https://doi.org/10.1093/nargab/lqab068

miRge3.0: a comprehensive microRNA and tRF sequencing analysis pipeline

Arun H. Patil and Marc K. Halushka*

Department of Pathology, Division of Cardiovascular Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received April 13, 2021; Revised July 02, 2021; Editorial Decision July 13, 2021; Accepted July 15, 2021

ABSTRACT

MicroRNAs and tRFs are classes of small non-coding RNAs, known for their roles in translational regulation of genes. Advances in next-generation sequencing (NGS) have enabled high-throughput small RNA-seq studies, which require robust alignment pipelines. Our laboratory previously developed miRge and miRge2.0, as flexible tools to process sequencing data for annotation of miRNAs and other small-RNA species and further predict novel miRNAs using a support vector machine approach. Although miRge2.0 is a leading analysis tool in terms of speed with unique quantifying and annotation features, it has a few limitations. We present miRge3.0 that provides additional features along with compatibility to newer versions of Cutadapt and Python. The revisions of the tool include the ability to process Unique Molecular Identifiers (UMIs) to account for PCR duplicates while quantifying miRNAs in the datasets, correct erroneous single base substitutions in miRNAs with miREC and an accurate mirGFF3 formatted isomiR tool. miRge3.0 also has speed improvements benchmarked to miRge2.0, Chimira and sRNAbench. Finally, miRge3.0 output integrates into other packages for a streamlined analysis process and provides a cross-platform Graphical User Interface (GUI). In conclusion miRge3.0 is our third generation small RNA-seq aligner with improvements in speed, versatility and functionality over earlier iterations.

INTRODUCTION

MicroRNAs (miRNAs) are a group of small non-coding RNA that act as master regulators of coding RNA translation (1). Altered miRNA expression affects cell stress signaling, cell proliferation and is central to human disease states (2). Each miRNA can target scores of mRNAs and have been found to regulate signaling pathways (3). In humans, the number of miRNAs is controversial with only 1111 miRNAs reported in miRGeneDB 2.0 (4) but 2656 miRNAs in the miRBase database (version 22) (5). Other groups list hundreds or thousands of additional small RNAs that are noncanonical miRNAs (6–8). The exploration of miRNA expression profiles has added valuable insights into mechanisms of cell processes in health and disease (2,9).

tRNA fragments and halves have more recently become of interest through their role in pathophysiology (10). Although generated by different processes and having different lengths, for the purpose of characterizing these entities as being derived from tRNAs, they will be collectively considered as tRFs. tRFs are thought to increase based on cell stressors and have variable expression patterns based on cell type and method of cell stress (11). Some functions overlap with miRNA activities. Other interactions with RNA binding proteins impact on diseases such as cancer (12).

Advances in genomics has enabled cost-effective high-throughput sequencing from small RNA libraries to study tissue (13,14) and cell (8,15) expression. Within small RNA-seq datasets, in addition to miRNAs and tRFs, other types of RNA such as rRNA, siRNA, snoRNA and mRNA fragments exist, some of whose expressions are variable in disease (16). In order to accurately identify and quantify sequence data, the reads must be aligned to reference sequences with appropriate parameters. Additionally, for miRNAs, a collection of nearly similar sequences, termed isomiRs, sum up to the reads of a particular miRNA (17).

In 2015, our laboratory published a small RNA-seq alignment tool, miRge, focused on miRNA expression (18). It was designed as a fast, smart alignment tool coded in Perl. An update, miRge2.0, coded in Python 2.7, introduced new features including novel miRNA discovery, isomiR description in the GFF3 format, tRNA fragment characterization and detection of A-to-I changes (19). With the deprecation of Python 2, with no new bug fixes and shift of python community support to a newer version, this could not be maintained. Here we present miRge3.0, coded in Python 3, as an even faster, more robust tool with greater functionality, a standardized format for reporting isomiRs, mirGFF3 (17), unique molecular identifier (UMI handling) and a new

*To whom correspondence should be addressed. Tel: +1 410 614 8138; Fax: +1 410 502 5862; Email: mhalush1@jhmi.edu graphical user interface (GUI) for both input and output data.

