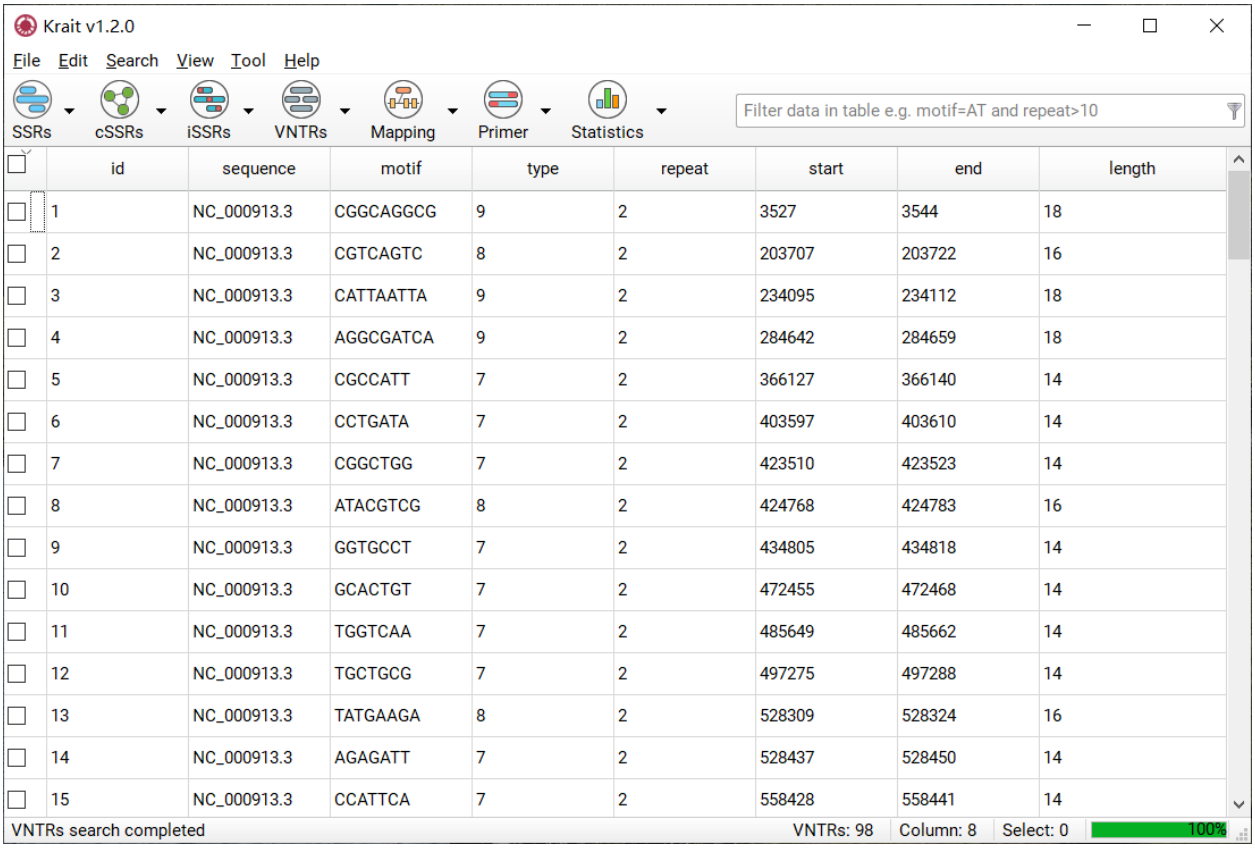
3.4.1 Start VNTR Search

1. Import fasta sequence file (See 2 Input Sequences).
2. Go to **VNTRs (toobar) -> Specify Search Parameters** to specify minimum and maximum length of motif and minimum repeats.
3. Click VNTR search bubtton to start search VNTRs.

Note: **Search Menu -> Search for VNTRs** and **VNTRs (toolbar) -> Search for VNTRs** will remove the previous searched VNTR results and then search for VNTRs again.

3.4.2 VNTR Search Results

After VNTR search finished, a table containing results will be displayed. An example was shown in Figure 7. The result table contains 8 columns.



id	sequence	motif	type	repeat	start	end	length
1	NC_000913.3	CGGCAGGCG	9	2	3527	3544	18
2	NC_000913.3	CGTCAGTC	8	2	203707	203722	16
3	NC_000913.3	CATTAATTA	9	2	234095	234112	18
4	NC_000913.3	AGGCGATCA	9	2	284642	284659	18
5	NC_000913.3	CGCCATT	7	2	366127	366140	14
6	NC_000913.3	CCTGATA	7	2	403597	403610	14
7	NC_000913.3	CGGCTGG	7	2	423510	423523	14
8	NC_000913.3	ATACGTCG	8	2	424768	424783	16
9	NC_000913.3	GGTGCCT	7	2	434805	434818	14
10	NC_000913.3	GCACTGT	7	2	472455	472468	14
11	NC_000913.3	TGGTCAA	7	2	485649	485662	14
12	NC_000913.3	TGCTGCG	7	2	497275	497288	14
13	NC_000913.3	TATGAAGA	8	2	528309	528324	16
14	NC_000913.3	AGAGATT	7	2	528437	528450	14
15	NC_000913.3	CCATTCA	7	2	558428	558441	14

