



[Main](#)

[Documentation](#)

[Download](#)

[Examples](#)

[Discussion](#)

[Help](#)

[Links](#)

Documentation

([download pdf version](#))

[1.Introduction](#)

[2.Download](#)

[3.Installation and configuration](#)

[4.Usage](#)

[5.Output](#)

[6.References](#)

1.Introduction

ABMapper is a portable, easy-to-use package for spliced alignment, junction site detection, and reads mapping. The core module was written in C++ and wrapped in PERL scripts.

License:

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 3 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.

2.Download

ABMapper could be downloaded from [here](#)

3.Install and configuration

Download ABMapper Binary and put "ABMapper","runABMapper.pl","libabmapper.PL" under the

same directory and run "perl runABMapper.pl", then you will get the usage information.

Please make sure to keep the working directory under where ABMapper located, otherwise, you should change some path manually.

1. change the require sentence in "runABMapper.pl", to a absolved path of libabmapper.PL

2.change the "\$path" variable in "libabmapper.PL" to the absolute path of ABMapper without the last "/".

If you want to compile it by yourself, please download the source code, use 'make' to compile the program.

4.Usage

1.runABMapper.pl: a PERL wrapper

Usage: perl runABMapper.pl <options>

Required:

-ref <string> : filename of reference list file which contains the number of reference files on the first line and the reference file names on the following lines.

-input <string> : reads file in fastq or fasta format with ".fa|fasta or .fq|fastq" suffix only, if '-sam' was set, this parameter will be ignored.

-sam <string> : SAM file generated by BWA(only BWA's SAM). '-input' will be ignored, if it was set.

Optional:

-error <int> : substitution errors allowed when doing alignment, default 2.

-seed <int> : seed length not larger than 15,default 10. Seed length <9 is not suggested.

-overlap <int> : maximum overlap of two fragments, with which motif is detected, default 15.

-type <int> : output type, when type=0, you can get minimal output with only a short brief summary in screen, and not file output; otherwise, when type =1 or 2, you can get file outputs and the number of outputs is specified by the following parameter "-multi". When type =1 (default), the number of outputs specified is the maximum of those for all the reference files; when type =2, the number of outputs specified is the maximum for each reference file.

-multi <int> :maximum output for every read, default 500. If set to 0, then all the outputs for all the reference files will be output, regardless type=1 or type=2.

-min_dist <int> : minimal distance between two fragments, default 10.

-max_dist <int> : maximum distance between two

fragments, default 400,000.

-output <string> : prefix of output, default is "tmp" concatenated with current second, minute, hour and day of month:tmp-{sec}-{min}-{hour}-{mday}.

SAM options:

-unhit : logic value, if set to true, unhit reads in BWA SAM would be extracted and used to do ABMapper mapping.

-repeat : logic value, if set true, repeat reads in BWA SAM would be extracted and used to do ABMapper mapping.

2.ABMapper: core program to do two-seed mapping

If the number of reads is very large, user can call ABMapper directly from command line. However, it would not produce as many output files as runABMapper.pl.

Usage:

./ABMapper <options>

Options:

Required:

-r<string>: filename of reference list file which contains the number of reference files on the first line and the reference file names on the following lines.

-i<string>: read file in format fasta(.fa or .fasta) or fastq(.fq, .fasta).

Optional:

-l<int>: seed length not larger than 15, default setting is 10.

You can choose proper integers, but seed length <9 is not suggested.

-m<int>: maximum output for every read, default 500. If set as 0, then all the outputs for all the reference files will be output.

-t<int>: default setting is 1 (the file outputs of all chrs <= number of output specified by -m).

You can choose 0 (no file output), 2 (the file outputs for every chr <= number of output specified by -m).

-o<string> : prefix of output files, default setting = 'temp'.

-O<int>: maximum overlapping of two fragments, with which motif is detected, default 15.

-d<int>: minimum exon distance, default 10.

-D<int> : maximum exon distance, default 400000.

-e<int>: maximum substitution errors allowed when doing alignment, default 2. You can specify an integer in 0-2.

Example of reference list file:

###start from next line#####

```
3
chr1.fa
/home/abmapper/chr2.fa
chr3.fa
###end#####
```

'3': means there are three chromosomes; chr1.fa is the name of the reference chromosome file. You can put the chromosome file in any readable directory, but we still suggest putting all chromosomes in one directory. It should be noted that, to specify a path in windows, it is required to use '/', such as D:/abmapper/data/chr2.fa.

5.Outputs

1. output of runABMapper.pl

user can set '-output' in parameter, default is tmp-{sec}-{min}-{hour}-{mday}, In the following, '*' represent the prefix defined by user or default.

- ***_tbl.txt**: tabular information which contains all putative hits and fragments
- ***_pairinfo.txt**: tabular information which contains both 'good' spliced alignment of two fragments extended from two seeds and exonic alignment.
- ***_sum.txt**: summary of alignment. It records hit number and the number of uncertain 'N' in read. Then it is followed by the statistical information
- ***.about**: details combined *_tbl.txt and *_pairinfo.txt together with header. The only difference is that canonical motif would be searched and position would be altered during the combination.
- ***.sam**: SAM format output of fully matched reads, user could use samtools to convert SAM file into BAM format.
- ***.js.bed**: bed file determined by canonical motif and distance limitation
- {input}.repeat.fa**: repetitive reads which have more than {max_occurrence} cutoff hits

2. output of ABMapper

- ***_tbl.txt**: tabular information with header, which contains all putative hits and fragments.
- ***_pairinfo.txt**: tabular information with header, which contains both 'good' spliced alignment of two fragments extended from two seeds and exonic alignment.
- ***_sum.txt**: summary of alignment. It records hit number and the number of uncertain 'N' in read. Then it is followed by the statistical information

6.References

Copyright (c) 2010. All Rights Reserved. Hong Kong Bioinformatics Center &
Department of Computer Science and Engineering, The Chinese University of Hong
Kong