MATERIALS AND METHODS

The revised small-RNA analysis pipeline miRge3.0 is implemented in Python (v3.8) and is designed to run in Linux, macOS, and Windows 10 with windows subsystems for Linux (WSL). miRge3.0 uses miREC for correcting erroneous single-nucleotide substitutions in miRNAs (https://github.com/XuanrZhang/miREC), Cutadapt (v3.0) (20) for adapter trimming on both ends of FASTQ reads and downstream quality control. Pandas (v0.25.3) libraries have been implemented to enable memory efficient annotations of small RNA molecules by reducing I/O-based operations. In addition, miRge3.0 requires Bowtie (v1.3.0) (21), ViennaRNA (v2.4.16) (22), SAMtools (v1.7) (23), biopython (v1.78), sklearn (v0.23.1), numPy (v1.18.4), SciPy (v1.4.1) and reportlab (v3.5.42) (https://pypi.org/project/reportlab/) for several functionalities including alignment operations, novel miRNA discovery, generating a PDF summary and graphical reports. In addition to this, miRge3.0 is designed to handle reads with Unique Molecular Identifiers (UMI) and further integrates DESeq2 (release 3.12) (24) function in R (v4.0) to compute differential expression.

Workflow and alignment steps of miRge3.0

The initial step of miRge3.0 is the removal of bad quality reads and adapter sequences using Cutadapt. miRge3.0 allows for a wide range of adapters (both 5′ and 3′ types) to be natively called and removed. At this step, UMIs, if present, can be trimmed for a fixed length specified by the user. The PCR duplicates can be removed by considering only the unique counts of UMI and read sequence combinations. After this initial step, identical reads are collapsed into a single read and the counts are captured in a Pandas data frame. The Pandas data frame will join reads and corresponding read counts for two or more FASTQ samples to create a complete data frame. This data frame is used for downstream alignments and for generating various summary results. The overview of the workflow is shown in Figure 1. The detailed workflow is described in Supplementary Data S1, Section S1.

Novel miRNA prediction

Novel miRNA detection is based on a machine learning algorithm and on a prediction model, built using support vector machine (SVM). The features predicted in our previous version (19) were rebuilt using sklearn (v0.23.1) to support python 3. The model and novel miRNA prediction is as described previously (19). In brief, the unaligned reads are aligned and clustered to the genome. The most stable region of each cluster is extracted as a putative mature miRNA and the coordinates are set based on the cluster where the frequency of each base over the total number of reads in that cluster is >0.8. Further, the pre-miRNA hairpins based on the fold energy of the sequence surrounding the cluster is determined using RNAfold. A SVM model is applied on the features of these pre-miRNAs and a probability value is calculated determining the significance of identified putative miRNA.

miRge3.0 graphical user interface

The miRge3.0 suite offers a GUI with substantial ease to install and use. The cross-platform GUI is developed with Node package manager (NPM, v6.14.4), Node.js (v12.16.3) and Electron (v1.4.13). For running miRge3.0, the GUI menu allows all features of the command line request (Supplementary Figure S1). All features related to switching between parameters and its execution is enabled by JavaScripts and HTML tags.

After the miRge3.0 run is complete, a dynamic summary output is generated (Supplementary Figure S2). The interactive charts are rendered using JavaScript and CSS obtained from High Charts (https://www.highcharts.com/), the icons in the miRge visualization HTML tabs are obtained from Font Awesome (v5.8.2, https://fontawesome.com/) and the JavaScript for interactive HTML table depicting novel miRNAs is obtained from Data Tables (https://datatables.net/). The specifics of GUI build are described in Supplementary Data S1, Section S2.

Speed testing

To benchmark miRge3.0 to other tools, we utilized the publicly available SRA files described in Supplementary Data S1, Section S3. The miRge3.0 pipeline uses standard small RNA libraries as a reference and, to compare the performance, two tools with similar features, Chimira (25) and sRNAbench (26), were chosen. Supplementary Table S1 describes some common technical features among the chosen tools. Speed comparisons were performed in two stages. At the first stage, the samples were run with default parameters for all tools except for specifying Illumina 3′-adapter sequence TGGAATTCTCGGGTGCCAAGGAACTCCAG. The samples were uploaded in FASTQ.gz format to Chimira and accession IDs were provided to sRNAbench to download from NCBI SRA and perform the annotation. The samples with FASTQ extension were run for miRge2.0 and miRge3.0. The annotations were timed from start to completion for all tools excluding the sample download and upload time. Chimira does not report time for Reaper and Tally steps in their log file; therefore, the timings for each sample were monitored manually. sRNAbench converts 'FASTQ' files to 'FASTQ.gz' using command 'gzip' before performing annotations. We have reported two time series including gzip conversion time (sRNAbench) and without (sRNAbench - gzip).