Figure 7: An example for VNTR search result

column	description
id	unique identifier generated by Krait
sequence	the name of sequence where SSR was found
motif	repeat unit of SSR
type	VNTR type, motif length
repeat	number of repeats
start	start position of SSR in original sequence
end	end position of SSR in original sequence
length	the length of SSR (bp)

3.4.3 Show VNTR Results

If you have searched VNTRs, you can click VNTR search button or go to **VNTRs (toolbar) -> Show VNTRs** or go to **View Menu -> Show VNTRs** to display VNTR results in table.

3.4.4 Remove VNTR Results

You can go to **VNTRs (toolbar) -> Remove VNTRs** or **View Menu -> Remove VNTRs** to remove searched VNTR results.

3.5 Settings for Search

The setting panel also can be opened by **Edit Menu -> Preferences**. The setting panel was shown in Figure 8.

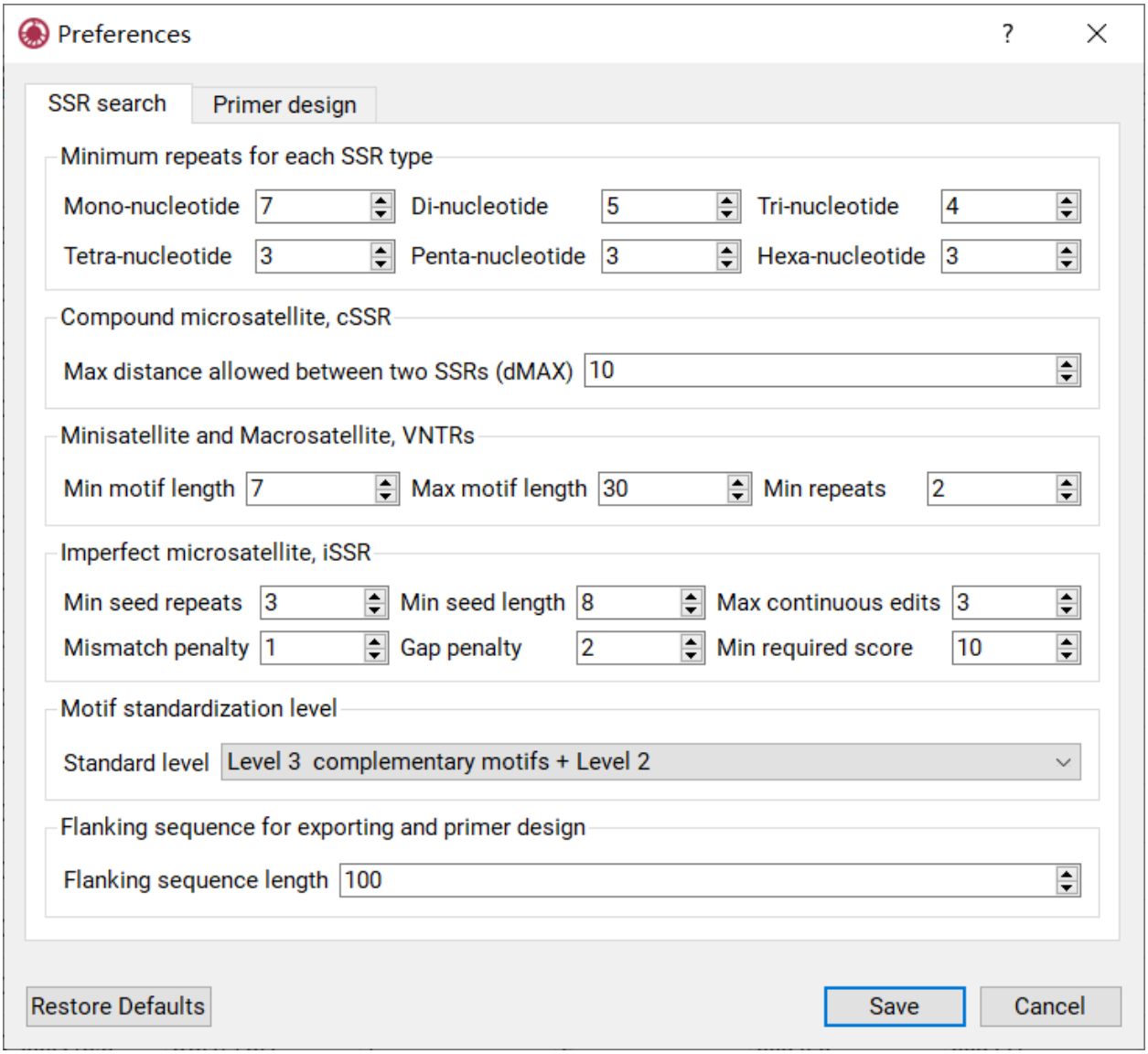


Figure 8: The setting panel for search

The setting panel contains 6 subpanels:

Minimum repeats Panel	specify minimum repeats for each type (di- tohexa-nucleotide) of perfect SSRs
cSSR Panel	specify maximum distance allowed between two perfectSSRs
VNTR Panel	set minimum and maximum length of motifs and minimumrepeats
iSSR Panel	set minimum repeats and length of extended seed, maximum consecutive edits allowed in an extension, mismatch penalty, indel penalty and minimum required score
Motif Standardization Level Panel	set the motif standardization level (0-4), the level only affects the result of perfect and imperfect SSRs search
Flanking Sequence Panel	set the flanking sequence length, used to design primer, export FASTA and display the details of SSR

3.5.1 Motif Standardization

Level 0	no standardization will be performed
Level 1	Similar motifs. For example, CA can be viewed as AC. ATGcan represent TGA and GAT
Level 2	Reverse complementary motifs, including Level 1. For example, CAT is a reverse complementary motif of ATG. ATG can represent TGA, GAT, CAT, ATC and TCA
Level 3	Complementary motifs, including Level 1 and Level 2. For example, TAC is a complementary motif of ATG. ATG can represent TGA,GAT, CAT, ATC, TCA, TAC, ACT and CTA
Level 4	Reverse motifs, including Level 1, Level 2 and Level 3. For example, GTA is a reverse motif of ATG. ATG can represent TGA,GAT, CAT, ATC, TCA, TAC, ACT, CTA, GTA, TAG and AGT

