miRge3 Release 0.0.1

Arun H. Patil and Marc K. Halushka

Nov 25, 2020

CONTENTS

1	Links	š	3
2	Table	e of contents	5
	2.1	Installation	5
	2.2	User guide	20
	2.3	miRge3.0 output	26
	2.4	MIT License	33

An update to Python package to perform comprehensive analysis of small RNA sequencing data, including miRNA annotation, A-to-I editing, novel miRNA detection, isomiR analysis, visualization through IGV, processing Unique Molecular Identifieres (UMI), tRF detection and producing interactive graphical output.

miRge3.0 is developed in python v3.8 and is a recent update of our previous version miRge2.0. This build includes command line interface (CLI) and cross-platform Graphical User Interface (GUI). For more details refer to documentation link below.

CHAPTER

ONE

LINKS

- Documentation
- Source code
- Report an issue
- Project page on PyPI

CHAPTER

TWO

TABLE OF CONTENTS

2.1 Installation

2.1.1 Linux OS

Welcome to installation protocol for Linux OS

Install python3.8 and R

This installation protocol is based on Ubuntu, please use the commands that suit your Linux distribution. For example, apt should be replaced with yum in Fedora/CentOS.

- · Search and start the terminal
- Follow the commands to update Ubuntu and install python 3.8 A password will be prompted when you type sudo, use the one you have set during Ubuntu (or your distro) installation.

```
sudo apt update
sudo apt install software-properties-common
sudo add-apt-repository ppa:deadsnakes/ppa
sudo apt install python3.8
sudo apt install python3-setuptools
sudo apt install python3-pip
sudo apt install r-base
```

Linux (Ubuntu 18.04) comes with python2.7 installed by default. To use python3.8, creating an alias in .bashrc would do the trick.

Use vim editor if you are familiar using this editor vi .bashrc or open the .bashrc using text editor by gedit .bashrc and add the following line at the bottom of the text. alias python=python3.8

Save and exit. After that type bash on the command line -Or- simply, close the terminal.

Installing miRge3.0

First install miRge dependenceis

• Search and start the terminal, execute the command below:

If you encounter a WARNING, like below:

```
WARNING: The script cutadapt is installed in '/home/arun/.local/bin' which is not on_

→PATH.

Consider adding this directory to PATH or, if you prefer to suppress this warning,

→use --no-warn-script-location.
```

Then, open a new terminal window or type cd to get to home directory. Add bin folder PATH to the .bashrc, as shown below: Example: export PATH=\$PATH:"/home/arun/.local/bin" Remeber to add your path /PATH_TO_USERS/bin.

Install miRge3.0 by this simple command

```
python3.8 -m pip install --user mirge3
```

To upgrade miRge3.0

python3.8 -m pip install --user --upgrade mirge3

Install additional C-libraries based tools

Install Bowtie

- Search and start the terminal
- Download bowtie

- unzip bowtie-1.2.3-macos-x86_64.zip
- cd bowtie-1.2.3-macos-x86_64
- pwd
 - /home/arun/software/bowtie-1.2.3-linux-x86_64
- Add these bowtie binaries to .bashrc as shown below:

export PATH=\$PATH:"/home/arun/software/bowtie-1.2.3-linux-x86_64"

• After that type bash on the command line -Or- simply, close the terminal.

Install Samtools

• Search and start the terminal, execute the below command: sudo apt install samtools

Install RNA Fold

- Search and start the terminal, execute the following commands:
- wget "https://www.tbi.univie.ac.at/RNA/download/sourcecode/2_4_x/ViennaRNA-2.4.16.tar.gz"
- cd ViennaRNA-2.4.16

sudo ./configure sudo make sudo make install

GUI requirements

Providing system wide access to miRge3.0, cutadapt, bowtie and bowtie-build, please type or (copy and paste) and submit each of the following commands on the terminal: **NOTE:** Make sure to change your path to python bin folder; Replace /home/arun/.local/with/Path on your computer/.

• Search and start the terminal, execute the following commands:

```
sudo ln -s /home/arun/.local/bin/miRge3.0 /usr/local/bin/miRge3.0
sudo ln -s /home/arun/.local/bin/cutadapt /usr/local/bin/cutadapt
sudo ln -s /home/arun/software/bowtie-1.2.3-linux-x86_64/bowtie /usr/local/bin/bowtie
sudo ln -s /home/arun/software/bowtie-1.2.3-linux-x86_64/bowtie-build /usr/local/bin/
$\impliesbowtie-build$
sudo ln -s /home/arun/software/bowtie-1.2.3-linux-x86_64/bowtie-inspect /usr/local/
$\impliesbowtie-inspect$
```

Downloading FASTQ files from NCBI:

- Search and start the terminal, follow the commands below:
- wget -c https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/2.10.8/sratoolkit.2.10.8-mac64.tar.gz
- tar -xvzf sratoolkit.2.10.8-mac64.tar.gz
- cd sratoolkit.2.10.8-mac64/bin
- pwd

- /home/arun/software/sratoolkit.2.10.8-ubuntu64/bin

· Add to .bashrc

- cd

- vi .bashrc or gedit .bashrc and add the following line at the bottom of the page

```
- export PATH=$PATH:"/home/arun/software/sratoolkit.2.10.8-ubuntu64/bin"
```

Save and exit. After that type bash on the command line -Or- simply, close the terminal.