Chimira and sRNAbench report modifications and isomiRs, respectively, along with basic annotations. In the second stage, although the miRge2.0 and miRge3.0 pipeline reports isomiR counts for each miRNA, for consistency, the miRge2.0 and miRge3.0 pipelines were supplemented with the '-gff' parameter to generate an isomiR GFF file. Further, Chimira and sRNAbench are hosted online and, while server specifications are unknown, they are expected to have large RAM capacity and numerous cores. sRNAbench uses 10 CPUs to download data from NCBI SRA and 4 CPUs for small RNA annotations. Thus, the miRge pipelines were

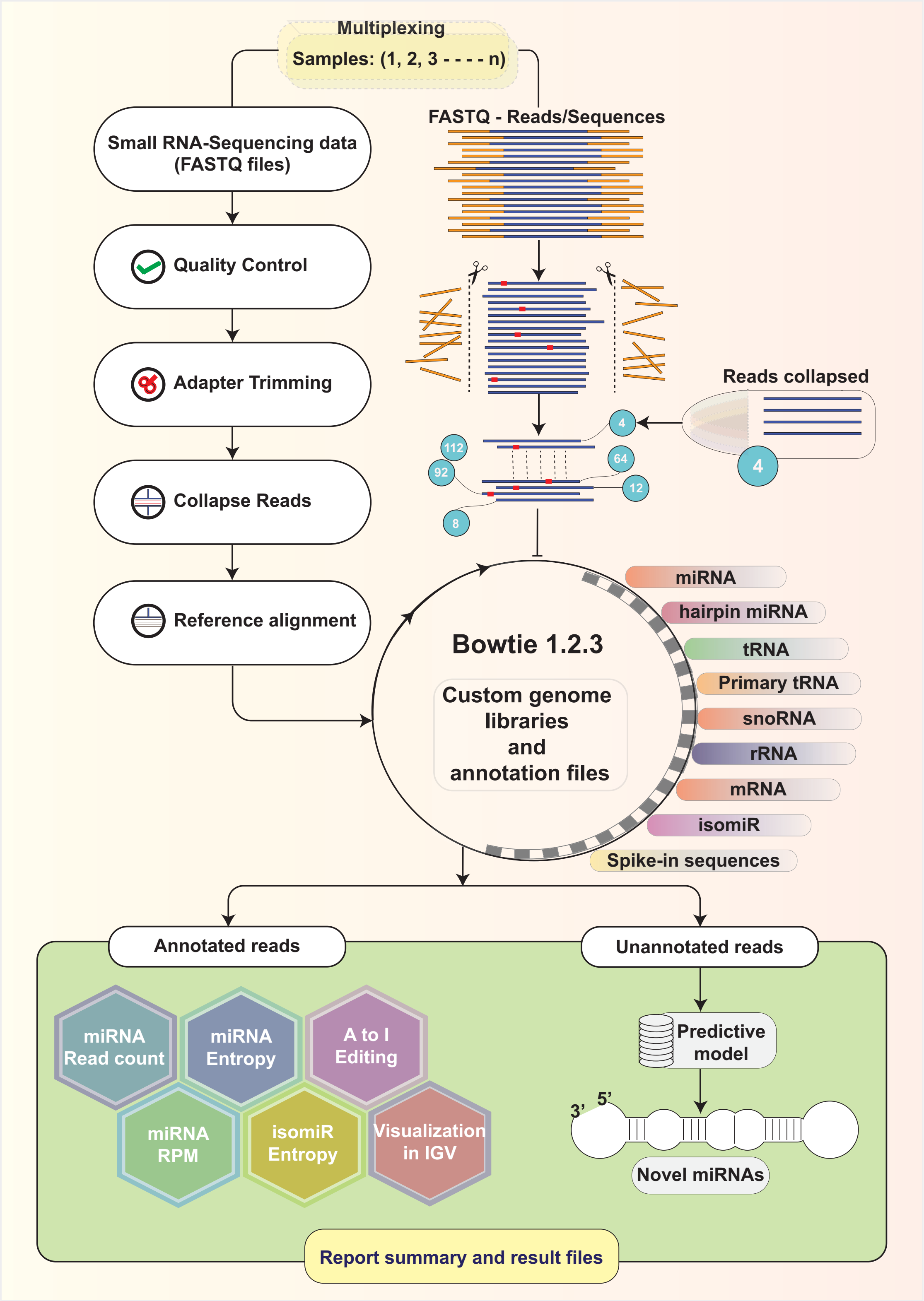

Figure 1. An overview of miRge3.0 workflow. A sample or samples (FASTQ, FASTQ.gz) are processed through a number of user-selected steps including quality control and adapter trimming. Identical reads are collapsed together and aligned to species-specific reference RNA libraries. Through multiple alignment steps, reads are assigned their appropriate RNA identity. Unaligned/unannotated reads can be sent to a predictive model to try to identify novel miRNAs. Annotated reads are outputted into a number of different files for downstream visualization and analysis.

run with 4 and 12 CPUs for each sample. Chimira has a maximum file upload limit of 1.6 GB precluding the running of sample SRR1028924.

Comparing 4N and Qiagen UMI output with or without deduplication

UMIs enable correction of PCR duplicates that arise during sample preparation. The workflow of UMI incorporated in miRge3.0 is described in Supplementary Figure S3. The usage parameters of UMI are further detailed in Supplementary Data S1, Section S4. Correcting for counts of the 4N ligation adaptor/UMI was performed on SRR6379839 (synthetic construct) (27), and SRR9115360 (NEXTflex, Human Brain) (28) and a Qiagen processed sample, SRR8557389 (29). The parameters for 4N method include -a TGGAATTCTCGGGTGCCAAG G (for adapter removal), -umi 4,4 (to trim 4 bases on both ends of the read after adapter removal). The parameters for Qiagen UMI include -a AACTGTAGGCACCATCAAT (for adapter removal), –umiqiagen (specifying Qiagen) and -umi 0,12 (To trim 0 bases at 5′ and 12 bases at 3′ after adapter removal). While the rest of the default parameters are kept constant for both runs, each dataset was run with and without the –umiDedup (remove duplicates) option. Other resources used in the development of miRge3.0 pipeline is provided in Supplementary Data S1, Section S5.

Differential expression analysis

miRge3.0 depends on DESeq2 (24), a Bioconductor R package, to estimate differential expression among the samples. A metadata of groups for control and condition should be supplemented in a file with tab-delimited text format, and miRNA read counts are used to perform the differential expression. If differential expression analysis is called, the analysis results appear in a text file reporting miRNA's and its corresponding log2 fold change values along with the *P*-value and adjusted P-value. Further, PDF files reporting a volcano plot of differentially expressed miRNAs and principal components analysis (PCA) plot of the input samples are automatically generated. An RData file is also generated for further integration with other tools and/or edit the output graphical format using R.

RESULTS

Features of miRge3.0

miRge3.0 is an updated/improved version over our previous version miRge2.0, with various changes being incorporated, the most significant ones among them are coding in Python 3, functionality with newer Cutadapt packages, miRNA error correction, differential expression determination, processing UMIs, and a GUI offering interactive graphical input and output. miRge3.0 also has a significant speed advantage over miRge2.0 due to efficient coding involving Pandas data frame, Cython implementation in Cutadapt and multiprocessing from python class ProcessPoolExecutor. miRge3.0 incorporates the small-RNA error correction tool, miREC. The choice of ideal parameters is tested and defined in Supplementary Data, Section S6. The command line version of miRge3.0 can be easily installed using pip and conda installation procedures (30).

