



(you can change application colors by clicking on bottom right colors)

Launch Analysis

Title	Path
-------	------

Alignment (BAM/SAM) + **ADD** x **DEL** ? **TEST**

Reference genome ☐ CRS ☒ rCRS ☐ Mouse ☐ Other

Results directory

Advanced parameters

eKLIPse temp directory

Downsampling threshold

Minimum read quality

Mapped part length

Blast %id threshold

Blast gap open cost

Min deleted mtDNA length

Number of Threads

Deletion position shift

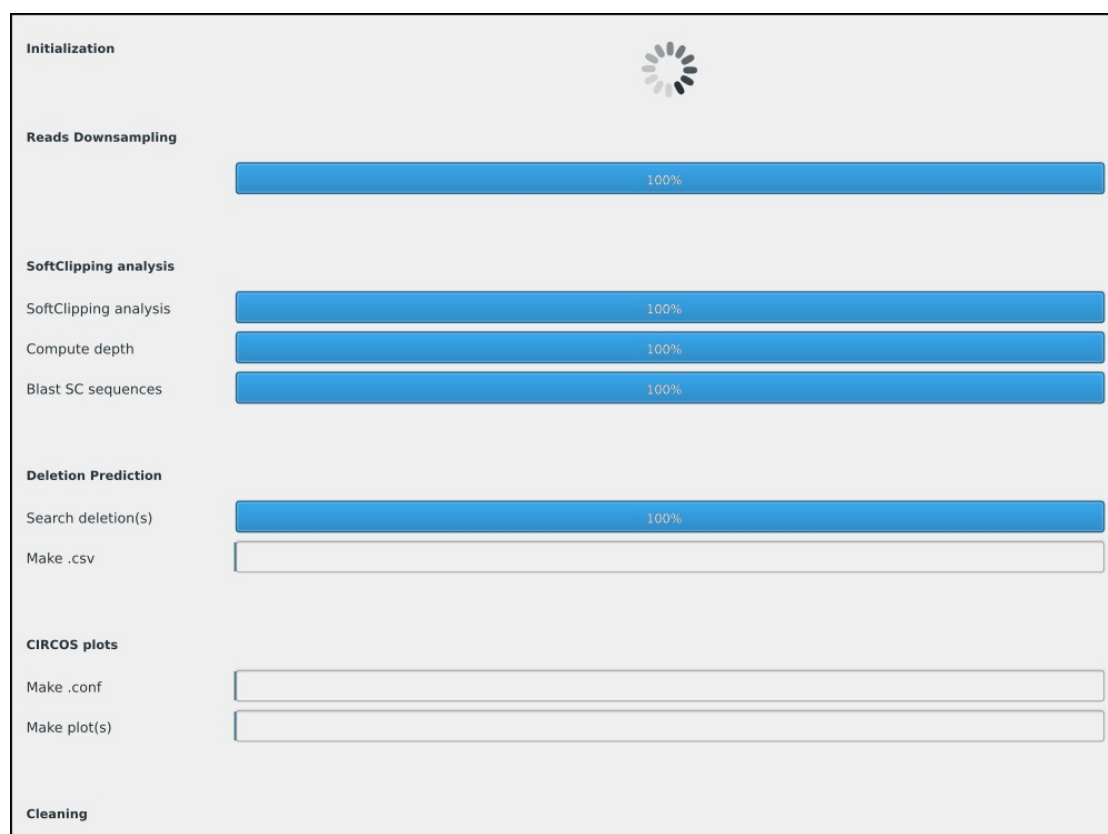
Soft-clip min length

Blast %cov threshold

Blast gap ext. cost

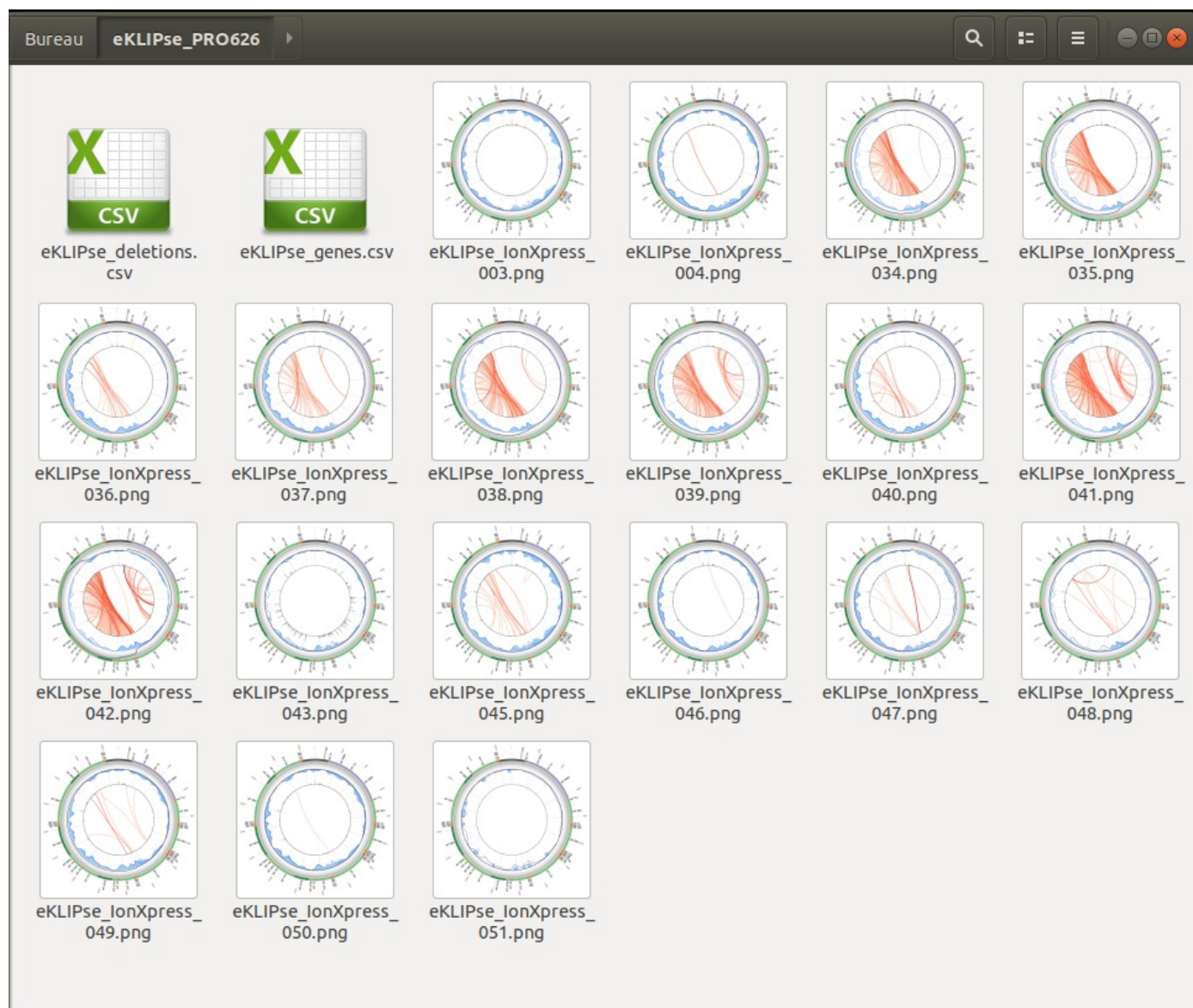
- 1 - To select your alignment files, click “ADD”. If required you can change alignments title by selecting corresponding cell.
- 2 - Select your reference genome. If you choose “Other”, browse to your own Genbank file by clicking on the folder icon.
- 3 - To change “results directory”, click on the folder icon.
- 4 - To modify “Advanced parameters” click on the expand icon. Please refers to “Parameters” section for further informations.
- 5 - Launch analysis by clicking “START”

Analysis in progress



eKLIPse analysis detailed progress can be followed on this window.

Results



Once the analysis is complete, the program automatically opens the results folder.

Testing

Two reduced alignments files are provided with the archive file.

Click "TEST" on the "Launch Analysis" windows before clicking "START".

Command Line Interface

Docker

A docker image is also available.

Follow building instruction [here \(https://docs.docker.com/get-started/part2/#build-the-app\)](https://docs.docker.com/get-started/part2/#build-the-app)

Linux

Requirements

Please install the following modules & tools:

- python 2.7
- [biopython \(https://github.com/biopython/biopython\)](https://github.com/biopython/biopython)
- [tqdm \(https://github.com/tqdm/tqdm\)](https://github.com/tqdm/tqdm)
- [samtools \(https://github.com/samtools/samtools\)](https://github.com/samtools/samtools)
- [blastn & makeblastdb \(http://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/\)](http://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) (>=2.3.0+)
- [circos \(http://circos.ca/software/download/\)](http://circos.ca/software/download/)

Testing

```
python eKLIPse.py --test
```

```
(*add "-samtools", "-blastn", "-makeblastdb" and "-circos" options if not in $PATH)
```

Running

```
python eKLIPse.py -in <INPUT file path> -ref <GBK file path> [OPTIONS]
```

[OPTIONS]