Obtaining and installing GUI application

• Download GUI for Linux

Uninstalling miRge3.0

To uninstall open the terminal and type:

```
python3.8 -m uninstall mirge3
```

2.1.2 macOS

Welcome to installation protocol for Mac OS

System prerequisites

- · Search and start the terminal, execute the following commands
- ruby -e "\$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/ master/install)"
- brew update
- brew install wget

Install python3.7

Please note, any version other than py3.7 causes error in Mac with multiprocessing, issues-1, issues-2. Download python 3.7.5 from python.org

• Search and start the terminal, execute the following commands

```
wget https://www.python.org/ftp/python/3.7.5/python-3.7.5-macosx10.9.pkg
sudo installer -pkg python-3.7.5-macosx10.9.pkg -target /
```

Mac comes with python2.7 installed by default. To use python3.7, creating an alias in .bash_profile would do the trick Open a new terminal window. Use vim editor if you are familiar using this editor vi .bash_profile or open the .bash_profile using text editor by open -e .bash_profile and add the following line at the bottom of the text.

```
alias python=python3.7
```

Save and exit. After that type source ~/.bash_profile on the command line -Or- simply, close the terminal.

Install R

· Search and start the terminal, execute the following command

brew install r

Installing miRge3.0

First install miRge dependenceis

• Search and start the terminal, execute the following command

If you encounter a WARNING, like below:

```
WARNING: The script cutadapt is installed in '/Users/loaneruser/Library/Python/3.7/bin
→' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this warning,
→use --no-warn-script-location.
```

Then, open a new terminal window or type cd to get to home directory. Add bin folder PATH to the .bash_profile, as shown below: Example: export PATH=\$PATH:"/Users/loaneruser/Library/Python/3.7/bin/" Remeber to add your path /PATH_TO_USERS/Python/3.7/bin.

Install miRge3.0 by this simple command

```
python3.7 -m pip install --user mirge3
```

To upgrade miRge3.0

python3.7 -m pip install --user --upgrade mirge3

Install additional C-libraries based tools

Install Bowtie

- · Search and start the terminal, execute the following command
- Download bowtie

- unzip bowtie-1.2.3-macos-x86_64.zip
- cd bowtie-1.2.3-macos-x86_64
- pwd
 - /Users/loaneruser/Software/bowtie-1.2.3-macos-x86_64
- Add these bowtie binaries to .bash_profile as shown below:

export PATH=\$PATH:"/Users/loaneruser/Software/bowtie-1.2.3-macos-x86_64/"

• After that type source ~/.bash_profile on the command line -Or- simply, close the terminal.

Install Samtools

• Search and start the terminal, execute the following command brew install samtools

Install RNA Fold

- wget "https://www.tbi.univie.ac.at/RNA/download/sourcecode/2_4_x/ViennaRNA-2.4.16.tar.gz"
- cd ViennaRNA-2.4.16

```
sudo ./configure
sudo make
sudo make install
```

Downloading FASTQ files from NCBI:

- · Search and start the terminal, execute the following command
- wget -c https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/2.10.8/sratoolkit.2.10.8-mac64.tar.gz
- tar -xvzf sratoolkit.2.10.8-mac64.tar.gz
- cd sratoolkit.2.10.8-mac64/bin
- pwd

```
- /Users/loaneruser/Software/sratoolkit.2.10.8-mac64/bin
```

- Add to .bash_profile
 - **-** cd
 - vi .bash_profile or open -e .bash_profile and add the following line at the bottom of the page
 - export PATH=\$PATH:"/Users/loaneruser/Software/sratoolkit.2.10.8-mac64/ bin"

Save and exit. After that type source ~/.bash_profile on the command line -Or- simply, close the terminal.

GUI requirements

Providing system wide access to miRge3.0, cutadapt, bowtie and bowtie-build, please type or (copy and paste) and submit each of the following commands on the terminal: **NOTE:** Make sure to change your path to python bin folder; Replace /Users/loaneruser/Library/ with /Path on your computer/.