Output files of miRge3.0

Multiple output files are generated in miRge3.0. There are both a file of mapped reads with counts for each of the small RNA types across samples processed and a similar file of unmapped reads. Separate files log read counts and reads per million (RPMs) for each miRNA species (miRBase or MirGeneDB). A log file is reported that includes the information of the executed command and parameters along with time stamp of that execution. Additional output files depend on user arguments/requirements. A Pandas data frame produces additional files such as isomiRs in mirGFF3 format, a BAM file for visualization in IGV, tRNA fragment annotations, A to I editing, and detection of novel miRNAs. Supplementary Figure S4 depicts the BAM tracks on IGV representing mapped canonical and isomiRs reads. At the end of the execution, miRge3.0 reports an HTML and a JavaScript file dynamically generated to report the summary of the analysis. By default, this summary includes an interactive graphic of percent read distribution for each sample as a horizontal stacked bar-graph (Supplementary Figure S5A), a histogram showing read length distribution (Supplementary Figure S5B) and a honeycomb (tile map) showing the top 40 abundant miRNAs (Supplementary Figure S6A). If specified, it will further include (i) a table of novel miRNAs identified across samples (Supplementary Figure S6B), (ii) cumulative isomiR variant type distribution in pie chart (Supplementary Figure S7A) and heat map showing read distribution of isomiRs for the top 20 abundant miRNAs (Supplementary Figure S7B) and (iii) a histogram showing the distribution of UMIs for each sample (Figure 2A–C). All of the summary file host interactive graphics allow users to view the data in a table format and enable the downloading of high-resolution images for presentation and/or publications.

UMI deduplication analysis

miRge3.0 has the ability to detect, remove and adjust for UMIs. Currently, two types of UMIs exist, those that act as ligation adaptors at the ends of the template (4N, Supplementary Figure S3A) and those within the adapter (Qiagen, Supplementary Figure S3B). The 12 bp Qiagen UMI, when collapsed with the entire read sequence, shows only rare duplications (Figure 2A). Due to ligation interactions of the shorter length 4N adaptor, the complete sequencing of short miRNA reads, along with the overall fewer nucleotides of the UMI ($N = 8$), many more copies of these 4N-style ligation UMIs can be seen in real (Figure 2B) or synthetic (Figure 2C) data. It is more likely the 'birthday paradox' is impacting on UMI distribution rather than true amplification error (see more details in Supplementary Data S1, Section S4). As a result of these differences, UMI removal, as performed for Qiagen-type longer UMIs has little effect on read counts (Figure 2D), whereas UMI removal on 4N samples can have a more significant effect (Figure 2E,F). Therefore, the adjustment for 4N UMIs, likely overcorrects and undercounts abundant miRNAs (28).