-out	<str>	: Output directory path	[current]
-tmp	<str>	: Temporary directory path	[/tmp]
-scsize	<int>	: Soft-clipping minimal length	[25]
-mapsize	<int>	: Upstream mapping length	[20]
-downcov	<int>	: Downsampling read number	[500000] (0=disable)
-minq	<int>	: Read quality threshold	[20]
-minlen	<int>	: Read length threshold	[100]
-shift	<int>	: Breakpoint sliding-window size	[5]
-minblast	<int>	: Minimal number of BLAST per breakpoint	[1]
-bilateral	<bool>	: Filter unidirectional BLAST	[True]
-mitosize	<int>	: Remove deleted mtDNA less than	[1000]
-id	<int>	: BLAST %identity threshold	[80]
-cov	<int>	: BLAST %coverage threshold	[70]
-gapopen	<int>	: BLAST cost to open a gap	[0:proton, 5:illumina]
-gapext	<int>	: BLAST cost to extend a gap	[2]
-thread	<int>	: Thread number	[2]
-samtools	<str>	: samtools bin path	[\$PATH]
-blastn	<str>	: BLASTN bin path	[\$PATH]
-makeblastdb	<str>	: makeblastdb bin path	[\$PATH]
-circos	<str>	: circos bin path	[\$PATH]
--test		: eKLIPse test	
--nocolor		: Disable output colors	

Parameters

Input file (-in)

eKLIPse accepts alignments in BAM or SAM format (require header) for both single and paired-end sequencing data. The input file is a simple tabulated text file as follow:

path_bam	title1
path_bam2	title2

mtDNA reference (-ref)

eKLIPse accepts any mtDNA reference genome in Genbank format.

rCRS (NC_012920.1.gb), CRS (J01415.2.gb) and *Mus musculus* (NC_005089.1.gb) are provided in “/data”

Downsampling (-downcov)

In order to reduce execution time, a downsampling option is available.

For singles deletions with low mutant load or multiples deletions, we advise to not downsample “-downcov 0”.

The obtained reads number must match to a sufficient mitochondrail genome coverage.

Sequencing & Alignment (-minq / -minlen)

According to your sequencing technology and library, you can adjust the minimum read length value (-minlen).

You can adjust minimum read quality (-minq), for example to consider multiple hits for a same read which reduce the minq.

Soft-clipping (-minq / -minlen)

For short read data, we advise to reduce minimal soft-clipping length (-scsize) and upstream mapping length (-mapsize).

For example, with 100bp reads, you could use “-scsize 15” and “-mapsize 10”.

Breakpoint sliding-window size could be modify if you expect a high number of homopolymers.

BLASTn (-id / -cov / -gapopen / -gapext)

BLASTn thresholds are mostly sequencing technology dependent.

Then according to your sequencing quality you could increase or decrease identity and coverage thresholds (-id / -cov).

Illumina is known to generate fewer errors and can therefore be more stringent on gap thresholds (-gapopen / -gapext).

For example, with illumina reads, you could use “-gapopen 5” and “-gapext 2”.

Filtering (-minblast / -bilateral / -mitosize)

According to your sequencing depth, quality and required stringency, you could modify filters.

Increasing the minimum number of BLAST per breakpoint increase the specificity but decrease the sensitivity (-minblast)

By default, eKLIPse filter out deleted mtDNA with a length under 1000bp.

But for example, if you're looking for sublimons you could reduce this length to 100bp.

eKLIPse is based on th search of bidirectionnal BLAST linking 5' breakpoint and 3' breakpoint.

It is therefore not recommended to disable this filter (“-bilateral False”).

Outputs

eKLIPse_deletions.csv

File containing all predicted deletions (bcp=breakpoint).

title	5'bpk	3'bpk	Freq	Freq for	Freq rev	5' Blast	3' Blast	5' Depth	3' Depth	Repetition
file1	7753	14601	3,46	0,38	6,55	2	23	1393	412	7754-GA-7755 14601-GA-14602
file2	7981	14955	7,40	4,28	10,51	2408	2506	7080	2544	7982-CT-7983 14955-CT-14956
file3	460	5243	7,24	13,72	0,76	7	1	72	197	458-CT-459 5242-CT-5243

eKLI^Pse_genes.csv

File summarizing cumulated deletions per mtDNA gene.

Gene	Start	End	Type	file3	file4	file5
MT-TF	577	647	trna	0,38	0,82	14,03
MT-RNR1	648	1601	rrna	2,27	14,42	14,03
MT-TV	1602	1670	trna	2,27	14,42	14,03
MT-RNR2	1671	3229	rrna	2,27	14,78	14,03
MT-TL1	3230	3304	trna	2,27	14,78	14,03
MT-ND1	3307	4262	protein	2,27	15,05	14,03

circos plot

One plot is created per input alignment. An example is shown below.