· Search and start the terminal, execute the following command

Obtaining and installing GUI application

• Download GUI for OSX

Uninstalling miRge3.0

To uninstall open the terminal and type:

```
python3.8 -m uninstall mirge3
```

2.1.3 Windows OS

Welcome to installation protocol for Windows OS

System prerequisites

- Require Windows 10
- Require WSL and Ubuntu 18

Install WSL

Please follow one of the following guidlines for installing WSL and Ubuntu 18.04 (recommended Ubuntu distribution)

- Quick and easy way
 - TopTechSkills: Watch the first 1:30 seconds, more info.
 - Patreon: Watch the first 4:04 seconds.
- Official windows page.
- Please remember the password prompted during ubuntu installation and use when prompted.

Install python3.8 and R



- E-llaw the common de te undete chunter and install without 2
- Follow the commands to update ubuntu and install python 3.8 A password will be prompted when you type sudo, use the one you have set during Ubuntu installation.

```
sudo apt update
sudo apt install software-properties-common
sudo add-apt-repository ppa:deadsnakes/ppa
sudo apt install python3.8
sudo apt install python3-setuptools
sudo apt install python3-pip
sudo apt install r-base
```

Linux (Ubuntu 18.04) comes with python2.7 installed by default. To use python3.8, creating an alias in .bashrc would do the trick

Use vim editor if you are familiar using this editor vi .bashrc or open the .bashrc using text editor by gedit .bashrc and add the following line at the bottom of the text. alias python=python3.8

Save and exit. After that type bash on the command line -Or- simply, close the terminal.

Installing miRge3.0

First install miRge dependenceis

· Search and start Ubuntu, execute the following command

```
python3.8 -m pip install --user cutadapt==2.7 reportlab==3.5.42 biopython==1.77 _

→scikit-learn==0.23.1 hypothesis==5.15.1 pytest==5.4.2 scipy==1.4.1 matplotlib==3.

→2.1 joblib==0.15.1 pandas==1.0.3 future==0.18.2
```

If you encounter a WARNING, like below:

```
WARNING: The script cutadapt is installed in '/home/arun/.local/bin' which is not on_

→PATH.

Consider adding this directory to PATH or, if you prefer to suppress this warning,

→use --no-warn-script-location.
```

Then, open a new terminal window or type cd to get to home directory. Add bin folder PATH to the .bashrc, as shown below: Example: export PATH=\$PATH:"/home/arun/.local/bin" Remeber to add your path /PATH_TO_USERS/bin.

Install miRge3.0 by this simple command

```
python3.8 -m pip install --user mirge3
```

To upgrade miRge3.0

python3.8 -m pip install --user --upgrade mirge3

Install additional C-libraries based tools

Install Bowtie

- · Search and start Ubuntu, execute the following command
- Download bowtie

- unzip bowtie-1.2.3-macos-x86_64.zip
- cd bowtie-1.2.3-macos-x86_64

• pwd

- /home/arun/software/bowtie-1.2.3-linux-x86_64
- Add these bowtie binaries to .bashrc as shown below:

export PATH=\$PATH:"/home/arun/software/bowtie-1.2.3-linux-x86_64"

• After that type bash on the command line -Or- simply, close the terminal.

Install Samtools

• Search and start Ubuntu, execute the following command sudo apt install samtools

Install RNA Fold

- wget "https://www.tbi.univie.ac.at/RNA/download/sourcecode/2_4_x/ViennaRNA-2.4.16.tar.gz"
- cd ViennaRNA-2.4.16

sudo ./configure
sudo make
sudo make install

GUI requirements

Providing system wide access to miRge3.0, cutadapt, bowtie and bowtie-build, please type or (copy and paste) and submit each of the following commands on the terminal: **NOTE:** Make sure to change your path to python bin folder; Replace /home/arun/.local/with/Path on your computer/.

· Search and start Ubuntu, execute the following command

Change Command Prompt Properties

One last thing to avoid an error The directory name is invalid:

• Type cmd in Windows search box, right-click on Command Prompt and select Open file location.

All Apps Documents W	/eb More 🔻		5 😨 🔊
Best match			
Command Prompt			64,
Арр	G Run as adminis	trator	_
Apps	🛛 Open file locati	ion	Command Prompt
Node.js command prompt	-⇔ Pin to Start		Арр
x86_x64 Cross Tools Comma Drompt for VS 2017	-⇔ Pin to taskbar		
		Ľ	Open
x64_x86 Cross Tools Comma Prompt for VS 2017	and >	5	Run as administrator
Developer Command Promu	at for VS		Open file location
2017	>	-17	Pin to Start
x86 Native Tools Command for VS 2017	Prompt >	-12	Pin to taskbar
Intel® Graphics Command	Center >		
Search the web			
℅ command - See web results	>		
Documents (9+)			
Settings (7+)			
℅ command Prompt			