CHAPTER

FOUR

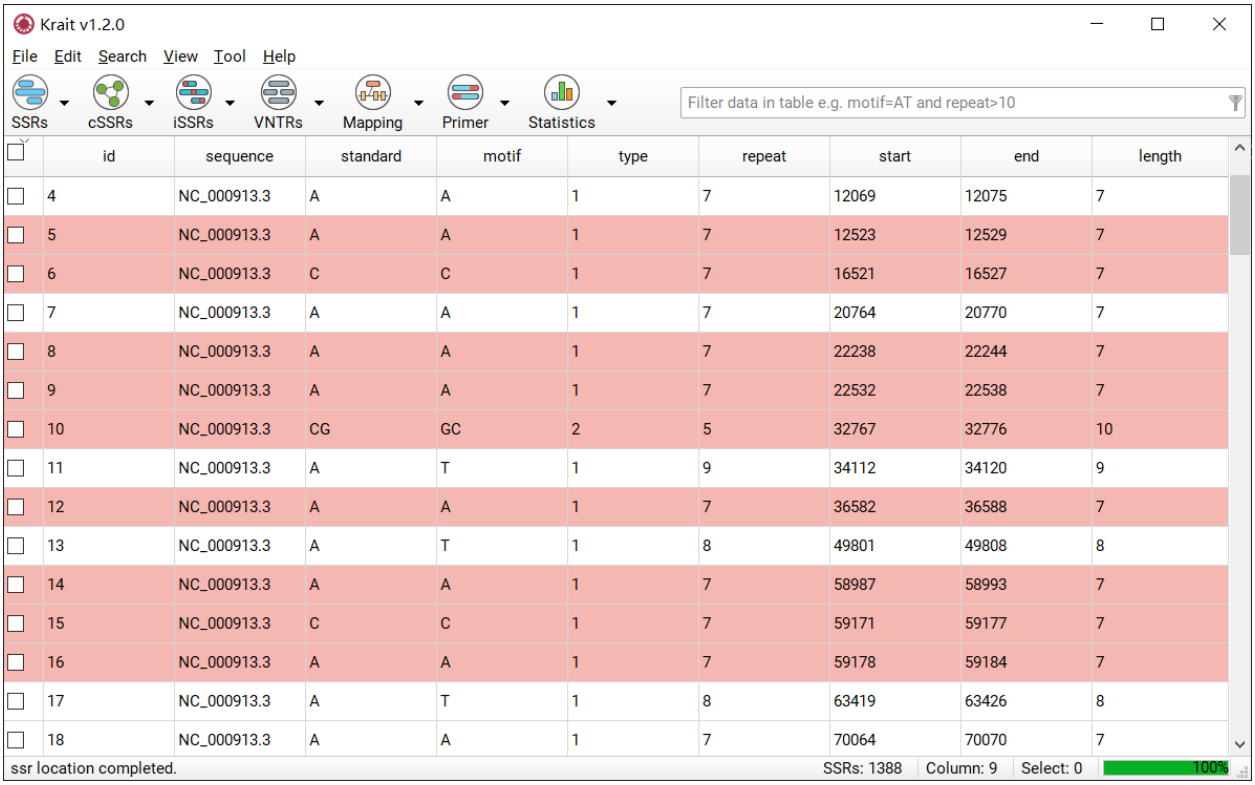
MAPPING REPEATS IN GENES

4.1 Start Mapping Repeats

1. Search SSRs, cSSRs, iSSRs or VNTRs (See 3 Search for Repeats)
2. Go to **Mapping (toolbar) -> Import Annotation File** to select a GTF or GFF file corresponding to the imported sequence file.
3. Click the mapping button to start mapping.

4.2 Locating Results

If mapping task finished, the repeats mapped in gene exon, intron, CDS or UTRs will be marked as different colors in table. An example was shown in Figure 9.



id	sequence	standard	motif	type	repeat	start	end	length
4	NC_000913.3	A	A	1	7	12069	12075	7
5	NC_000913.3	A	A	1	7	12523	12529	7
6	NC_000913.3	C	C	1	7	16521	16527	7
7	NC_000913.3	A	A	1	7	20764	20770	7
8	NC_000913.3	A	A	1	7	22238	22244	7
9	NC_000913.3	A	A	1	7	22532	22538	7
10	NC_000913.3	CG	GC	2	5	32767	32776	10
11	NC_000913.3	A	T	1	9	34112	34120	9
12	NC_000913.3	A	A	1	7	36582	36588	7
13	NC_000913.3	A	T	1	8	49801	49808	8
14	NC_000913.3	A	A	1	7	58987	58993	7
15	NC_000913.3	C	C	1	7	59171	59177	7
16	NC_000913.3	A	A	1	7	59178	59184	7
17	NC_000913.3	A	T	1	8	63419	63426	8
18	NC_000913.3	A	A	1	7	70064	70070	7

Figure 9: An example for mapping result

4.3 Select by Color

This function allows selection of repeats on the basis of their color i.e. the region in where the repeat located. You can go to **Mapping (toolbar) -> Show SSRs in CDS, Exon, UTR or Intron** to screen as shown in Figure 10.

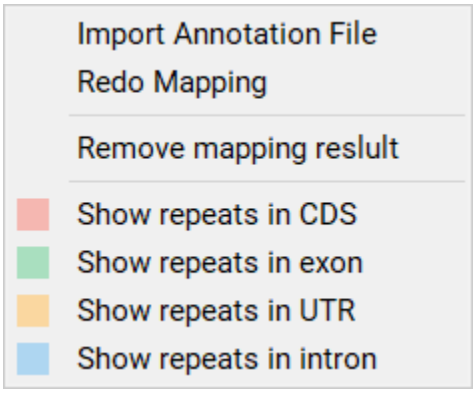


Figure 10: Selection by Colors

4.4 Remove Mapping Results

You can go to **Mapping (toolbar) -> Remove mapping result** to remove mapping results.