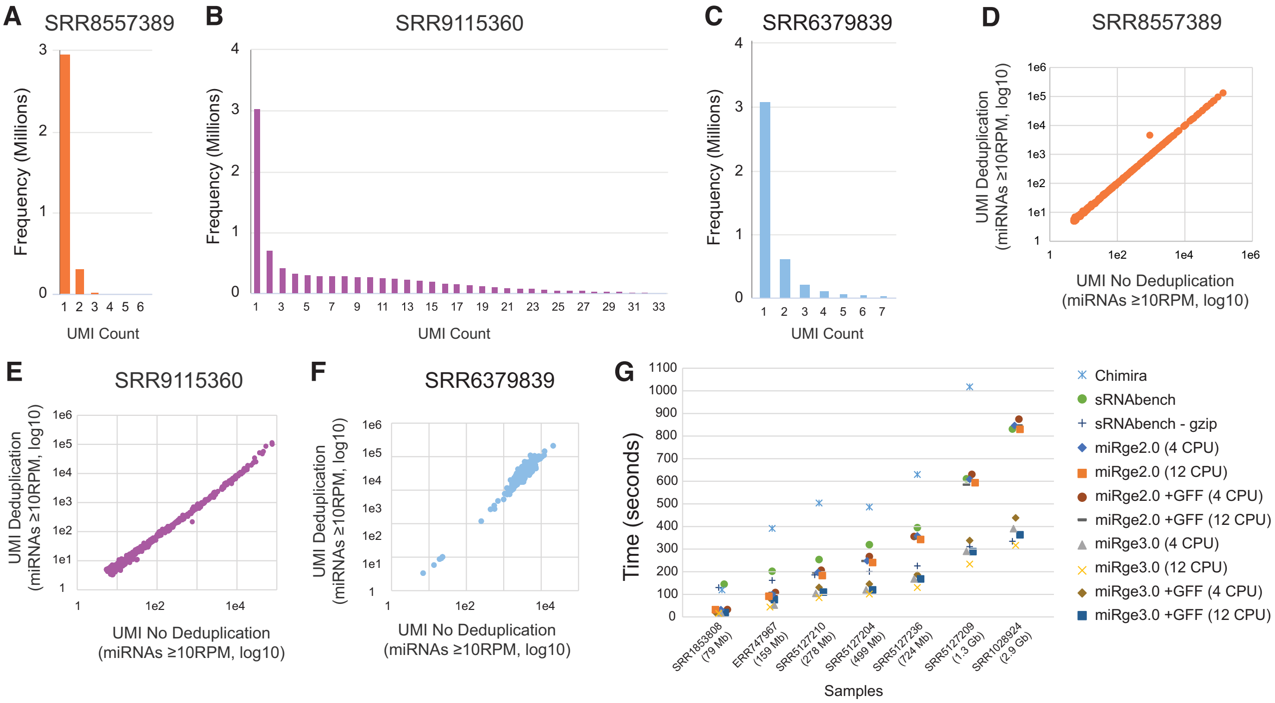

Figure 2. Comparison across tools and UMI analysis. (**A**) A FASTQ file with Qiagen UMIs showing few examples of duplication. (**B**) A FASTQ file with the 4N ligation adaptors showing significant counts of duplicated UMIs. (**C**) A FASTQ file of synthetic data showing some increase in UMIs. (**D**) The correlation between deduplicated and non-deduplicated miRNA RPM counts for Qiagen UMIs was very similar between repeated miRge3.0 runs (r^2 = 0.9996). (**E** and **F**) The correlation was lower for samples with 4N UMIs (r^2 = 0.95 and 0.81, respectively). (**G**) Run speed for seven samples ranging from 79 MB to 2.9 GB file sizes across four tools.

Speed comparison and abundance estimation across other alignment tools

Performance speed of miRge3.0 was compared with miRge2.0, Chimira and sRNAbench. SRR datasets with different read depths were used to monitor processing speed across these tools. As sRNAbench converts a raw 'FASTQ' file into a 'FASTQ.gz' file in an initial step, we provide time data for both options (sRNAbench and sRNAbench – gzip). miRge2.0 and 3.0 times were reported using 4 CPU or 12 CPU with or without the +GFF feature. Although all of these tools are fast relative to most other miRNA aligners, overall, miRge3.0 with 12 CPUs consistently had the best execution speed (Figure 2G). Non-default parameters, such as isomiR GFF reporting will compromise miRge3.0 speed. The miRNA abundance is expressed in terms of raw read counts or as normalized abundance (RPM). Table 1 shows the number of unique miRNAs expressed and those with ≥10 RPM for seven samples across the four alignment tools. sRNAbench, miRge2.0 and miRge3.0 all show similar trends in the reporting of miRNA counts. Chimira has much higher detection of miRNAs, which is partly the result of it not having the same false-positive controls employed by the other tools such as minimal read counts or percent canonical read controls.

DISCUSSION

miRge3.0 is a significant advance from our earlier, popular miRge and miRge2.0 alignment tools. The major enhancements are increased speed due to additional multithreading, the development of GUI assistance for both setting up the alignment parameters and for viewing the output, handling of UMIs, and usability on MacOS, Linux or Windows 10 platforms. This is in addition to useful features that were already in miRge2.0 which include a tRF detecting tool, native utilization of the GFF3 isomiR format, novel miRNA detection, and A-to-I editing detection. As a result of these features, we are confident that miRge3.0 is among the most useful small RNA aligners currently available.

A key aspect of all versions of miRge has been the use of specialized RNA libraries and iterative alignments. Aligning miRNAs and tRFs is challenging due to isomiRs, short lengths, modifications and multiple related sequences duplicated in the genomes (10,18). Tradeoffs have to be considered in setting up alignment parameters as no one method can accurately assign all sequences. We have found that a 'one size fits all' alignment method to the species' genome results in the poorest understanding of the sequenced material with the most error. Alignments are improved when reducing the search space by focusing on only the transcribed parts of the genome. The second improvement we take is to perform multiple sequential searches of the collapsed RNA sequences to specific RNA species in which the alignments start with very 'tight' parameters and ultimately become 'loose' to account for isomiRs.