	Manage	Mana	ige	Windows System
View	Shortcut Tools	Applicatio	n Tools	
PC → Lo Nar	ocal Disk (C:) > Us	ers > arun	> App[Data > Roaming > Microsoft > Windows
77 77 78 78 78 78 78 78 78 78 78 78 78 7	Command Promp Control Panel File Explorer Run This PC Windows Adminis	t trative	Open Run wit Open fi Add to Add to Compre Run as a Share w Pin to S Edit wit Scan wi	th graphics processor file location o archive o "Command Prompt.rar" ress and email ress to "Command Prompt.rar" and email administrator with Skype Start ith Notepad++ vith Windows Defender
			Restore Send to Cut Copy Create s Delete	e previous versions o
			Rename	ne
			Propert	ties
9 KB				
		. 🧿	шì	

- Right-click on Command Prompt and click on Properties.
- Under the Shortcut tab, replace Start in option by changing the value <code>%HOMEDRIVE%%HOMEPATH%</code> to

	🚰 Commar	nd Pro	ompt P	roperties			×
	Terminal		Sec	urity	Details	Previous \	/ersions
	General	Sho	ortcut	Options	Font	Layout	Colors
	C:\	Cor	mmand	Prompt			
	Target type		Applica	tion			
	Target loca	tion:	system	32			
	Target:		%windi	r%\system3	32\cmd.exe		
				pupel.			_
	Start in:		%WIN	DIR%			
	Shortcut ke	ey:	None				
	Run:		Norma	l window			\sim
	Comment:		Perform	ns text-base	ed (command-	line) functions.	
	Open F	ile Lo	cation	Chan	ge Icon	Advanced	
				OF	((Cancel	Apply
lick OK							

- Reference 1. Stellarinfo 2. Microsoft

Obtaining and installing GUI application

• Download GUI for Windows 10

Double	click	miRge3.0.exe	to	install	miRge.	3.0	windows	GUI	application.
Name				Date modified		Туре		Size	
📙 .icon	i-ico			10/25/2020 5:0	6 PM	File fold	er		
📙 win-	unpacked			10/25/2020 5:0	6 PM	File fold	er		
📄 build	der-effectiv	e-config.yaml		10/25/2020 5:0	6 PM	YAML Fi	le	1 KB	
📄 miRg	ge3.0 Setup	0.0.1.exe.blockmap		10/25/2020 5:0	7 PM	BLOCKN	1AP File	38 KB	
🔣 miRg	ge3.0.exe			10/25/2020 5:0	7 PM	Applica	tion	34,726 KB	

<u>9</u> Rg	miRg	e3.0	Setu	р
-	_			

License Agreement

Please review the license terms before installing miRge3.0.

Press Page Down to see the rest of the agreement.

MIT License

Copyright (c) 2020 Arun H. Patil and Marc K. Halushka

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install miRge3.0.

miRge3.0 0.0.1 ---

I Agree

Can

• Click Next to complete miRge3.0 installation

Uninstalling miRge3.0

• Step 1: To uninstall open the terminal and type:

python3.8 -m uninstall mirge3

• Step 2:



– Under

off



- Then select Uninstall by clicking Ok. Done.

2.2 User guide

2.2.1 Parameters

To view command-line parameters type miRge3.0 -h:

```
usage: miRge3.0 [options]
miRge3.0 (Comprehensive analysis of small RNA sequencing Data)
optional arguments:
 -h, --help show this help message and exit
  --version show program's version number and exit
Options:
                             list of one or more samples separated by comma or a
 -s,
        --samples
→file with list of samples separated by new line (accepts *.fastq, *.fastq.gz)
 -db, --mir-DB
                             the reference database of miRNA. Options: miRBase and
→miRGeneDB (Default: miRBase)
                            the path to miRge libraries
 -lib, --libraries-path
 -on,
        --organism-name
                            the organism name can be human, mouse, fruitfly,
→nematode, rat or zebrafish
 -ex, --crThreshold
                            the threshold of the proportion of canonical reads for
\rightarrowthe miRNAs to retain. Range for ex (0 - 0.5), (Default: 0.1)
 -phr, --phred64
                            phred64 format (Default: 33)
                             switch to annotate spike-ins if spike-in bowtie index.
 -spk, --spikeIn
→files are located at the path of bowtie's index files (Default: off)
       --isoform-entropy switch to calculate isomir entropy (default: off)
 -ie.
 -cpu, --threads
                            the number of processors to use for trimming, qc, and
→alignment (Default: 1)
 -ai, --AtoI
                            switch to calculate A to I editing (Default: off)
 -tcf --tcf-out
                            switch to write trimmed and collapsed fasta file
\hookrightarrow (Default: off)
 -qff
       --qff-out
                           switch to output isomiR results in gff format (Default:
⇔off)
 -bam
        --bam-out
                            switch to output isomiR results in gff format (Default:
\hookrightarrow off)
 -trf
        --tRNA-frag
                            switch to analyze tRNA fragment and halves (Default:
⇔off)
 -0
        --outDir
                             the directory of the outputs (Default: current_
→directory)
```