CHAPTER

FIVE

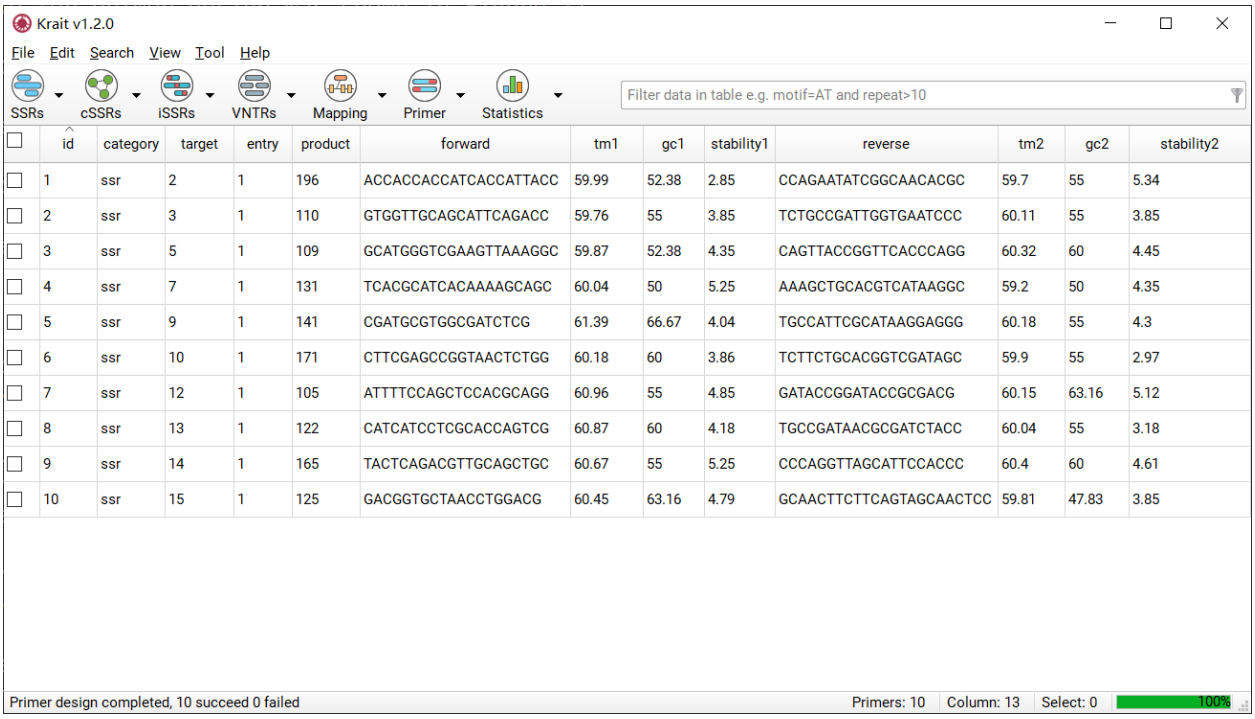
PRIMER DESIGN

5.1 Start Primer Design

1. Select tandem repeats in table that you want to design primer.
2. Go to **Primer (toolbar) -> Specify Primer3 Setttings** to set parameters for primer3.
3. Click design primer button to start design primer for selected tandem repeats.

5.2 Primer Design Results

After primer design, a table contains primer details will be display. The result table contains 12 columns. An example for primer design was shown in Figure 11.



	id	category	target	entry	product	forward	tm1	gc1	stability1	reverse	tm2	gc2	stability2
☐	1	ssr	2	1	196	ACCACCACCATCACCATTACC	59.99	52.38	2.85	CCAGAATATCGGCAACACGC	59.7	55	5.34
☐	2	ssr	3	1	110	GTGGTTGCAGCATTCAGACC	59.76	55	3.85	TCTGCCGATTGGTGAATCCC	60.11	55	3.85
☐	3	ssr	5	1	109	GCATGGGTCGAAGTTAAAGGC	59.87	52.38	4.35	CAGTTACCGGTTCACCCAGG	60.32	60	4.45
☐	4	ssr	7	1	131	TCACGCATCACAAAAGCAGC	60.04	50	5.25	AAAGCTGCACGTCATAAGGC	59.2	50	4.35
☐	5	ssr	9	1	141	CGATGCGTGGCGATCTCG	61.39	66.67	4.04	TGCCATTCGCATAAGGAGGG	60.18	55	4.3
☐	6	ssr	10	1	171	CTTCGAGCCGGTAACTCTGG	60.18	60	3.86	TCTTCTGCACGGTCGATAGC	59.9	55	2.97
☐	7	ssr	12	1	105	ATTTTCCAGCTCCACGCAGG	60.96	55	4.85	GATACCGGATACCGCGACG	60.15	63.16	5.12
☐	8	ssr	13	1	122	CATCATCCTCGCACCAGTCG	60.87	60	4.18	TGCCGATAACGCGATCTACC	60.04	55	3.18
☐	9	ssr	14	1	165	TACTCAGACGTTGCAGCTGC	60.67	55	5.25	CCCAGGTTAGCATTCCACCC	60.4	60	4.61
☐	10	ssr	15	1	125	GACGGTGCTAACCTGGACG	60.45	63.16	4.79	GCAACTTCTTCAGTAGCAACTCC	59.81	47.83	3.85

Figure 11: An example for primer design

id	unique identifier generated by Krait
category	the type of repeats, ssr, issr, cssr or vntr
target	the id of the tandem repeat
entry	the primer number for tandem repeats. A tandem repeats can contains multigroup primers
product	the product size of primer
forward	the sequence of forward primer
reverse	the sequence of reverse primer
tm1,tm2	the temperature of the forward primer and reverse primer
gc1,gc2	the GC content of the forward primer and reverse primer
stability1,2	the end stability of the forward primer and reverse primer

5.3 Show or Remove Primer Design Results

If you have designed primers, you can go to **Primer (toolbar) -> Show Designed Primer** to display primer results. The primer results can be removed by **Primer (toolbar) -> Remove Designed Primer**.

5.4 Settings for Primer Design

The settings for primer design shown in Figure 12 are the same with Primer3 tags. Click the setting name will redirect to the corresponding tags in Primer3 manual.