Overall, miRge3.0 is a more user-friendly tool with advantages over our previous version and other tools. However, there are a few limitations. The 4N-style UMI analysis can benefit from a computational method to adjust UMI

Table 1. miRNA annotations across tools

Tissue/Cell	SRA references	Alignment tool	miRNA reads	Unique miRNAs	miRNAs > 10 RPM
T Cell CD8+ (Neonatal)	SRR1853808	Chimira	1 011 733	620	250
		sRNAbench	996 785	396	190
		miRge2.0	1 004 944	317	216
		miRge3.0	1 004 438	318	217
Platelets	ERR747967	Chimira	2 644 772	970	280
		sRNAbench	2 030 951	504	222
		miRge2.0	2 652 547	423	256
		miRge3.0	2 561 109	423	259
Retinal pigment epithelium	SRR5127210	Chimira	5 951 602	1045	427
		sRNAbench	5 274 077	911	318
		miRge2.0	5 802 230	777	372
		miRge3.0	5 742 333	777	371
Cortical neuron	SRR5127204	Chimira	15 447 871	1433	325
		sRNAbench	15 076 370	1033	243
		miRge2.0	15 353 234	873	296
		miRge3.0	15 316 966	875	295
Cardiac fibroblast	SRR5127236	Chimira	7 759 550	1427	410
		sRNAbench	7 489 281	989	320
		miRge2.0	7 674 927	844	367
		miRge3.0	7 661 528	849	367
Renal proximal epithelium	SRR5127209	Chimira	35 894 433	1830	395
		sRNAbench	34 961 340	1382	306
		miRge2.0	35,748,118	1171	359
		miRge3.0	35 697 299	1178	360
Islet alpha cell	SRR1028924	Chimira	–	–	–
		sRNAbench	43 746 104	1177	240
		miRge2.0	43 880 787	910	279
		miRge3.0	43 770 112	913	279

correction relative to the expected number of UMIs based on the overall short lengths. This and other miRNA alignment tools could experience speed enhancements by incorporating GPUs instead of CPUs for computational steps. These improvements are planned in additional iterations of miRge3.0.

In conclusion, miRge3.0 is a significantly improved version of our miRNA alignment tool that we believe has the strongest complement of speed, usability, accuracy and features in this software category.

DATA AVAILABILITY

miRge3.0 source code, bioconda package, PyPi and GUI are available at https://github.com/mhalushka/miRge3.0; https://anaconda.org/bioconda/mirge3; https://pypi.org/project/mirge3/ and https://sourceforge.net/projects/mirge3/files/ respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

The authors thank the Pathology Department IT department for providing a Mac computer for the developing and testing of miRge3.0 GUI on Mac OS (OS-X) and the use of the Ares Linux server (Pathology Informatics). The authors thank Kate Halushka with graphic assistance.

FUNDING

National Institutes of Health [1R01HL137811, R01GM130564, P30CA006973 to M.K.H.]; American Heart Association [17GRNT33670405].

Conflict of interest statement. None declared.

REFERENCES

1. Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.
2. Mendell,J.T. and Olson,E.N. (2012) MicroRNAs in stress signaling and human disease. *Cell*, **148**, 1172–1187.
3. Bushati,N. and Cohen,S.M. (2007) microRNA functions. *Annu. Rev. Cell Dev. Biol.*, **23**, 175–205.
4. Fromm,B., Domanska,D., Hoye,E., Ovchinnikov,V., Kang,W., Aparicio-Puerta,E., Johansen,M., Flatmark,K., Mathelier,A., Hovig,E. *et al.* (2020) MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Res.*, **48**, D1172.
5. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
6. Backes,C., Fehlmann,T., Kern,F., Kehl,T., Lenhof,H.P., Meese,E. and Keller,A. (2018) miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Res.*, **46**, D160–D167.
7. Londin,E., Loher,P., Telonis,A.G., Quann,K., Clark,P., Jing,Y., Hatzimichael,E., Kirino,Y., Honda,S., Lally,M. *et al.* (2015) Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc. Natl. Acad. Sci. USA*, **112**, E1106–E1115.
8. McCall,M.N., Kim,M.S., Adil,M., Patil,A.H., Lu,Y., Mitchell,C.J., Leal-Rojas,P., Xu,J., Kumar,M., Dawson,V.L. *et al.* (2017) Toward the human cellular microRNAome. *Genome Res.*, **27**, 1769–1781.
9. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.