(continues on next page)

(continued from previous page)

```
-shh
        --quiet
                             enable quiet/silent mode, only show warnings and errors_
\leftrightarrow (Default: off)
Data pre-processing:
 -a, --adapter
                             Sequence of a 3' adapter. The adapter and subsequent_
→bases are trimmed
        --front
                             Sequence of a 5' adapter. The adapter and any preceding_
 -a.
→bases are trimmed
 -u,
        --cut
                             Remove bases from each read. If LENGTH is positive,
\rightarrowremove bases from the beginning. If LENGTH is negative, remove bases from the end
 -nxt, --nextseq-trim NextSeq-specific quality trimming (each read). Trims
⇔also dark cycles appearing as high-quality G bases
 -q, --quality-cutoff
                           Trim low-quality bases from 5' and/or 3' ends of each
→read before adapter removal. If one value is given, only the 3' end is trimmed
                             If two comma-separated cutoffs are given, the 5' end is,
⇔trimmed with the first cutoff, the 3' end with the second
      --length
 -1.
                             Shorten reads to LENGTH. Positive values remove bases_
\mathop{\hookrightarrow}\!\mathsf{at} the end while negative ones remove bases at the beginning. This and the following
                             modifications are applied after adapter trimming
 -NX,
        --trim-n
                             Trim N's on ends of reads
                           Discard reads shorter than LEN. (Default: 16)
        --minimum-length
 -m,
                             Removes PCR duplicates and trim UMI of length by
 -umi, --uniq-mol-ids
\rightarrow specifying two comma-separated cutoffs as 5' cutoff,3' bp from both ends of the
→read. eg: 4,4 or 0,4
 -udd, --umiDedup
                             Specifies argument to removes PCR duplicates (Default:
-False); if TRUE it will remove UMI and remove PCR duplicates otherwise it only.
→remove UMI and keep the raw counts
 -umiq, --umiqiagen
                             Removes PCR duplicates of reads obtained from Qiagen,
⇔platform (Default: Illumina; "-umi x,y " Required)
Predicting novel miRNAs:
 The predictive model for novel miRNA detection is trained on human and mouse!
 -nmir, --novel-miRNA
                          include prediction of novel miRNAs
                             the minimum length of the reatined reads for novel.
 -minl, --minLength
→miRNA detection (default: 16)
 -maxl, --maxLength
                             the maximum length of the reatined reads for novel.
→miRNA detection (default: 25)
 -c, --minReadCounts the minimum read counts supporting novel miRNA.
→detection (default: 2)
 -mloc, --maxMappingLoci the maximum number of mapping loci for the retained.
→reads for novel miRNA detection (default: 3)
 -sl,
       --seedLength
                             the seed length when invoking Bowtie for novel miRNA.
\rightarrow detection (default: 25)
 -olc, --overlapLenCutoff the length of overlapped sequnce when joining reads.
⇒into longer sequences based on the coordinate
                             on the genome for novel miRNA detection (default: 14)
                             the maximum length of the clustered sequences for novel.
 -clc, --clusterLength
→miRNA detection (default: 30)
Optional PATH arguments:
                            the path to system's directory containing bowtie binary
 -pbwt, --bowtie-path
 -psam, --samtools-path
                            the path to system's directory containing samtools
⇔binary
 -prf, --RNAfold-path
                            the path to system's directory containing RNAfold binary
```

2.2.2 miRge3.0 libraries

miRge3.0 pipeline aligns the raw reads against a set of small-RNA annotation libraries. The libraries specific to the organism of interest can be obtained from SourceForge. Downloading the libraries on terminal:

Command-line Interface (CLI)

We recommend to create a directory miRge3_Lib and download using wget as shown below,

```
mkdir miRge3_Lib
cd miRge3_Lib
wget -O human.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/human.
→tar.gz/download"
wget -O mouse.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/mouse.
→tar.gz/download"
wget -O rat.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/rat.tar.
⇔gz/download"
wget -O nematode.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/

→nematode.tar.gz/download"

wget -O fruitfly.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/
⇔fruitfly.tar.gz/download"
wget -O zebrafish.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/
⇔zebrafish.tar.gz/download"
wget -O hamster.tar.gz "https://sourceforge.net/projects/mirge3/files/miRge3_Lib/

→hamster.tar.gz/download"
```

Users can download only what is necessary. Unzip the files once downloaded by the following command:

tar -xzf human.tar.gz

Replace human with the organism of interest. If you want to extract all the files at once, you could use tar -xzf *.tar.gz instead.

Graphical User Interface (GUI)

We recommend to create a folder miRge3_Lib and download the libraries directly from SourceForge. Once downloaded, extract/unzip the compressed files.

Building new libraries

If you are interested in creating specific library for an organism that is not part of this set then please refer to miRge3_build.