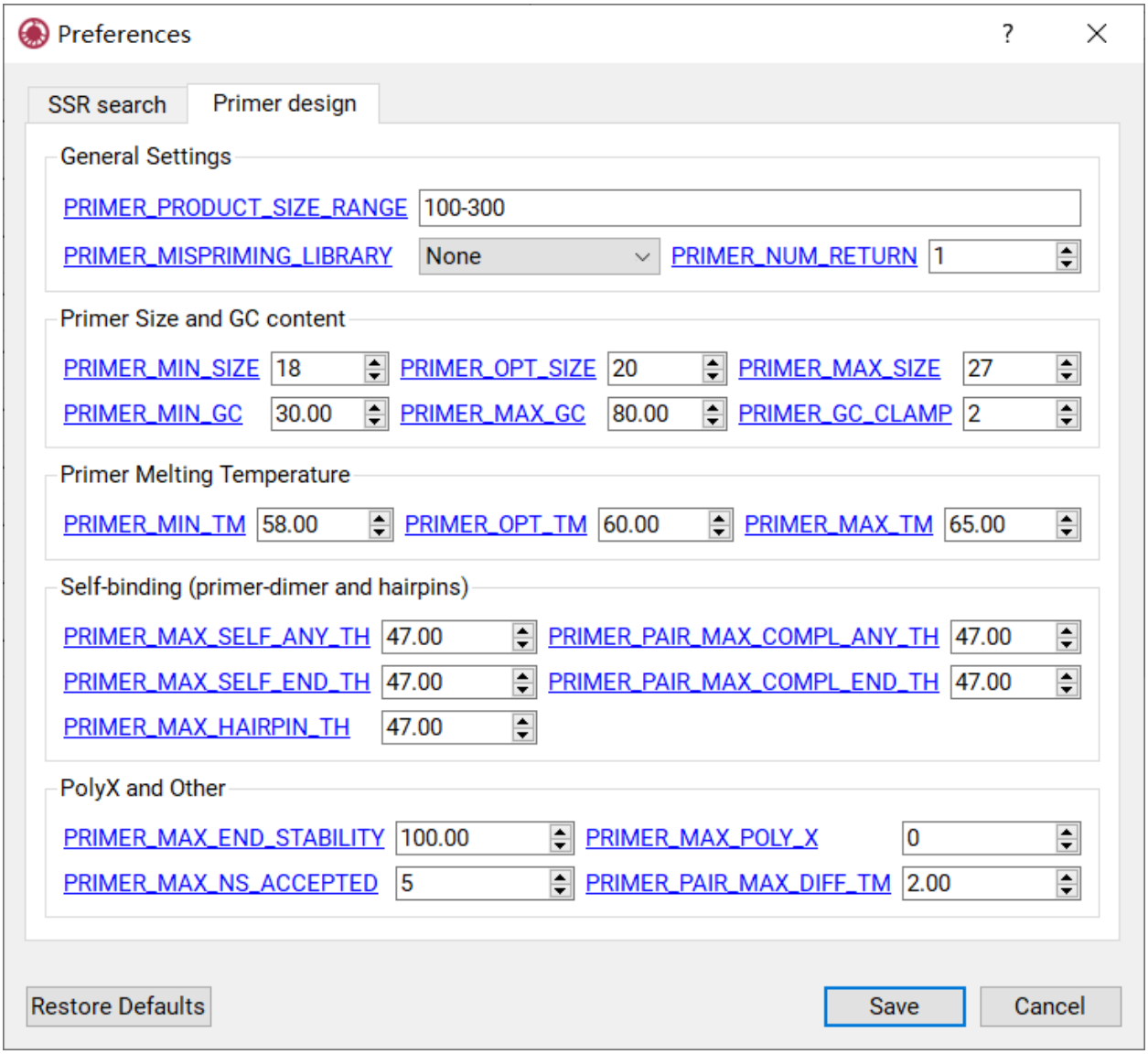


Figure 12: Setting panel for primer design

CHAPTER

SIX

FILTER DATA IN TABLE

6.1 Start Filter Data

1. Verify that the current display is a table with rows.
2. Input the conditions into input box as shown in Figure 13.
3. Press **Enter** to filter rows in table.