10. Magee,R. and Rigoutsos,I. (2020) On the expanding roles of tRNA fragments in modulating cell behavior. *Nucleic Acids Res.*, **48**, 9433–9448.
11. Looney,M.M., Lu,Y., Karakousis,P.C. and Halushka,M.K. (2020) Mycobacterium tuberculosis infection drives mitochondria-biased dysregulation of host tRNA-derived fragments. *J. Infect. Dis.*, **223**, 1796–1805.
12. Goodarzi,H., Liu,X., Nguyen,H.C., Zhang,S., Fish,L. and Tavazoie,S.F. (2015) Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell*, **161**, 790–802.
13. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
14. Cheng,W.C., Chung,I.F., Tsai,C.F., Huang,T.S., Chen,C.Y., Wang,S.C., Chang,T.Y., Sun,H.J., Chao,J.Y., Cheng,C.C. *et al.* (2015) YM500v2: a small RNA sequencing (smRNA-seq) database for human cancer miRNome research. *Nucleic Acids Res.*, **43**, D862–D867.
15. de Rie,D., Abugessaisa,I., Alam,T., Arner,E., Arner,P., Ashoor,H., Astrom,G., Babina,M., Bertin,N., Burroughs,A.M. *et al.* (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.*, **35**, 872–878.
16. Su,H., Xu,T., Ganapathy,S., Shadfan,M., Long,M., Huang,T.H., Thompson,I. and Yuan,Z.M. (2014) Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene*, **33**, 1348–1358.
17. Desvignes,T., Loher,P., Eilbeck,K., Ma,J., Urgese,G., Fromm,B., Sydes,J., Aparicio-Puerta,E., Barrera,V., Espin,R. *et al.* (2020) Unification of miRNA and isomiR research: the mirGFF3 format and the mirtop API. *Bioinformatics*, **36**, 698–703.
18. Baras,A.S., Mitchell,C.J., Myers,J.R., Gupta,S., Weng,L.C., Ashton,J.M., Cornish,T.C., Pandey,A. and Halushka,M.K. (2015) miRge - a multiplexed method of processing small RNA-Seq data to determine microRNA entropy. *PLoS One*, **10**, e0143066.
19. Lu,Y., Baras,A.S. and Halushka,M.K. (2018) miRge 2.0 for comprehensive analysis of microRNA sequencing data. *BMC Bioinform.*, **19**, 275.
20. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*, **17**, 2.
21. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
22. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
23. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
24. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
25. Vitsios,D.M. and Enright,A.J. (2015) Chimira: analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics*, **31**, 3365–3367.
26. Aparicio-Puerta,E., Lebron,R., Rueda,A., Gomez-Martin,C., Giannoukakos,S., Jaspez,D., Medina,J.M., Zubkovic,A., Jurak,I., Fromm,B. *et al.* (2019) sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Res.*, **47**, W530–W535.
27. Giraldez,M.D., Spengler,R.M., Etheridge,A., Godoy,P.M., Barczak,A.J., Srinivasan,S., De Hoff,P.L., Tanriverdi,K., Courtright,A., Lu,S. *et al.* (2018) Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat. Biotechnol.*, **36**, 746–757.
28. Wright,C., Rajpurohit,A., Burke,E.E., Williams,C., Collado-Torres,L., Kimos,M., Brandon,N.J., Cross,A.J., Jaffe,A.E., Weinberger,D.R. *et al.* (2019) Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics*, **20**, 513.
29. Decmann,A., Nyiro,G., Darvasi,O., Turai,P., Bancos,I., Kaur,R.J., Pezzani,R., Iacobone,M., Kraljevic,I., Kastelan,D. *et al.* (2019) Circulating miRNA Expression Profiling in Primary Aldosteronism. *Front Endocrinol (Lausanne)*, **10**, 739.
30. Gruning,B., Dale,R., Sjodin,A., Chapman,B.A., Rowe,J., Tomkins-Tinch,C.H., Valieris,R., Koster,J. and Bioconda,T. (2018) Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods*, **15**, 475–476.