2.2.3 CLI - Example usage

Example command usage:

```
miRge3.0 -s SRR772403.fastq,SRR772404.fastq,SRR772405.fastq,SRR772406.fastq -lib_

→miRge3_Lib -on human -db mirgenedb -o output_dir -gff -nmir -trf -ai -cpu 12 -a_

→illumina
```

Output command line:

bowtie version: 1.2.3 Samtools version: 1.7 RNAfold version: 2.4.14 Collecting and validating input files... miRge3.0 will process 4 out of 4 input file(s). Cutadapt finished for file SRR772403 in 2.5358 second(s) Collapsing finished **for** file SRR772403 **in** 0.0126 second(s) Cutadapt finished **for** file SRR772404 **in** 7.3542 second(s) Collapsing finished **for** file SRR772404 **in** 0.2786 second(s) Cutadapt finished **for** file SRR772405 **in** 11.0667 second(s) Collapsing finished **for** file SRR772405 **in** 0.8585 second(s) Cutadapt finished **for** file SRR772406 **in** 3.5771 second(s) Collapsing finished **for** file SRR772406 **in** 0.8677 second(s) Matrix creation finished in 0.3838 second(s) Data pre-processing completed in 27.2443 second(s) Alignment **in** progress ... Alignment completed in 15.8305 second(s) Summarizing **and** tabulating results... The number of A-to-I editing sites for is less than 10 so that no heatmap is drawn. Summary completed in 71.4691 second(s) Predicting novel miRNAs Performing prediction of novel miRNAs... Start to predict Prediction of novel miRNAs Completed (104.83 sec) The analysis completed in 222.2487 second(s)

2.2.4 miRge3.0 GUI

• The application is cross platform, the image below is a screenshot of the software from MacOS



• The software is easy to use with default parameters. The parameters are tabulated into four groups such as basic, trimming parameters, novel miRNA prediction and other optional parameters.

	•••					miRge3.0			
		Home	About C	Contact				Johns Hopkins Universi	ty School of Med
		Compr Doc Uplo	ehensive ana cumentation ad files ar	alysis of small RNA seque	ncing data. ers below:				
		Ba			Optional				
		Sele	ct reference d	database:	miRGeneDB	•			
		Sele	ct the organis	sm name:	Human	- 0			
		Path	to miRge libr	raries:	Choose Folder	No folder choosen	0		
		Inpu	t sequence file	le(s):	Choose File(s)	No file(s) choosen	0		
		Path	to output dire	rectory:	Choose Folder	No folder choosen	0		
		Num	ber of CPUs t	to use:	1	0			
		Exec	cute analysis p	pipeline:	Su	Ibmit Reset			
Screenshot with basic parameters	28		3	🎜 🤜	1	1	🚺 🔗 🕻	9 🗅 😨	