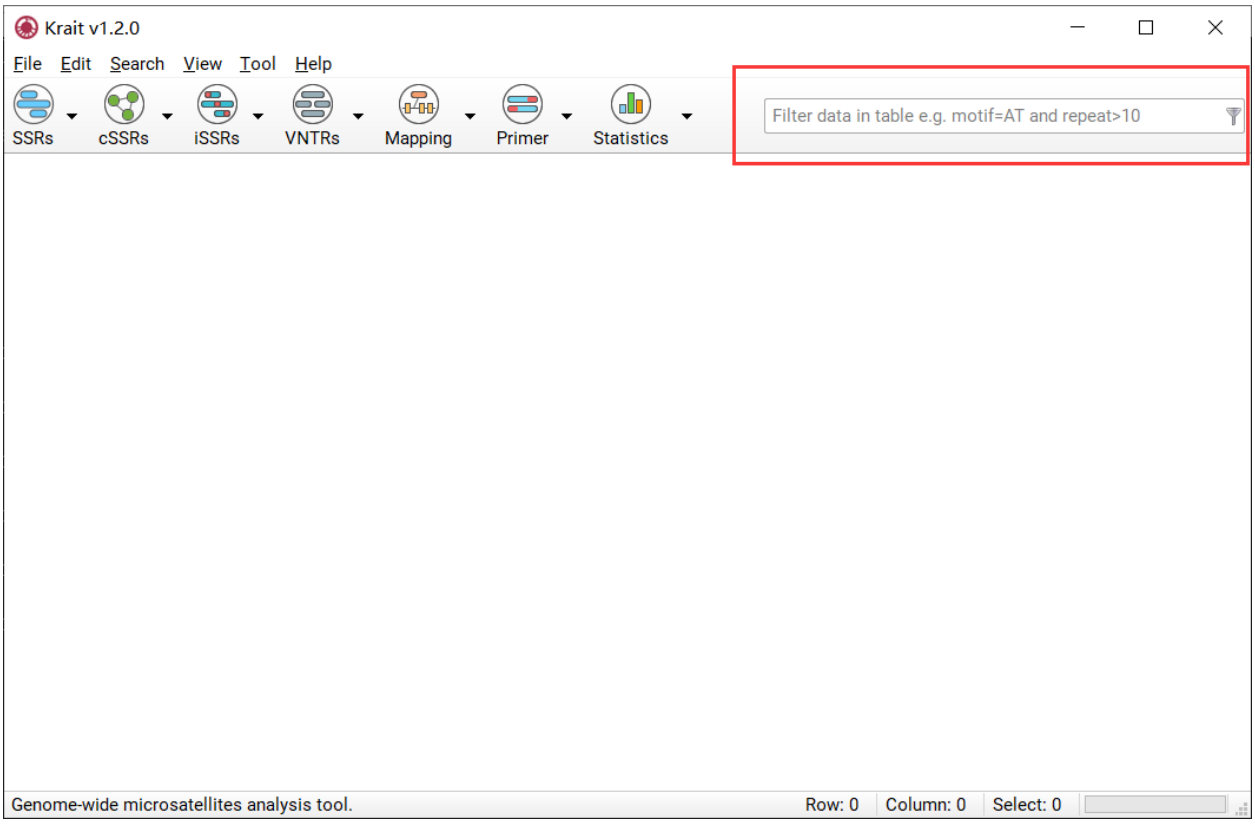


Figure 13: Filter condition input box.

6.2 Construct Filter Conditions

Krait provides comparison operators and logical operators that allow you to construct the condition. The comparison operators include **=, >, <, >=, <=**; the logical operators include **AND**, **OR**. There is a special operator **IN** can be used. The condition is case-insensitive. Thus you can use column name and operators to construct a condition to filter table rows. The numeric column can be allowed to use all comparison operators and the text column only be allowed to use **=** and IN. Examples:

1) Suppose you want to find SSRs whose motif is ATG, you can construct a condition:

```
motif=ATG
```

2) If you want to find SSRs whose motif is ATG and repeats greater than 10, you can construct a condition:

```
motif=ATG and repeat>10
```

3) If you want to find SSRs whose motifs are ATG, AT, AAAG and repeats between 10 and 15, you can construct a condition:

```
motif in (ATG,AT,AAAG) and repeats>=10 and repeats<=15
```

4) If you want to find SSRs whose length less than 12 or length greater than 20, you can construct a condition:

```
length<12 or length>20
```

