LICENSE

AASRA

Copyright (C) 2016-2020 Chong Tang, Yeming Xie

This program is free software: you can redistribute it and/or modify it

under the terms of the GNU General Public License as published by the

Free Software Foundation, either version 1 of the License, or (at your

option) any later version.

This program is distributed in the hope that it will be useful, but

WITHOUT ANY WARRANTY; without even the implied warranty of

MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General

Public License for more details.

You should have received a copy of the GNU General Public License along

with this program. If not, see <http://www.gnu.org/licenses/>.

SYNOPSIS

AASRA: a comprehensive solution for small RNA annotation

CITATIONS

If you use AASRA in your work, please cite one of the following:

INSTALL

Dependencies

All dependencies must be executable and findable in the user's PATH

python (version 2.7.x or higher): Generally installed in linux and mac machines by

default. Expected to be installed at /usr/bin/python

bowtie2 (version 2.1.0 or higher): Free from

https://sourceforge.net/projects/bowtie-bio/files/bowtie2/ note: requires bowtie2

... NOT bowtie!

Bowtie2-build: Free from

https://sourceforge.net/projects/bowtie-bio/files/bowtie2/

featureCounts

Install

Download AASRA from the Download section of the Github site. By adding the AASRA directory to your PATH environment variable, you ensure that whenever you run AASRA-index.sh -h or AASRA.sh -h from the command line, you will get the version you just installed without specifying the entire path. Follow the instruction of your operating system to add the directory to your PATH. If you prefer to install AASRA by copying the AASRA files to an existing directory in your PATH, make sure you copy all the AASRA files, including AASRA.py, AASRA AASRA-index.py and AASRA-index.

USAGE

The AASRA-index indexer:

Usage: AASRA-index [options] -i <input\_file> -l <5’\_anchor\_sequence> -r <3’\_anchor\_sequence> -s <output\_SAF\_file>

Default command: AASRA-index -i index.fa -l CCCCCCCCCC -r GGGGGGGGGG -s index.saf

TEST

Test data and brief instructions are available at

http://github/data/AASRA-index\_TestData/

OPTIONS

-h : print a help message

-v : print AASRA-index version

-i : input reference file must be a fasta file (.fasta or .fa)

-l : 5’ end nucleotide anchor sequence.

-r : 3’ end nucleotide anchor sequence.

-s : file name of the output SAF file generated by AASRA-index.

The AASRA aligner:

Usage: AASRA [options] -f <fasta\_input> -p <thread\_number> -i <input\_file> -l <5’\_anchor\_sequence> -r <3’\_anchor\_sequence> -b <anchored\_bowtie2\_index>

Default command: AASRA -p 4 -i sample.fastq -l CCCCC -r GGGGG -b anchored\_index.fa

OPTIONS

-h : print a help message

-v : print AASRA version

-f : the reads input file is fasta file (.fasta or .fa)

-i : input reference file must be a fasta or fastq file (.fasta, .fa, .fastq or .fq)

-l : 5’ end nucleotide anchor sequence.

-r : 3’ end nucleotide anchor sequence.

-b : file name of the output SAF file generated by AASRA-index.

SYSTEM RECOMMENDATIONS

AASRA was developed on devices running Ubuntu 12.04.5 LTS, 64-bit. It has also been tested on Apple Mac OSX and CentOS. At least 4G memory is suggested. Alignments benefit from multiple processing threads, via specifying the -p option. The AASRA-index portion is single-threaded. At least 50G of hard disk space is recommended to be available, due to the generation of possible large size of the temporary alignment files.

The total time of analysis depends on genome size, number of reads analyzed, and your equipment. Excluding building bowtie index, we generally have observed run times for alignment runs to take between 20 minutes and 1 hours using default AASRA settings.

ALIGNMENT METHODS

Details of alignment methods and performance testing

For full details on AASRA’s alignment methods and the results of

performance testing, see Chong et al. (2016). This is a pre-print of a manuscript

that is under peer review as of this writing (March 17, 2016).

Reads pre-processing

Reads file formats

Small RNA index to be aligned must be in fasta formats. The fasta sequence for each small RNA must be in a SINGLE line (NOT multiple lines of sequence for one small RNA).

No paired-end support

There is no support for paired-end reads in AASRA. Small RNA data

are assumed to be single-ended, and represent the 5'-->3' cDNA sequences

of cloned RNAs.

Unique read names required

The small RNA reads must all have unique names within a given file. If

this requirement is not met, alignments will be completely unreliable

due to errors in interpreting and handling of multi-mapped reads.

Adapter trimming

AASRA assumes your reads are already trimmed. Trimming

simply looks for the right-most exact match to the given apdater

sequence, and when found, chops it off. If a read is smaller than 15nts

after trimming, it is discarded. For more sophisticated adapter

trimming, consider cutadapt or trimmomatic.

Alignment overview

AASRA2 uses bowtie2 to align reads. It first anchors the reference index and generate bowtie2 index accordingly. It then anchors the sample reads file and aligned anchored sample to anchored bowtie2 index. The final output is a single .sam formatted alignment file.

mismatches

By default, AASRA allows up to 1 mismatch for a valid alignment.

This helps with sequencing errors and SNPs. If a read has some

alignments with 0 mismatches, and some with 1, only those with 0

mismatches are kept. The option --mismatches controls this threshold,

and can be set to 0, 1, or 2.

OUTPUT FILES

All output files are all put in the same location as the input genome and sample reads files.

Results file

The file Results.txt is a plain-text tab-delimited file that contains

the core results of the analysis. The columns are labeled in the first

row, and are:

sam file

The standard sam formatted alignment will be generated in the AASRA output for downstream analysis.

SAF files

If -s option is used in AASRA-index, a standard SAF will be generated according to the reference index (fasta). The generated SAF file could serve as the featureCount input reference file to count the alignment results.

TEST

Test data and brief instructions are available at

http://github/data/AASRA/testData/