		miRge3.0	
	Home About Contact		Johns Hopkins University School
	miRoe		
	3.0		
	Comprehensive analysis of small RNA sequen	cing data.	
	Documentation »		
	Upload files and select paramete	rs below:	
	Basic Trimming Novel miRNA	Optional	
	Adapter trimming parameters ($oldsymbol{arepsilon}$	Restore default)	
	Unique Molecular Identifiers (UMI):	• No UMI 🔷 Illumina UMI 🔷 Qiagen UMI	
	Trim - Sequence of a 3' adapter	TGGAATTCTCGGGTGCCAAGGAACTCCAG	3
	Trim - Sequence of a 5' adapter		0
			-
	Enable NextSeq-specific quality trimming	20 🚺	
	Enable low-quality bases trimming	20 3	
	Shorten reads to LENGTH	21 3	
		0	
	Trim N's on ends of reads	0	
ing parameters	and the second		
· 1			
		miRge3.0	
	Home About Contact		Johns Hopkins Univer
-			
	miBro		
	miRge		
	Comprehensive analysis of small RNA	sequencing data.	
	Comprehensive analysis of small RNA	sequencing data.	
	Comprehensive analysis of small RNA and the Documentation »	sequencing data.	
	Comprehensive analysis of small RNA a Documentation »	sequencing data. meters below:	
	Comprehensive analysis of small RNA and Documentation » Upload files and select param	sequencing data. meters below:	
	Comprehensive analysis of small RNA s Documentation » Upload files and select paran Basic Trimming Novel mIRM	sequencing data. meters below: A Optional	
	Comprehensive analysis of small RNA s Documentation » Upload files and select paral Basic Trimming Novel miRM	sequencing data. meters below:	
	Comprehensive analysis of small RNA so Documentation > Upload files and select param Basic Trimming Novel miRNA Novel miRNA prediction para	sequencing data. meters below: VA Optional ameters (& Restore defaut)	
	Comprehensive analysis of small RNA a Documentation » Upload files and select param Basic Trimming Novel miRNA Novel miRNA prediction para	sequencing data. meters below: VA Optional ameters (C Restore default) Off	
	Comprehensive analysis of small RNA as Documentation >> Upload files and select parama Basic Trimming Novel miRNA Novel miRNA prediction parama	sequencing data. meters below: VA Optional ameters (C Restore default) Off	
	Comprehensive analysis of small RNA s Documentation » Upload files and select parat Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length:	sequencing data. meters below: VA Optional ameters (C Restore defaut) Off 16 3	
	Comprehensive analysis of small RNA s Comprehensive analysis of small RNA s Documentation » Upload files and select parar Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length:	sequencing data. meters below: VA Optional ameters (2 Restore defaut) Off 16 0 25 0	
	Comprehensive analysis of small RNA : Documentation » Upload files and select paran Basic Trimming Novel miRNA Detect novel miRNAs: Minimum Length:	sequencing data. meters below: VA Optional ameters (C Restore default) Off 16 25 6	
	Comprehensive analysis of small RNA s Comprehensive analysis of small RNA s Documentation » Upload files and select parar Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Maximum Length:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1	
	Comprehensive analysis of small RNA as Documentation » Upload files and select parar Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Maximum Length:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 2 6	
	Comprehensive analysis of small RNA is Documentation » Upload files and select paran Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Minimum Length: Minimum read counts:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 2 3 6	
	Comprehensive analysis of small RNA : Documentation » Upload files and select parar Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Maximum Length: Maximum mapping loci: Seed length:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 3 6 25 6 3 6	
	Comprehensive analysis of small RNA : Documentation » Upload files and select parar Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Maximum Length: Maximum mapping loci: Seed length:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 2 3 6 25 6 3 6	
	Comprehensive analysis of small RNA : Documentation » Upload files and select parara Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Maximum Length: Maximum mapping loci: Seed length: Overlap length cutoff;	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 2 3 6 25 6 14 6	
	Comprehensive analysis of small RNA : Documentation » Upload files and select parara Basic Trimming Novel miRNA Novel miRNA prediction para Detect novel miRNAs: Minimum Length: Maximum Length: Maximum mapping loci: Seed length: Overlap length cutoff:	sequencing data. meters below: A Optional armeters (C Restore default) Off 16 25 6 2 3 6 25 6 14 6 6 7 7 7 7 7 7 8 7 7 7 8 7 8 7 8 7 8 7 8 8 7 8 8 8 8 8 8 8 8 8 8 8 8 8	
	Comprehensive analysis of small RNA : Documentation » Upload files and select parara Basic Trimming Novel miRNA Detect novel miRNA prediction parara Detect novel miRNAs: Minimum Length: Maximum mapping loci: Seed length: Overlap length cutoff: Cluster length:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 3 3 2 14 14 30 0 1 4	
2NA predictions	Comprehensive analysis of small RNA : Comprehensive analysis of small RNA : Documentation » Upload files and select parara Basic Trimming Novel miRNA Detect novel miRNA prediction parara Detect novel miRNAs: Minimum Length: Maximum Length: Maximum mapping loci: Seed length: Overlap length cutoff: Cluster length:	sequencing data. meters below: A Optional ameters (C Restore default) Off 16 25 6 2 3 6 25 6 14 14 6 30 6	



2.2.5 Resources

- Lu, Y., et al., miRge 2.0 for comprehensive analysis of microRNA sequencing data. 2018. BMC Bioinformatics. PMID.
- Baras, S. A., et al., miRge A Multiplexed Method of Processing Small RNA-Seq Data to Determine MicroRNA Entropy. 2015. *PLoS One*. PMID.

2.3 miRge3.0 output

2.3.1 Command and sample run with UMI datasets

```
miRge3.0 -s SRR8557389.fastq,SRR8557396.fastq,SRR8557398.fastq,SRR8557399.fastq -lib_

→miRge3_Lib -on human -db miRGeneDB \

-o temp -a AACTGTAGGCACCATCAAT -udd --qiagenumi -umi 0,12 -cpu 12 -q 20 -NX -

→nmir -minl 16 -maxl 25 -c 2 \

-mloc 3 -sl 25 -olc 14 -clc 30 -gff

bowtie version: 1.2.3

cutadapt version: 2.7

Samtools version: 1.7

RNAfold version: 2.4.14

Collecting and validating input files...

miRge3.0 will process 4 out of 4 input file(s).
```

(continues on next page)