CHAPTER

SEVEN

STATISTICAL ANALYSIS

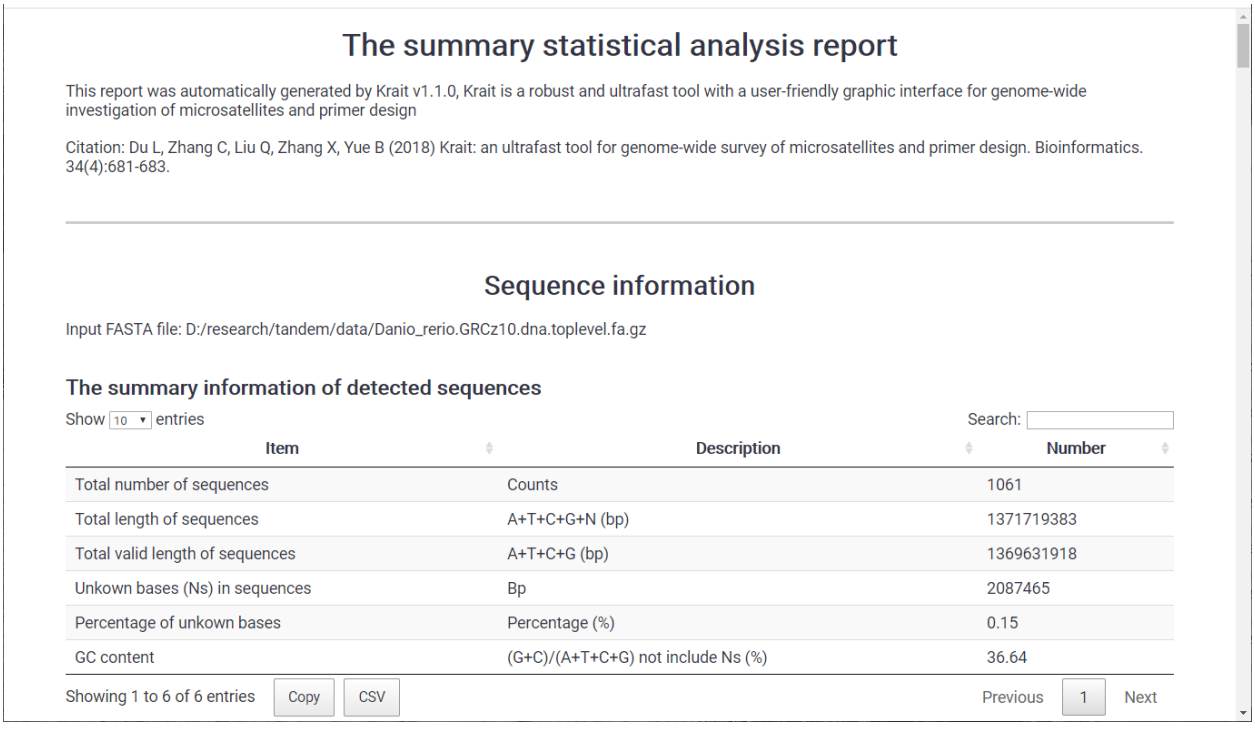
7.1 Perform statistical Analysis

If you have searched SSRs, cSSRs, iSSRs or VNTRs, you can just click the statistics button to start statistical analysis.

If you have performed statistical analysis, click the statistics button will display the statistical results. If you want to perform statistical analysis again, you can go to **Statistics (toobar) -> Statistical Analysis**.

7.2 Statistical Analysis Results

Once statistical analysis finished, a statistical report contains several tables and graphs will be generated and save to a html file. You can use browser (Chrome or firefox) to view html report.



The summary statistical analysis report

This report was automatically generated by Krait v1.1.0, Krait is a robust and ultrafast tool with a user-friendly graphic interface for genome-wide investigation of microsatellites and primer design

Citation: Du L, Zhang C, Liu Q, Zhang X, Yue B (2018) Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics. 34(4):681-683.

Sequence information

Input FASTA file: D:/research/tandem/data/Danio_rerio.GRCz10.dna.toplevel.fa.gz

The summary information of detected sequences

Show 10 entries Search:

Item	Description	Number
Total number of sequences	Counts	1061
Total length of sequences	A+T+C+G+N (bp)	1371719383
Total valid length of sequences	A+T+C+G (bp)	1369631918
Unkown bases (Ns) in sequences	Bp	2087465
Percentage of unkown bases	Percentage (%)	0.15
GC content	(G+C)/(A+T+C+G) not include Ns (%)	36.64

Showing 1 to 6 of 6 entries Copy CSV Previous 1 Next

Figure 14: A part of statistical report

You can also go to **File Menu -> Export Statistical Report** to perform statistics and export html formatted statistical report.

7.3 Refresh Statistical Results

If you have performed statistical analysis, you can go to **Statistics (toolbar) -> Refresh Statistical Analysis** to regenerate statistical results.

CHAPTER

EIGHT

EXPORT OUTPUT FILES

Krait allows user to export Table (CSV, GTF, GFF) and FASTA formatted files.

8.1 Export Table

1. Select rows in table that you want to export.
2. Go to File Menu -> Export Selected as Table and select a format (CSV, GTF, GFF or TEXT) and provide an output file name.
3. Click **Save** to export to output file.

8.2 Export FASTA

1. Select rows in table that you want to export.
2. Go to **File Menu -> Export Selected as Fasta** and provide an output file name.
3. Click **Save** to export to output file

CHAPTER

NINE

FEEDBACK

Report issues: https://github.com/lmdu/krait/issues.

Contact email: adullb@qq.com

CHAPTER

TEN

INDICES AND TABLES

- genindex
- modindex
- search