(continued from previous page)

```
Cutadapt finished for file SRR8557389 in 21.0854 second(s)
Collapsing finished for file SRR8557389 in 0.0699 second(s)
Cutadapt finished for file SRR8557396 in 10.305 second(s)
Collapsing finished for file SRR8557396 in 0.6016 second(s)
Cutadapt finished for file SRR8557398 in 10.891 second(s)
Collapsing finished for file SRR8557398 in 0.911 second(s)
Cutadapt finished for file SRR8557399 in 14.2126 second(s)
Collapsing finished for file SRR8557399 in 1.1292 second(s)
Matrix creation finished in 0.4788 second(s)
Data pre-processing completed in 62.762 second(s)
Alignment in progress ....
Alignment completed in 16.9863 second(s)
Summarizing and tabulating results...
Summary completed in 7.8131 second(s)
Predicting novel miRNAs
Performing prediction of novel miRNAs...Start to predictPrediction of novel miRNAs_
→Completed (220.35 sec)
The analysis completed in 310.7281 second(s)
```

2.3.2 Output tree structure

An output directory is created for each run such as miRge.2020-10-9_1-35-53, where the name is followed by date time format miRge.yy-dd-mm-hr-mm-ss.

```
The following output is in general, however, the resultant output files are based on_
⇔the options selected during miRge3.0 execution.
miRge.2020-10-9_1-35-53
  - run.log (Gives the detailed log of miRge3.0 execution)
  - unmapped.log (Gives the detailed log of novel miRNA prediction)

    mapped.csv (CSV file with read counts across each smallRNA library)

    unmapped.csv (CSV file with unaligned/mapped reads)

 — annotation.report.csv (Basic annotation report with small RNA distribution in CSV,
\rightarrow format)
 - annotation.report.html (Basic annotation report with small RNA distribution in_
→HTML format)
sample_miRge3.gff (GFF file with reads with isomiRs across one or more samples,
\rightarrow if -gff option selected)
miR.Counts.csv (miRNA raw read counts across samples)
  - miR.RPM.csv (miRNA Read Per Million - RPM counts across samples)
 — *_umiCounts.csv (Counts for each unique UMI for each sample)

    index_data.js (Javascript file with data generated for visualization)

  - miRge3_visualization.html (HTML for data visualization)
 - FOLDER_novel_miRNAs

    - *.pdf (novel miRNA structure in PDF format for each miRNA)

    \hookrightarrowCSV format)
 — a2IEditing.detail.txt
  - a2IEditing.report.csv
  - a2IEditing.report.newform.csv
```

(continues on next page)

(continued from previous page)

2.3.3 miRge - interactive visualization

miRge3.0 produces several interactive visualization graphics as follows

			miRge3.0: 0	Comprehensive analysis	of small RNA sequencing Data.
• Screenshot of the miRge visualizati	on HTML tab	M SmallRNA distribution	Read Length	isomiR results	Abundant miRNAs
Servenshot of the hintege visualizati					
		≡			
	View in ful	l screen			
	Print chart				
-	Download I	PNG image			
	Download I	IPEG image			
	Download I	PDF document			
	Download S	SVG vector image			
-	Download (CSV			
	Download 2	XLS			
Chart view and download options	View data t	able			



Read distribution

• Screenshot of the smallRNA read distribution for each sample

SRR8557389: Read Length Dis



- Screenshot of the read length distribution for each sample
- Screenshot of the tile map representing top 40 high abundant miRNAs for each study



• Screenshot of the variant distribution for all samples combined (isomiRs) Cumulative isomiR variant type distribution of the samples



• Screenshot of the heatmap representing variants for each sample for the top 20 high abundant miRNAs (isomiRs)



• Screenshot of the histogram representing UMI counts across each sample



Show	10 v entries		Se	arch:	
id	Name	Probability	Chr	Start pos.	En Pos
1	SRR8557389_novel_miRNA_1	0.9999987516742271	chr2	134127130	13412
2	SRR8557389_novel_miRNA_2	0.9999915571402658	chr1	172138858	17213
3	SRR8557389_novel_miRNA_3	0.999982270847492	chr3	113594918	113594
4	SRR8557389_novel_miRNA_4	0.9963371234941661	chr8	27433413	274334
5	SRR8557389_novel_miRNA_5	0.9938801463496213	chr17	81125886	811259
6	SRR8557389_novel_miRNA_6	0.9908373862619209	chr17	74748663	74748
7	SRR8557389_novel_miRNA_7	0.9905492996218249	chr12	69584745	69584
8	SRR8557389_novel_miRNA_8	0.9781204295092701	chr22	20086072	20086
9	SRR8557389_novel_miRNA_9	0.9723790135555249	chr5	141849784	141849
10	SRR8557389_novel_miRNA_10	0.9655899863443722	chr19	13836292	138363
Showir	ng 1 to 10 of 26 entries			Previous	1

• Screenshot of a list of novel miRNAs identified across samples

2.3.4 Resources:

The graphics for miRge3.0 visualization is enabled with javascripts and CSS obtained from the following:

- Interactive charts from HighCharts
- Icons from Font Awesome
- Interactive HTML table

2.4 MIT License

Copyright (c) 2020 Arun H. Patil and Marc K. Halushka

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PAR-TICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFT-WARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.